THE PHCNOLOGY AND PHONETICS OF ENGLISH INTONATION

by

JANET BRECKENRIDGE PIERREHUMBERT

A.B., Harvard University
(1975)

SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE
DEGREE OF

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 1980

Signature of Author ______________________
Department of Linguistics and Philosophy
September 9, 1980

Certified by ______________________
Morris Halle
Thesis Supervisor

Accepted by ______________________
Samuel Jay Keyser
Chairman, Departmental Committee

THE PHONOLOGY AND PHONETICS OF ENGLISH INTONATION

by

Janet Breckenridge Pierrehumbert

Submitted to the Department of Linguistics and Philosophy
on 9 September 1980 in partial fulfillment of the requirements for
the Degree of Doctor of Philosophy

ABSTRACT

This thesis develops a system of underlying representation for English intonation. It gives an account of what different tunes are possible and how they are aligned with different texts. It characterizes the rules which map the underlying representations into phonetic realizations.

The different tunes are described as structured strings of L and H tones generated by a finite-state grammar. The strings consist of one or more pitch accents, which are aligned with stressed syllables on the basis of the metrical pattern of the text, plus two additional tones which characterize the intonation of the end of the phrase. The pitch accents are either a tone, or a pair of tones on which a strength relation is defined. The two additional tones are the boundary tone, found at the end of the phrase regardless of the metrical structure of the text, and the phrase accent, which follows immediately after the pitch accent on the main phrase stress and controls the intonation from there to the boundary.

Local context-sensitive rules map the string of tones into the quantitative values which determine the fundamental frequency contour. These rules apply left to right, and include downstep and upstep rules resembling those which have been studied in African tone languages. A transform of the fundamental frequency domain which makes these rules linear is proposed on the basis of experimental data. Evidence is presented that superficially nonlocal intonational characteristics, such as the overall trend of the contour, really arise from local rules. The thesis also reviews other experimental results and explains how they are accommodated within the framework proposed.

Thesis Supervisor: Morris Halle

Title: Ferrari P. Ward Professor of Modern Languages and Linguistics

ACKNOWLEDGMENTS

Above all, I am grateful to Mark Liberman, from whom I learned how to think about the relation between phonetics and phonology. His work on intonation was the starting point for the work presented here, and his insights into where the problems lay have made a crucial contribution to my progress. During my stays at Bell Labs, he always made time not only to discuss theoretical issues, but also to teach me practical skills for doing phonetics.

I am also grateful to Morris Halle, Paul Kiparsky, and Osamu Fujimura for their contributions to my work. Morris Halle and Paul Kiparsky faced my swarms of numbers with equanimity, and helped me see how to fit my results into the broader picture. Osamu Fujimura has encouraged my interest in phonetics since I was an undergraduate, and was instrumental in providing the facilities which made this work possible.

I would like to thank Lauri Carlson for his useful remarks about intonational contrasts, and Bruce Hayes for helpful discussions about the relation of intonation and metrics. The alertness and patience of my main speakers, Mary Breckenridge, Kathy Galotti, Mark Liberman, and David Shipman, made a large contribution to the success of my investigations. The development of my ideas also benefited greatly from the chance to present them at the Sloan seminar on intonation, held at Stanford in March. For this, I would like to thank Will Leben and all the participants.

Lastly, I'd like to thank my typists, Anne Clee and Abby McLaughlin, whose cheerful and expeditious work made finishing the the thesis as painless as possible.

Work on the thesis was supported by Bell Telephone Laboratories, a National Science Foundation Graduate Fellowship, and an American Association of University Women Dissertation Fellowship.

TABLE OF CONTENTS

Page No.

Chapter 1

OVERVIEW

1.1 Introduction

One of the more intractable problems in phonology has been the description of English intonation. It is clear that the same sentence, with the same stress pattern, can be said with many different melodies in English, and that these melodies have an important role in its expressive force. For example, Figure 1 shows five different melodies for "Anna," as determined by computer tracking of the fundamental frequency of the speech waveform. (Fundamental frequency, or FØ hereafter, is the physical correlate of pitch.) In all five of these utterances, the first syllable is stressed and the second is unstressed. In Figure 1 A, there is an FØ peak on the first syllable followed by a fall to the bottom of the speaker's range. This would be a typical pattern when "Anna" was used as the answer to a question. Figure 1 B is similar to 1 A, except that the FØ rises again at the end. This pattern could also be used as the answer to a question; it contrasts with 1 A in carrying an implication that the answer is incomplete. The pattern in 1 C could be used for calling out to Anna. The fall in this FØ contour contrasts with that in 1 A in stopping far short of the bottom of the speaker's range. Figure 1 D shows an FØ contour which is rather similar to that in 1 B. However, while 1 B started at a fairly high level, and rose from the beginning, 1 D starts at a much lower level and maintains this level for a little while before rising. Two common uses of the pattern in 1 D are to convey incredulousness, and to imply that the speaker is giving only one of many possible examples. Lastly, in Figure 1 E, we see a contour which is very low on the stressed syllable and then rises up to the end. This is a common melody for a question: "Is it Anna?"

The number of different melodies multiplies for phrases with more stressed syllables. For example, Figure 2 shows three different patterns for the phrase "another orange," all of which end with a contour on "orange" like that on "Anna" in Figure 1 A. In 2 A, the FØ peak on "orange" is preceded by another slightly lower peak on the stressed syllable of "another." This is a quite neutral way to say this phrase; it might serve as the answer to "What's this?" In Figure 2 B, the peak on "another" is much higher than that on "orange," even though "orange" still has the main phrase stress. One use of this pattern is to convey judiciousness. In 2 C, the stressed syllable of "another" has a very low FØ. This pattern is often used to convey surprise, or to imply that the speaker is repeating something he really should not have to repeat. Figure 3 gives a similar set of examples in which the pattern on "orange" is like the question pattern on "Anna" in Figure 1 E. In 3 A, "another" has an FØ peak, while in 3 B, it has a low FØ value as it did in 2 C.

Not only can the same text have many different melodies, the same melody can occur on many different texts. Figure 4 gives two additional examples of the melody shown on "another orange" in Figure 3 C. In 4 A, the sentence is, "It's really too good to be true"; in 4 B, it is, "That's a remarkably clever suggestion." In all three examples, the FØ starts out relatively high, falls to a low value on the main stress of a word ("really" in 4 A and "remarkably" in 4 B), rises gradually to a peak on the main stress of the phrase, and then falls to the bottom of the speaker's range. The expressive force of the melody also seems to be similar in all three cases; all would do well as surprised exclamations.

Pursuing this line of observation, we note that the same melody can be aligned with a given sentence in several different ways, which correspond to the different options in assigning phrasal stress. This point is illustrated in Figure 5 and 6. Figure 5 A shows an FØ contour for the sentence, "Legumes are a good source of vitamins." The main stress of the phrase is on the work "vitamins," and this word has the fall-rise contour which was originally exemplified in Figure 1 B. (The vertical dotted line marks the onset of the syllable with the main phrase stress here and in Figure 6.) Figure 5 B shows an FØ contour for the same sentence, with "good" under focus: "Legumes are a <u>good</u> source of vitamins." Now, "good" carries the main stress in the phrase, and the fall-rise pattern is stretched out over "good source of vitamins." Note that the peak remains on the main phrase stress, the end of the rise remains at the end of the phrase, and it is the bottom of the configuration which is stretched out to cover the additional material. Figure 5 C shows an FØ contour for "<u>Legumes</u> are a good source of vitamins." Here, "Legumes" has the main phrase stress, and the fall-rise pattern covers the whole sentence.

Figures 6 A through 6 C show a similar series for, "Are legumes a good source of vitamins?", produced with the intonation originally introduced in Figure 1 E. One of the interesting points about these contours is that what appears in 6 A as a rise frum the nuclear stress to the end of the phrase resolves itself into two parts when spread out over more material: a rise which immediately follows the low FØ on the nuclear stress, and a rise at the very end of the phrase. In between, we see an FØ plateau.

One main aim of this thesis is to develop an abstract representation for English intonation which makes it possible to characterize what different patterns a given text can have, and how the same pattern is implemented on texts with different stress patterns. The second aim is to investigate the rules which map these phonological representations into phonetic representations. These two aims go hand in hand, since we seek the simplest possible underlying representation by determining what properties of the surface representation can be explained by rules applying during the derivation instead of being marked in the underlying form.

The phonological characterization of intonation has three components. The first is a grammar of allowable phrasal tunes. This grammar generates sequences of L and H tones, with structure which we will discuss shortly. The second component is a metrical representation of the text. For this, we will use the metrical grid developed in Liberman (1975) and Liberman and Prince (1977). The grid tells us which syllables are stressed and which are unstressed, and also describes the relationships in strength among the stressed syllables. The strongest stress in the phrase, the nuclear stress, will have a particularly important role in the description of intonation. Lastly, we have rules for lining up the tune with the text. The complete phonological representation for intonation is thus a metrical representation of the text with tones lined up in accordance with the rules. In other languages, rules which alter tonal values or delete tones can apply to such a representation. English appears to lack such rules, with the result that the underlying and derived phonological representations of intonation are identical. The rules of interest are thus the rules which assign phonetic values to tones and construct the FØ contour between one tone and the next.

What will be used here as the phonetic representation, or the output of these implementation rules, is the FØ contour. The choice of this representation as against a fine transcription in the character of IPA segmental transcription is theoretically motivated. One of the main themes of the work presented here is that interesting language specific rules can be found all the way down to a quantitative description of speech. There is no well-defined level of description which is more concrete than a derived phonological representation, yet still linguistic rather than quantitative in character, at which the linguist may leave off and turn his work over to the physiologist. As a result, many of the rules presented below are schemata for computing quantitative values for tones, while other rules make use of these values. The rules will be shown to explain both quantitative features of the FØ contour, and also characteristics which one might at first take to be qualitative, such as the overall shape. We will argue against alternative descriptive frameworks in which a level corresponding to the phonetic transcription is posited.

On the other hand, the choice of the FØ contour as the phonetic representation over a description framed in terms of the motor control has purely practical motivations. We have no doubt that various regularities in how the transition between one tonal value and the next is carried out will eventually be traced to the constraints of laryngeal control. Regularities in the scaling of tonal values may be related to the interaction of laryngeal and respiratory control. The possibility that the H and L are articulated differently even for the same FØ value also needs to be considered. In view of these possibilities, there would be no justification for taking the view that the acoustic characterization of intonation is theoretically primary. Rather, the reason to concentrate on the characterization of FØ contours is that these are the most accessible data which are relevant to a quantitative description of intonation. FØ contours can be obtained in quantity with the aid of a computer program for pitch tracking; articulatory data on intonation must be obtained by much more difficult and painstaking techniques, such as electromyographic studies of the laryngeal muscles, and tracheal punctures. It would thus appear sensible to proceed as far as possible on the basis of FØ contours, given that they incorporate regularities which either an acoustic or articulatory theory of intonation should be able to account for. We hope that the results of our study of FØ contours will be useful in identifying the problems to which experimental work on the production of intonation should be addressed.

The phonetic representation of intonation, as realized in the FØ contour, has particular importance as evidence about the underlying representation, because of the failure of methods for accessing a more abstract representation which have been useful elsewhere in linguistics. For example, segmental transcription by ear has been a useful tool even though the resulting representation may not be a well-defined level in the theory. Transcription by ear of intonation is the source of so many gross errors in the literature that we feel it cannot be relied on. Judgments of what is a variant of the same word, and what represents a different (though perhaps synonymous) word have often been of use to phonologists. The capability for introspection about the form of intonation patterns seems to be far inferior to that which native speakers can bring to bear on lexical items. At best, one can obtain judgments about whether the meaning conveyed the intonation is similar or dissimilar, but judgments about sameness or difference of form are not available. Linguists have relied heavily on informants' judgment that linguistic forms are not well-formed in their language. As Maeda (1976) has shown, it is possible to obtain such judgments about intonation patterns. However, the investigator has such difficulty producing wrong intonation patterns reliably that one must rely, as Maeda did, on computer synthesis to generate the nonoccurring patterns. This means that it is not practical to obtain such judgments at an early stage in the investigation, but only after arriving at strong hypotheses about how intonation is represented and mapped into FØ contours.

The primary difficulty in using the F0 contour as the phonetic representation of intonation is the extent to which it is affected by the speech segments. Not only is it disrupted by unvoiced sounds, but also there are substantial effects of the segments on the F0 during voiced parts of the signal. The F0 at the onset of the vowel after an unvoiced consonent is considerably higher than after a sonorant consonant. There is a sharp dip in F0 in the vicinity of voiced obstruents and glottal stops. High vowels raise the F0; the difference between a high vowel and a low vowel in the same intonational context can be as much as 25 Hz. Such effects on F0 have been the object of much study in their own right (see for example, Peterson and Barney, 1951; House and Fairbanks, 1953; Lehiste, 1970; Lea, 1973; Ohala, 1978). They bear on questions about the acoustics and physiology of speech, and the consonantal effects also play a part in linguistic theory as a historical source of tones (Hombert, Ohala and Ewan, 1979). However, from the point of view of characterizing the intonation system synchronically, they are a source of noise which must be factored out. While the qualitative features of segmental effects are fairly well understood, we have no good quantitative theory describing their relation to stress, intonation pattern, and overall range. As a result, it is sometimes difficult to separate F0 excursions arising from the intonation system from ones which arise from segmental effects, and it is often difficult to determine precisely the location of tones with respect to the text. These difficulties can be in some measure addressed by using textual material which is entirely voiced, but it is not possible to compose natural texts in which they are entirely controlled for.

Figures 1 through 6 illustrate some features found in FØ contours which we will want the theory to account for. First, we will want to describe the options in the treatment of metrically strong syllables. For example, in Figure 2 A, the stressed syllable of "another" has a peak. In 2 C, the same stressed syllable has a very low FØ value. These and other options will be described by positing an inventory of tonal patterns which may be assigned to metrically strong syllables. These are the pitch accents. In Figure 2 A, the pitch accent on "another" is a H tone in 2 C, it is a L tone. We will also see that pitch accents may consist of an ordered pair of two tones. In Figure 1 D, for example, the pitch accent on "Anna" is L+H. The H tone is responsible for the peak in the FØ contour, while the L is responsible for the low onset that we noted above. A bitonal accent, H+L, will also be invoked to account for the contrast of the contour in Figure 2 B to the contour in 2 A, which has two H accents. In this case, however, the existence of the L tone in the underlying representation is less obvious in the FØ contour and must be motivated indirectly. The complete inventory of pitch accents is reviewed in Section 2, and described in detail in Chapters 2, 4, and 5.

A second concern will be to explain the characteristics of the FØ contour at the end of the phrase, after the last pitch accent. Here, there are a number of options which relate to the interpretation of the expressive force of the phrase as a whole. For example, Figure 5 shows several instances of an FØ pattern which is frequently used on an incomplete statement. The last pitch accent, on the main stress of the phrase, is H; after the peak corresponding to the H, the F∅ drops and stays low until the very end, where it rises. Figure 6 shows an F∅ pattern which is commonly used on yes/no questions. Here, the last pitch accent is L; the F∅ contour rises, makes a plateau, and then rises again. An important contrast between the tonal characteristics of the end of the phrase and those found earlier in the phrase is that F∅ movements at the end of the phrase do not line up with metrically strong syllables. In both of the examples just given, there is an F∅ excursion at the end of the phrase even though the last syllable in the phrase is not metrically strong. This behavior is accounted for in Liberman (1975) by positing tones which align with the phrase boundary, and this solution is adopted here. In both Figures 5 and 6, the boundary tone is H; the tone mapping rules will account for the fact that the H in 6 is so much higher than that in 5. The behavior after the last pitch accent but before the phrase boundary will be accounted for by positing an additional tone, the phrase accent, as Bruce (1977) did in his elegant description of Swedish. The phrase accent is L in Figure 5 and H in Figure 6. The phrase accent is positioned near the end of the word with the nuclear stress, in our data, and thus controls the course of the F∅ contour immediately following the nuclear pitch accent. The separate influence of the phrase accent and the boundary tone can be seen clearly when the nuclear accent is some distance from the end of the phrase, as in Figures 6 B and 6 C. The first rise in these contours is the rise from the L nuclear accent to the phrase accent, and the second rise is from the phrase accent to the boundary tone. The L phrase accent is also evident when it is sandwiched between two H's, as in Figures 1 B and 1 D.

A third area of interest will be the characteristics of the FØ contour in between the tones. Three types of cases are found. First, the FØ contour may take a direct course between two tones, as in Figures 2 C and 4, where the FØ starts rising immediately after the L tone and ends its rise at the H tone. Second, one finds cases where the FØ continues level and then rises at the last instant, as in the phrase accent/boundary tone sequence in Figures 5 B, 5 C, 6 B, and 6 C. Third, we will see cases in which the FØ dips down between two H tones, even without an intervening L tone, as in Figure 2 A. These features of the FØ contours will be accounted for by phonetic rules for interpolating between tones. In particular, the level stretches will be accounted for by a tone spreading rule. We will show this rule to be phonetic rather than phonological in character, in the sense that its applicability is controlled by quantitative relations rather than the underlying tonal description.

The last major area of concern in our description of FØ contours will be to explicate the relation of tonal pattern to control of pitch range. English makes considerable expressive use of pitch range, with the result that what is clearly the same basic intonation pattern can be produced in many different pitch ranges. The reader can persuade himself of this by calling out to someone he imagines to be across the room, and then across the street. Chapter 3 will show that examination of how tonal relations are scaled quantitatively under changes in overall range will yield considerable insight into the form of the tone mapping rules. We will also be concerned with distinguishing instances of tonal differences from instances of pitch range differences. As we introduce more complex intonation patterns, it will become evident why this has proved to be a difficult and controversial problem in the past.

Our approach to this problem has two main elements. First, our tone mapping rules alter the phonetic values assigned to the two tones as a function of their tonal context. For example, we already mentioned that the boundary tone is raised after a H phrase accent. H is also lowered (or downstepped) after a H+L accent; this rule is responsible for the relative peak levels in Figure 2 B. Because of these rules, a H at the end of the phrase can end up lower than a L earlier in the phrase, and a L at the end can be higher than an earlier H. The results thus contrast with those of a system in which some fixed portion of the overall range for each phrase is allocated to each tone. The context sensitive tone mapping rules carry the major burden of making a two-tone description of English intonation possible; without them, it is doubtful that a description with any reasonably small number of levels could succeed. Second, the system permits a new choice for expressive use of pitch range at each new pitch accent. The choice is made on the basis of the stress subordination in the phrase and the speaker's desire to highlight particular information. Sections 3 and 4 elaborate this idea. In contexts where both the first and the second of these influences on pitch range are applicable, the two seem to interact multiplicatively in determining the FØ value of a given tone. Chapter 4 presents an account of tonal evaluation which makes this interaction seem natural.

1.2 Well-formed Tonal Sequences

The thrust of our observations about Figures 1 through is that tunes are linguistic entities, which have independent identity from the text. Tunes and texts cooccur because tunes are lined up with texts by linguistic rules. These two ideas were among the major results of Liberman (1975). One of the questions which they lead to, and which was answered incompletely there, is: What are the well-formed tunes of English? To answer this question, we need to know what counts as one tune and what counts as several tunes in a row, and we need a characterization of the internal makeup of a tune.

Like other researchers, we will take the melody for an intonation phrase to be the "tune" whose internal makeup is to be described. As a rule of thumb, an intonation phrase boundary (transcribed here as %) can be taken to occur where there is a nonhesitation pause or where a pause could be felicitously inserted without perturbing the pitch contour. Figure 7 shows an F0 contour in which the intonation phrase boundaries are marked with pauses. However, in normal speech, one finds many cases where the boundary is not marked by a pause, but only by lengthening of the last syllable in the phrase. Such a case is shown in Figure 8. There is a considerable literature about how an utterance is broken up into intonation phrases (see Halliday, 1967; Downing, 1970; and Bing, 1979). This is really a problem in the relation of syntax and semantics, which is outside the scope of this thesis. So, we will not attempt to give a rigorous theory of where phrasing breaks occur, or to what communicative purpose, but only make a few observations to give the reader some understanding of what we are referring to.

First, it is known that some constructions are obligatorily set off as a separate phrase. These include nonrestrictive relative clauses and noun phrases which have undergone left dislocation out of nonsubject position:

1) The new version % which is much easier to use % will be on the market in 1981

2) That hot pepper oil % you shouldn't put too much of it on.

Parentheticals are ordinarily a separate phrase:

3) This wine, as you might grudgingly call it, came from Century Vineyards.

Many proposed adverbial constructions are also:

4) In spite of what he said % I don't think we should do it.

However, in many or perhaps most cases, a sentence can be run together as one phrase or chopped up into several at the speaker's discretion. For example, Figures 8 and 9 show two different phrasings for the sentence, "Does Manitowoc have a library?" Figures 10 and 11 show a similar pair for, "Anna came with Manny." Figure 10 would be a likely answer to the question, "Who did Anna come with?" Figure 11, containing two intonation phrases, would be a more likely answer to a double barreled question: "Who came? And with who?" When the phrasing is optional, there are still constraints on where phrase boundaries can occur. For example, the phrasing indicated for 5) is rather bad:

5) Three mathematicians % in ten derive a lemma

When uttered with the indicated phrasing, as in Figure 12, the meaning conveyed is not 5) but 6):

6) Three mathematicians % intend to rival Emma.

Similarly, the a) version of the following sentence is far better than the b) version:

7a) Both the bumper and the bashed-in fender % will have to be replaced.

b) Both the bumper and the bashed-in % fender will have to be replaced.

It is a matter of dispute whether the phrasings we have indicated to be bad fail because the resulting phrases are not syntactic constituents, because they are not semantic units. There are many cases in which acceptable phrasings involve what appear to be syntactic non-constituents:

8) This is the cat % that ate the rat % that stole the cheese . . .

In some of these cases, the intonation phrases are syntactic constituents provided that we countenance string vacuous applications of independently motivated syntactic readjustments. 8), for example, could arise from extraposition of the relative clause by the same rule which applies in 9) to extrapose the relative clause from its underlying position in the subject noun phrase to the end of the sentence.

9) That exterminator called that was supposed to do something about the carpenter ants.

This makes a syntactic account of phrasing more plausible than examples like 8) might at first suggest. The existence of the bitonal accents also directly improves the prospects for constructing a syntactic account of phrasing. In particular contexts, these accents generate FØ configurations which are quite similar to configurations arising from a phrase accent and a boundary tone. This means that observations in the literature that a sentence can be uttered in a way that seems to have an intonation break at location x cannot really be taken at face value. To demonstrate that the theory must allow an intonation break at location x, it is necessary to find an intonation pattern for the sentence which would not admit an alternative analysis based on bitonal accents.

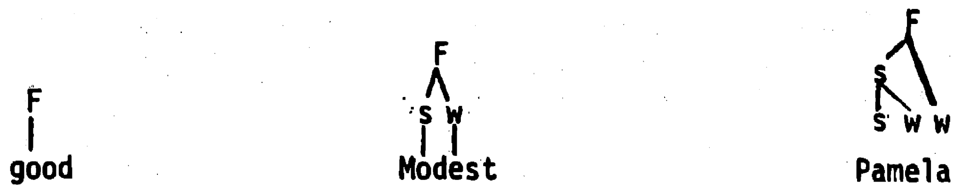
The thread of our observations is that a syntactic account of phrasing is put on the defensive by examples in which surface nonconstituents appear to be intonation phrases, but that it may yet be salvaged by finding alternative interpretations for these examples.

The well-formed tunes for an intonation phrase are comprised of one or more pitch accents followed by a phrase accent and then a boundary tone. (There is also a leading boundary tone after a pause.) The pitch accents themselves consist of either a tone or a pair of tones. We have already pointed out the L and H accents in Figures 2 A and 2 C. Figure 10 showed an FØ contour with a two tone pitch accent, L+H. The FØ starts low and then rises to a higher value, even though there is no second stressed syllable to carry another pitch accent. (The subsequent fall and rise are accounted for by the phrase accent, which is L, and the boundary tone, which is H). When this same pitch accent is not crowded with other tones onto such a brief text, the H generally falls on a syllable following the accented syllable and only the L is on the accented syllable. An example illustrating this point is given in Figure 13, where L is found on "rig" in "rigamarole," and H does not occur until "ma." This pitch accent contrasts with the L+H accent shown in

Figure 14, in which the H falls on the accented syllable and the L falls on an immediately preceding syllable. Thus, pitch accents can differ not only in the tones which make them up, but also in a feature controlling alignment with the text. The notation which will be used to refer to such contrasts is an extension of the notation in Goldsmith (1976); the tone which falls on the accented syllable will be marked with a star. The tone which leads or trails the starred tone will be marked with a raised hyphen. Hence, the transcription for the accent in Figure 13 will be L*+H^{-}, and that for the accent in Figure 14 will be L^{-}+H*. There is also a H^{-}+L*, which occurs on "bring" in Figure 15, and a H*+L, which occurs in Figure 2 B on "another," though not transparently. In Chapter 5, the H*+H^{-} accent will be introduced. This accent contrasts with a plain H* accent because the H^{-} is subject to tone spreading and generates FØ plateaus which would not arise from H*. The theoretically possible accents L*+L^{-}, L^{-}+L* and H^{-}+H* do not exist for what we will suggest are systematic reasons: the implementation of accents is such that there are no contexts in which they would be contrastive.

The starred/unstarred relation in pitch accents may be compared to the stressed/unstressed relationship within the metrical foot, an entity which will also play a role in our discussion of text/tune association. The metrical foot, as discussed by Selkirk (1980) and Hayes (1980) is comprised of a stressed syllable and associated unstressed syllables, which are organized hierarchically. (The hierarchical organization is represented in Hayes (1980) and Selkirk (1980) by a tree structure, but could in principle be represented as a domain in the metrical grid developed in Liberman and Prince (1977) and discussed below.) If a word has only one stressed syllable, it has only one foot. Here are some examples of words consisting of a single foot:

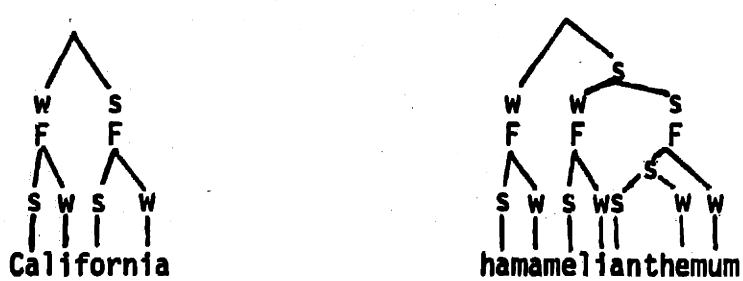
10)



In 10b) and c), s and w indicate the stress relations among the syllables. 's' means "relatively stronger," and 'w' means "relatively weaker"; the syllable which is not dominated by any w's is thus the strongest one in the foot. A general term for the element with the greatest metrical strength in a domain is the "designated terminal element." For example, "Pam" is the designated terminal element of its foot, and a syllable with the main stress in the phrase is the designated terminal element of the phrase.

Words which have more than one stress are composed of a number of feet, which are themselves organized hierarchically. The s w labelling above the level of the foot indicates the stress subordination among the feet.

11)



Distinguishing the foot tree from the word tree, as proposed in Selkirk (1980), gives a way of describing the contrast in stress pattern between pairs like "modest"/"gymnast" and "banana"/"bandanna." These contrasts can be described as follows:

12)

13)

This description relies on the claim that the designated terminal element of a foot has the same status whether the foot is monosyllabic or polysyllabic, and as a result gets full vowel quality, longer duration, and so on, in either case. Thus, once the description of stress uses feet, the feature [± stress] used in Liberman and Prince (1977) is no longer necessary.

The descriptions of English word stress given in Liberman and Prince (1977) and Hayes (1980) carry over the claim advanced in Chomsky and Halle (1968) that English stress is largely predictable. The factors which enter into this prediction in the most recent version, Hayes (1980), are universal constraints on foot structure and tree structure, and language particular rules for constructing structure on strings of syllables.

The bitonal accents resemble bisyllabic feet in that they consist of two elements ordered in time on which a strength relationship is defined; the starred tone is the stronger one, and the unstarred tone is the weaker one. The single tone accents are the counterparts to the monosyllabic feet. We noted that the one syllable in a monosyllabic foot has the same status as the strongest syllable in a longer foot. Similarly, a single tone accent behaves like a starred tone in that it lines up with the accented syllable rather than leading or trailing. Thus, the single tone accents will be represented as L* and H*. Because the pitch accents can have only one or two tones, they can have only these two basic metrical structures; there are no tonal structures corresponding to the more complex derived English foot structures found, for example, in "banana" and "Pamela" above. We will see in the next section that the hierarchical structure for the entire tune is also impoverished compared to the structures found for texts. Lastly, we see from the coexistence of H^*+L^- and L^*+H^- with H^-+L^* and L^-+H^* that the metrical structure for pitch accents is not predictable. Instead, the bitonal accents with their associated structure appear to be lexical items in the intonational system.

The end of the intonation phrase has distinctive tonal characteristics, apart from those attributable to the pitch accents. These characteristics will be attributed to the existence of two extra tones following the pitch accent on the nuclear stress of the phrase. These are the boundary tone, which occurs at the phrase boundary regardless of the preceding stress pattern, and the phrase accent, which is placed shortly after the nuclear accent regardless of how soon the phrase boundary occurs. In Figures 5 and 6, a H boundary tone is responsible for the rise which stays at the end of the phrase as the nuclear accent is shifted. In Figures 1 A and 2, the boundary tone is L and the FØ contour ends at the bottom of the speaker's range. The phrase accent in

Figure 5 is located near the end of the word carrying the nuclear stress, and spreads to the right by a rule formulated in Chapter 5 to create the sustained low value found in Figures 5 B and 5 C. Similarly, in Figures 6 B and 6 C, the H phrase accent is responsible for the rise which starts right after the nuclear L* accent; the spreading rule is responsible for the plateau which spans the rest of the phrase up to the final rise.

The phrase accent and the boundary tone can also be seen clearly in Figures 1 B and 1 D. In these contours, there is only one stressed syllable to carry a pitch accent. A pitch accent can have at most two tones, but the contour in 1 B must clearly be described using three tones while that in 1 D requires four. Thus, these contours must be described using a phrase accent and boundary tone in addition to the pitch accent. It is only in phrase final position that FØ contours of this complexity can be assigned to textual material with just one stressed syllable.

In Figure 6 A, the rise from the nuclear accent to the phrase accent and that from the phrase accent to the boundary tone are compressed onto a small amount of material, and so the phrase accent is not evident as a corner in the FØ contour. It still plays an important role in how the contour is computed, however. In Chapters 2 and 4, we will discuss a tone mapping rule which raises the phonetic value of a boundary tone after a H phrase accent. It is because of this rule that the H boundary tone in Figures 6 B and 6 C is so much higher than the H phrase accent, and it is also involved in computing the value of the boundary tone in Figure 6 A.

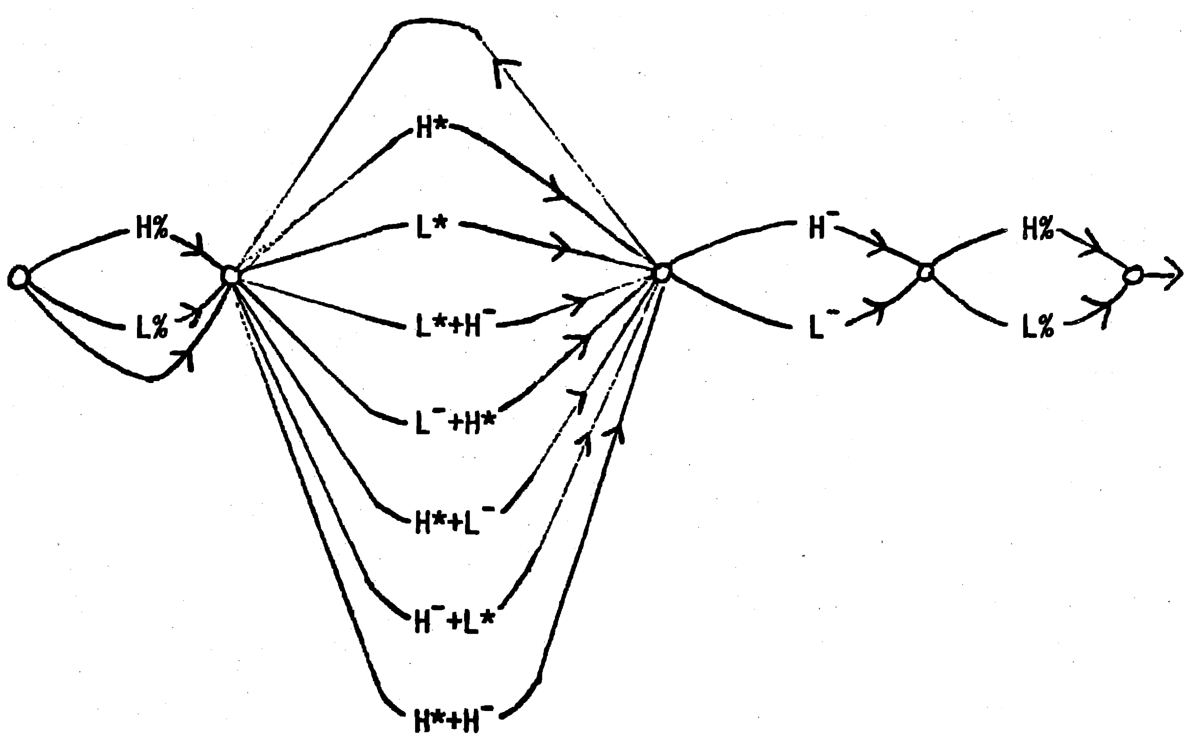
The phrase accent and the boundary tone are each a single tone; neither can be bitonal like a pitch accent. All four possible combinations of the two tones are found. Observations made by Trager and Smith (1951), Ladd (1978), and Carlson (MS) suggest that a meaning for the boundary tone can be identified, and so there is no reason to view the combination of phrase accent and boundary tone as a lexical unit. We will also see that all possible combinations of nuclear pitch accent, phrase accent, and boundary tone correspond to well-formed F0 contours in English. The appendix to the Figures gives a schema for the F0 contour which would result by the phonetic rules developed in the thesis from each of the possible combinations, and also an F0 contour exemplifying the schema.

Our notation system uses %, the symbol for the intonation phrase boundary, as a diacritic for the boundary tones. That is, the two boundary tones will be referred to as L% and H%. Once the boundary tone has been singled out, the phrase accent and the unstarred tones in bitonal accents group together as tones which line up neither with metrically strong syllables nor with the phrase boundary. Chapter 5 will show that the tone spreading rule also treats these tones as a class; these are the only tones which can undergo spreading. Thus, we will extend the used of the raised hyphen to mark phrase accents as well as unstarred tones in pitch accents. With this notation, the F0 pattern in Figures 1 B and 5 C is transcribed as H* L^- H%; that in 10, as L*+H^- L^- H%; that in Figures 1 E and 6 C, as L* H^- H%. The symbols L and H will be reserved for referring to any L or H tone, regardless of whether its diacritic is *, %, or $^-$. Similarly, T*, T%, and T^- will be used to refer to tones with the indicated diacritic, whether L or H.

It is important to note that the diacritics *, %, and $^-$ are unrelated to tonal value. That is, the difference between H* and H% is not to be compared to the difference between H and M in some other language; both H* and H% are equally H tones, but they differ in how they are associated with the text.

Given these observations, it is possible to formulate a grammar which generates the set of well-formed tonal sequences for an intonation phrase. Our hypothesis is that this grammar is finite state, and thus can be represented by the following transition network:

14) Boundary Tone | Pitch Accents | Phrase Accent | Boundary Tone



It is clearly an idealization to allow tonal specifications of indefinite length; in our experience, a phrase most commonly has two or three pitch accents, and phrases containing more than five are quite rare. However, the claim that pitch accents occur in any combination, and with any phrase accent and boundary tone, appears to be better motivated. As we mentioned, all sequences of nuclear accent, phrase accent, and boundary tone are well-formed in English. We have also found examples of many different combinations of pitch accents within the phrase; we have not been able to identify any cooccurrence restrictions among the pitch accents, though it is, of course, possible that future research will.

There are both important similarities and important differences between this characterization of the English melodies and proposals made in Liberman (1975) and Bolinger (1958). As in Liberman's account, the tunes are taken to be well-ordered strings of tone levels. Theories of English melody which could not be described in this way include theories in which there are several levels of tonal specification which add up in some way, and theories in which the FØ contour is decomposed into a sequence of FØ changes rather than into a sequence of levels at crucial points, which are connected up. In the course of the following chapters, we will attempt to motivate the choice of Liberman's general approach over these alternatives. A number of more specific features of Liberman's account are not taken over here. Liberman has four tones, L, LM, HM, and H, but our theory has only two, L and H. In Liberman's account, the tunes are entries in an intonational lexicon. In ours, they are productively generated by a grammar. Our impression is that the meaning of the contours is in general compositional from the pitch accents, phrase accent, and boundary tone; the tunes in Liberman's lexicon may then be characterized as intonational idiom chunks. The tunes in Liberman's lexicon all exhibit what would count as single tone pitch accents in our theory. He offers no theory of the types of intonation which will be described here using bitonal accents. Lastly, the form of our finite state grammar for tunes means that they have no internal bracketing above the level of the pitch accent. In Liberman's account, tunes have a metrical tree structure in which what count here as the nuclear pitch accent and the phrase accent appear as a constituent.

Our theory shares with Bolinger's the idea that pitch accents are morphemes and that a phrase can contain a mixture of different pitch accents. We take cognizance of his observation that a pitch accent can impose a particular relationship between the FØ on the accented syllable and the immediately preceding or following FØ value, independent of the existence of any other accents. In his theory, all pitch accents are like this, and they are accordingly described in terms of FØ changes. In our theory, the bitonal accents have this property and there are also two single tone accents which do not. Furthermore, since our system has no tritonal accents, we would claim that a pitch accent cannot constrain the FØ both to the right and to the left of the accented syllable.

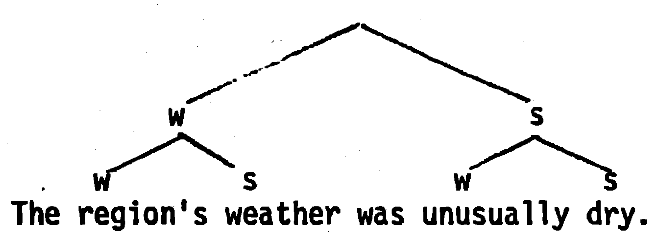
In the brief discussion of phrasal intonation in Bolinger (1958), it is presented as a problem which is separate from the problem of describing the pitch accents, and which is still unsolved. A description in which the phrasal intonation appears in the overall direction of the pitch accents is suggested. In our theory, there is no separate layer of phrasal intonation. It appears that some melodic contrasts which Bolinger would attribute to differences in phrasal intonation are generated by the phrase accent and boundary tone in our theory. Some contrasts which he would describe as involving the same accents but a different overall trend arise in our system as pitch accent differences. In particular, the bitonal accents which trigger downstep affect the overall trend. Differences in overall trend can also arise in our system through expressive use of pitch range. Our distinction between expressive use of range and accent differences takes up one of the main points in Bolinger (1951). Nonetheless, it is unclear whether this distinction would be applied to particular cases in his theory in the same way as in ours.

1.3 Association Rules

The association of the boundary tone with the text is straightforward. The boundary tone is found at the end of the phrase, regardless of the metrical structure of the phrase. In a theory in which the structure of the text is described by a hierarchical structure of domains rather than using boundary symbols, this means that the boundary tone aligns with the right edge of the intonation phrase. In our data, the phrase accent is found near the end of the word with the nuclear stress even when this is not a metrically strong position. There is a certain amount of variation in its placement, but it seems unlikely to us that this variation is linguistically significant. Thus, the interesting problem in text/tune association is where the pitch accents go. The basic observation is that pitch accents are assigned to metrical feet on the basis of the metrical structure for the entire phrase. The outcome is that the designated terminal element of an arbitrary metrical foot may, but need not necessarily, carry a pitch accent. Since all outputs of the grammar of allowable tonal sequences have at least one pitch accent, the well-formed text-tune associations have a pitch accent on the nuclear stress, or designated terminal element of the phrase.

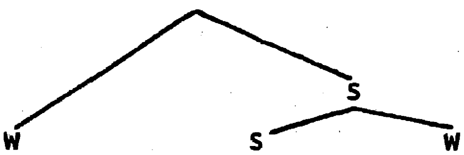
These observations can be made more precise using the metrical grid notation for stress developed in Liberman (1975) and Liberman and Prince (1977). The metrical grid is a device for interpreting the metrical tree, which was introduced above for describing word stress. Here, we will be interested in the use of these representations to describe phrasal stress subordination. An example illustrating the metrical tree at the phrase level is given in 15):

15)



The internal structure of the metrical feet is omitted. The form of the tree is the same as the syntactic structure, except that "the" and "was" are assumed to have been cliticized, or in other words incorporated into a foot with a word stress. As above, s and w are used to label the stronger and weaker nodes of each pair by the Nuclear Stress Rule (Chomsky and Halle, 1968; Liberman and Prince, 1977). The Nuclear Stress Rule appears to be a default case at the phrase level. The labelling at any level of the tree may be reversed to highlight particular information in the phrase, and in some locutions s w labelling is actually more common than the w s labelling which would result from the Nuclear Stress Rule. For example, a typical stress contour for, "We're looking for an apartment to rent," would be:

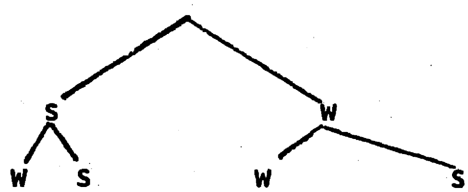
16)



We're looking for an apartment to rent

The stress contour which would result from assigning the nuclear stress of the phrase to "rent," as the Nuclear Stress Rule would, comes off as a correction or an implicit contrast. Similarly, the most normal rendition of most sentences involving "even" and "only" has the nuclear stress on the constituent they modify:

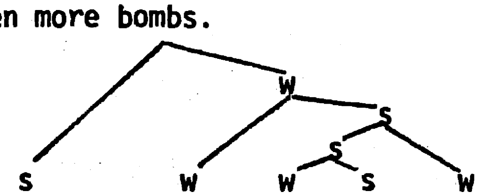
17)



Even John has agreed to contribute.

However, as example 18) shows, this is only a tendency; in the right circumstances, the constituent modified by "even" can be less prominent:

18) -- I think we should build even more bombs.



-- Nonsense! It would be a disaster to build even more bombs.

There has been a considerable controversy about the relative roles of syntax and pragmatics in controlling stress subordination.

Noteworthy work on this topic includes Bresnan (1971), Berman and Szamosi (1972), Bolinger (1972), Bresnan (1972), Schmerling (1974), and Ladd (1978). We have reached three basic conclusions from this literature. First, pragmatics has a very important role in determining stress subordination. Second, one must distinguish the typical stress subordination in a sentence from the stress subordination computed by syntactic rule, because the most normal pragmatics for the sentence may predict the exceptional stress pattern. Third, the syntactically based Nuclear Stress Rule appears to be needed to supply default values for stress subordination. However, how stress subordination arises is not our main concern. For our purposes, it is enough to have a good representation of the outcome, which can be applied in describing the intonation system.

The metrical tree in itself defines only a partial ordering among the stressed syllables. For example, in the tree for, "The region's weather was unusually dry," in 15) above, the labelling for the bottom level of branching says that the designated terminal element of "dry" is stronger than that of "unusually," and that of "weather" is stronger than that of "region's." The labelling for the top level of branching says that the designated terminal element of "was unusually dry" is stronger than that of "The region's weather." In short, "dry" is stronger than "wea." Without additional conventions, the relative strength of the stressed syllables in "unusually" and "region's" goes undefined.

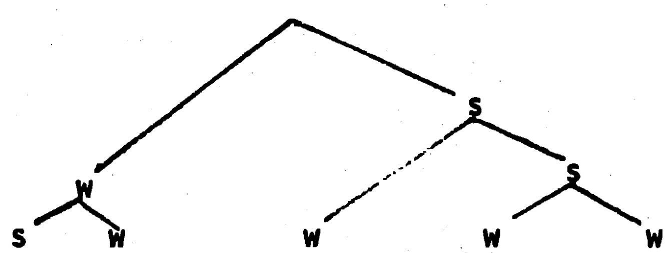
The position taken in Liberman and Prince (1977) is that there is a level of description, the metrical grid, at which the relative strength of such syllables is defined. The metrical grid contrasts with the tree in lacking bracketing and in defining a complete rather than partial ordering of relative metrical strength. The degrees of freedom left over by the tree structure are filled in as optional variants in the construction of the grid. To be precise, any grid is possible which meets the following convention:

19) Relative Prominence Projection Rule

In any constituent on which the strong-weak relation is defined, the designated terminal element of its strong subconstituent is metrically stronger than the designated terminal element of its weak subconstituent (p. 316).

With regard to 20) for example, 19) says that "worms" is stronger than any of the preceding syllables with word stress but does not define the relation of these to each other.

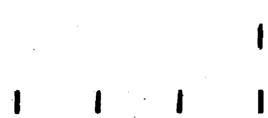
20)



It's organized on the model of a gallon of worms.

As a result, this sentence can have a large number of different metrical grids, including the following (where tick marks are used to represent the relative prominence of each word stress, and syllables with lesser stress are omitted).

21a)



```
b)                  |
      |       |     |
      |   |   |     |

c)                  |
          |   |     |
      |   |   |     |

d)                  |
      |   |         |
      |   |   |     |

e)                  |
      |             |
      |   |         |
      |   |   |     |

f)                  |
          |         |
          |   |     |
      |   |   |     |
```

The choice of grid for the sentence seems to depend on which words in the sentence the speaker wants to highlight. Alternating configurations seem to be preferred, but nonalternating configurations are also possible in particular contexts.

Given the grid representation of metrical strength, the assignment of accents to the text can be characterized in the following way:

22a) If a foot has a pitch accent, any foot of equal or stronger metrical strength in the phrase also has a pitch accent

<u>except that</u>

b) There are no pitch accents after the nuclear stress of the phrase.

Let us consider how these observations apply in some simple cases. Produced in isolation, a word containing two feet with primary stress on the right may have either one pitch accent on the nuclear stress, or a pitch accent on both feet:

23a) California
H* L^- L%

b) California
L* H* L^- L%

When a word with two feet and primary stress on the left is produced in isolation, it can have only one pitch accent

24a) Abernathy
H* L^- L%

b) *Abernathy
H* H* L^- L%

The b) form is impossible because it has a pitch accent after the nuclear stress. However, if a word with this stress pattern occurs in prenuclear position, it can carry two pitch accents:

25) It's perambulating Peter!
H* L* H* L^- L%

In Figures 5 and 6, the nuclear stress was moved through the phrase. In every case, the phrase accent controlled the FØ contour from the nuclear accent up to the boundary tone at the end of the phrase. Thus, when the nuclear accent was early in the phrase, the stressed syllables following it did not carry pitch accents of their own but instead took their FØ contours from the phrase accent.

The sentence shown in 20) can have a number of different patterns of accentuation, depending on which of the different options in constructing the grid is taken. For example, we can have:

26a) It's organized on the model of a gallon of worms
H* H* L^- L%

```
b)  It's organized on the model of a gallon of worms
                           |                    |
                           H*                   H* L- L%

c)  It's organized on the model of a gallon of worms
         |                 |                    |
         H*                H*                   H* L- L%

d)  It's organized on the model of a gallon of worms
         |                 |          |         |
         H*                H*         H*        H* L- L%
```

In all cases, any foot which has a pitch accent is more prominent than any other prenuclear foot which does not. Also, in cases like 26), where there is more than one accent in the phrase, the subordination among the accents is reflected in their phonetic values. A H* accent on a syllable with more stress corresponds to a higher FØ value than a H* on a syllable with less stress. For example, in Figure 2 A, the H* on "or" in "orange" in on the nuclear stress of the phrase, and so it is higher than the H* on "another." For L* accents, the situation is reversed; a L* accent on a syllable with stronger stress is lower than that on a less stressed syllable. The two L* accents in Figure 3 B reflect this regularity.

The amount of difference in phonetic value between one accent and another like accent which is metrically subordinated to it is continuously variable. In an intonation with a H* prenuclear accent and a H* nuclear accent, the nuclear accent could be anywhere from not significantly higher than the prenuclear accent to a great deal higher. What controls this variation is something like "amount of emphasis." As Figure 16 shows, intonation patterns with only one pitch accent can be produced with difference amounts of emphasis, with consequent variation in the height of the accent. It is not surprising that this kind of variation also plays a role where there are several accents.

In the subsequent chapters, the term "prominence" will be used to refer to the aggregate of metrical strength and emphasis, as it pertains to the control of tonal values. We will assume that each pitch accent has an associated prominence value, that prominence is continuously variable, and that the prominence of a metrically stronger accent is at least as great as that of a weaker accent, though not necessarily greater. We will not attempt to explain where prominence values come from, but will leave this task to pragmaticists and semanticists.

The different possible accentuation patterns for, "It's organized on the model of a gallon of worms," preclude an alternative rule for interpreting the metrical tree as a grid which is explored in Liberman and Prince (1977) and also rejected there. This is the convention which would generate the same complete ordering of the stressed syllables in the phrase as the stress rules in Chomsky and Halle (1968) do:

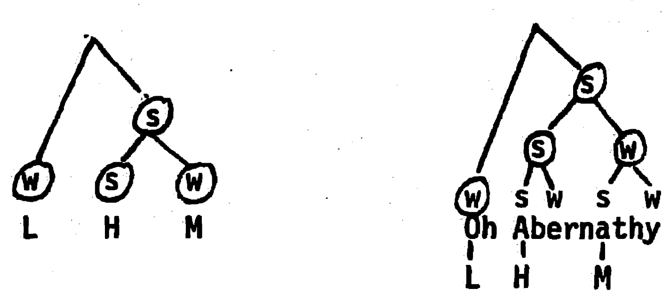
27) If a terminal node <u>t</u> is labelled <u>w</u>, its stress number is equal to the number of nodes that dominate it, plus one. If a terminal node <u>t</u> is labelled <u>s</u>, its stress number is equal to the number of nodes that dominate the lowest <u>w</u> dominating <u>t</u>, plus one.

The stress contour that this interpretation of the tree generates for sentence 20) is 2 3 4 1, or the same as the metrical grid 21e). In combination with 22), which implies that any prenuclear accented syllable is metrically stronger than any prenuclear unaccented syllable, this rule would incorrectly predict that "model" in 20) can carry a pitch accent only if "organized" does, and "gallon" can, only if both "model" and "organized" do. When all carried accents, the prediction is that the pitch range for the implementation of the first three accents would reflect metrical strength decreasing from the first to the third. In fact, given the preference for alternating patterns, such configurations are relatively rare.

It would be nice to find an account of text-tune assignment from which the two points in 22) fell out as corollaries instead of appearing as unrelated observations. Such a theory is attempted in Liberman (1975), but not entirely successfully. We will discuss this theory and its problems, and then lines along which improvements might be attempted.

In Liberman (1975), the sequence of pitch accents is given a metrical tree structure, like the tree structure for the text, and the accents are then assigned to the text by matching tree structures. Matching is carried out from the top of the tree down, with some complications which we will go into shortly. A tone in the tune shows up on the designated terminal element of the constituent it is matched to. 28) illustrates how congruence between text and tune is established for one example. The correspondence between nodes in one tree and those in the other is indicated by circling them.

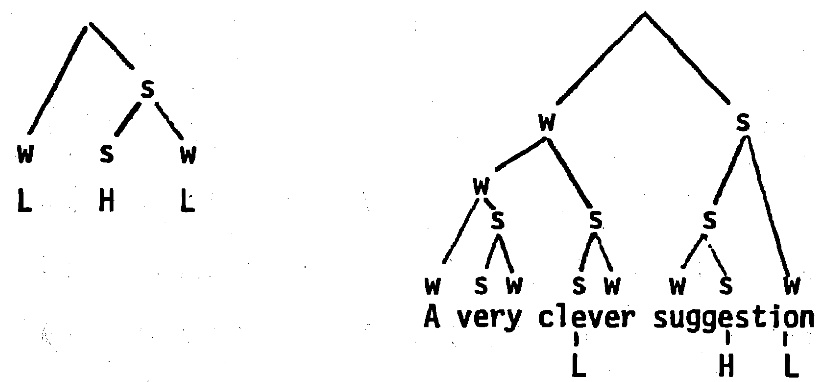
28)



(In the transcription for this contour in our theory, an H tone in a context for downstep replaces the M here.)

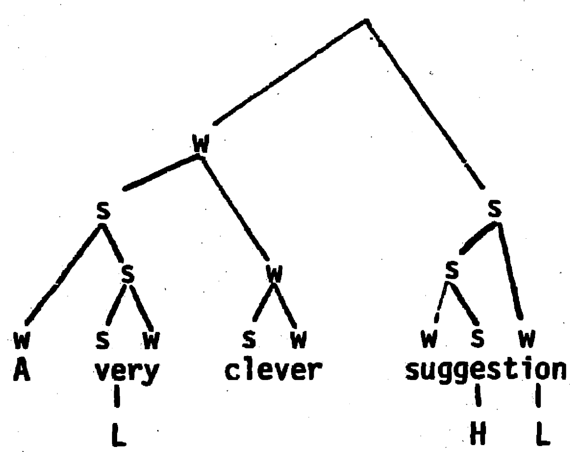
Clearly, Liberman's hypothesis has the consequence that a weaker node cannot be assigned a tone when a stronger node is not. In 29), for example, the prenuclear L tone ends up on the stressed syllable of "clever" because this is the designated terminal element of "a very clever."

29)



"Ver" in "very" would receive the prenuclear L only if the node labelling for "a very clever" has been reversed as in 30), so that it was the designated terminal element

30)



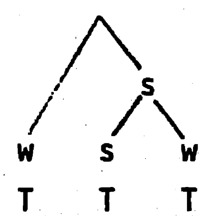
The theory also offers an explanation for the lack of postnuclear accents.

Setting aside the boundary tone, all phrasal tunes in Liberman (1975) are right branching trees in which what count here as the nuclear tone and the phrase accent are a constituent. If such a tree is labelled by the Word Rule from Liberman and Prince (1977), given here as 31), the results take the form in 32).

31) <u>Word Rule</u>

In a pair of sister nodes N_1 N_2, N_2 is <u>s</u> iff it branches (p. 268).

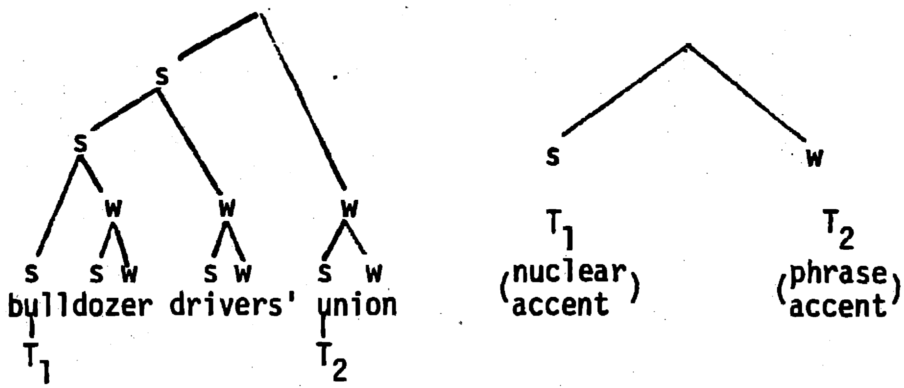
32)



Labelling by the Word Rule makes sense in a theory where tunes are lexical items, although it must count as idiosyncratic once we have concluded that tunes are syntactic objects. The results of this labelling is that there is just one tone, the phrase accent, appearing after the designated terminal element of the tonal tree. Given how congruence is established between the tone tree and the text tree, this means that only the phrase accent follows the nuclear stress.

The most important difficulty with Liberman's proposal is that it misaligns the phrase accent with the text. For example, the Appendix to the Figures gives FØ contours for all combinations of nuclear accent, phrase accent, and boundary tone on "bulldozer drivers' union." Given the following trees for the text and the tune, Liberman's alignment principle predicts that the phrase accent would land on the stressed syllable in "union":

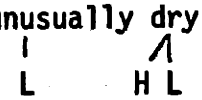
33)



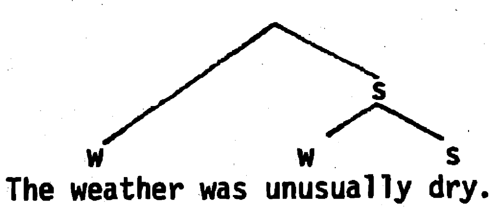
Instead, in every case where the phrase accent is on a different level from the nuclear accent and can be seen in the F0 contour, it is located near the end of the word "bulldozer." In Chapter 2, we will present further evidence that the phrase accent is not aligned with the metrically strongest syllable available. Instead, it is positioned with a fair amount of variation near the end of the word with the nuclear accent.

Liberman also discusses two complications which arise in his account. First, in some cases, congruence between the textual tree and the tune tree can only be established on the assumption that the textual tree has been readjusted from the original syntactic structure. For example, consider 34), as an outcome from the underlying structure shown in 35) and the tune in 36)

34) The weather was unusually dry.

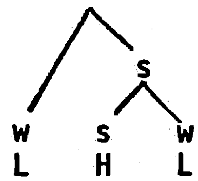


35)



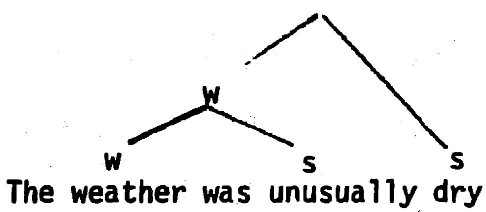
(Foot structure is omitted.)

36)



Clearly, the principle of matching tune to text from the top of the tree down does not generate the output in 34); without modification, the theory predicts that the L tone could only end up on "weather." This location is possible but it is not the only possibility. To generate 34) using the tree congruence principle, we need another tree structure for the sentence, 30). Such a tree structure would arise by some kind of extended cliticization.

37)

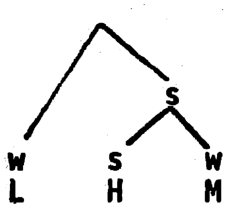


A second complication is that the congruence principles must in some cases skip over material in the textual tree. Liberman's example is 38):

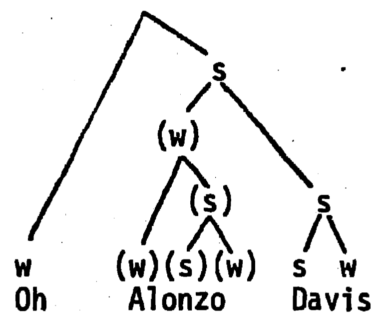
38) Oh, Alonzo Davis
L H M

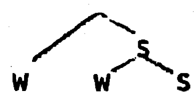
This is a well-formed alignment of 39) and 40) only on the assumption that the parenthesized nodes in 40) are disregarded in establishing congruence, because the topmost s branches to w s in the text and s w in the tune.

39)



40)



Note that both of these cases involving right branching text trees of the form 

. The interpretation of the tree as a metrical grid already takes the two w nodes to be freely ordered. This means that any form of text/tune assignment which was formulated on the grid and captured the generalizations in 22) above could generate both 34) and 38) without rebracketing or other additional complications. More generally speaking, complications seem to arise because the principles for establishing congruence between trees require that not only metrical strength but also bracketing be matched. Bracketing mismatches have to be circumvented by complications of the system. As Liberman notes, it would be surprising to find a tune which could not be aligned with texts of a particular structure. This means that the rules for establishing congruence between trees must have enough loopholes to circumvent all possible bracketing mismatches. In these circumstances, it is unclear how we would find evidence that bracketing was relevant, apart from its influence on relative metrical strength. It is our belief that tunes do not, in fact, have bracketing above the level of the pitch accent, and that the difficulties with Liberman's original proposal point towards text/tune alignment being controlled by metrical grids rather than metrical trees. We leave it as an open question whether alignment is computed from the textual grid alone, or whether the tune also has a grid which is matched to the textual grid. One could seek to explain post-nuclear deaccenting by marking the nuclear accent in tune grid, and formulating congruence principles under which the nuclear accent would align with the nuclear stress of the text. We know of no factors which would favor subordination among pitch accents in the tune in addition to this. On the contrary, positing additional subordination would seem to lead to difficulties with "mismatches" of the same general character as those noted for trees.

1.4 Tonal Implementation

In Section 1, we noted that two kinds of rules are involved in mapping the tune into a phonetic representation. One kind of rule evaluates tones phonetically, and a second constructs the F0 contour between one target value and the next.

Our hypothesis about how tones are evaluated is based on the results of an experiment reported in Chapter 3, which investigated how two intonation patterns were scaled as subjects varied their overall pitch range. The two intonation patterns studied are the responses in the following dialogues, and are illustrated in Figures 17 and 18, respectively.

41) -- What about Anna? Who did she come with?

-- Anna % came with Manny.
| |
H* L^- H% H* L^- L%

42) -- What about Manny? Who came with him?

-- Anna % came with Manny.
| |
H* L^- L%   H* L^- H%

Each of these responses has two intonation phrases, H* L^- L% and H* L^- H%. The order of the two phrasal tunes in 42) is reversed from the order in 41). The H* L^- L% is found on the phrase which is foregrounded by the context, and H* L^- H% is found on the information which is backgrounded. As a result, the phrases differ not only in tonal transcription but also in the prominence of the peak; the peak in H* L^- L% is higher.

The two basic patterns were produced by the subjects in a wide variety of pitch ranges, in response to a number indicating the degree of overall emphasis with which the pattern was to be produced. Measurements of the values of H* and L% in the H* L^- L% phrase, and of all three tonal values in the H% L^- H% phrase were taken for each contour. The relationships among these values under changes in pitch range was then investigated.

It was found that the lowest FØ values, corresponding to L%, remained fixed for each speaker as higher values varied. These values may be taken as defining the bottom of the speaker's range, or the lowest value he is disposed to produce at the given location in the utterance. The course of this bottom of the range over the utterance will be referred to as the baseline. This and other studies have found that the baseline falls slightly through the utterance. The major result of the experiment was that the declining baseline controls the scaling of FØ values higher in the pitch range. This is the case even in intonation patterns which do not have enough low values for the baseline to be seen clearly in the FØ contour; in short, values are scaled with reference to the baseline defined as the hypothetical bottom of the speaker's range, whether or not nearby values instantiate baseline. In particular, the finding for the contours shown in Figures 17 and 18 is that the height of the foregrounded and backgrounded peaks are in a constant ratio when their value is taken to be the following transform of the FØ value:

43) $$\hat{P} = \frac{P-B}{B}$$

Here, P is the FØ value of the peak, B is the FØ value of the baseline at the location of the peak, and $\hat{P}$ is the transformed value. Using this transform, it was possible to arrive at an estimate of the baseline for each speaker on the basis of the measurements at the peaks. This estimate was independently confirmed for each subject by measurements of L%.

The results for these particular contours suggests two hypotheses about how the phonetic values of tones might in general be determined. First, we will suppose that 43) is the relevant transform for all tones, or, in other words, that the unit for the phonetic value of tones is baseline units above the baseline. This supposition defines a graph paper for the phonetic value of tones, which is illustrated in Figure 19. Second, we will suppose that the target values corresponding to tones are related as ratios of baseline units above the baseline. Figure 20 shows how these assumptions apply to explain the peak relations in Figure 17. We will see in Chapter 4 that they can serve as the basis for a good description of English F0 patterns involving downstep. One such pattern is illustrated in Figure 21. The general character of this pattern, which is exponential and asymptotic to the baseline, is explained by taking each level to be a constant ratio in baseline units above the baseline of the level which precedes it. More specifically, the tonal transcription for Figure 21 is:

44) I really believe Ebenezer was a dealer in magnesium.

H* H^-+L* H^-+L* H^-+L* H^-+L* L^- L%

Each H following H+L is lowered relative to the preceding H by a factor k (with $k < 1$). This lowering readjusts the value H would have on the basis of its prominence relation to the preceding accent. Thus, like tones which are not downstepped, downstepped tones can have higher or lower values depending on their prominence. In H+L H, the L's are related to the H in the same accent by the same factor, k, which controls downstep. As a result, the last two tones are on the same level if the prominence is the same. We will also see downstepped contours of the form H L+H in which the downstepped H remains higher than the level of the H preceding it.

The exponential form of the F0 in Figure 21 can clearly be generated by evaluating tones iteratively left to right. We will assume that the series of phonetic values of the tonal sequence is initialized by the speaker's choice of value for the first H tone, for expressive purposes. Given that the value of a downstepped H is lowered by a factor of k relative to the preceding H, a chain of downstepped H's then results in a sequence of phonetic values of the form $k/H_1/$ $k^2/H_1/. . . k^n/H_1/$. (/T/ will be used to refer to the phonetic value of a tone, in baseline units above the baseline.) We will see in Chapter 4 that this general approach to tonal evaluation can be motivated for the other rules which evaluate tones. That is, once the sequence of tonal values for the phrase is initialized, the value for each new tone, T_{i+1}, is computed as a function of its prominence and of the phonetic and phonological values of tones to the left. The need to refer to the phonetic value of a tone to the left is obvious in the example of downstep just discussed. It is also necessary to refer to the phonological representation, because downstep is only brought into play when H_{i+1} is in the contexts H+L ____ or H L+____. The rule mentioned above which raises the value of the boundary tone after a H^- phrase accent also refers to both phonetic and phonological values; it applies only after H^-, and raises the value of the boundary tone by the phonetic value of the phrase accent. Similar observations can be made about the other rules for evaluating tones which will be presented in Chapter 4.

Although tone mapping rules can refer both to the input and the output of other tone mapping rules, the theory is saved from being global by restrictions on the use of such information. First, rules can refer to the phonetic values of tones already mapped, but not change them. This contrasts with ordinary phonological rules, which can effect changes anywhere in the domain they have access to. Secondly, the tone evaluation rules are local; the rules we will propose for English have no right context and refer no further back than the pitch accent preceding the tone being evaluated. Both the phonetic and the phonological values of tones outside of this window are inaccessible to the rule evaluating the current tone. Without further work, it is unclear what universal constraints can be placed on the size of the window for such rules. A different formulation of how far back the rules may refer will be needed for languages in which tones are not organized into pitch accents. Facts from a dialect of Zulu discussed in Cope (1970), Clements (1980), and below, suggest that reference to the phonological context to the right of the tone being mapped is allowable. However, we are unaware of any cases in which reference to phonetic context to the right is necessary, or in which the context for a tonal implementation is nonlocal.

In our model, the interpolation between one target and the next is carried out when the value of the target on the right becomes available. In constructing the F0 contour between two targets, the interpolation rules make reference to their value and to their location in time; it also appears that they can make reference to the underlying tones, since interpolation between L and H is handled differently than interpolation between two H's in cases where the first H is lower. We will have some observations about the general character of the interpolation in particular contexts. However, a good theory of interpolation probably waits on a better understanding of the motor control for intonation. In particular, we would expect the constraints of the motor system to be important when tones are compressed onto a small amount of material. One possibility which needs to be kept in mind is that the interpolation process may result in undershooting or overshooting of the phonetic value of tones when the tones are sufficiently crowded. A description of this character is suggested in Bruce (1977) for cases in which the pitch accent and the phrase accent are crowded together in Swedish. Here, we have concentrated on contours in which the tones are well separated and so we do not have comparable observations for English.

One consequence of our account of tonal implementation is that there is no level of systematic phonetic representation for intonation such as was suggested for segmental phonology in Chomsky and Halle (1968). This point can be made clear by considering the situation when the tone evaluation rules have gotten half way through implementing the tonal sequence for a phrase. To the right of the current window are the remaining unevaluated tones, still represented in the same form as in the underlying representation. To the left of the window is the F∅ contour computed thus far (or a motor representation of it). The tonal sequence underlying this contour is entirely unaccessible; specifically, the types, locations, and phonetic values of the tones are not accessed. Within the current window, evaluation and interpolation rules can access type, location, and phonetic value of tones. Clearly, in such a system, there is no well-defined level of representation in between the underlying representation as it is before any rules apply and the F∅ contour which is output. The strongest candidate for such a level, the sequence of target values computed by the tonal evaluation rules, is not a systematic phonetic representation for two reasons. First, it appears that the interpolation rules make reference to tonal type as well as tonal value;

thus, the sequence of target levels is not in itself sufficient to determine the FØ contour. Second, the units in a systematic phonetic representation are bundles of linguistic features. These features are not necessarily binary valued, but are presumed to be n-ary valued for some small n. The target values output by the tone evaluation rules do not have this character; they are values of a continuously valued physical parameter. Thus, the target values bear more similarity to the durations, formant values, and the like which are presumably computed from a systematic phonetic representation than to the systematic phonetic representation itself.

1.5 FØ Levels versus FØ Changes

One of our aims in developing a two tone theory of English intonation is to steer a true course between previous theories with four tone levels, such as Trager and Smith (1951), Pike (1945), and Liberman (1975), and theories which have sought to explicate intonation in terms of FØ changes rather than FØ levels, such as Bolinger (1951 and 1958), Ladd (1978) and Clark (1978). To our mind, a theory framed in terms of target levels is attractive because it affords good facilities for describing how the same intonation pattern lives up with different texts; the crucial points in the contour, the FØ targets, can be lined up with crucial points in the text, with stretches in between computed accordingly. The behavior of a given contour under changes in pitch range can be modeled in a similar fashion, by transforming the target points. Chapter

3 reports experimental data for which such a model was found to give a better quantitative fit than the best competing models framed in terms of F0 changes. The chapter also discusses a case where the two approaches make different claims about the status of the relationship between two F0 contours. The contours in question are the declarative terminal fall, shown in Figure 22 and represented here as H* L^- L%, and the vocative contour, shown in Figure 23 and represented here as H*+L^- H^- L%. The surface difference is that the terminal fall falls to the baseline, whereas the fall in the vocative stops well above the baseline. In a theory in which intonation patterns are described in terms of F0 changes, these two contours count as two instances of the same type, a smaller fall and a larger fall. In a theory framed in terms of target levels, the two contours differ in that the final target in the first is the baseline, and in the second, above the baseline. The theory does not rule out this difference being counted as a difference in type. In our view, these two F0 contours are as good as illustration of a difference in type as any, and the existence of this difference in type tends to support a target level theory.

In order to maintain that intonation patterns are decomposed into sequences of target levels, however, it is necessary to answer the objections of F0 change theorists to previous target level theories. A careful examination of these objections shows that difficulties noted arise from the postulation of four phonemically different tones, and that they can be circumvented by a system which has two tones and an appropriate phonetics. Bolinger (1951) noted that the four tone theory confounds differences in tonal type with differences in pitch range, with the result that surface forms as analyzed in the theory are chronically ambiguous. The problem arises through the interaction of two factors: first, tonal specification is relatively sparse in English, with the result that many or most English intonation phrases would not contain instances of all four tones in a four tone theory. Secondly, pitch range is used expressively in English. It is obvious that a phrasal pattern, such as the vocative just discussed, can be produced in many different pitch ranges; we also suggest that pitch range is varied for expressive reasons within the phrase, between one pitch accent and the next. The consequence for the four tone theory is that it would be impossible for the listener to decide whether the contour in Figure 2 C, for example, was an instance of L* H* L^- L% produced in a moderate pitch range, or L* LM* L^- L% produced in a larger pitch range. The difference between the two contours in Figure 24 could accordingly be either a tonal difference or a range difference. Thus, even if pitch range were used expressively only at the phrasal level, there would be large classes of putative intonational distinctions generated by the four tone system which could not in principle be distinguished. Given that pitch range is also used expressively at the pitch accent level, the amount of ambiguity under the four tone system becomes much more severe. Our two tone system does not suffer from this difficulty because L* H* L^- L% is the only analysis for a contour of the shape shown in Figure 24. Differences in the vertical scale of such a contour arise from expressive use of pitch range, and not from differences in tonal assignment. It should be noted that Bolinger's observations do not constitute an argument that four tone systems are in general impossible; obviously, four tone systems exist. Presumably, greater tonal density or restrictions in the expressive use of pitch range help to make tone values recoverable. In four tone systems where tone is lexical, contextual disambiguation of words may also help to establish reference points for tonal level.

A second objection to four tone theories of intonation is raised in Ladd (1978) and (1978a). He notes that there is a semantic correlate to the difference between intonation patterns which rise at the end of the phrase and ones which are level; the level ones, he claims, are "stylized," carrying the connotation that the utterance is ritualized or rhetorical, while the rising ones do not carry this connotation. In a four tone theory of intonation, it is impossible to capture the idea of "rise," since there is no feature decomposition for four tones under which L H, L HM, L LM, LM H, LM HM, and HM H form a natural class which excludes all level and falling tone pairs. This problem is circumvented in a two tone description. In Chapter 2, we will argue that the semantic difference Ladd describes can be viewed as the difference between contours ending in H% and those ending in L%, and that this view provides a more coherent account than Ladd's does of a number of intonation patterns in addition to those he describes. Other categories of rise which can be treated naturally in the framework here include rises arising from L H, and rises arising from H H with greater prominence on the second H. There appears to be no need for a mechanism which would group together all three of these types of rise.

The FØ contour shown in Figure 21 would be a possible basis for an argument in favor of a theory framed in terms of FØ changes, although we are not aware of such an argument being made in the literature. In such a theory, this intonation pattern can be described as alternating "level" and "fall." A similar example in Bolinger (1958) is described in a series of C pitch accents, where each C accent consists of a fall into a level portion. Within his framework, there is nothing to prevent the number of steps in such a sequence from being quite large, as indeed it is in Figure 21. This intonation pattern presents a problem for a theory of intonation with a small number of levels and a phonetics in which each level occupies a constant place in the total range employed. It is possible to create a pattern which steps downwards, but the number of levels in the system puts a limit on the possible number of steps. To describe the contour in Figure 21, it would be necessary to have six different levels. This number has never been proposed for English, and the existence of languages with even five levels is contested (Yip, 1980).

This problem is circumvented here by positing a more elaborate phonetics, which can change the location in the range which corresponds to a particular tone. Specifically, we propose a downstep rule which lowers the location of H after a H+L pitch accent. When this rule applies iteratively to a series of H+L accents, it creates the configuration shown in Figure 21. Such rules are well known from African tone languages, where an underlying analysis with level tones is well-motivated and generally agreed on. The analysis will also be extended to downstep the H in L+H in the configuration H L+H. Figure 25 shows an FØ contour in which the downstep of H in this context is apparent. The ability to generalize to this additional set of cases gives the target level theory an edge over FØ change theories: In an FØ change theory, Figure 25 would be analyzed as a series of rise/falls. There is no obvious reason that this contour should downstep like the level-fall contours, where downstep is generated automatically, rather than failing to downstep like other rising and falling contours.

1.6 Intonational Meaning

We would like to conclude this chapter with some remarks about intonational meaning. In the literature, one can distinguish two approaches towards the problem of establishing which intonation patterns are linguistically distinct and which count as variants of the same pattern. One approach attacks the problem by attempting to deduce a system of phonological representation for intonation from observed features of FØ contours. After constructing such a system, the next step is to compare the usage of FØ patterns which are phonologically distinct. The contrasting approach is to begin by identifying intonation patterns which seem to convey the same or different nuances. The second step is to construct a phonology which gives the same underlying representation to contours with the same meaning, and different representations to contours with different meanings.

The work presented here takes the first approach, in fact, it stops at the first step in the first approach. While we hope that the system of phonological representation proposed here will be useful in investigating intonational meaning, we do not offer such a theory here. In some cases, rough descriptions of a meaning or usage of a particular contour are suggested. We include these only to help the reader picture what type of intonation is under discussion; there is no representation that they are a complete description of the meanings of the contour in question, nor that they are expressed in the correct terms for a theory which could provide such a complete description.

The second approach meets with obstacles at two levels. First, similarity of meaning is not in general a good argument for similarity of form; if we learn that a creature may be called either a pangolin or an anteater, we do not conclude that the segmental transcription of these two words is the same. Forgetting this point is apt to lead the investigator to construct a theory which counts intonation patterns of extremely different form as instances of the same thing, and which as a result lacks resources for describing differences. On the other hand, keeping the point in mind means that judgments about meaning are no longer sufficient. Instead, it is necessary to ask the informant, "Do these two intonation patterns carry the same meaning and are they instances of the same form?" As we suggested in Section 1, such judgments are hard to come by because they overtax the native speaker's powers of introspection. This means that the inventory of possible difference representations has to be established by indirect means. These include comparison of their surface manifestations, and also experiments such as Bolinger (1951) and Nash and Mulac (1980), in which a judgment about the acceptability of some inference is used as evidence for an underlying linguistic distinction.

The second obstacle to beginning by making an inventory of intonational meanings is that they are extremely context dependent. The impression one gets from the intonation varies depending on the semantic content of the text, the presumed aims of the speaker, and whether or not there is a change of speaker. The result is something that is commented on in Liberman (1975): the meaning of a given intonation pattern can be startlingly exact in one context and startlingly different in another. For example, the pattern in Figure 4, which is transcribed here as H% L* H* L^- L%, is often used with an implication like, "I've told you this before -- how can you be so stupid as to need reminding" (Sag and Liberman, 1975). In this usage, the pattern seems disgruntled and overbearing. On the other hand, the implication that what the sentence says is obvious can also be addressed to oneself rather than the listener. The message is then something like, "Heavens! This should have been obvious, but I only now noticed." In this usage, the intonation pattern can seem polite and involved. The same pattern can be used on a wh-question as a way of making a denial, or on a greeting. Similar observations about the context dependence of intonational meaning are made in Gunter (1974), Hirst (1974), and Carlson (MS). Literature on the pragmatic character of intonational meaning is reviewed in Ladd (1978).

These observations do not mean that attempting to characterize intonational meaning is futile. They do mean, however, that intonational meanings cannot be observed directly in simple cases. Instead, they will have to be figured out by unravelling the contributions of the various parts of the linguistic and paralinguistic systems to the interpretation of large numbers of examples. The problem is a very difficult one, which is far from being solved. One obstacle to its solution has been the lack of a good theory of pragmatics, which would point to what variation in context need to be considered and to what effects on interpretation context might have. A second obstacle has been the lack of a phonological account of intonation, which would give an idea of what is to be counted as the same intonation on different texts. Thus, it is our hope that the phonological account given here will be helpful in developing a theory of intonational meaning.

It has been proposed in Lieberman (1967) that some intonation meanings are grammatical rather than pragmatic, and that it is just the grammatical intonation distinctions which are properly of interest to linguists. As the above discussion would suggest, we cannot agree with this position. As far as we have been able to determine, the meaning conveyed by choice of tune is always pragmatic, and we feel pragmatics is properly of interest to linguists. An example due to Mark Liberman illustrates the fact that even the strongest candidates for intonation patterns with grammatical meaning, the yes/no question patterns, are pragmatic markers rather than grammatical markers. Supposing that he shows up for an appointment but is not entirely sure that he is in the right office, he might say to the receptionist:

45) My name is Mark Liberman
$$\text{H* } \text{H}^- \text{ H\%}$$

This pattern is like the pattern in Figures 1 E and 6, except that the nuclear accent is on the same level as the phrase accent. It means something like, "My name is Mark Liberman, and is it familiar to you?" The intonation solicits a response, but the response sought is not information about whether the sentence is true or not, but rather of a much more general character. In contrast, when subject auxiliary inversion is used to mark a question, exactly the truth value is in question. As a result, the grammaticalized relative of 45) would seem very bizarre in the same circumstances:

46) Is my name Mark Liberman?

A similar example was overheard while leaving a movie theatre:

47) I thought it was good
                          |
                          H* H⁻ H%

This conveys something like, "I thought the movie was good, but I don't want to say anything too definite about it until I've heard what you have to say." In contrast, "Did I think it was good?" would be possible only as an echo question, since one knows better than other people what one thinks.

In view of such examples, one concludes that choice of tune can at most interact with other contextual factors to favor one grammatical interpretation over another. A similar position is taken in Bolinger (1957-1958).

Chapter 2

SOME BASIC INTONATIONAL PHENOMENA

2.1 Introduction

Among the observations introduced in Chapter 1 were the ideas that the text is divided into intonation phrases, whose tunes are the basic units of description for intonation, and that these tunes can be analyzed in terms of melodic correlates of stress and of phrasing. We noted that the melodic correlates of stress, or pitch accents, are found on at least one but not necessarily all stressed syllables in the phrase, and that they come in several different types.

Our choice of domain for the description of intonation is not controversial. Our intonation phrase corresponds to the domain of the intonational work in Liberman (1975), to the domain of the "tone unit" in Crystal (1969), to the "sense group" in Armstrong and Ward (1926) and Vanderslice and Ladefoged (1972), and to the "tone group" in Ashby (1978) and Halliday (1967). It appears to be the same as the "breath group" in Lieberman (1977). As Section 4 will show, a H* nuclear accent followed by a L^- phrase accent and L% boundary tone generates the FØ fall characteristic of the end of his "unmarked breathgroup," while a contour involving a H phrase accent and/or boundary tone has an FØ rise, like the end of his "marked breath group."[1] Our general observations about the melodic correlates of stress are also found in many previous works. Bolinger (1958) distinguishes the pitch accents from phrasal intonation, and describes English as having three different pitch accents. He views stress as the abstract potential for carrying accent, and notes that accentable syllables are under various circumstances deaccented. Ladd (1978) and Bing (1979) present modifications of his proposals. In the British tradition, Ashby (1978) gives a particularly clear picture of how some but not all stressed syllables are accented and of how accents can be of different types. Vanderslice and Ladefoged (1972) use the feature [± heavy] to refer to syllables which would here be called stressed. They describe [+ heavy] syllables as [± accent], and present proposals about the distribution of [+ accent]. They observe that different kinds of pitch accents exist, but do not make a serious effort to describe them phonologically. Liberman (1975) does not use the pitch accent as a phonological entity. However, his text/tune association procedure, which was discussed in Chapter 1, assigns tones to some but not necessarily all metrically strong syllables, and the tones assigned can be of different types.

In Chapter 1, we made a proposal about how to characterize the melodic correlates of stress and phrasing. We claimed that both the pitch accents and the correlates of phrasing can be characterized using two tones, L and H. The pitch accents consist either of a tone, or of a pair of tones with relative strength defined. The correlates of phrasing follow the last pitch accent (which falls on the nuclear stress of the phrase), and consist of a phrase accent, either L or H, and a boundary tone, also either L or H. Content sensitive rules implementing the tonal string play a crucial role in determining the shape of the FØ contour from such a tonal specification.

These proposals relate melodic correlates of stress and phrasing differently than other theories have done. For example, in Jackendoff (1972), Ladd (1978), and Bing (1979), the sequence of H* nuclear accent, L^- phrase accent, and H% boundary tone is taken to be a type of pitch accent. (In Jackendoff, this is the B accent, while in Ladd and Bing, it is the A-rise accent.) Under such an account, the phrase does not carry tones apart from the pitch accents; the special melodic features of the end of the phrase are handled by restricting some pitch accents to phrase final (that is, nuclear) position. In Section 4, we will suggest how our decomposition improves on such an account. On the other hand, some theories, such as Bolinger (1958) and Thorsen (1980), take the melodic correlate of phrasing to be the overall shape of the contour, on which the pitch accents ride. Our theory does not have two different layers of specification which add up in this fashion; the underlying representation of intonation is a well-ordered sequence of tones, and the tones which are associated with the phrase as a whole rather than with particular pitch accents are confined to the end of the phrase. This means that differences in overall shape of the contour arise indirectly, through the rules for implementing pitch accents. The reasons for this choice of description are given in Chapter 4, where the character of the tonal implementation rules is investigated in more detail. The two accounts which are closest to to ours in how they relate the correlates of stress and phrasing are Liberman (1975) and Bruce (1977). Both of these descriptions take the underlying representation of intonation to be a well-ordered sequence of tones. In Liberman, the boundary tone has a special status as a phrase-final marker, but the pitch accents and the phrase accent do not differ in status. Bruce's description of Swedish develops the idea of the phrase accent as a tone which is generated in addition to the pitch accents, which follows the nuclear accent, and which provides information about phrasal intonation. This idea makes possible a very clean account of the Swedish accent differences and their implementation in nuclear and prenuclear position.

This chapter covers the phenomena which are the natural constituency for a characterization of intonation as a string of L and H tones organized into pitch accents, phrase accent, and boundary tone. That is, it discusses intonational contrasts and qualitative features of FØ contours which reflect this characterization in a straightforward fashion. Section 2 covers the phonetic characteristics of the L* and H* accents, and discusses what differences may in general be found between L and H. The topic of Section 3 is the L^*+H^- accent, which is perhaps the most transparent case of a bitonal accent. Section 3 discusses the implementation of the phrase accent and boundary tone, and also the motivation for this decomposition of the phrase final contour. Section 5 suggests how the grammar of allowable tonal sequences given in Chapter 1 might be expanded to account for the intonation of tag expressions. The relation between the basic features of our theory of intonation and a number of results from the experimental literature is discussed in Section 6.

2.2 L*, H*, and the Difference between L and H

Two of the pitch accents introduced in Chapter 1 were the single tone accents, L* and H*. Figures 1 and 2 show how these two accents contrast before H*; Figures 3 and 4 show how they contrast before L*. Figures 5 through 8 show H* and L* in nuclear position, before the H^- and L^- phrase accents.

We can identify three surface differences between H* and L*. First, there is a paradigmatic distinction in level: in all of the figures just mentioned, L* is lower than H* in the same context. Second, as we mentioned in Chapter 1, the phonetic value of L* decreases if its prominence is increased. In the same circumstances, the phonetic value of H* increases. This point is illustrated in Figure 9, where a H% L* H* L^- L% contour is shown as produced with two different degrees of overall emphasis. In the contour with greater emphasis, L* is lower and H* is higher.[2] There are two tricky points about this second observation. First, the lowering of L*s and the raising of H*s under increased prominence are not completely symmetric, because the L and H stand in different relations to the extremes of the speaker's range. As mentioned in Chapter 1, the baseline, or the hypothetical bottom of the speaker's range, plays an important part in the intonational system. It is involved in computing how prominence is reflected in the phonetic values of H* accents, and points on the baseline often occur in the FØ contour. Now, a L* pitch accent is lower than H* in the same context. Often the distance from H* to the baseline is not great: this limits the extent to which L* can be lowered to reflect prominence, given that the L* cannot be realized any lower than the baseline. The lowering of L's under prominence is thus prone to saturation. In emphatic speech, it is common to find a series of L*s which are on essentially the same level, instead of reflecting stress subordination in relative FØ levels. The situation with the H*s is quite different. There is surely some hypothetical ceiling to the speaker's range, but it does not appear to be important in practice, either as a value that speakers use in producing intonation, or as a hypothetical entity with a role in the intonation system. In the experiments reported in Chapter 3, subjects produced some FØ contours with FØ ranges as much as an octave larger than their normal speaking ranges. Even in these cases, there was no tendency for FØ to saturate as a way of marking prominence; the prominence relations were reflected in the FØ contour in the same way as in contours with more moderate FØ ranges.

The second tricky point is this. In order to give a precise answer to the questions, "How high a H? How low a L?", we need to know what counts as "zero high" or "zero low." Chapter 3 will argue that "zero high" is the baseline; "How high" means "How high above the baseline." The baseline cannot correspond to "zero low," since the baseline is the most, not the least, low. It is at present not clear what counts as "zero low," and depending on what does, the lowering of L under prominence could work out in very different ways. Further experimental work on the scaling of L*s in a variety of contexts is needed to answer this question.

The third difference between L* and H* is that they are treated differently by the interpolation rules. Between two H*s, one finds the dipping shown in Figures 10 and 11. The interpolation between L* and any other tone (either L or H) is monotonic. For example, in Figure 2, the F0 takes a direct course between L* and H*; in Figure 12, the F0 contour is flat between the two L*s, instead of displaying the hump which would count as a counterpart to the dipping between H*s.

This complication in the interpolation rules is in some ways unattractive, and we have made a serious attempt to get rid of it by developing an account under which the dip in contours like Figures 10 and 11 arises from a L tone. It does not appear possible to do this without considerably changing the form of the theory. The major obstacle is that the L tone in contours like 10 and 11 would have to be the unstarred tone of a bitonal accent, since we have no other source of tones that can fall on metrically weak syllables phrase internally. Specifically, it would have to be either H^*+L^- or L^-+H^* in order to explain the lower F0 values in between the peaks. However, both of these accents are already used in accounting for the downstepped contours described in Chapter 4. If one of these accents is taken to be the source of the dipping intonation, then a sequence of H* accents must be taken to be the source for one type of downstepped contour.

This change would considerably complicate the formal statement of the downstep rule. It would also make the rule an exceptional one cross-linguistically, since downstep is ordinarily found in sequence with alternating tonal types.

There is also one phonetic regularity which falls out naturally from the claim that interpolation between H*s is nonmonotonic. A phonetic characterization of dipping says essentially that the FØ falls until it is time to start aiming for the next H* level; this characterization predicts that the amount of dipping would be less for H*s which were closer in time, and could disappear for H*s which were sufficiently close together. Examination of O'Shaughnessy's corpus (O'Shaughnessy, 1976) supports this prediction. The corpus includes 89 FØ configurations of one or the other of the types shown in Figure 13. Of the 51 which have a dip between the two peaks, as in Figure 13 A, 48 have one or more unstressed syllables between the two H*s. Of the 38 which have no dip, as in Figure 13 B, 36 have no unstressed syllables between the two targets.[3] Pierrehumbert (1979a and 1980) describes a successful computer program for synthesizing neutral declarative intonation which makes use of this principle. The program computes a local minimum between two H* accents as a function of their separation in time and frequency.

The nonmonotonic interpolation rule for H*s means that they typically show up as peaks in the FØ contour. Specifically, in our data, the FØ target corresponding to the H tone is ordinarily located near the end of the accented syllable, and so the F0 contour on the stressed syllable is a rise with a local maximum at the end. (In the case where H* shares a phrase final syllable with a L^- phrase accent, the accented syllable is lengthened, the peak is earlier in the syllable, and the F0 contour during the latter part of the syllable is falling.) L*s next to H*s show up as local minima in the contour; for example, in Figure 2, the L* on "remarkably" is readily located as the point where the rise to the following H* starts. However, L*s next to other L tones do not show up as inflections in F0, because the monotonic interpolation for L's means that the local contour is as close to flat as the relative level of the L's permits. Figure 14, for example, shows a contour taken from tapes of spontaneous speech made during an experiment on personal interaction,[4] which we would analyze as L* L^- L%. Figure 15 shows an example of the "contradiction contour" discussed in Liberman and Sag (1974), in which the L* nuclear accent is between a L^- phrase accent and a L* prenuclear accent. Cases like this are a major source of ambiguity in the intonational system; in cases where the L tones are not on different levels, the location of accents is not readily recovered from the F0 contour. In Figure 15, the prenuclear L* is as low as the nuclear L* and so the contour is the same as if "advantage" carried the nuclear accent and the L^- phrase accent was responsible for the low F0 level from there to the end of the phrase. At best, the listener may infer location of the nuclear accent from information about phrasal stress subordination carried by amplitude, duration, and segmental characteristics.

As we pointed out in Chapter 1, not all metrical feet receive a pitch accent. The FØ contour on unaccented feet is determined by interpolation between the flanking tonal specifications, in the same way that the FØ contour on unstressed syllables within the foot is determined. Thus, the interpolation rules tell us what types of FØ contours we can expect to find on unaccented feet. In Figure 2, for example, "clever" is unaccented. Its FØ contour is determined by the interpolation between L* on the left and H* on the right, and is thus rising throughout. Figure 12 illustrates, for comparison, the case in which "clever" has a L* accent of its own. A similar comparison is shown in Figures 16 and 17. In 16, "good" is deaccented between L* and H*, while in 17, it has a H* accent. If an unaccented foot is found between two H*s, we expect its FØ contour to be generated by the same kind of dipping which was found in Figures 10 and 11. Figures 18 and 19 show two such examples. In 18, "book" is unaccented and it carries the falling part of the interpolation between H* on "which" and H* on "mean." In 19, taken from O'Shaughnessy (1976), the verb is unaccented. As O'Shaughnessy notes, deaccenting of verbs in running speech is very common, even when they are not really predictable from the context.

We have noted three differences between the surface reflexes of L* and H*: there are differences in level, in behavior under changes in prominence, and in interpolation behavior. How will these observations hold up when we consider the L and H tones in general? The difference in level will hold up exactly as stated: in every context, L is lower than H would be in the same context. (Here, context is taken to include prominence relations.) This is an interesting result, because it means that tones are not neutralized in English. As we will see in Chapter 4, there is nothing about our account of tone mapping rules which prevents tones from being neutralized phonetically, and it appears that they can be in languages with total downstep (that is, languages that lower H to the level of a preceding L). It is important to note that the difference in level does not hold up across contexts. The result of downstep and upstep is that a H can be lower than a L earlier in the phrase, and a L can be higher than H earlier in the phrase.

Our observation about interpolation between a L* accent and another tone also generalizes. All L tones contrast with H* in requiring monotonic interpolation to an adjacent tone. For example, Figure 20 shows a case of H^{-}+L* H* in which H* has been downstepped to the level of the preceding L. The interpolation between the two starred tones is straight, even though they are separated by two unstressed syllables. This is a particularly clear illustration of this regularity, because the L and H are both starred tones. Since only unstarred tones undergo spreading, the alternative hypothesis that the plateau arose from spreading can be eliminated. Whether dipping is found between any two H tones is another question. Unstarred H tones are subject to rightward tone spreading when the next tone is phonetically equal or higher. Dipping would come into question only when the context for spreading is not met. This rather special set of cases would include H* H$^-$+L* and also L*+H$^-$ H$^-$+L* with H$^-$+L* under less prominence than L*+H$^-$. We are not sure whether dipping is found in these cases or not.

The difference between L* and H* under changes in prominence does not generalize to L and H. H tones all go up under increases in prominence, but not all L's go down; in some cases, H's drag L tones upward with them. Specifically, a L tone in a bitonal accent goes up as a ratio of the value of the H. This point is illustrated in Figure 21. We will see in Chapter 3 that the L$^-$ phrase accent after H* also goes up as increasing prominence raises the H*.

2.3 The L*+H$^-$ Pitch Accent

The clearest example of a pitch accent involving two tones is the L*+H$^-$ accent. The contrast between the L*+H$^-$ and the H* accents is illustrated in Figures 22 and 23. In Figure 22 A, the nuclear accent is H*, and falls on "le," which has the primary stress in "legumes." In Figure 22 B, the nuclear accent is L*+H$^-$: the primary stressed syllable has a very low FØ, indicating the presence of a L tone, and the FØ peak, corresponding to the H tone, does not occur until the following syllable. Figure 23 shows that the same contrast can be implemented on a monosyllable. In Figure 23 A, there is a H* pitch accent, L$^-$ phrase accent, and H% boundary tone on the syllable "Anne"; in 23 B, the pitch accent is L*+H$^-$ rather than H*, and the phrase accent and the boundary tone are the same as in A). One appropriate use of the intonation patterns in 22 B and 23 B would be to indicate incredulousness. The L^*+H^- pitch accent can also be used less emphatically to indicate that the speaker views his reply as incomplete. An example due to Lauri Carlson makes this point nicely (Carlson, MS). In answer to the question, "Who wants tea?", a polite person might say, "I do," using a L^*+H^- pitch accent. This answer carries the implication that perhaps someone else present might also want tea. It is also possible to answer the question using a H* pitch accent on "I". However, in this case, the implication is something like, "The answer to your question is, I want tea"; there is a note of rude disregard for the possible wishes of other people present. Other authors who have noted the contrast illustrated in Figures 22 and 23 include Sledd (1956), Bolinger (1958), Vanderslice and Ladefoged (1972), and Ladd (1978). Crystal's (1969) "rise fall" and "rise fall rise" involve a nuclear L^*+H^-; his "spiky head" appears to involve prenuclear L^*+H^- accents.

The L^*+H^- accent provides the first example in which the unstarred tone of a pitch accent is subject to tone spreading. The tone spreading rule, which is developed in Chapter 5, spreads T^- to the right when the next tone is phonetically equal or higher. An FØ contour in which the rule has applied to L^*+H^- H* is shown in Figure 24. This contour contrasts with H* H*, since the first accented syllable has a low FØ value. It also contrasts with L* H*, because the FØ jumps up quickly and then makes a plateau instead of rising gradually.

We noted in Chapter 1 that a strength relationship is defined on the two tones of bitonal accents, and that it is the stronger tone which lines up with the accented syllable. In Figures 22 and 25, the L* in L*+H$^-$ is associated with the stressed syllables "le" and "rig." The lccation of the H$^-$ is only derivately governed by stress, in that it follows immediately after the L*. In order to clarify how the unstarred tone in a bitonal accent is aligned with the text, a corpus of utterances with L*+H$^-$ accents was collected. The stress pattern on the material immediately following the accented syllable was varied systematically. Two speakers were used, and a total of 108 FØ contours resulted in which it was clear that a L*+H$^-$ rather than H* had been produced.

The hypothesis suggested by this corpus of FØ contours was that the H$^-$ is located at a given time interval after the L*, regardless of the stress pattern on the material following the accented syllable. For speaker KXG, the mean time was 19.1 centiseconds ($\sigma = 1.8$). For speaker MB, it was 20.2 centiseconds ($\sigma = 3.9$). One might suppose that these intervals arise as the amount of time needed to execute the change in FØ level specified by the pitch accent. If this is so, it would be likely that the interval would increase with overall pitch range. Our corpus did not have enough variation in range to test this prediction; the possibility of variation with speech rate was also left for future investigation.

The picture predicted by this hypothesis comes out clearly in a tabulation of where the H$^-$ fell in examples with different types of metrical structure.

The words in the corpus which had the longest syllable following the accented syllable were the compounds "newsreel," "windmill," "hoosegow," and "headwind." In the utterances where the $L*+H^-$ accent fell on these words, the H was in the first half of the second syllable, in some cases falling before the beginning of the vowel. The H fell near or at the end of the post-accented syllable in the words where this syllable was of moderate length: "Lieberman," "mothersinlaw," "Brobdingnag," "Kelloggs," "motherwort," and so on. In words where the post-accented syllable was an extremely reduced CV syllable, H fell on the next syllable after that: "cardamon," "Alamo," "hedebo," "catamount," "rigamarole," and so on.[5] A number of apparent exceptions to this picture turned out not to be exceptions when sufficiently detailed phonetic analysis was performed. For example, in three out of four cases, the H^- fell on the second syllable in "Manidae." Examination of the waveform showed that the speaker had not flapped the /n/ in these cases, perhaps because the word was unfamiliar. As a result, the post-accented syllable was significantly longer than in words like "Alamo" and "hebedo," where the dental was flapped. Similarly, in one of the four utterances involving "Brobdingnag," the H^- occurred on "nag." In this case, the speaker had reduced the post-accented syllable to a dental plus a syllabic nasal, whereas in the other cases, this syllable was pronounced with an unreduced vowel.

In our corpus, almost all of the cases in which H^- was found on the second rather than the first syllable over from the accented syllable are cases in which one of two metrical rules discussed in Kiparsky (1977) would be applicable. These are the Victory Rule, which deletes an unstressed vowel medially before a sonorant followed by an unstressed vowel in words like "opera" and "victory," and the Resolution Rule, which permits a VCV sequence to count as a single metrical position (either strong or weak) in the poetry of Chaucer and Shakespeare. The Resolution Rule is stated in 1) and exemplified in 2).

1) V ↕ M   CV ↕ ∅   (where M = metrical S or W)

2a) And spends / his <u>prodi</u>/gal wits / in boot/less rhyme

b) Come to / one mark, / as <u>many</u> / ways meet / in one town (p. 236)

Under our account of how H^- is aligned with the text, this correspondence is not surprising. The syllable following the accented syllable is skipped just when it is extremely short. It is exactly extremely short syllables that are on the road to deletion by processes like the Victory Rule. A rule like Resolution would be apt to incorporate only very short syllables in with the preceding one, and its application might trigger shortening. However, these observations also suggest that H^- alignment might be handled metrically, by an alignment rule ordered after Resolution and the Victory Rule. The metrical alignment rule would associate the H tone phonologically with the next metrical element after the metrical element corresponding to the accented syllable. This would ordinarily be the next syllable after the accented syllable, but when this syllable had been subjected to the Resolution Rule or the Victory Rule, it would be the syllable after that.

Three considerations lead us to conclude that our original proposal accounts for the facts better than the alternative metrical proposal. First, the timing of H^- with respect to the syllable it is on differs from that of H*, which we know to be associated by a phonological rule. Except when H* is crowded by other tones, the peak corresponding to it is found at the end of the syllable the tone is attached to. H^-, as we noted, occurs earlier or later in the syllable according to the length of the syllable it is on. This difference is hard to explain if H^-, like H*, is phonologically associated with a syllable; we would expect the phonetic rules for timing to treat $\begin{matrix}\sigma\\|\\H\end{matrix}$ uniformly regardless of how it arose. The variable timing of H^- in L*+H^- is, however, easily explained in our account, in which H^- is separated from L* by a fixed time interval, without regard to the segmental or syllabic character of the material following the accented syllable. Second, a phonetic account of the timing of L* and H^- is needed to account for the FØ contour in Figure 23 A, where both fall on a single syllable. Once we have such an account, additional metrical rules to account for the behavior of L*+H^- on polysyllabic material would appear to be superfluous. Third, the metrical account relies on Resolution to account for the timing of H^- in a large number of examples where the Victory Rule could not apply (Amityville, hedebo, monograph, etc.) Whereas the Victory Rule is a productive rule of fast speech in English, Resolution was a poetic rule which was lost after Shakespeare. Without further motivation, we are reluctant to regard it as a productive rule of modern English as the metrical account of unstarred tone alignment would require.

Our account of unstarred tone alignment generalizes readily to account for where H* contrasts with H^*+H^-, an accent which is discussed in Chapter 5. We would take the lag of H^- after L* in L^*+H^- to arise from the need to execute an FØ change. In H^*+H^-, there is no FØ change and hence we suppose there to be no lag. This means that H^*+H^- contrasts with H* only if tone spreading applies to H^- to generate an FØ plateau where the FØ contour after H* would dip. This is in cases where the next tone is equal or higher phonetically. The natural generalization of the metrical theory of alignment, however, would be to assign any T^- in T^*+T^- to the metrical element following the accented syllable. This would incorrectly predict a contrast between H^*+H^- and H* even when the next tone is lower; for H*, the fall to the lower tone would start at the end of the accented syllable, whereas for H^*+H^-, the onset of the fall would be delayed. Avoiding this consequence would require a complication of the tone alignment rules which does not appear to have a natural basis.

2.4 Tonal Correlates of Phrasing

Chapter 1 presented the idea that the underlying description for a melody is a sequence of tones, comprised of tones contributed by pitch accents and tones marking the borders of the intonation phrase. Of the latter, there were two types: the boundary tone (T%) and the phrase accent (T^-). This section deals with how these concepts serve to characterize the FØ configurations found at phrase borders. We will first present additional examples and how we propose to describe them, and then justify the description on the basis of the natural classes it engenders.

A boundary tone, as Liberman (1975) observed, occurs right at the phrase boundary, regardless of the stress pattern of adjacent material. This point is illustrated in Figure 26. Figure 26 A shows an FØ contour ending with a continuation rise, or H% boundary tone. Here, the last syllable in the phrase, which carries the peak corresponding to this tone, is stressed. In 26 B, the stressed syllable is the fourth syllable back from the end, but peak is still at the end. In Figure 26 C, the FØ contour has an utterance internal H%, which is located at the boundary between the two phrases.

A boundary tone can occur not only at the end of a phrase, as in Figure 26, but also in utterance initial position, where it determines how the FØ contour begins. In particular, a H% initial boundary tone adds a note of vivacity to an intonation pattern; as Liberman (1975) notes, the H% boundary tone in a tone level theory corresponds to the "high prehead" in O'Connor and Arnold (1961). The FØ contour in Figure 27 illustrates how the H% initial boundary tone stays at the phrase boundary even when the first syllables in the phrase are unstressed.

It is not hard to find F∅ contours which are plausibly analyzed as having a L% initial boundary tone. One such contour is shown in Figure 20. However, there are many onset levels in between the two extremes, which might in principle count as either high values for L, or as low values for H. We have not found a basis for transcribing a leading boundary tone in such cases. A possibility to keep in mind is that they do not have a leading boundary tone; the F∅ onset may be some kind of neutral value. An observation which favors this possibility is made in Liberman (1975). As he notes, it is difficult to have a H% initial boundary tone when the first syllable in the phrase is an accented syllable with a L* accent. He attributes this difficulty to ill-formedness resulting from the boundary tone falling on the metrical beat. If this is so, then the occurrence of any boundary tone when the first syllable was accented would be impossible. Utterances beginning with accented syllables exist, of course, and the F∅ onset is determined somehow. Thus, Liberman's observation implies that F∅ contours can be begun without a phrase initial boundary tone.

The occurrence of phrase initial as well as phrase final boundary tones raises the question of whether it is possible to have both at an intonation phrase boundary utterance internally. It is our impression that such contours are possible, provided the phrase boundary is also marked with a pause. An example of this type is provided in Figure 28. The first phrase ends with H% and the second begins with another, higher, H%.

The phrase accent occurs after the nuclear pitch accent, and before the phrase final boundary tone. There are thus two extra tones at the end of the phrase. In Figure 23 B, it was clear that two post-nuclear tones were needed in order to generate a fall-rise following the nuclear L*+H$^-$. Other examples where it is clear that two post-nuclear tones are needed are shown in Figure 29 and 30. In Figure 29, the nuclear accent is a L*, and it occurs early in the phrase because of the focus on "Manitowoc." There is a H% boundary tone, which is responsible for the F0 maximum at the end of the phrase; however, the contour does not rise smoothly from the L* to this H%, but rather rises, forms a plateau, and then rises again. Thus, an additional H tone is needed to define the corner in the contour. This is the H$^-$ phrase accent; it is lower than the H% because T% is subject to upstep after H$^-$. This F0 contour has a typical form for a yes/no question; in particular, Sag and Liberman (1975) and Rando (1980) also note that the rise-plateau-rise configuration is found when the distance from the nuclear accent to the end of the phrase is great enough for it to be observed.

A contour in which both a L$^-$ phrase accent and a L% boundary tone can be seen is shown in Figure 30. The F0 falls quickly and then levels out. Thus, the tonal specification for this contour must be H* L$^-$ L%, with the L$^-$ phrase accent placed at the end of the word "Monarch." We suggest that the description of intonation is considerably simplified by the assumption that the tonal specification for every intonation phrase ends in a phrase accent and a boundary tone. Under this assumption, the underlying representation for the end of the contour in Figure 27, for example, is also H* L^- L%. However, here the L^- does not shop up as a corner in the contour; instead, it appears that in such close quarters, a coalescence of the two like tones takes place. It is unclear whether this should be handled by deleting one of the two tones phonologically, or by a phonetic implementation which leaves the two barely separated in time. We lean towards the second solution.

The boundary tones align with the text in a particularly straightforward fashion. It is much less clear what principles control the alignment of the phrase accent with the text. In order to investigate this question, a corpus of 350 FØ contour with focus early in the utterance was collected.

For each sentence in the corpus, the subject (MB) produced three nuclear intonation patterns: H* L^- H%, L*+H^- L^- H%, and L* H^- H%. The sentences were designed to vary the stress pattern immediately following the nuclear accent, and the length of the word carrying the nuclear accent, on the hypothesis that the phrase accent might show some affinity for the next stressed syllable or for the word boundary.

This set of FØ contours was instructive, but still left many questions unanswered. The major observation was that the distance of the phrase accent from the syllable with the nuclear stress varied considerably for all three intonation patterns, in a way which appeared to be systematic. This point is illustrated in Figure 31, where the L* H^- rise is shown on three words with a secondary stress following the primary stress, and in Figure 32, where the H* L^- fall is shown for three words of the same character. Such data made it possible to rule out the claim in Ashby (1978) that the execution of these two contours displays a fixed time course. Independent disconfirmation of this claim had been found in working on the intonation synthesis program mentioned above, where a fall suitable for phrase final monosyllabic words with short intrinsic duration, such as "bit," had been found to be too quick for polysyllabic words like "rival" (Pierrehumbert, 1980). It is also possible to rule out the hypothesis that the phrase accent occurs on the next stressed syllable after the syllable with the nuclear stress. The type of contour on which this conclusion was based is shown in Figure 33. Here, the next stressed syllable is "bel" in "gorbelly," but the H^- does not occur until the following syllable, which is unstressed. (All eight utterances where "gorbelly" carried the nuclear accent of a question exhibited this pattern exactly.)

The strongest hypothesis was that the phrase accent is placed at the boundary of the word carrying the nuclear accent, regardless of stress pattern. We present this hypothesis with a few caveats, however. In FØ contours like 22 B, in which a L^*+H^- was assigned to a bisyllabic word, the L^- phrase accent did not fall right at the end of this word, but rather further to the right. It is in fact hard to imagine that the fall from H^- to L^- in Figure 22 B could have been produced any faster than it was. Secondly, due to the curvature of the interpolation between H^- and L^-, and to the interference of segmental effects, it was very difficult to decide where the L^- was located. There were many cases where a person with a different theory could have located the L^- a syllable to the left or to the right of the word boundary. The interpretation that would first jump to mind varied for repetitions of the same sentence. This was less of a problem with the L* H^- sequence, but even here there would be a good number of examples in which the location of the H^- could be legitimately disputed. Third, this conclusion applies only in the case where the word with the nuclear stress is not phrase final. Figure 23 already showed that when the word with nuclear stress is phrase final and the boundary tone is higher than the phrase accent, the phrase accent is pushed back into the word. It is only by compressing the intonation pattern onto the material available that the speaker preserves its contrast with other patterns. The phrase accent also occurred before the end of a phrase final accented word in many cases where the word was long, even when the boundary tone was of the same tonal type as the phrase accent. One such case is shown in Figure 34. Lastly, it has been observed that the phrase accent in chanted calling contours falls most naturally on a metrically strong syllable. For instance, what is transcribed in Liberman (1975) as 3), and here as 4), has the phrase accent on the designated terminal element of the second foot.

```
3) Abernathy
   |/   |/
   H    M
```

```
4) Abernathy
    |    |
 H*+L-   H- L%
```

In Liberman (1975), this is treated as the stylized case which reveals the general pattern. Our impression is that this behavior is only characteristic of chanted speech. In our corpus of normal speech, the phrase accent did not fall on the metrically strongest syllable after the nucleus, as Liberman's account would specifically predict; more generally, we could not identify any tendency for it to prefer a metrically strong syllable to a weak one.

The positive face of our observations is that there was a contrast between the behavior of the phrase accent and the pitch accents in this corpus. The starred tone of the pitch accent was always found on the accented syllable; the phrase accent showed no special affinity for stressed syllables, let alone for more prominent stressed syllables over less prominent ones. For this reason, it was clear to us from this experiment that the phrase accent is unstarred, in the same sense that the unstarred tone in a bitonal pitch accent is. Thus, it is natural that it shares the ability of the floating tone in a pitch accent to spread.

There is, however, an interesting contrast between the phrase accent and the unstarred tone in a pitch accent. As we have seen, H^- in L*+H^- occurs about 20 cs after the L*; in Figure 27, we see that the H^- phrase accent after L* can be considerably more delayed. This means that L* H^- and L*+H^- H^- can be distinguished when there is enough syllabic material to carry the distinction. The L*+H^- H^- has a very sharp rise in pitch after the accented syllable, whereas the L* H^- has a rise which spans the whole word carrying the accented syllable. This difference is illustrated in Figure 35. Figure 35 A shows the normal question, involving a $L^* H^-$ contour; 35 B shows the $L^*+H^- H^-$ contour, where the use of the L^*+H^- accent leaves the listener with the impression that the question was incredulous. Recall that this was also one of the usages of the L^*+H^- accent before a H^* or L^-.

The implementation of the phrase accent and the boundary tone are affected by two rules. First, the phrase accent, like the H^- in a L^*+H^-, is subject to spreading. In Figure 26, the L^- in the sequence $H^* L^- H\%$ spreads from its location at the end of the word with the nuclear stress to the end of the phrase. The result is that the rise to the H% occurs only at the very end of the phrase, rather than being spread over the post-nuclear material. As Figures 30 A and 31 show, the H^- in the sequence $L^* H^- H\%$ behaves in the same way: its value is maintained up until the end of the phrase, where the rise to the H% occurs.

The context for spreading of T^- is when the next tone is phonetically equal or higher. The motivation for this formulation of the rule depends on results in Chapters 3 and 4, and is thus put off until Chapter 5. The consequence for the implementation of the phrase accent is that H^- spreads before either H%, which is higher, or L%, which is equal after upstep. L^- spreads before H%. However, L^- does not spread before L%; spreading is blocked because L% is on the base-line and L^- is somewhat higher. This means that $L^- L\%$ is implemented as a gradual FØ fall rather than having an FØ plateau like the other cases.

The second rule affecting the implementation of the phrase accent and the boundary tone is the rule responsible for making the H% in Figures 30 A and 31 higher than the preceding H^-. The motivation for this rule depends on two additional observations. First, as noted by Sag and Liberman (1975) and also Rando (1980), a question need not have a final rise; the level of the phrase accent can instead by continued up to the end. The contrast between these two ways of ending a question is shown in Figure 36; 36 A, which has the final rise, is the normal form for a question soliciting information, while the contour in 36 B, lacking the rise, is often used for rhetorical questions. (It is also used in list intonations.) The second observation is that the contours shown in 36 C and 36 D are impossible, on any analysis with a nuclear L* and H^- phrase accent as marked. That is, there is no contour in which a H^- phrase accent is followed by a boundary tone which causes the FØ to fall back to the baseline.[6] This distribution of possibilities suggests that Figure 36 B represents the case where the H^- phrase accent is followed by L% boundary tone. Thus, the target values corresponding to both L% and H% after H^- are shifted upwards by comparison to their target values after L^-. Like downstep, upstep is a rule which readjusts the phonetic value of a tone in a particular context. The formulation of the rule is discussed in Chapter 4, with an eye towards the lessons learned from a phonetic investigation of downstep. The formulations considered both realize L% at the level of the preceding H^-. H% is realized roughly as much higher than H^- as it is higher than L^-. The two formulations differ about whether the difference between H% and T^- would be exactly the same in both cases. We take the upstep rule, like other tonal implementation rules, to be a local rule. That is, we are not led to suggest that questions have an overall rising and expanding pitch range, as Bing (1979) does. Figure 38 shows a comparison between a longish declarative sentence and the same sentence ending in a question rise. The overall shape of the question is slightly downdrifting, just like the statement. It is only the phrase final sequence H^- H% (which is compressed onto a single syllable) that generates an FØ value standing out above the generally falling pattern. We will pick up this issue again in Chapter 4.

In the analyses just given, we have recognized three contributions to the FØ contour from the nuclear stress to the end of the intonation phrase: the nuclear pitch accent, the phrase accent, and the boundary tone. What is accomplished by breaking down these FØ configurations in this way, instead of viewing them as holistic units? First, we note that separating the phrase accent from the nuclear pitch accent has the same advantages in an analysis of English as it did in Bruce's (1977) analysis of Swedish. Being able to refer to the phrase accent phonologically makes it possible to account for observed variation in the distance between the phrase accent and the nuclear stress. This separation also makes it possible to claim that the inventory of nuclear and prenuclear pitch accents is the same; the claim that there is an extra tone after the nuclear accent explains why there is additional FØ movement on or next to the syllable with the nuclear accent.

In Stockholm Swedish, according to Bruce, all pitch accents are H+L and the phrase accent is always H. In English, there are a number of different pitch accents, and the phrase accent and boundary tone may both be either H or L. This makes it possible to find additional motivation for the decomposition into nuclear accent, phrase accent, and boundary tone, beyond what exists in Swedish.[7]

First, this decomposition makes it possible to predict what different FØ configurations are possible over the part of the phrase from the nuclear stress to the end. This point may be made by contrasting the sequences of nuclear accent, phrase accent, in the present framework with the "nuclear tones" in British work. The nuclear tones are coextensive with our sequences,[8] but as Crystal (1969) points out, are usually considered to be single phonological entities. While this tradition has given us some of the best descriptive work on intonation, the assumption that the nuclear tone is an unanalyzable entity forces this approach to stop at description. It offers no basis for explanations of why many FØ configurations which would contrast perceptually do not also contrast linguistically. For instance, let's return to the question of why English intonation does not have a three way contrast among rising, level, and falling FØ after a H^- phrase accent, as shown in Figure 36. In the framework here, such a contrast is impossible because the two tone description of English intonation provides only two alternatives for the boundary tone slot, whereas describing this contrast would require three. Such an explanation relies crucially on taking the boundary tone to be a theoretical entity, and would not be available if the types sketched in the Figures 36 C and D were viewed as unanalyzable wholes. In the same vein, the claim that the same inventory of pitch accents serves for both nuclear and prenuclear position constrains the forms of the phrase final configurations. In fact, it will be possible to claim that all combinations of pitch accent, phrase accent, and boundary tone generate well-formed FØ contours. The Appendix to the Figures summarizes the possibilities and how they arise. In Crystal, prenuclear material is treated separately as the prehead and head of the phrase; under such assumptions, the way FØ marked stress in the head could in principle be completely unrelated to the form of the nuclear tones.

The decomposition into nuclear accent, phrase accent, and boundary tone can also be motivated by the possibilities for cross-classifying contours that it affords. A number of examples in which such cross-classification is exploited have already come up. The spreading rule for the phrase accent has the result that the final H% rise after a H^- was timed in the same way as the H% rise after a L^-. The claim that the association rule for the phrase accent places it at the end of the word with the nuclear stress has the consequence that the H^- in a L* H^- sequence is located at the same place as the L^- in a H* L^- sequence on the same material. The upstep rule applies to T% after H^-, regardless of what type of nuclear accent precedes. Thus, it not only predicts the possible terminations of the contours in Figure 36, but also that of the superficially different contour in Figure 6. In Chapter 4, we will see that the same rule also makes correct predictions for the vocative contours, which have a partial fall rather than a rise after the nucleus. (Examples are given in Figure 39.) This prediction will fall out because the vocative contours will be analyzed as H*+L$^-$ H$^-$ T%, with T% subject to upstep after the H$^-$. In the domain of meaning, it was suggested that the L*+H$^-$ accent can add a note of incredulousness to an utterance, whether the following phrase accent is H$^-$ or L$^-$. Ladd (1978 and 1978a) discusses an additional case in which an intonational nuance can be pinned on a shared feature of two different nuclear contours. The contours he discusses are the "low rise" nucleus, which is illustrated in Figure 40 A and which is analyzed here as L* L$^-$ H%, and the "high rise" nucleus which is illustrated in Figure 40 B, and has the analysis H* H$^-$ H%. Each of these contours is related phonologically and semantically to what he calls a "stylized version." The stylized version is level instead of rising, and is appropriately used when the utterance is predictable or stereotyped. The stylized version of 40 A is 40 C, analyzed here as L* L$^-$ L%. The stylized version of 40 B is 40 D, which involves a H* H$^-$ L%. Thus, in the present framework, the meaning difference between the plain contours and the stylized contours is the meaning difference associated with the H% as against the L%. This analysis makes a further prediction for a case which Ladd does not discuss. It predicts that the relationship of L* H$^-$ H% and L* H$^-$ L%

would be that of "plain" to "stylized." This prediction seems correct, since as we have observed, L* H^- L% is an appropriate contour for rhetorical questions. Ladd's analysis, unlike the analysis proposed here, does not carry over to this new case, however. In his view, the difference between "plain" and "stylized" is the difference between a level nucleus and a rising one. Now, both L* H^- L% and L* H^- H% are rising, since the sequence L* H^- generates a rise. Thus, Ladd's descriptive terminology does not make it possible to identify the difference between L% and H% after L* H^- with the difference in cases where the nuclear accent and phrase accent are both L or both H.[9]

2.5 Tags

An interesting problem which is discussed in Liberman (1975) and Bing (1979) is how to describe the intonational contrast found between sentences like 5) and 6).

5) Sam struck out my friend.

6) Sam struck out, my friend.

One rendition of the contrast is shown in Figures 41 and 42. In Figure 41, an FØ contour for 5), "friend" has nuclear stress in the phrase. In 42, an FØ contour for 6), "out" has a nuclear accent and the FØ falls to its lowest level by the end of "out." As Bing points out, intonation patterns like 6 occur not only on vocatives, but also on other tag expressions, including polite expressions, expletives, epithets, tag questions, quotative and epistemic verbs, and sentence adverbials. The melody shown in Figure 42, for example, would also be appropriate on any of the following sentences.

7) That's enough, thank you.

8) It's broken, damn it.

9) He won't do it, the bastard.

10) He's sorry, isn't he.

11) "Good heavens," Joe muttered.

12) He forgot it, unfortunately.

13) This is my sister, Mary.

One of Bing's observations about these expressions is that they lack FØ marking of prominence; in our terms, they do not carry pitch accents. Two types of examples bring this point home. First, an expression in this class can have a large number of metrical feet without becoming eligible for the FØ inflections which would arise from pitch accents. A normal pronunciation of 14), for example, has the same melody as shown in Figure 42, with the low level section of the FØ contour expanded to cover the additional material:

14) It's time to get up, you good-for-nothing lazybones.

Secondly, sentences like 7) through 13) can take on strikingly different interpretations if a pitch accent is assigned to the tag expression. For example, in either 15) or 16), "Mary" is no longer a vocative but rather an appositive:

```
15) This is my sister Mary
                 |      |
                 H*     H* L- L%

16) This is my sister     Mary.
                 |         |
                 H* L- L%  H* L- L%
```

17) no longer means that Joe muttered, "Good heavens." Instead, it is an expression of astonishment that Joe muttered something:

17) Good heavens, Joe muttered.
H* L^- H% H* L^- L%

Similarly, 18) is no longer a courteous way of saying, "That's enough." Rather, a rude "That's enough," is followed by an unrelated "Thank you," which is probably to be interpreted as a dismissal.

18) That's enough. Thank you.
H* L^- L% H* L^- L%

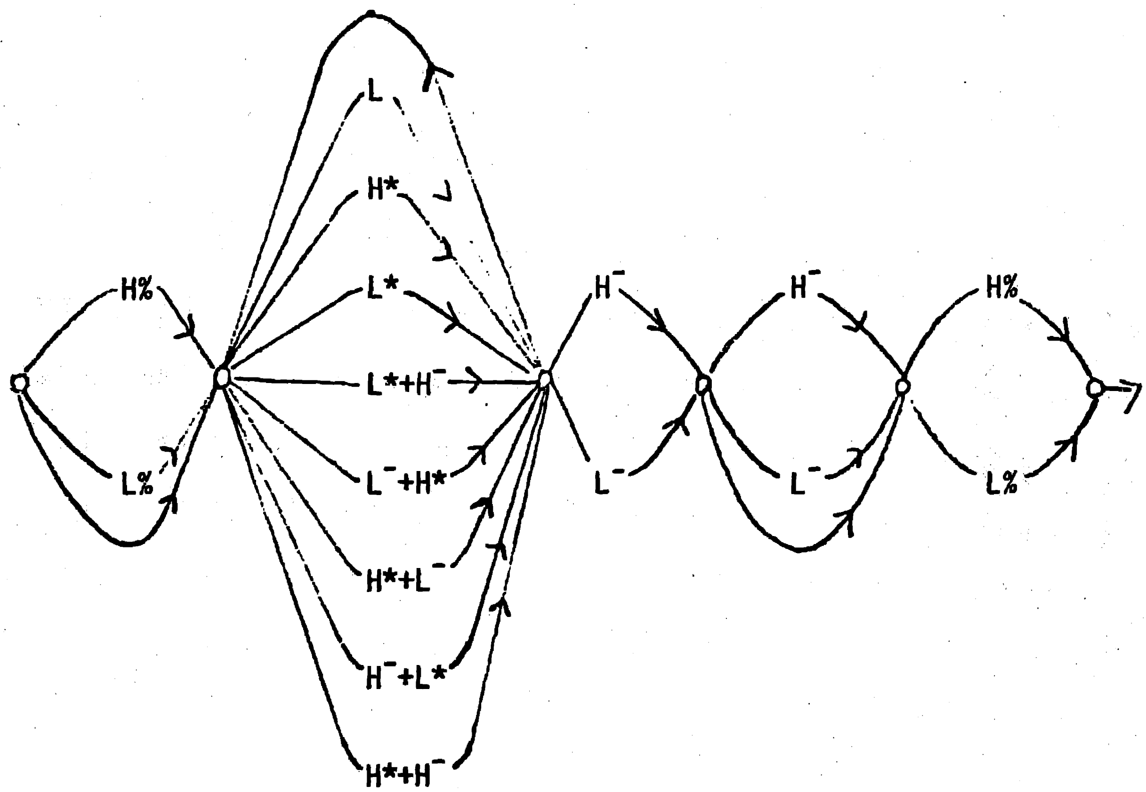
In some cases, there seems to be no alternative interpretation available when a pitch accent is assigned to a tag expression, and so the resulting intonation is bad. In particular, most of sentences 7) through 13) cannot be rendered as a single phrase with the nuclear accent on the tag expression.

19) * It's broken damn it.
H% L* H* L^- L%

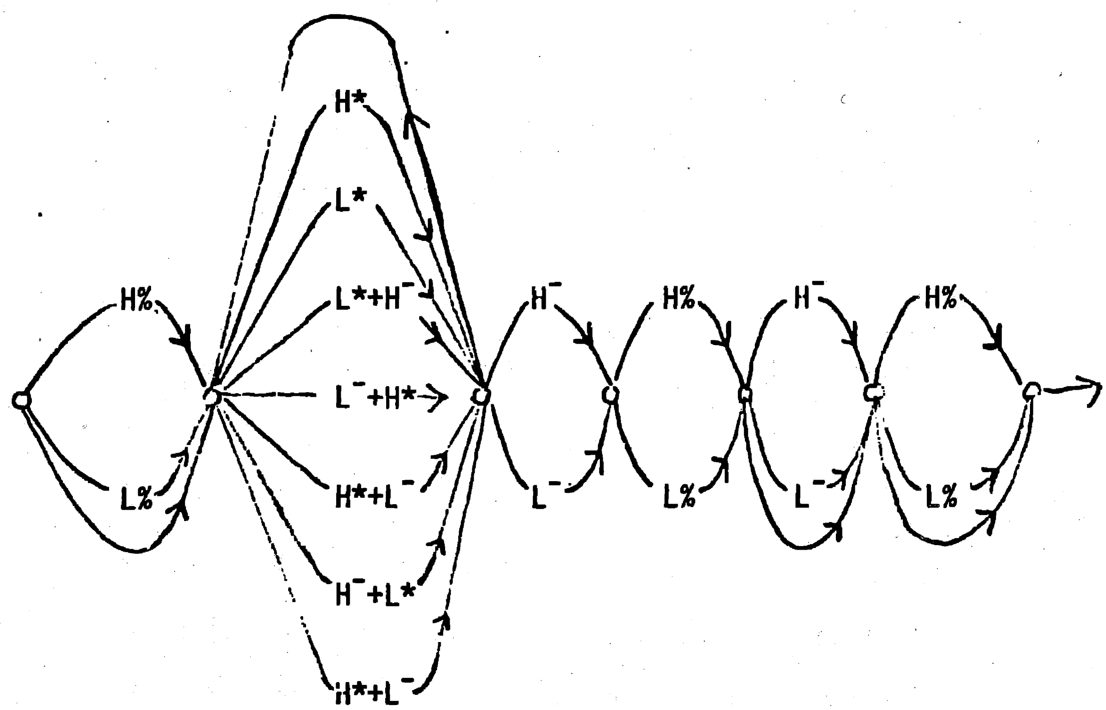
The difference between the tag and nontag intonations is not only a difference in accentuation. Liberman notes that if sentence 5) is produced with nuclear stress on "out", and "friend" deaccented, the FØ contour looks like Figure 43. The salient difference between this contour and 42 is the timing of the fall; in 42, it is completed by the end of "out" whereas in 43, it continues on "my." The same difference can be found in contours ending in H%. This is shown in Figures 44 and 45.

Our hypothesis is that the intonation of tag expressions is to be accounted for by one of two expansions of the grammar of tonal sequences for the intonation phrase given in Chapter 1. Either the tag expression carries a second phrase accent, which follows the first phrase accent before the boundary tone, or else the tag carries both a second phrase accent and a second boundary tone. Under the first proposal, the grammar of allowable tonal sequences would look like 20), and under the second, it would look like 21).

20)



21)



These two alternatives have a number of features in common. First, both do not assign pitch accents to the tag expression. Second, assigning the first phrase accent to the end of the main clause explains why the FØ in Figures 42 and 43 drops so quickly: in these contours, both H* and L^- are found on "out." Third, taking tag intonation to arise from an expansion of the postnuclear part of the tonal sequence explains why the same intonation is not found on the preposable tag expressions when they are preposed. As examples 22) through 25) indicate, there is no intonation pattern for preposed tag expressions which shares the distinctive lack of pitch accents found in the postposed tags. Instead, preposed tags are produced with the intonation of independent phrases, or else treated as part of the following phrase,

22) Mary, this is my sister

H* L^- H% H* L^- L%

23) Mary this is my sister.

H* H* L^- L%

24) Unfortunately, he forgot it

H* L^- H% H* L^- L%

25) Unfortunately he forgot it.

L* H* L^- L%

Fourth, it is our impression that postposed tag expressions are not felicitously set off from the rest of the phrase by a real pause. If there is a real pause, we find the interpretations which arise from assigning two phrasal tunes, as in 16) through 18). This follows from an account in which they are part of the same intonational phrase as what comes before. The point is somewhat delicate, because there is substantial lengthening of the last syllable in the main clause, before the tag. This lengthening would arise with or without an intonational phrase boundary, because of the influence of the sentence boundary.

The alternative accounts summarized in 20) and 21) differ in the number of melodies for tags they are capable of generating. 20) gives us four different combinations of two phrase accents. On the assumption that the rule which upsteps a boundary tone after a H^- phrase accent also raises a second phrase accent, these four possibilities correspond to the four phrase accent configurations shown in Figures 46 through 49. Figure 50 provides an additional and somewhat clearer illustration of the double upstep found in sequences of the form

H^- H^- H%. Altogether, 20) gives us eight different configurations from first phrase accent to the end of the phrase, since as Figure 42 showed, the boundary tone may be L% instead of H% as it is in Figures 46 through 49.

The extra boundary tone separating the two phrase accents in 21) provides an extra degree of freedom, so 21) generates distinctions which 20) does not. For example, this grammar predicts a distinction between L^- L% H^- and L^- H% H^-. L^- L% H^- would according to our phonetic rules have a slight fall terminating on the baseline at the end of the main clause, followed by a rise to H^- at the beginning of the tag. L^- H% H^- would have the rise beginning at the end of the main clause, so that the higher level would already be reached at the beginning of the tag. It is not clear that so fine a distinction is actually used in the intonation of tags. On the other hand, clear examples of the contour which would be transcribed L^- H% L^- H% can be found. An example, taken from Liberman (1975), is shown in Figure 51. One question which needs to be answered is whether 20) is the grammar of intonation patterns for some well-defined class of cases, or whether melodies like those in Figures 45 through 49, which look like outputs of 20), are really the outputs of 21) in which the second phrase accent matches the level of the preceding boundary tone. Another question is whether 21) is needed in its full power, or whether a more restricted version of it is the right one.

The intonation of tags is interesting not only as a phonological problem, but also as a problem in how intonation is related to syntax and semantics. There are very few cases in which a particular type of expression is so strongly related to particular intonational features; for the most part, as we have seen, choice of text and choice of tonal specification can be viewed as independent. Both the representation and the interpretation of the relationship between the postponed tag and its main clause are unclear. More specifically, it is unclear how the representation of the tag and the tonal association rules conspire to line up the extra tones correctly with the text. It is not obvious what prevents the tag from being set off as a separate intonational phrase, given that the main clause ends in a sentence boundary and intonation breaks can in general occur at such major syntactic breaks. The relationship between the tag and its clause may have a special syntactic status which does not allow a phrasing break, or it may be that splitting the construction into two phrases conveys a message about the information structure which is at odds with the informational interpretation of a tag. A related question is why tags can not carry the nuclear stress in a phrase which includes also the main clause.

2.6 The Experimental Literature

In the description of intonation proposed in the last sections, the underlying representation of the F0 contour for a phrase is a string tone, comprised of pitch accents, a phrase accent, and a boundary tone. These are lined up with the text on the basis of the prominence relations and the location of the intonational phrase boundary. The F0 between any two tones is determined by phonetic rules, on the basis of what the tones are and how they are related in time and frequency.

This cluster of features has a number of ramifications which are subject to experimental confirmation. Here, we discuss the relationship of these features of the description to the results of experiments on FØ as a perceptual cue for stress, on FØ as a cue for the location of boundaries, and on the categorical perception of intonation patterns.

In the wake of Fry's classic study (Fry, 1958), the impression grew up that FØ can be viewed as a transducer of stress: the higher the stress, the higher the FØ (or the greater the FØ movement). In the framework outlined here, the relation of FØ to stress is not as direct as this. Rather, a word with a given stress pattern could have any of a number of different FØ contours, depending on the intonation pattern that was being used. A given FØ pattern could be compatible with more than one conclusion about the location of stress, if more than one assumption about where the accent is located was consistent with a well-formed intonational analysis for the contour. However, some FØ contours do not display this kind of ambiguity, but instead permit only one conclusion about the stress pattern. It is only in the second kind of case that FØ can serve as a cue for stress. In fact, this general picture is supported by experimental work since Fry (1958), and by Fry's study itself. Morton and Jassem (1965) report that either lowering or raising the FØ locally can produce the impression that a syllable is stressed. This means that the perception system does not translate FØ height directly into stress level. We would expect this result, since a stressed syllable may have a L* or H* accent. Nakatani and

Aston (1978) report that FØ was not a cue for stress on a noun following a focused adjective. As we mentioned in Chapter 1, no pitch accents are assigned, even to stressed syllables, after the nucleus. Since the focused adjective in Nakatani and Aston's experiment carried the nuclear stress, FØ could not be used to mark stress on the following noun.

Fry (1958) studied how FØ and duration influence perception of stress on the word "subject," which has initial stress as a noun and final stress as a verb. The interaction of duration with sixteen different FØ contours was examined. He found that some FØ patterns overrode duration as a cue for stress; that is, for these patterns, subjects gave the same stress judgment more than half the time, regardless of the relative duration of the two syllables. The patterns which best overrode duration as a cue for stress appear to be those for which one intonational analysis would be highly preferred. For example, the two patterns involving a falling FØ on the first syllable followed by a low FØ on the second syllable would most readily be interpreted as instances of a H* L^- L% pattern on the noun "subject."[10] By contrast, the pattern with a high FØ on the first syllable and a low and then rising FØ on the second syllable was judged to be a noun when the first syllable was long and a verb when the second syllable was long. The tabulated results for this contour have 51% noun judgments, suggesting the FØ contour did not bias stress judgments in either direction. This result does not seem surprising, since the FØ pattern bears a fair resemblance to either a H* L^- H% assigned to the noun, or a H% L* H^- H% assigned to the verb. The interpretation of results for contours which would not be acceptable in English for either stress pattern is rather unclear. Fry himself suggests that a syllable with FØ inflection will be perceived as stressed over a level syllable, regardless of the linguistic system. One cannot, however, take this to be proven by his experiments. For one, he does not in any way control for effects of the linguistic system on judgments. Secondly, the contours he examined are not a systematic sample of the set of possible contours: for example, he includes results for one contour with an inflected first syllable and a high level second syllable, but results for four contours with a low level first syllable and an inflected second syllable. Given that the results for two contours he included do not support his conclusion, it seems possible that a different selection of contours would have resulted in different averaged results.

In our description of intonation, the intonational phrase boundary is the only boundary which has a surface reflex in the FØ contour, namely, the surface reflexes of the phrase accent and the boundary tone. The FØ contour does not in any comparable way mark the word or syntactic phrase boundaries within the phrase; the phonetic rules for interpolating between tones are blind to the structure of the concurrent textual material. Thus, we predict that the possibilities for FØ to serve as a perceptual cue for boundaries are quite limited. Three recent experiments tend to confirm this prediction.

Wales and Toner (1979) studied what kinds of ambiguous sentences may be disambiguated using intonation. The three categories of ambiguities they examined were lexical ambiguities, exemplified in sentence 26; deep structure ambiguities, exemplified in sentence 27; and surface structure phrasing ambiguities, exemplified in sentence 28.

26) Isn't that what a ruler is for?

27) Flying planes can be dangerous.

28) He carried nothing to indicate that he was one of the group.

The only successful disambiguations in their study involved sentences like 28) with the two possible surface structure bracketings. Homonyms as in 26) could not be disambiguated using intonation, nor could deep structure ambiguities which did not have a correlate in surface structure bracketing. Our interpretation of this result is that the speaker for the experiments used intonational phrase boundaries to disambiguate some surface structure bracketings. She was unable to disambiguate sentences like 26) and 27), because the intonation system provides no way of distinguishing readings which do not differ in stress or phrasing.

Experiments by Streeter (1978) confirm that the tonal correlates of the intonational phrase boundary are effective perceptual cues. She studied disambiguation of the phrase, "A plus E times O," and found that subjects were able to use FØ in determining whether the speaker's intention was "(A plus E) times O" or "A plus (E times O)." Unlike Wales and Toner, she reports on what characteristics of the FØ contour permitted disambiguation. The two speakers in the study both used an intonational phrase boundary to mark the bracketing; one speaker used a H* L^- H% on the last (or only) variable in the first phrase; the other speaker used a contour with a L* pitch accent and a H% boundary tone (it is not clear what the phrase accent was for this speaker).

Nakatani and Schaffer (1978) report experiments on the perception of word boundaries in reiterant speech (speech in which the speaker has been asked to replace some or all syllables of an utterance with the same syllable, here ma, preserving the prosodic pattern of the model). They found, as we would predict, that FØ is not a cue for word boundary location when the stress contour is fixed. That is, subjects were unable to use FØ to decide whether the ma-ma imitiations of the underlined words in 29) and 30) represented "mama ma" or "ma mama."

29) The noisy dog kept everyone up all night.

30) The bold design kept everyone's attention.

(Duration differences due to the lengthening of monosyllabic content words could be used with some effectiveness.) FØ was an effective cue for word boundary location only in cases in which it marked a stress pattern which was compatible with only one location of the word boundary (given the contextual constraints). The 1 1 0 stress pattern in 31), for example, would only be possible for "ma mama" and not for "mama ma."

31) The near future is not yet determined for her.

The present approach predicts these results; the pitch accents provide a way of marking stress, and given the stress pattern, the subject would in some cases be able to infer where the word boundary is. However, because the interpolation between pitch accents is insensitive to word boundaries, the location of the word boundary cannot be inferred from the FØ contour except as it marks stress.

We would like to stress that the Nakatani and Schaffer result follows from characteristics of the English intonation system rather than from universal principles. Nothing excludes the possibility that some other language assigns tones is a way which permits word boundaries to be recovered from the FØ contour. For example, Nakagawa and Sakai (1979) report experiments in which Japanese subjects were able to use FØ to determine the number of words in speech which had been resynthesized using white noise and damped sine waves in order to remove segmental information.

A third ramification of the description of intonation proposed here is that the listener should be able to perceive qualitatively different intonation patterns; the difference between H and L in intonation is given a status similar to the difference between, say, [+ coronal] and [- coronal] in the segmental domain. Two experiments in this area seem worth discussing. Hadding-Koch and Studdert-Kennedy (1964) did a comparative study of perception by Americans and Swedes of FØ contours ending in a rise. Subjects were asked to judge whether what they heard was a statement or a question, and some evidence for a qualitative difference between the statement contour and the question contour was found. In the framework here, there is evidence for a qualitative distinction between contours ending in L^- H% and contours ending in H^- H%. This result might have been stronger if the stimuli had been designed in closer conformity to the facts of English intonation. Perhaps because the plan of the experiment required the use of FØ contours which are acceptable in both Engligh and Swedish, all the stimuli had a peak-fall-rise pattern. Some tended towards the question contours we have seen above more than others. However, none were really accurate representations of these contours, which do not have falling FØ anywhere past the nuclear stress.

An experiment by Nash and Mulac (1980) investigated the perception of the contrast between the L^*+H^- and the H^*. These pitch accents were implemented on "thought" in "I thought so"; the FØ on "I" was also varied. Subjects were asked to judge whether the completion "...and I was right" or "...and I was wrong" was more appropriate. No discourse context was provided. The assumption underlying the experiment was that "... and I was wrong" is the more acceptable completion when the L^*+H^- is used, and "... and I was right," when the H^* is used.

Three aspects of the results are of interest here. First, statistically significant differences in the direction predicted were found. That is, listeners were successful in distinguishing between the L^*+H^- accent and the H^*, and in relating this distinction to the semantic difference between the two choices for a response. Second, listeners were somewhat inconsistent in their responses: in particular, the results were strongest for first presentations and weaker for subsequent repetitions. Third, the listeners nonetheless did not make much use of the "can't tell" option on the response sheet; as Nash and Mulac say, they preferred "to impose definite, albeit contradictory, interpretations, rather than to recognize inherent ambiguity."

Nash and Mulac conclude from this experiment that English may have a lexical tonal distinction between the patterns studied. We agree with this conclusion; it remains to explain the degree of inconsistency which was found. One point, which Nash and Mulac bring up, is that in the later repetitions, comparison to preceding stimuli affected responses. A second point is that the response alternatives were less directly related to the distinction in intonation pattern than the authors seem to have realized. Which of the two suggested completions is appropriate is a complex function of the pitch accent used and the discourse context of the sentence. The L*+H$^-$ can be completed with "... and I was right" in the following kind of context:

32) Well, I thought so, but I didn't feel I could tell her that. It
L*+H$^-$
turned out I was right, though.

Conversely, it is possible to use the H* in the context "... and I was wrong":

33) --- But doesn't your book say semantic interpretation is done on deep structure?
--- Well, at that time, I thought so. But I was wrong.
H*

The result that subjects did not view the stimuli as ambiguous even when their interpretation had changed suggests that they recognized the intonation patterns, but changed their minds about the discourse contexts in which the patterns were to be imagined. This hypothesis also explains why the responses of some subjects (an average of 24% per stimulus) did not fit into the expected pattern from the onset. These subjects may have right away based their replies on a different sort of context than the majority.

Footnotes to Chapter 2

1) We would like to stress a point which Lieberman makes himself: The characteristic configurations of the end of a "breath group" have been regularized as markers of grouping, and are often produced by the speaker without actual inhalation at the boundary between groups. This point will be important in Chapter 3 in the discussion of declination, which characteristically has as its domain a series of intonation phrases.

2) Liberman uses a four tone system to transcribe these contours, and the difference in range is attributed to difference choices of tones. As Bolinger (1951) points out, such a description confounds range differences and tonal types in a way that leads to chronic ambiguity. In a two tone theory, the contrast illustrated in Figure 9 could only arise from expressive use of range. This explanation predicts that the phonetic values of the tones are continuously variable along the dimension of contrast in the figure, and this prediction seems to be true.

3) The generalization that F0 contours with no dip also lack unstressed syllables between the peaks obtains only because O'Shaughnessy's subjects used H* pitch accents in their neutral reading intonation. As Chapter 5 will show, the sequence $H^{*}+H^{-}$ H* generates a contour with no dip regardless of how many unstressed syllables there are between the accents. For some speakers (apparently including O'Shaughnessy's subjects), the $H^{*}+H^{-}$ accent is a rather marked one,

used in exclamations and rhetorical wh-questions, for example. For others, including JBP, it is much more usual and could well show up in neutral reading.

4) We are grateful to Mike Wish of Bell Labs for providing these tapes.

5) In the cases under the discussion, the medial syllable was not actually deleted; the syllabic nucleus could clearly be identified in the speech waveform.

6) The intonation pattern in Figure 36 C, which is impossible in English, occurs as a question intonation in Czech, which lacks the upstep rule for the L boundary tone. Figure 37 illustrates the forms this intonation pattern takes as the nuclear stress (marking the focus of the question) is moved through the phrase.

7) One type of motivation for this decomposition which exists in Swedish is still missing in English. Bruce and Gårding (1978) show that the analysis of the contour into accents plus sentence accent can be used to explicate superficially complex dialectal variation in Swedish. It would be interesting to know what dialectal variation exists in English, and how it may be characterized.

8) In saying the nuclear tones are coextensive with our sequences of nuclear accent, phrase accent, and boundary tone, we are setting aside one serious problem with a nuclear tone analysis which is not carried over into the present framework. In our framework, every intonation phrase ends with an accent, phrase accent, and boundary tone. If the accent falls on the last syllable of the phrase, the phrase accent and boundary tone are also crowded onto the same

syllable; otherwise, they are strung out over syllables following the nuclear stress. In some cases, crowding neutralizes distinctions which would be possible when more material is available. This solution was arrived at in the spirit of Liberman's (1975) observation that one makes progress in the study of intonation by considering tune separate from text. Crystal takes a different approach. He breaks down the end of the intonation phrase into the nucleus (the syllable with nuclear stress) and the tail (consisting of syllables following the tail, if any). Then, he considers what features the FØ on the nucleus and tail can display. Obviously, since the nucleus can be in absolute phrase final position, the FØ contour it carries can be the full sequence of accent, phrase accent, and boundary tone. Thus, Crystal is able to identify as nuclear tones many of the phrase final configurations given here. His discussion of the tail is inconclusive, however: he notes that it ordinarily continues the direction of a unidirectional nucleus, but in some cases displays linguistically significant variation.

It appears to us that studying the FØ on the nucleus and tail separately and then attempting to combine them allows many regularities to fall between two stools. For instance, we predict that there would be extremely strong cooccurrence restrictions between Crystal's nuclear tones and his linguistically significant tails, but Crystal does not discuss this question. Also, considering the nuclear tone in its most compressed form to be basic results in lack of attention to how the parts of complex nuclear tones are aligned with the text, and causes some intonational contrasts to be missed.

9) In his discussion of the plain and stylized contours, Ladd (1978) also included a plain and stylized fall. The plain fall is illustrated in contours like Figure 5, and is described here as H* L^- L%. The stylized fall is the vocative contour, which is analyzed in Chapter 4 as having the underlying form H*+L^- H^- L%. Figure 39 A shows an example in which this tune is used to call out to someone; it is also used for various kinds of admonitions and rhetorical remarks, as in 39 B.

In Ladd's view, the salient difference between the H* L^- L% and the H*+L^- H^- L% is that the first shows continuous FØ movement while the second is produced as a series of two levels, as when chanted. Vocative contours recorded in which the A B and B A contours discussed in Chapter 3 showed that this is not really correct; the H* L^- L% and the H*+L^- H^- L% are consistently distinguished by the level of the FØ at the phrase boundary. The H* L^- L% has an FØ fall all the way to the bottom of the speaker's range, while the H*+L^- H^- L% has only a partial fall. Because of this difference, the two can be distinguished even when the H*+L^- H^- L% is produced with continuous FØ movement rather than being chanted. The difference is shown in Figure 39 B.

In view of these facts, viewing the difference between H* L^- L% and H*+L^- H^- L% as another instance of the phonological distinction illustrated in Figures 40 and 41 would not appear to be possible.

Even under Ladd's assumptions, it is quite unclear how this can be counted as another case of the same phonological type. Ladd does not discuss, for example, why the fall should be stylized as a step down, when the rises are stylized as a level FØ rather than a step up. He also does not note the existence of the variant of the vocative contour with a rise at the end, which is shown in Figure 39 C. Under our assumptions, this contour represents the H*+L$^-$ H$^-$ H%, and thus has the same relation to the H*+L$^-$ H$^-$ L% in 39 A that Figure 41 has to Figures 40. The relation of H* L$^-$ L% to H*+L$^-$ H$^-$ L% is different; these two differ in nuclear accent and phrase accent. By taking the H*+L$^-$ H$^-$ L% to be the stylization of the H* L$^-$ L%, Ladd precludes treating it as the stylization of the H*+L$^-$ H$^-$ H%. This would appear to be a wrong move, since H*+L$^-$ H$^-$ H% is closer semantically to the H*+L$^-$ H$^-$ L%, as well as exhibiting a phonetic difference more parallel with Ladd's other cases of stylization.

10) The two alternative interpretations, H$^-$+L* H$^-$ L% and H% L* L$^-$ L%, are both quite unusual patterns which would not seem natural without a discourse context which strongly motivated them.

11) Nash and Mulac describe the accents in Bolinger's framework. Our conclusion about how the stimuli would be described in the present framework was made on the basis of their FØ contours for the stimuli.

Chapter 3

DECLINATION AND THE CONTROL OF PITCH RANGE

3.1 Introduction

The last two chapters have discussed how tonal value and prominence interact in determining the FØ contour. This chapter is concerned with the third important determinant of the FØ contour, declination, and with the interaction of declination with tonal value and prominence. Declination is a gradual downdrift and narrowing of the pitch range, which occurs within the body of the intonation phrase and frequently over the course of several intonation phrases. It does not arise from tonal specifications, but rather is a factor in determining how tonal specifications are mapped into FØ values. This effect has been most extensively studied in Dutch (Collier and 't Hart, 1971; 't Hart and Cohen, 1973; Collier, 1975), and English (Maeda, 1976; Breckenridge and Liberman, 1977; Pierrehumbert, 1979b; Sternberg, et al., 1980; Sorensen and Cooper, 1980; Olive, 1974). It has also been reported for Japanese (Fujisaki, et al., 1979), French (Vaissierre, 1971), Finnish (Hirvonen, 1970), Danish (Thorsen, 1980), Swedish (Bruce, 1977). Phenomena reported in Meyers (1976) and Schachter and Fromkin (1968) suggest it is also found in Hausa and Akan. Bolinger (1978) suggests that the phenomenon may be universal.

An understanding of declination is important for two reasons. First, it is needed to model FØ contours accurately and to account for how the listener recovers tonal values and prominence relations from the F0 he hears. It is known that listeners make allowances for declination in judging the relative height of H tones (Breckenridge and Liberman, 1977; Pierrehumbert, 1979b). However, many features of this normalization process still aren't understood. Secondly, an understanding of declination is important to a theory of what categorial distinctions are possible in an intonation system. This is the case because declination continuously varies the graph paper on which tones are evaluated. If the character of this graph paper is predictable, then the effects of the changing graph paper can be factored out, and a rich system of distinctions in tonal value and prominence can be recovered by the listener. If, on the other hand, the character of the graph paper is itself subject to meaningful variation, then there are limits to how rich a system of distinctions in tonal value and prominence can be simultaneously maintained. If the tonal system is too rich, recovering the declination and the tonal and prominence distinctions from the F0 contour can become a mathematically underdetermined problem.

3.2 The Scaling of H* Values

The experiment reported here (and abstracted in Liberman and Pierrehumbert, 1979) investigated how declination and prominence interact in determining the phonetic value of H tones. While only two specific intonation patterns were studied, the results suggest a model which can be extended to many other cases as well.

The two patterns studied, shown in Figures 1 and 2, were selected because they made it possible to examine how the implementation of a given prominence relation between two nuclear pitch accents was affected when their position relative to the declination was switched. The pattern in Figure 1 is produced as the answer in the following dialogue:

1) What about Manny? Who came with him?

```
---- Anna          came with Manny.
     |                       |
     H* L⁻ L%                H* L⁻ H%
```

Figure 2 shows the answer in the following dialogue:

2) What about Anna? Who did she come with?

```
---- Anna          came with Manny.
     |                       |
     H* L⁻ H%                H* L⁻ L%
```

Note that each of these intonation patterns has two phrases. Thus, they contrast phonologically with the single phrase answers which would also be appropriate for the same questions, shown in Figures 3 and 4. The two phrasal intonation patterns involved in 1) and 2) are $H^* L^- L\%$ and $H^* L^- H\%$. These occur in both orders. In addition to the difference in tonal specification between the two phrased patterns, there is a difference in prominence: the H* in $H^* L^- L\%$ is more prominent than that in $H^* L^- H\%$, because it is used on the main answer to the question, while $H^* L^- H\%$ is used on the background information. We will follow Jackendoff's (1972) usage and refer to the $H^* L^- L\%$ configuration, with its greater prominence, as "A", and to $H^* L^- H\%$, with its lesser prominence, as "B". A convenient mnemonic is, "A for answer, B for background." Obviously, A and B are not pitch accents, since they incorporate an accent, a phrase accent, and a boundary tone as well as a prominence feature. (They do not correspond to Bolinger's A and B accents, a point on which Jackendoff is confused.)

For the experiment, four subjects were recorded in a sound treated booth. Two were the authors, and two were Bell Labs employees who were not involved in the study of intonation. Each subject was presented with a stack of file cards, on which were written:

- A background question.
- The sentence to be read.
- A number from 1 to 10, indicating the degree of "overall emphasis" to be used in reading the sentence.

The recording session included several intonation patterns in addition to those just described, but the only question/sentence pairs which will interest us here are those in 1) and 2). The subject read out the question, and then the answer with the indicated degree of emphasis. With only a small amount of practice, even the naive subjects were able to vary emphasis and intonation pattern orthogonally as the experiment required. The instructions for varying the degree of emphasis were very effective in eliciting a wide variety of pitch ranges. We should note that amplitude and duration also varied with emphasis; in particular, the maximum interpeak durations for each subject was about twice the minimum. We will see below that there are reasons not to be too concerned with the effects of these duration differences.

The stimuli were randomized in sets consisting of all combinations of intonation pattern and degree of emphasis. For the first subject, 8 such sets were recorded. The data turned out to be clean enough that for subsequent subjects the number of repetitions was lowered to 6.

The FØ contour of each utterance was determined using an LPC pitch tracker due to Bishnu Atal. The voiced/voiceless decision in the program was suppressed because it made errors on some low amplitude parts of the signal which were of interest. The revised program computed an FØ value everywhere: voicing decisions were made manually on the basis of the scatter in the computed FØ values and the periodicity of the waveform.

The regularities in the data which our model seeks to explain are illustrated in the graphs for subject JBP (the least consistent subject) in Figures 5 through 10. Figure 5 is a plot of the value of the L% in A against the value of the H*, when A precedes B. (This is in sentences of type 1.) Figure 6 is the corresponding plot for the BA case, or sentences of type 2; the value of the H* in A is plotted against the value of the L%. (In all plots, first peak values, if used, are on the horizontal axis and second peak values, if used, are on the vertical axis.) Clearly, in both plots, the value of H* varies considerably but the value of L% varies little by comparison. Furthermore, the value of L% appears to be uncorrelated with the value of H*. Table I shows the slopes and correlation coefficients of lines fit to these scatter plots, and to the comparable plots for the other subjects. The results for the BA order strengthen the observation made in Maeda (1976) and Boyce and Menn (1979) that the terminal FØ value for the unmarked

Table I

SLOPES AND CORRELATIONS FOR THE RELATION OF L% TO H*

Subject	AB order		BA order	
	Slope	r^2	Slope	r^2
JBP	.13	.43	-.03	.02
MYL	.05	.11	.01	.04
KXG	.2	.55	-.06	.08
DWS	.27	.32	.06	.23

declarative utterance appears to be an invariant for a given speaker's voice. These studies were corpus studies in which pitch range was not varied systematically, and now we see that the observation holds up when a more thorough examination of the relation of L% to the total range is made.

The relationship between L% and H* shown in Figures 5 and 6 contrasts with the relationships shown in Figures 7 and 8. Figure 7 is a plot of the value of the L^- in B against the value of the H*, for the BA case. Figure 8 is the corresponding plot of L^- against H* for the AB case. In both cases, the value of L^- increases with the value of H*. The slopes and correlation coefficients in Table II show that this increase was sizable and significant for all speakers. Similar results were found for the relation between H% and H* in B, and also for the relation between H* in B and H* in A which will be discussed further shortly.

Table II

SLOPES AND CORRELATIONS FOR THE RELATION OF THE L^- PHRASE ACCENT TO H*

Subject	AB order		BA order	
	Slope	r^2	Slope	r^2
JBP	.42	.69	.25	.62
MYL	.10	.37	.16	.63
KXG	.59	.83	.36	.82
DWS	.70	.83	.25	.59

The regularities in Tables I and II can be accounted for by the theory that each speaker has a floor for FØ, which is held constant when the overall range is increased and FØ values above the floor are scaled up. We will refer to this floor as it changes over the utterance as the baseline. We take L% to be on the baseline, and accordingly its value does not increase with overall range. L^-, H%, and H* are all above the baseline, and so their values do increase with overall range. Since L% is the only tone whose value is on the baseline in the patterns investigated, the time course of the baseline cannot be inferred from the FØ counter for any given utterance. On the assumption it is invariant, however, its behavior can be inferred by comparing different contours. Comparing the values of L% in the AB and BA cases shows that the baseline declines during the utterance. The difference between the value of L% in first position and its value in second position averaged 14 Hz for the male speakers and 33 Hz for the female speakers, and was statistically significant for all. This result corroborates the reports of a declining baseline in the references given in Section 1.

Analysis of the data showed that an estimate of the baseline can also be derived by comparing the relation of the two H* peaks in the AB order to their relation in the BA order. This was so, because the baseline was found to control the scaling of the intonation contour throughout the pitch range. Our model for this scaling is motivated by regularities in the H* data for each subject which are exemplified in the scatter plot for subject JBP shown in Figure 9.

In Figure 9, the distance along the horizontal axis represents the height of the first peak, in Hz. The vertical axis represents the second peak. X's represent the AB data points, I's represent the BA data points. The observations about this plot which we will wish to explain are the following:

• The relation between the A and B peaks appears to be linear, in both orders.

• The configuration of the AB and BA data points suggests that lines fitted through them would intersect at point O in the figure, which represents the baseline (x = the median of L% values in AB, y = the median of L% values in BA). This makes sense intuitively: given that the pitch range above the baseline decreases as the plotted peak values decrease, it means that the AB and BA cases are neutralized when the pitch range is zero.

• The AB and BA data points are not symmetric around the line y = x; the whole plot is tilted, so that the slope of each set of points is less than if this were the case. This means the declination shows up not only in the baseline, but also in how the peaks are scaled under the two orders. In both orders, B is backgrounded relative to A. In the AB order, declination adds to the effect of backgrounding, resulting in a B peak which is very much lower than the A peak. In the BA order, by contrast, declination lowers the A peak relative to the B peak, offsetting the effect of backgrounding B. This means that B is at most slightly lower than A.

The model which accounts for these features of the data can be expressed, in its strongest form, as follows:

• Each speaker has a declining FØ baseline which is a characteristic of his voice. This baseline is invariant, in the sense that the onset level and the total drop remain the same as overall pitch range and utterance length are varied. The slope does vary in inverse proportion to length, since the drop remains fixed. The baseline represents the lowest FØ value the speaker would be disposed to reach at any given point in the utterance. It is not instantiated in every phrase, since the tonal specification need not include any tones whose value is on the baseline.

• FØ peaks are scaled as the peak-to-baseline difference (in Hz), divided by the baseline value at the location of the peak. That is:

$$\hat{P} = \frac{P - b_p}{b_p} \qquad (3)$$

This means that baseline declination determines how tones are scaled throughout the pitch range.

• The A peak and the B peak are related by a constant ratio, in the scaled domain. That is, regardless of pitch range,

$$\hat{A} = c\ \hat{B} \qquad (4)$$

where c is a constant greater than one. Figure 10 shows how this model determines the relative peak heights in the AB and BA cases.

Three parameters are involved in fitting this model to the first-peak/second-peak data points: b_1, the value of the baseline at the first peak, b_2, the value at the second peak, and c, the constant relating the A and B peaks. Specifically, algebraic manipulation shows that the equation of the value second peak (P_2) in Hz as a function of that of the first peak (P_1) is (5) for the BA order:

$$P_2 = c(b_2/b_1)\ P_1 + (1-c)b_2 \qquad (5)$$

In the AB order, where A is the first peak and B is the second peak, the equation is the same except that 1/c is substituted for c:

$$P_2 = (1/c)(b_2/b_1)\ P_1 + (1-1/c)b_2 \qquad (6)$$

For each subject, values for b_1, b_2, and c were found which minimized the mean absolute perpendicular deviation of the data points from the predicted values. Absolute perpendicular deviation was selected as the measure of fit for two reasons: First, the line fitted on this basis is the same when computed for x against y as for y against x, like the principal components line and in contrast to ordinary regression. Second, absolute perpendicular deviation is the most perspicuous measure which overcomes the problem with the quadratic weighting used in fitting a principal components line -- excessive weighting of outliers. Optimal values for the three parameters were found numerically, using a computer.

The results of fitting the model to the first peak-second peak data are shown in Figures 11 through 14. The upper and lower lines in each figure are the predicted values for the peak relations in the BA case and the AB case, respectively. The middle line represents what the model predicts to be equally prominent peaks. Note that the intersection of the fitted lines (at (b_1, b_2)) is in every case very close to the O representing the baseline for the subject. This is a striking confirmation of the model, since the measurements of L% on which the estimate of the baseline is based played no part in fitting the model parameters c, b_1, and b_2 to the peak data. Table III lists for each subject estimates of c, b_1, and b_2, measured L% values, and mean absolute deviation of observed from predicted values.

The patterns in the data which led us to propose the present model also permit us to reject a number of other hypotheses about how H tones are scaled. If the scaling function were

$$\hat{P} = \frac{P}{b} \qquad (7)$$

Table III

THE RESULTS OF FITTING THE MODEL FOR EACH SUBJECT

Subject	Fitted parameter values			Median of Measured L% Values		Mean deviation of observed data points from those predicted
	c	b_1	b_2	b_1	b_2	
JBP	1.4	158	134	156	137	12.3
MYL	1.5	89	75	87	77	6.6
KXG	1.5	142	105	143	103	12.3
DWS	1.4	116	101	124	101	7.3

(For JBP and KXG, the fits reported are slightly sub-optimal; accepting a fit in which the mean deviation was less than 0.05 worse than the optimal fit in these two cases gave values of b_1 and b_2 which were noticeably closer to measured values.)

instead of (3) above, the lines fitted to the first-peak/second-peak data points would go through the origin. For all the subjects, lines fitted by least absolute deviation separately to the AB and BA data points did not go through the origin. Jackknifing (Mosteller and Tukey, 1977) was used to show that the difference of the intercepts from zero was statistically significant. If the scaling function were

$$\hat{P} = P - b \qquad (8)$$

then the AB points and the BA points would be symmetrically disposed around the line y = x. As Figures 11 through 14 show, this was clearly not the case for any of the subjects.

A more interesting possibility is that peaks are scaled not by a hypothetical baseline, but by L's actually found in the FØ contour. Under such a theory, the reflex of prominence is not target level but amount of FØ change. The intonation patterns investigated provide an opportunity to test this hypothesis, because the L% in A is on the hypothetical baseline, but the L^- in B and the L preceding the first peak in both patterns are not. This means that the relationship between the scaled A peak and the scaled B peak would come out differently under this hypothesis than under our model.

A number of different versions of this hypothesis were investigated, including models using the logarithmic musical scale rather than the linear Hz scale. We restrict our attention here to the two best fitting ones, which preserve the scaling function (P-b)/b for P in Hz. In one version of the hypothesis, the values of b were the low values preceding peak values in the same contour; in the second version, the values for b were the low values following. The only free parameter in both cases is the constant c relating the A and B peaks. Values for c were found which minimized the mean absolute perpendicular deviation under both versions, using the same methods as above.

Table IV compares the mean deviations found under these assumptions to those which result under our model. Scaling by actually occurring low values gives a worse fit for all subjects, and for some subjects, the fit was considerably worse. The subject for which the

Table IV

A COMPARISON OF THE MODEL IN WHICH PEAKS ARE SCALED BY A HYPOTHETICAL BASELINE TO TWO MODELS IN WHICH PEAKS ARE SCALED BY ACTUALLY OCCURRING LOW VALUES. VALUES ARE MEAN ABSOLUTE PERPENDICULAR DEVIATIONS.

Subject	Hypothetical Baseline	Low after Peak	Low before Peak
MYL	6.6	9.1	12.3
DWS	7.3	17.1	27.1
JBP	12.3	38.8	62.1
KXG	12.2	30.4	31.3

alternative hypotheses fit best, MYL, was also the subject who showed the least tendency to raise L^- above the baseline. The comparison in Table IV is not completely conclusive, because noise in the production or measurement of the observed lows may be working against models which scale actual rises and falls. The effect of such noise was less in the baseline estimates against which b_1 and b_2 of our model were compared, because a summary statistic, the median, was used. However, we feel the results in Table II suggest that the data are better handled in terms of relations in target level rather than relations in size of rise or fall.

3.3 Is the Baseline Invariant?

The model just presented crucially assumed that the baseline was invariant as the overall emphasis was varied. If this were not the case, we would not have been justified in collapsing measurements of b_1 and b_2 taken from utterances with different degrees of emphasis, and we would not expect that the first-peak/second-peak data for the larger pitch range utterances would point to the same baseline values as the data for the smaller pitch range utterances. A close examination of Figures 11 through 14 suggests that the relation between the A and B peaks is actually slightly curved, so that higher values point to a somewhat different baseline than lower values. However, a linear approximation works quite well -- perhaps surprisingly well in view of fact that changes in overall pitch range were accompanied not only by changes in duration but also by changes in average amplitude and therefore presumably in subglottal pressure. We are not aware of any other experimental results bearing on the relation between average amplitude and amount of declination. However, there is a body of work which suggests that amount of declination remains invariant under changes in utterance length. These results add plausibility to our assumption that the baseline in our study did not vary significantly.

The claim that the FØ drop exhibited by the baseline is constant was first put forward by Maeda (1976) on the basis of his corpus study. His materials included both isolated sentences and paragraphs. He fit the baseline by eye to the lowest points in the contour. In extended material, points where the declination appeared to have reset were identified, and a series of baselines was fit. A typical result is shown in Figure 15, which is taken from Maeda. Maeda concluded that the baseline drops a constant amount, regardless of the length of the material it subtends. This, of course, implies that the baseline slope is in inverse proportion to length. The standard deviations for drop he reports are on the order of 10% of the drop, with insignificant covariance with length.

One prediction of our model is that taking the baseline to be a measurable feature of individual FØ contours would inflate the variance in observed slope, because low values can vary through factors other than variation in declination. This would reduce the chances of identifying any relationship between amount of declination and length which did exist. However, given that the standard deviations for drop Maeda reports are not large, his results at least suggest that any effect of length on drop is not large. Two subsequent experiments have provided additional evidence for Maeda's original claim.

Sternberg, et al. (1980) examined declination in lists of two to five numbers produced in an experiment on motor latency. Subjects had been instructed to speak as quickly as possible, to put equal stress on each number, to avoid phrasing, and to speak in a monotone. The measure of declination was the sequence of medians of the FØ values in stressed syllables. Under these circumstances, the total amount of declination was found to be constant regardless of the length of the list. Although the FØ medians of the stressed syllables were almost certainly above the baseline, they should provide an accurate reflection of relative amounts of baseline declination, given the experimental design. The absolute amounts of declination reported cannot be meaningfully compared to the amounts found in our experiment without examining the FØ contours the subjects produced.

Maeda's claim is also confirmed by a perception experiment reported in Pierrehumbert (1979b). The experiment examined how the perception of relative peak height in a nonsense sentence was affected by the separation between the peaks. The two FØ peaks which the subjects compared were separated variously by one unstressed syllable, by two unstressed syllables, and by three syllables including a medial stress with pitch accent. It was found that subjects made a correction for declination in all three of these cases: if two peaks were equal in Hz, the second sounded higher, and the second had to be lower for the two to sound equal. The amount of this correction was the same for all three types of stimuli.

A study which challenged Maeda's conclusion is reported in Sorensen and Cooper (1980). In this study, declination in paired sentences of 8 and 16 syllables was examined. A curve fit through the peaks in the sentence was used to measure declination. The slope of declination was found to be less for longer sentences, but the decrease was less than in proportion to sentence length. Hence, the total drop was greater in longer sentences. This seemed to be related to the fact that the first FØ peak in the longer sentences was about 6% higher than in the shorter sentences. The apparent discrepancy between these results and Maeda's can be explained in terms of our model. As Figure 10 showed, our model defines a kind of graph paper for mapping tones. The lines on the graph paper define what value an H tone must have later in the utterance to count as equally prominent as an H tone earlier in the utterance. Given how these lines are defined, they tilt more in the FØ domain, the higher they are in the range. Thus, if speakers for whatever reason began a longer sentence at a higher FØ value, we predict a steeper decline and a greater total drop for declination line fit through the peaks. This would be true even if the FØ drop of the baseline itself was not affected by sentence length. The fact that the longer sentences began higher than the shorter ones is not surprising. It is known that paragraph initial declarative sentences have a higher first peak than paragraph internal sentences (Enkvist and Nördstrom, 1978, Lehiste, 1975). So, it appears that a larger pitch range is used in general to signal the onset of a larger semantic unit.

If this explanation of Sorensen and Cooper's result is to hold up, we must also explain why Sternberg, et al. (1980) did not find a similar pattern. Their measure of declination, the FØ median, is also above the baseline and would be expected to drop more if a greater pitch range was used. In fact, Sternberg, et al., report that the onset FØ of the first number in the list was unaffected by the length of the list. One possible explanation of the contrast between this result and Sorensen and Cooper's is that the instructions to the subjects in Sternberg, et al. suppressed expressive use of pitch range. Or, it is possible that the effect Sorenson and Cooper found only exists in longer utterances than

Sternberg, et al. were interested in.

In the Pierrehumbert (1979b) experiment on the effect of interpeak separation on perception of relative peak height, the first peak was not varied. Assuming that subjects' judgments were made on the basis of the first peak actually heard, we would not under the assumptions of our model expect a greater correction for declination for the longer sentences. However, we would expect a greater correction if pitch range were varied, with or without variation in utterance length. An experiment on the effect of pitch range on perception of relative peak height is reported in Pierrehumbert (1979b). A greater correction for declination was found for the wider pitch range stimuli in this experiment, as predicted by our model.

3.4 Hypotheses about the Implementation of Intonation

The experimental results discussed in Sections 3.1 and 3.2 suggest a number of hypotheses about the structure of the intonational system which will play an important part in subsequent chapters. Our aim here is to lay out what these hypotheses are and why they are plausible. We do not mean to suggest that they are proven by the experimental results presented. Rather, we feel that they are in some measure justified by the results above and by the part they will play below in providing a coherent picture of English intonation. We look to future experimental work to provide further justification, or to uncover how they need to be corrected.

Our first hypothesis, which was also discussed above, is that the speaker's baseline, or the hypothetical bottom of his range to which tonal values are referenced, is a quite invariant feature of his voice. It follows that differences in the overall configuration of the F0 contour arise not from differences in declination, but from differences in tonal specification and prominence. This hypothesis interacts with our claim that the rules which compute the F0 contour from a representation of prominence relations with associated tones have a narrow window as their domain. Taken together, these two claims restrict the range of intonational distinctions which can be described in the theory.

Our second hypothesis is that the model we have worked out for describing prominence relations between two phrases also applies within the phrase. This means that a gradually declining baseline is defined within the phrase, and that the phonetic value of tones is computed in baseline units above the baseline. The notation /T/ will be used to represent the phonetic value of a tone as expressed in these units. One of the lessons of Figures 11 through 14 is that phonetic value is continuously variable. We will also assume that a prominence relation between two H tones is expressed as a ratio between their phonetic values. The idea that the tone mapping rules are formulated in terms of ratios of baseline units above the baseline will be extended to cover the phrase accent and the boundary tone. It will be crucial in explaining the behavior of the downstepped tones discussed in Chapter 4.

Our third hypothesis is that the baseline plays a role in perception as well as production. In particular, it is our impression that a listener can tell whether the FØ contour has reached the speaker's baseline or not. One possible basis for such a decision would be a normalization process based on a small sampling of speech, similar to that which permits listener's to normalize for a speaker's vowel space (Ladefoged and Broadbent, 1956). The correlation between vocal tract size and size of larynx may be good enough to permit the listener to infer from formant values where the bottom of the speaker's range would be. Or, FØ values near the baseline may be produced with a characteristic source spectrum. Whatever the basis of the decision, the ability to make it affects what categorial decisions can be made in the intonation system. Consider, for example, the contrast between the vocative and the declarative terminal fall shown in Figure 16. From the point of view of the speaker, the difference between these two is that the terminal fall goes all the way down to the baseline, while the fall in the vocative stops well above the baseline. We believe that this difference is recoverable by the listener. It will be described in Chapter 4 as a categorial difference between H* L^- L% and H*+L^- H^- L%, in which the H^- is downstepped but remains above the baseline. If the tonal values are not referenced to the baseline in this way, the two contours in Figure 16 differ not in type but in size: one is a larger fall and one is a smaller fall. They are related in the same way as a more emphatic and a less emphatic instance of H* L^- L%. This approach is taken in Ladd (1978), who does not recognize a difference between a declarative and a vocative in which the FØ changes continuously.

The results just presented do not lead us to draw conclusions about whether the implementation system for intonation should be framed in terms of acoustic parameters or articulatory ones. Obviously, the results are consistent with an account in which the speaker computes an FØ value as the implementation for a tone, and determines a series of motor commands which will enable him to attain the FØ target. However, it is equally possible that the observed regularities are byproducts of an implementation system which maps tones directly into motor commands. Suppose, for example, that declination is generated by a gradually declining subglottal pressure curve, as suggested in Collier (1975) and Bolinger (1978), that tones are implemented laryngeally, and that subglottal pressure and laryngeal parameters interact multiplicatively in controlling FØ. If this is so, the data would be accounted for without positing FØ target values anywhere in the system. This picture is no doubt too simple. However, a close comparison of our results with Maeda's suggests that it is more plausible than Maeda's results would at first seem to suggest. On the basis of a review of the literature on subglottal pressure and its effect on FØ, Maeda estimates the contribution of subglottal pressure to the amount of declination to be 15 Hz for a male subject. In our study, the baseline fit for DWS exhibited a 15 Hz drop, while the baseline fit for MYL had a drop of 14 Hz. Maeda rejects the hypothesis that subglottal pressure completely accounts for declination, since he found drops of 20 to 40 Hz for the speakers he studied. However, his method of determining the baseline almost certainly inflated the observed drop relative to what it would be in our model. His example F$\emptyset$ contours show that his corpus had a high percentage of the downstepped intonation patterns discussed in the next chapter. A baseline fit by eye to such contours declines sharply, since the tonal specification generates a drop in the contour in addition to the drop arising from declination. Pilot results discussed in the next chapter suggest that a baseline fit by Maeda's procedure is not relevant to the description of these contours. Instead, the values of the peaks in such a contour are computed with reference to a baseline which is not seen in the F$\emptyset$ contour and which can be identified with the baseline for the AB and BA contours. Maeda's decision to regard L% as below the baseline may also have contributed to the larger drops found in his study. Given that he fit baselines to low points remaining when L% was set aside, his baseline is higher in the speaker's range than the baseline in our model. As we just pointed out with regard to Sorensen and Cooper's results, we expect to find more declination the higher in the range its measure is taken.

Chapter 4

DOWNSTEP, UPSTEP, AND LEFT-TO-RIGHT TONAL IMPLEMENTATION

4.1 Introduction

Chapter 2 gave no account of a family of intonation patterns which are very common in both American and British English. In Figures 1 through 4, the four members of this family are shown on the same phrase, "There are many intermediate levels". Figures 5 through 8 give additional examples. The interesting property shared by these contours is their steeply falling configuration. Because they fall much faster than declination would account for, the last peak in each has a lower phonetic value than the preceding ones, even though it occurs on the nuclear stress and is therefore underlyingly the most prominent. The contours differ in the features of the FØ contour which are local to each accented syllable. In Figures 1 and 5, the FØ level on each accented syllable is sustained over subsequent unaccented syllables, with a sharp fall at the next accented syllable. This pattern is reported by Kingdon (1958) and Crystal (1969) to be the most common in British English. In Figures 2 and 6, there is a gradual fall from one accented syllable to the next, so that the FØ on an accented syllable is a local maximum with respect to material immediately following it. Figures 3 and 7 have a relatively low FØ on each stressed syllable with a rise immediately following. This suggests that the accents in these figures are the familiar L^*+H^-. Figures 4 and 8 show a similar rising-falling pattern, except that the peaks occur on the accented syllables and the valleys immediately precede.

One might contemplate resolving the conflict between phonetic value and prominence seen in these contours by increasing the English tonal inventory. Under such a proposal, the contour in Figure 2, for example, might be analyzed as a sequence of High, High-Mid, and Low-Mid, with the low phonetic value of the tone on the nuclear stress attributed to its phonological character. This move is inadvisable, however, because the number of distinct levels in contours of the type we are discussing can be considerable. The F0 contour in Figure 5 has six distinct levels. There are no known cases of a language with six phonemic tone levels (Maddieson, 1978); in fact, the existence of languages with even five level tones is disputed (Yip, 1980).

In the description of these intonation patterns proposed here, the claim that English has only two tones is maintained. However, we will not suppose that each of these tones corresponds to some fixed part of the overall pitch range. Instead, we will propose that the location in the range corresponding to a particular tone is computed by context sensitive rules, and thus changes between one tonal location and another. More specifically, the underlying descriptions for the four intonational types shown in Figures 1 through 8 all involve bitonal pitch accents. For Figures 1 through 4, the accents are respectively H^{-}+L*, H*+L^{-}, L*+H^{-}, L^{-}+H*. The tone implementation rule which is responsible for the descending pattern lowers the target corresponding to the H on the right in the sequences H L+H and H+L H. In a sequence where the context for the rule is met more than once, such as H L+H L+H, the second H is lowered relative to the first, and the third is lowered relative to the second, so that descending terraces result. We will refer to a rule of this type as a downstep rule. As we observed in Chapter 2, English also has an upstep rule, which raises the target corresponding to either L or H after a H$^-$ phrase accent. Because of this rule, the boundary tone after H$^-$ is either at the same level as H$^-$ (if it is L%), or else higher (if it is H%). Figure 9 illustrates how this upstep rule is reflected in the FØ contours for H* H$^-$ H% and H* H$^-$ L%.

In our account of downstep and upstep, the decomposition of the intonation pattern into pitch accents, phrase accent, and boundary tone plays a crucial role. The bitonal form of the pitch accents is responsible for describing contrasts in the local behaviour of the FØ around the stressed syllables. The claim that a sequence of such accents generates the terracing pattern means that there is a ready account of FØ contours in which terracing occurs in only part of the phrasal tune. Two such contours are shown in Figures 10 and 11. In the framework here, the contour in Figure 10 is readily described as H* H*+L$^-$ H* L$^-$ L%, and the one in Figure 11, as L* H*+L$^-$ H* L$^-$ L%. This advantage is shared by the descriptive framework in Bolinger (1958), in which stepping configurations are also described as a series of accents. In Crystal (1969), by contrast, the stepping configurations are viewed as a type of head; the existence of heads which start off one way and continue another is noted as a puzzle. A third consequence of our decomposition of the intonation pattern is that it raises the possibility of the H$^-$

phrase accent being downstepped after a H+L nuclear pitch accent. This possibility will be exploited to describe the vocative pattern shown in Figure 12. This pattern has a partial fall from a peak on the nuclear stress to the phrase accent; it contrasts both with the sequence $H^*\ H^-$, which has a sustained high level FØ, and the sequence $H^*\ L^-$, which falls closer to the bottom of the speaker's range. Superficially, this pattern would suggest the need for a mid tone, as in the description in Liberman (1975). Taken together, the downstep rule and the separation of the pitch accent and the phrase accent make a two-tone description possible. Furthermore, we will see that the downstep rule interacts with the upstep rule to predict the behaviour of the boundary tone in vocatives.

The introduction of tonal implementation rules as powerful as downstep and upstep raises many questions: What contextual information can such rules in general make use of? What mathematical relations may they express? What conventions control their interaction? These questions are the topic of this chapter. The downstepped patterns will play a central role in the discussion, because they suggest two different approaches to describing the representation and implementation of intonation which can only be evaluated after detailed examination of the phenomena. Under one approach, the FØ contour results from the interaction of a global specification of the intonation pattern with local specifications. Under this approach, Figures 1 through 4 would exhibit a generally descending pattern on which the L and H tones rode. Two different versions of this approach will be discussed. In one version, due to Lea (1973) and Thorsen (1978, 1979 a,b,c and d, 1980), the global and local specifications are essentially independent, and superimpose. In another, developed in Clements (1980) and Huang (1979), the tonal specifications are the bottom level of the global characterization, and are one of several factors which determine its form. Neither of these approaches will be adopted here. Instead, we will suggest that the overall contour arises only as a byproduct of the application of local tonal implementation rules. Figure 5 and 13 illustrate one observation which makes this approach seem plausible. Both of these contours have a large number of downsteps, so that it is easy to see the exponential character of the implementation. More precisely, Section 4 will show that each level is a constant ratio of the previous level, in baseline units above the baseline. This is just the type of overall configuration which can arise through a local rule which computes each value as a function of the value of immediately preceding tones. That is, the FØ contour can be computed by local rules applying interatively left to right. If it had turned out instead that the step size depended on the number of upcoming steps, or the number of steps so far, then a non-local implementation would have been required.

A comparison between Figures 13 and 14 illustrates another correct prediction of this approach. Under the local theory, the level which follows target level x in a downstepped sequence should be the same, regardless of what tones or target levels are found elsewhere in the phrase, and regardless of how many downsteps precede or follow. To a good approximation, Figures 13 and 14 are in line with these predictions.

In Figure 13, the H^- on "Ebenezer" is the second downstep in a sequence of L^*+H^- accents. In Figure 14, the same tone is the first downstep, since its left context is L^*+H^- H^*. However, due to the smaller initial pitch range in Figure 14, the value of H^* there is about the same as that of the H^- originating on "believe" in Figure 13, even though the latter is downstepped and the former is not. Consequently, $/H^-/$ on "Ebenezer" is also the same in both cases, and the implementation of the following L^*+H^- on "dealer" is likewise the same.

Section 3 surveys phenomena which suggest that all F0 contours are computed from the underlying tonal sequence by local context sensitive rules. We already suggested in Chapter 2 that upstep is a local rule. We will also identify contextual variants of L, and provide evidence suggesting that H's are evaluated with reference to immediately preceding tones.

The technical formulation of our hypothesis about downstep and its relation to other tone mapping rules has the following features:

. The value of the first pitch accent in the phrase is a free choice, governed by pragmatic or expressive factors.

. Subsequent tones are scaled in relation to immediately preceding tones, taking prominence relations into account. For a series of H^*'s, the relation is:

1) $$/H^*_{i+1}/ = /H^*_i/ \frac{\text{Prominence } (H^*_{i+1})}{\text{Prominence } (H^*_i)}$$

We will show in Section 3 that this rule generalizes to govern the relation H's in any two pitch accents which contain a H, and also of the H^-

phrase accent to such an accent. The general form of the rule is thus:

2) In $H_i(+T)\ (T+)H_j$: $/H_j/ = /H_i/ \dfrac{\text{Prominence } (H_j)}{\text{Prominence } (H_i)}$

Where H_j is the phrase accent following H_i, the prominence ratio is apparently constrained to be one, and so $/H_j/ = /H_i/$. Note that 2) does ~~not cover $H^*\ L^-\ H\%$, where $/H\%/$ is typically less than $/H^*/$. As stated,~~ the rule does cover $H^-\ H\%$. It is difficult to evaluate its correctness in this case, however, since /H%/ varies considerably for expressive reasons.

. L tones, like H tones, are also scaled in relation to preceding tones; however, a single rule does not cover the L* accent, the L in a bitonal accent, and the L phrase accent and boundary tone. As a result, the value of two successive L's can differ even without a change in prominence. Rules which will play a part here are:

3) In H+L: $/L/ = k/H/ \quad 0 < k < 1$

4) In H(+T) L+: $/L/ = n/H/ \dfrac{\text{Prominence } (H)}{\text{Prominence } (L)} \quad 0 < n < k$

(The value of H in L+H is then computed by rule 2.)

5) In $H(+T)\ L^-$: $/L^-/ = p\ /H/ \quad 0 < p < k$

(Here, we mean L^-, the phrase accent, and not L^-+)

6) $/L\%/ = 0$ (or as a ratio, $/L\%_{i+1}/ = 0\ /T^-_i/$)

Rules 3) through 6) are motivated in Section 3.

We also speculate that the rule for a L pitch accent or phrase accent following a L* accent is:

7) $/L_{i+1}/ = /L^*_i/ \dfrac{\text{Prominence } (L^*_i)}{\text{Prominence } (L_{i+1})}$

Using the inverse of the prominence ratio used in rule 2) governing H's means that $/L_{i+1}/ < /L^*_i/$ if Prominence $(^*L_{i+1})$ > Prominence (L^*_i), as observed in Chapter 2. However, rule 7) still guarantees that $/L/ > 0$ for finite prominence. On the assumption that the prominence of the phrase accent is the same as that of the nuclear accent, rule 7) assigns a L^- phrase accent the same value as a L* nuclear accent.

• /H/ as determined by rule 2) is subject to readjustment by downstep and upstep. The downstep rule is:

8) In H+L H_i and H L+H_i: $/H_i/ = k/H_i/$

A tentative formulation of the upstep rule is:

9) In H^- T: $/T/ = /H^-/ + /T/$

(Here, we mean H^-, the phrase accent, and not $+H^-$.)

It is important to note in interpreting rules 8 and 9 that "=" is an assignment operator, as in rules 1) through 7), and not a logical operator. That is, 8) and 9) are not equations to solve, but rules which assign a new value to a tone on the basis of its old value. Rules 8) and 2) together mean that downstep interacts multiplicatively with prominence to determine the value of downstepped H tones. Rules 3) and 4) interact with 8) to lower successive L tones in downstepped sequences. k in rule 3) is the same as k in 8), and so H+L H exhibits total downstep. n in 4) is a smaller fraction than k, and so H L+H displays partial downstep.[1]

• Rules 1) through 9) are part of the package of rules which map tonal specifications and prominence relations into F0 contours. This package of rules applies iteratively left to right, and includes the rules for interpolating between target values discussed in Chapter 2 and further in Chapter 5. The increment for this iteration process is one tone. In computing the value for the new tone, rules can refer back at most as far as the previous pitch accent. Values computed under previous iterations are not subject to modification. Figure 15 gives a step by step derivation for a series of accents with downstep, as it would be implemented under these assumptions. The implication of the description is that tones can be mapped into F0 contours by a finite state machine.

The downstepping observed in English has substantial similarities to downdrift and downstep as they have been studied in African tone languages. Both downdrift and downstep rules shift downwards the location in the pitch range at which a particular tonal type is implemented; in the paradigmatic case, H is lowered after L. Traditionally, downdrift has been distinguished from downstep by its surface transparency. Downdrift is viewed as an automatic assimilation of tone to its predecessor, and so in a two tone language a downdrifted H is found only after a preceding L. In languages with downstep, by contrast, there are surface distinctions between the sequences H H and H !H (where !H represents a downstepped H). In many cases, it has been possible to motivate analyses in which sequences like H !H arise from an underlying representation in which the two H's are separated by a L which fails to appear on the surface (Stewart 1965; Schachter and Fromkin 1968; McCawley 1970). Clements and Ford (1979) characterize these tones which fail to appear on the surface as floating tones, that is, as tones which have not been associated with syllables in the text. They argue that all cases of downstep should be derived synchronically from representations with underlying floating tones. A language will exhibit both downdrift and downstep if the lowering rule is insensitive to whether the triggering tone is floating or attached. If it is triggered only by a floating tone, then the language has downstep but not downdrift. Although in the typical case H is lowered after L, in more complex systems, other tones may cause lowering or be lowered. The main thrust of Clements and Ford's proposal is that downstep and downdrift are really the same thing. In view of this result, we will make no distinction between the two in our discussion of tonal implementation in English and African languages; we will use the term "downstep" to refer to cases where the underlying tonal representation shows up transparently on the surface as well as to cases where it does not.

The main similarities between English downstep and the classic cases of downstep in African tone languages are that the tonal value for H is lowered after L; that the new value for H governs not only the downstepped H, but also the value for any H's to the right; and that lowering due to downstep is over and above lowering due to declination. The first point repeats an observation made above. The second point is illustrated in Figure 16A, which shows a schematized FØ pattern for a two tone language with simple downdrift. The value for H is lowered after L, and the new value is continued on subsequent H tones which do not themselves follow a L. We will see in Section 3 that similar phenomena are found in English. Figure 5 illustrates the third point for English. The first peak in this contour is about 1 2/3 baseline units above the baseline, and the drop from this peak to that on the nuclear accent is 207 Hz. The results in Chapter 3 suggest that the maximum baseline decline likely for a speaker with this pitch range is about 40 Hz. Given our model, the effect of declination at 1 2/3 baseline units aboe the baseline would thus be at most 67 Hz. This means that the nuclear accent is 140 Hz lower than it would be if it has the same phonetic value as the first peak. The literature also contains two types of evidence for distinguishing declination and downstep in African tone languages, although to our knowledge this distinction has not been made. Phrase internally, declination can show up in the FØ of successive like tones, where downstep does not apply. Meyers' (1976) instrumental study found that the second of two like tones in Hausa is lower than the first, though not as much as if it had been subject to downstep. FØ contours in Silverstein (1976) corroborate this finding. Hombert (1974) reports a similar result for H tones in Shona. Earlier transcriptions by ear (e.g., Hodge and Hause (1944) reported that the second of two like tones in Hausa maintained the level of the first; the contrast between this report and Meyers' finding suggests that non-instrumental observations on this point in other languages are probably not reliable. A second way that declination can show up is by affecting the FØ values in successive phrases, as it did in Chapter 3. Schachter and Fromkin (1968) formulate downdrift in Akan as a rule which applies within each phrase, relating the value of later tones in the phrase to the value of the first. However, they point that "there is also a kind of downdrift within the sentence as a whole, such that the pitch of the first [-tone] (L) or [+tone] (H) segment of each successive phrase is somewhat lower than that of the first similarly-valued segment of the preceding phrase". Thus the overall pattern is as sketched in Figure 16B. We would analyze this pattern as involving declination over the whole sentence, with downdrift within each phrase superimposed on the overall pattern of declination.

There are also important differences between English downstep and downstep as it has been observed in tone languages. First, English downstep is conditioned by the morphological organization of the intonation; it takes place in sequences of the form H+L H and H L+H, but not in other alternating tonal sequences. Tonal organization comparable to the organization into pitch accents appears to be lacking in African tone languages, and therefore plays no role in tonal implementation. Second, English appears to use pitch range expressively within the phrase to an extent which is not paralleled in African tone languages. Downstepped tones are themselves subject to expressive variation in level. Thus one of the problems which English presents is how to describe the interaction of relative prominence and downstep in controlling tonal value.

In view of these similarities and differences, it is interesting to compare the formulation proposed here for downstep in English to the formulations which have been proposed for African tone languages. There have been two major groups of proposals. One, exemplified by Schachter and Fromkin (1968), Fromkin (1972), Peters (1973) and Meyers (1976), generates downstepped sequences through a process which applies iteratively left to right, as here. A second approach, proposed in Clements (1980) and Huang (1979), generates downstep from a hierarchical representation similar to the metrical trees discussed in Chapter 2. Our reasons for selecting iterative rule application over the hierarchical representation are discussed in Section 9. However, we share with Clements and Huang the view that tonal values are determined relationally. The difference is that in our system, these relations are defined locally on tones near to each other while in Clements and Huang, they are represented hierarchically. Our approach contrasts both groups of formulations in having no level of representation between the underlying sequence of tones and the FØ contour. In the other approaches, a level of representation is proposed which shares with the phonetic representation the property of encoding steps overtly, while sharing with the underlying representation the property of being invariant for different speakers or choices of pitch range. All authors presuppose but do not develop phonetic rules which compute an FØ contour from the intermediate representation. Our reason for dispensing with such a level of representation is that in all proposals it encodes how many downsteps have preceded any given downstep. There is nothing to prevent the phonetic rules from mapping use of this information in arbitrary ways, and so the form of the FØ contours generated by downstep is quite unconstrained. In our system, this information is not available; each downstep is computed as a purely local relationship. The only possible outcome is an exponential decay asymptotic to the baseline. Section 4 will present evidence that this outcome is the correct one for English.

4.2 The Qualitative Behaviour of the Downstepping Accents

Figures 1 through 8 introduced the four bitonal accents which trigger downstepping: H^-+L*, H*+L^-, L*+H^-, and L^-+H*. Of these, all but H*+L^- occur transparently in the FØ contours. In Figure 1, for example, the target level on "med" in intermediate is low relative to that on the metrically weaker syllable to its left, and so we infer a H^-+L* accent. In Figure 3, "intermediate" has the familiar FØ contour resulting from a L*+H^- accent, with a relatively low target on the stressed syllable followed by a higher target on the subsequent metrically weaker syllable. Figure 4 shows the same basic pattern, but shifted over. Here, the higher target is on the stressed syllable[2], and the lower target precedes so, the accent must be L^-+H*. In Figure 2, the unstarred tone of the pitch accent does not show up in the same obvious way; we see a gradual fall from one accented syllable to the next instead of the abrupt drop followed by a plateau which we would expect for H*+L^-. The argument for analyzing this contour as arising from H*+L^- pitch accents is thus of the form "Fit the only remaining peg into the only remaining hole". It is clear that the overall shape is that of a downstepped contour. Given Clements and Ford's result, this implies that the underlying representation has alternating tonal values; the only possible non-alternating description, H* H* H*, is in any case already used up for a different type of intonation. Our claims about the character of pitch accents as well as the clear desirability of generating Figure 2 with the same downstep rule which applies in 1, 3, and 4, thus force us to look for a characterization of 2 using a bitonal accent with opposite valued tones. H^*+L^- is the only such accent which is not already in use. Below, we will see additional evidence for the existence of the H^*+L^-, and we will show how the surface non-appearance of the L^- can be described in the present framework.

The topic of this section is the qualitative features of English intonation patterns which are related to these four accents and the formulation of the downstep rule. Specifically, we will show that H+L downsteps any following H and that the H in L+H is downstepped after any preceding H, as rule 8) states. We will show that downstep does not occur in other alternating tonal sequences, and that the H+L and L+H accents occur distinctively even in non-downstep contexts. The section ends with an account of the fate of L^- in H^*+L^-.

It is clear from Figures 1, 2, 5 and 6 that a H+L pitch accent downsteps the H in a following H+L accent. To establish that rule 8) is correct, it is necessary to establish that H+L also downsteps other H's. Figure 17 shows an example in which H^-+L^* downsteps a following H*. The nuclear H* on "Ebenezer" is downstepped to the level of L* on "lieve" in "believe", and shows up clearly in the FØ contour because the pitch level shared by the two tones is maintained well into "nez" before the fall to the L^- phrase accent begins. For comparison, Figure 18 shows an FØ contour in which "Ebenezer" has a H^-+L^* accent instead of H*. Here, the FØ is already falling on "be", and it falls throughout "nez". Figure 19 shows a contour in which H^-+L^* has downstepped the H^- phrase accent. (There is a L% boundary tone, which is upstepped to the level of

H^- by rule 9). The result is a contour in which the pitch level on the nuclear stress is maintained until the end of the phrase. This level is well above the baseline; the contour shows levels out at about 190 Hz, or about 50 Hz above the terminal baseline point for this speaker. This contour may be compared to the example of H^-+L* L^- H% shown in Figure 20, where it is clear that L^- is lower than the L* in the nuclear accent, as rules 3) and 5) imply. A prediction of the two tone theory is that there is no third distinctive level for the phrase accent after H^-+L*, higher than that in Figure 19. And in fact, the type of contour shown in Figure 21 seems to be impossible.

Downstepping a H^- phrase accent after H*+L^- results in the vocative contour illustrated in Figures 12 and 22. This contour is characterized by a fall from a peak on the nuclear stress partway to the baseline. The extent of the fall in Figure 22 may be compared to that in the H* L^- L% shown in Figure 23. In previous accounts, the vocative contour has been described as a H M sequence (Liberman 1975; Leben 1976). The introduction of the downstepped H makes it possible to dispense with the M tone. The claim that the phrase accent is H^- also interacts with the upstep rule to predict the possible levels for the boundary tone in vocative contours. As Figures 22 and 24 indicate, a vocative contour can either level out above the baseline, or else end in a rise. The level of the downstepped H^- phrase accent in Figure 24 may be compared with that of the L^- phrase accent in the H* L^- H% contour shown in Figure 25. There is no variant of the H*+L^- H^- contour with a fall to the baseline at the end. Just as in the question intonations discussed in Chapter 2, we take L% to be the boundary tone in the contours which end level, and H% to be the boundary tone in the contours which end in a rise. The two tone theory and the upstep rule together explain why no variant which falls at the end is found.

The existence of the H*+L^- accent with its non-appearing L^- is crucial to our explanation of Figures 22 through 25 as contours generated by downstepping a H^- phrase accent. The nuclear accent in these contours could not be H^-+L* or L*+H^-, because these would produce a local minimum rather than a peak on the nuclear stressed syllable. L^-+H* is not possible, because we know independently that L^-+H* H is not a context for downstep.

Turning now to downstep in the L+H accents, we observe that Figures 3, 4, 7, and 8 show that the H in such an accent is downstepped if the preceding accent is of the same type. In Figure 8, we also see that H* in L^-+H* is downstepped after a H* accent. The F0 contour on "believe Ebenezer" in 14 makes the same point for H* L*+H^-. As stated, rule 8) predicts downstep in the context H% L+H. The contour in Figure 26 appears to confirm this prediction. It would also seem plausible to interpret the contour in Figure 27 as involving downstep in H% L*+H^-, particularly in view of the fact that the F0 value at the H^- is raised by the /k/ in "remarkably". (/b/ was also partially devoiced in this utterance, so the F0 is also raised at its release). These observations are presented with some caution, however, because we do not have much data on the scaling of H%; we lack the basis of comparison which the scaling of the H* accents provides for downstep triggered by H in a pitch accent.

A conspicuous contrast between cases of H^-+L* H and cases of H L+H is that the former has total downstep, with H on the right lowered to the same level as the preceding L for the same prominence, while the latter has partial downstep, with the H higher than the preceding L even after downstep. (In H*+L^- H, the lack of L^- on the surface makes this distinction moot.) We have opted to describe this difference by formulating the downstep rule to compute the same relation between /H_i/ and /H_j/ in H_i L+H_j and H_i+L H_j, and then treating the L's separately. In H_i+L H_j, /L/ is related to /H_i/ by the same coefficient k which related /H_i/ and /H_j/; in H L+H, /L/ comes out lower. Our reason for taking this approach is that the exponential decay in a downstepped sequence seems to occur at the same rate whether the pitch accents involved are H+L or L+H. This point can be seen by examining Figures 1 through 4, and also by comparing Figures 5 and 13. If the contrast between total downstep in the H+L cases and partial downstep in the L+H cases were handled by computing the value of the downstepped H as two different functions of the value of the preceding L, this regularity would be purely coincidental.

One might consider explaining the contrast between total downstep in H^-+L* H and partial downstep in H L+H by proposing that /L/ is in both cases related by the factor k to the /H/ for H in the same pitch accent. This proposal gives the right qualitative result, since ignoring prominence differences, the output for H^-+L* H would be /H^-/ k/H^-/ k/H^-/, while for H L+H it would be /H/ k^2/H/ k/H/. However, the proposal is not quantitatively correct. It predicts that in sequences of the form

H L_i+H L+H_j, /L_i/ = /H_j/. In Figures 3, 4, 7, 8 and 13, we see that /L_i/ < /H_j/ in sequences of this sort. This means that the constant relating /L/ and /H/ in L+H, or n/k in the present statement of the rules, is less than k.

So far, we have reviewed cases where rule 8) predicts that downstep would occur, and it does. The correctness of the rule also needs to be supported by supplying cases where the rule predicts downstep will not occur, and it does not. Figures 27 and 28 show two FØ contours with H L H tonal configurations which could in principle trigger downstep but do not because they are not organized into pitch accents in the manner specified by the English downstep rule. In Figure 27, the H* on "suggestion" is not downstepped in the sequence L*+H^- L* H*. In Figure 28, we do not find downstep in H* L* H*. Rule 8) also predicts that we would not find downstep in sequences of the form H% L* H*. With the same caveats about the scaling of H% as before, this appears to be the right interpretation of Figure 29.

All four of the downstepping accents also occur distinctively in non-downstep contexts. For instance, all cases of L*+H^- in Chapter 2 were not in downstep contexts. Figures 30 and 31 illustrate the contrast between L^-+H* and H* after L*; in L* H*, there is a gradual rise from L to H whereas in L* L^-+H*, a level stretch in the FØ contour is followed by a sudden rise. Figure 20 provided one example in which H^-+L* occurred distinctively in a non-downstep context, after L*. H^-+L* is also distinctive after H*; it shows up either as a corner in the FØ contour, as in Figure 1, or if it has sufficient relative prominence, it shows up as a peak, as in Figure 32. Even though the L^- in H^*+L^- is not evident in the FØ contour, this accent still contrasts with H* before L+H. The contrast arises because downstep applies in H L+H but fails to apply in H+L L+H. An example illustrating the lack of downstep in H^*+L^- L+H is shown in Figure 33. This FØ contour may be compared to a contour with downstep for the same sentence, shown in Figure 34.

Our transcription of Figure 33 may seem counterintuitive, since it means that before L+H, H* and H^*+L^- exchange the descriptive roles they had before H. However, the hypothesis that H^*+L^- is the trigger for downstep both for H and for L+H is quickly rejected by considering Figures 3 and 4. Taking H+L L+H to be the structural description for downstep would make it impossible to generate these contours, which transparently exhibit downstep in a L+H L+H sequence. The structural description H(+L) L+H would generate these contours, but would wrongly predict downstep in L+H after both H+L and H, leaving no way to generate Figure 33. Thus the conclusion that downstep applies in H L+H sequence and that Figure 33 represents H+L L+H still stands.

We have noted that the L^- in H^*+L^- fails to appear in the FØ contour after triggering downstep of the H to the right. Similarly, in Clements and Fords' (1979) account of downstep, floating tones fail to appear on the surface after triggering downstep. These proposals are on the surface inconsistent with the claim that tonal implementation rules apply iteratively left to right, and cannot change anything computed on the last iteration. To see why this is so, consider the situation for $H^*_i+L^-$ H^*_{i+1} at the beginning of the iteration on which $/H^*_{i+1}/$ is to be computed. L^- must still be present in the representation at this point in the derivation, because downstep of H^*_{i+1} would fail to occur if it had already been deleted. If L^- is present, then its value was computed on the last iteration, and an interpolation between this value and that of the preceding H^*_i was constructed. Thus deleting L^- on the H^*_{i+1} iteration entails deleting not only a computed target value, but also a stretch of FØ contour.

The theory described here leaves an out for handling this case which still leaves in place the general restriction. Since, as we have suggested, the value of downstepped H's in English appears to be referenced to the value of the H in the preceding pitch accent rather than that of the intervening L, the only possible evidence that L was evaluated at all would be its occurrence on the surface with some value. Such evidence is exactly lacking in the case of H^*+L^-. This observation opens the way to an account which is consistent with our general claims about the form of the theory, and actually involves fewer rules than an account with L^- deletion. Specifically, we propose that rule 3) above be restricted to apply only to H^-+L^* and not to H^*+L^-:

10) In H^-+L^*: $/L^*/ = k/H^-/$

As a result, no value for L^- in H^*+L^- can be computed on the iteration on which it is the target tone. Since the interpolation rules necessarily refer to two target values between which an interpolation is to be constructed, it follows that no interpolation between H* and L^- is constructed on this interation. On the next iteration, $/H^*_{i+1}/$ is computed with reference to $/H^*_i/$ and L^-, but without reference to $/L^-/$.

At this point, the target value preceding $/H^*_{i+1}/$ is $/H^*_i/$, and an interpolation between these two H targets is constructed, just as between two H*s. The large difference in levels generally means that the interpolation is monotonic but nothing excludes sagging between sufficiently separated targets. The failure of L^- to appear in the FØ contour thus follows without the necessity for a deletion rule.

This type of solution is also available for the tone languages described in Clements and Ford (1979), in which a floating tone triggers downstep of a following tone but fails to appear on the surface. There are two alternatives, depending on whether regularities in the phonetic output are found to require that evaluation of the downstepped tone make reference to the phonetic value of the floating tone, or only to its phonological value. In the latter case, the situation is the same as that found in English. We can propose that a target value is not computed for the floating tone: as a result, the interpolation rules fail to construct an FØ contour between it and the tone on either side. It is skipped over when an interpolation is constructed between the tone to its left and that to its right. The floating tone affects the FØ contour only indirectly, through its effect on the phonetic value of the downstepped tone. If the system must refer to the phonetic value of the floating tone, the situation is not identical to that found in English but can still be handled in the present framework. Since the interpolation rules construct an FØ contour as a function of the location in time and phonetic value of two tones, the rules would also skip over a tone which had no location in time. Floating tones by definition are not associated with the text, and so the interpolation rules would skip over them whether or not they had a phonetic value. Either of these alternatives is a more precise version of Clements and Foud's proposal that floating tones are not transmitted to the articulatory component. However, we do not treat the L^- in H^*+L^- as a floating tone, because the unstarred tones in English are in general assigned a location in time, and the only ways of excepting the L^- from this assignment appear to be ad hoc.

Our treatment of these cases in which a tone fails to appear after triggering downstep is not equivalent to a theory in which tone deletion rules in general may be ordered after phonetic rules. Our treatment predicts that only limited context could be relevant to the non-appearance of the tones in question. If tone sandhi rules proper can make more extensive use of context, then it would also be possible for a language to have tone deletion rules which are not limited in this way. Such deletion rules, however, would presumably be ordered amongst the phonological rules rather than among the rules which map the phonological representation of tone into a phonetic representation. The constraints on tone sandhi rules have not been a topic of discussion here, because as far as we know English does not have any.

4.3 Left-To-Right Tonal Implementation in Other Contexts

In the case of downstep, the exponential character of the result strongly suggests that an iterative process is at work. In this section, we present the facts which lead to the inference that the values of all tones, whether downstepped or not, are computed in relation to the immediately preceding tonal values. These are the facts which motivate rule 2) for H tones and rules 3) through 6) for L tones. The first observation is that the value of a H tone is carried through onto a following H tone of equal prominence, with differences in prominence scaled accordingly. (The value of the following H may then be readjusted by downstep). One consequence is that prominence relations are reflected in FØ values in the same way, whether or nor the first tone was itself downstepped. The second observation is that the value of L tones is also computed in relation to the values of immediately preceding tones. However, the L in a bitonal accent, the L^- phrase accent, and L% are all related differently to a preceding H; this means that the value of successive L tones can differ even without a change in prominence.

The situation in English may be compared to that reported for African tone languages with downstep or downdrift. As Figure 16A and 16B indicated, the value of a downstepped H is carried through on subsequent H's until the next downstep. The value of the first L following the downstepped H is also shared by subsequent L's until the next downstep. (Huang 1979; Clements 1980; Meyers 1976; Schachter and Fromkin 1968). Such a system of implementing a bitonal distinction is redundant compared to that found in English. A paradigmatic distinction between two tones can be conveyed if the value for one is consistently different from the value for the other at the same location. This does not require that the amount of the difference be the same across locations. Thus additional information can be carried by varying the amount of the difference. In English, the extra information carried is information about phrasing and prominence.

The end of this section reviews how the tone mapping rules discussed support the hypothesis that the left to right mapping of tones into tonal values has an increment of one tone and permits reference back as far as the previous pitch accent.

Figures 35 and 36 illustrate one consequence of rule 2), which computes the value of a H tone in relation to that of a preceding H tone. The transcription for the FØ contour in Figure 35 is H*+L$^-$ H*+H$^-$ H*. (The H*+H$^-$ accent, which is discussed in Chapter 5, scales like H* but generates an FØ plateau instead of a dipping configuration). The medial pitch accent in this sequence is downstepped in the context of the first. The nuclear H* accent is not itself in a context for downstep. Nonetheless, its phonetic value is still less than that of the less prominent H*+L$^-$. This is so because its phonetic value is computed in relation to the value of the accent to the left, without regard for whether that accent was downstepped or not. Thus, the prominence difference between the nuclear H* and the prenuclear H*+H$^-$ is implemented in the same way in Figure 35 as it is in Figure 36 where H*+H$^-$ was not downstepped. This means that the effects of downstep are carried over onto subsequent non-downstepped H tones.

In Figure 36 as in all patterns with H*+H$^-$ accents, /H$^-$/ = /H*/. On the assumption that both tones in the same pitch accent have the same prominence, this follows from rule 2).

We also proposed that rule 2) computes a value for H tones in downstep contexts, which downstep then readjusts. The rules were formulated in this way to capture the fact that downstepped tones are subject to variation on the basis of prominence in the same way that non-downstepped tones are. Figure 37 illustrates this point. The figure shows two FØ contours for H*+L^- H* L^- L% which share the same phonetic value for H* in H*+L^-. The values for the nuclear H* differ, because the word with the nuclear stress was somewhat more emphasized in one utterance than in the other. Our observations about the interaction of downstep and prominence are only qualitative, however. Further experimentation is needed to determine whether this interaction is really multiplicative, as rules 2) and 8) would predict.

In all our examples of a H^- phrase accent after a pitch accent ending in H, H^- continues the previous level. Thus, not only would it appear correct to compute the value of H^- by rule 2), but also it seems that the phrase accent is taken to have the same prominence as the nuclear accent. This is an interesting result, because the value of the H boundary varies independently from that of the nuclear accent according to how non-final the speaker views his utterence. In principle, it would be perfectly possible for the phrase accent to vary in a similar fashion. For H^- after H^-+L*, rule 2) sets the value of the phrase accent equal to that of the H^- in the nuclear accent. As a result, the phrase accent has the same value as the L* after downstep. This can be seen in Figures 19 and 38. Figure 38 makes the same point for H^-+L* H^- that Figures 35 and 36 made for H*+L^- H* H*. Here, the H^- in the nuclear H^-+L* has itself been downstepped because of the H^-+L* to its left. Rule 2) carries this downstepped value onto the H^- phrase accent, which then undergoes an additional downstep. If the value of the phrase accent were computed on the basis of the prominence of the nuclear accent, for instance, instead of taking its phonetic value from the nuclear accent, then it would not have downstepped exactly to the level of the L*.

According to rule 2), the H in a bitonal accent is related to other H tones in the same way that a H* is. We have already made use of this assumption in giving an account of Figure 38. A further prediction of this claim is that the step size for equally prominent accents is the same in H*+L$^-$ H* as in H*+L$^-$ H*+L$^-$, for example, and the same in H* L$^-$+H* as in L$^-$+H* L$^-$+H*. The relations between the contours in Figures 7 and 8, 13 and 14, and 17 and 18 appear to be in line with this prediction.

Rule 2) assigns equally prominent H tones equal phonetic value, whether they are in a bitonal accent or not. According to rule 7), equally prominent L*s are also assigned equal phonetic values. This behaviour is not shared by all L tones; instead, the value of the L tone is computed differently in different contexts, so that differences in phonetic value arise without a plausible source in prominence differences. Figure 20 provides one illustration of this point. The transcription for this contour is L* H$^-$+L* L$^-$ H%. Here, the L* of the nuclear accent is higher than the prenuclear L*. This situation arises because of rule 3), which scaled L* in H$^-$+L* as a fixed ratio of the phonetic value of H$^-$ without reference to preceding L tones. If the value of L tones were computed by a generalization of rule 7) similar to rule 2), which relates two like tones disregarding unlike tones in the same pitch accents, the outcome would be different; the L* in the bitonal nuclear accent would be equal to or lower than the L* accent, since L* accents become lower under increased prominence. The behaviour of sequences of H$^-$+L* accents under changes in overall pitch range also reflects rule 3). This behaviour is illustrated in Figure 39. When the pitch contour is increased, the entire F0 contour is raised proportionately; the value of the L*s as well as the H$^-$s goes up. This follows when /L*/ is computed as a ratio of /H$^-$/, and contrasts with the behaviour of L* when it is not in a bitonal accent.

Returning to Figure 20, one also notes that the L* of the nuclear H$^-$+L* is higher than the L$^-$ phrase accent. The situation thus contrasts with the cases of L* L$^-$ shown in Figure 40, where the L$^-$ is assigned the same phonetic value as the preceding L* by rule 7. Limited observations suggest that the L$^-$ phrase accent after H$^-$+L* is treated the same way as it would be after H*; its value is slightly above the baseline. Similar observations can be made for L*+H$^-$ L$^-$. These observations are covered by rule 5). We would predict, but have not confirmed, that the value of L$^-$ in these contexts increases with the value of the nuclear H tone, as it does in H* L$^-$.

Section 1 suggested that the tone mapping rules only refer back as far as the preceding pitch accent. The tonal value relations we have just reviewed all fit this hypothesis. The main rule propagating the value of H tones, rule 2) above, computes the value of H* or the H in a bitonal accent in relation to that of the H in an immediately preceding accent. When rule 2) is applied to a H$^-$ phrase accent, the immediately preceding accent is the nuclear pitch accent and it is to the nuclear accent that H^- is related. The downstep rule, 8), does not refer to phonetic values of tones in the preceding pitch accent, but it does refer to the tonal description of the pitch accent. H is downstepped only after H-L, and H in L-H, only after H.[3] Rule 5) for L^- after H(+T) computes L^- as a ratio of the value of the preceding accent, the nuclear accent. Rules 3), 4), 6) and 7) also look no further back than the previous accent.

Our hypothesis obviously permits the value of the boundary tone to be computed from the value of tones in the nuclear accent. It is not clear whether this possibility is used in the system. The rule L% = 0 is non-relational; if H% after H^- is handled by rule 2), its underlying value is computed in relation to the phrase accent rather than the nuclear accent. The upstep rule for L% and H% after H^- is likewise insensitive to the character of the nuclear accent. Thus the only case in which there is a real possibility of the boundary tone being mapped with reference to the nuclear accent is H% after L^-. The FØ contours gathered for the experiment in Chapter 3 showed a strong relationship between H% and H*. This relationship is shown for a typical subject in Figure 41. It could arise from a rule computing /H%/ as a function of /H*/. However, it could also arise indirectly from $/L^-/ = m\ /H^*/$ and $/H\%/ = n\ /L^-/$, given that $/L^-/$ shows $m \neq 0$. If the relationship does arise in this fashion, when we would expect H% to scale the same way after L* L^- as after H* L^-. We do not know if this is the case, since that we have no data relating scaling of H after L to its scaling after H, with prominence equated.

The increment for iteration of the tone mapping rules was hypothesized to be one tone. Furthermore, we stipulated that tonal values computed on one iteration cannot be changed on the next. A two tone increment would in most cases give the same power as a one tone increment with the permission to recompute the tone preceding the current tone. However, in sequences of $T_i^* \;\; T_{i+1}^* \; T_{i+2}^*(-T)$, using a two tone increment would be inconsistent with our hypothesis that tone mapping rules only refer to the last accent: when i was incremented to i+2, the last accent would be T_{i+1}, whose value is not yet computed. This situation would leave no basis for the computation of $/T_{i+1}/$ or $T_{i+2}/$. This argument can obviously be extended to rule out that possibility that the increment is any greater.

4.4 The Scaling of Downsteps

According to the downstep rule formulated in Section 4.1, successive H values in a downstepped sequence are related as follows:

11) $H_{i+1} = k\ /H_i/ \quad 0 < k < 1$ (indexing here omits L tones)

This means that the total sequence of values can be described as $V(k^n)$ where $V = /H_1/$, n is the index, k is the downstep coefficient, and the value of tones is expressed in baseline units above the baseline.

This section reports the results of a pilot experiment to test the predictions of this formulation of downstep. The model predicts that the size of the first step is constant, regardless of the number of steps in the phrase. This follows because the model has no look-ahead. It also predicts that the phonetic value of the nuclear accent is lower, the longer the phrase. Since the step size is claimed to be constant as measured in baseline units above the baseline, it should decrease as measured in Hz or semitones.

Two alternative hypotheses will be entertained. One is that downstep has a look-ahead mechanism, which the speaker uses to divide his total range by the number of steps. This hypothesis predicts that the size of the first step is smaller, the longer the phrase, but that the phonetic value of the nuclear accent is invariant, for a given /H_i/. A second possibility is that the steps are computed by local recursive rules, as proposed, but not in the space of baseline units above the baseline. Under such a hypothesis, the step size would be invariant with length of phrase and position in the phrase, but only when expressed in Hz or semitones rather than in baseline units above the baseline.

The results of the pilot experiment confirmed our porposal over both of the alternatives.

The subject for the experiment (MB) recorded the following sentences, with the indicated intonation pattern. There were twelve or more repetitions of each,

12) I really believe him.

H* H^-+L* L^- L%

13) I really believe Ebenezer.

H* H^-+L* H^-+L* L^- L%

14) I really believe Ebenezer was a dealer.

H* H^-+L* H^-+L* H^-+L* L^- L%

15) I really believe Ebenezer was a dealer in magnesium.

H* H^-+L* H^-+L* H^-+L* H^-+L* L^- L%

The sentences were constructed in this fashion so that steps at a given location would be comparable across sentences. Also, all accented syllables had a high front vowel in order to minimize the effect of differences in vowel height on FØ, so that earlier and later steps in the same sentence would be as comparable as possible. The sentences were written on cards and randomized together, with additional sentences that will not interest us here interspersed. The subject was trained to produce the desired intonation pattern using phrases that had only one step (H* H^--L*), and was quite successful in extending this pattern to longer phrases. Four utterances were eliminated from analysis because of misplacement of nuclear stress and a number were also eliminated because they did not have the desired H^-+L* accents. However, there were at least 10 correct utterances for each sentence type. FØ values were measured at the FØ peak in the case of the H*s, and at the amplitude maximum of the vowel in the case of the L*s in H^-+L*s. (The L*s of course do not have a peak on the accented vowel).

The results of the experiment are summarized in Figure 42. Each data point is the mean of the FØ values for the given word in the same sentence. Data points from words in the same sentences are connected by lines. Vertical bars indicate the standard error for each data point.

A visual examination of the graph confirms our downstep rule over the look-ahead hypothesis. The size of the first step is quite insensitive to phrase length[4], and the nuclear accent is lower, the longer the phrase. The step size within a phrase is clearly not constant in Hz, although visual examination does not make it clear whether it might be constant in semitones.

More detailed analysis is necessary to determine whether exponential character of Figure 42 can be generated by a constant ratio in baseline units above the baseline, as proposed above, or whether it involves a constant musical interval, as some researchers have claimed for downdrift in African tone languages. Table I gives the intervals for each step shown in the graph.

Clearly, the steps are not equal in this domain; the step size decreases as the phrase progresses. To make our case, we have to show that a baseline units model can be fit in which the step does not decrease in this way. Table II gives ratios of each accent to its predecessor, in a hypothesized domain of baseline units above the baseline. The hypothetical baseline falls from 150 Hz to 140 Hz; the terminal value was estimated from declarative FØ contours for this subject, while the initial value was arrived at hit or miss. This transform removes the trend observed in Table I. As in Table I, the third entry in the bottom row is an outlier; otherwise, early values for step size are comparable to later values.

It is important to note in comparing Table I and Table II that the amount of variability in the step size values is not a measure of relative goodness of fit of the two models. An intelligible measure of relative goodness of fit is the comparison between deviations of observed values from values predicted by the two models. These deviations are tabulated in Tables III and IV, on the following pages. Table III shows the deviations of the data observed from the values predicted by the equal

Table I: <u>Step Size in Semitones</u>

Sentence	believe/really	Ebenezer/believe	dealer/Ebenezer	magnesium/dealer
12)	-6.4			
13)	-5.2	-4.7		
14)	-5.7	-4.1	-3.4	
15)	-5.9	-3.9	-1.6	-2.4

Table II: <u>Step Size as a Ratio of Estimated Baseline Units above the Baseline</u>

Sentence	believe/really	Ebenezer/believe	dealer/Ebenezer	magnesium/dealer
12)	.57			
13)	.60	.54		
14)	.58	.61	.54	
15)	.55	.61	.80	.62

Table III: <u>Deviations of Observed Values from Predicted Values for the Equal Interval Model</u>

A - In Semitones

Sentence	believe/really	Ebenezer/believe	dealer/Ebenezer	magnesium/dealer
12)	1.9			
13)	1.0	1.5		
14)	1.3	-0.3	-1.1	
15)	1.6	1.1	-1.6	-3.5

Mean absolute deviation 1.5

B - In Hz

Sentence	believe/really	Ebenezer/believe	dealer/Ebenezer	magnesium/dealer
12)	30.7			
13)	16.5	18.3		
14)	21.9	-4.0	-11.5	
15)	27.2	14.9	-17.9	-32.4

Mean Absolute deviation 19.5

Table IV: Deviation of Observed Values from Predicted Values for the Baseline Model

A - In Semitones

Sentence	believe/really	Ebenezer/believe	dealer/Ebenezer	magnesium/dealer
12)	0.4			
13)	0	0.6		
14)	0.3	0.1	0.5	
15)	0.7	0.4	-1.1	-0.9

Mean absolute deviation 0.5

B - In Hz

Sentence	believe/really	Ebenezer/believe	dealer/Ebenezer	magnesium/dealer
12)	6.0			
13)	0	7.4		
14)	4.8	1.3	4.9	
15)	11.3	5.4	-12.1	-8.6

Mean absolute deviation 6.2

musical interval model. The value for "really" in each sentence was taken to be the mean of the observed values for that sentence. Subsequent values were computed taking the step interval to be the mean of the observed intervals, 4.34 semitones. The mean absolute deviation of the observed data from the predicted values was 19.5 as computed in Hz and 1.5 as computed in semitones. As one would expect from examining Table I, taking the step size to be a constant musical interval resulted in predictions which were too high for the left side of the table, and too low for the right side. Table IV shows the deviations of the data observed from the values predicted by downstep rule 8), presupposing a baseline running from 150 Hz to 140 Hz. The step size was taken to be 0.60, the mean of the observed steps measured as a ratio of baseline units above the baseline. The mean absolute deviation is 6.2 in Hz, and .49 in semitones. Thus, even a rough estimate of the baseline enables the downstep rule operating in the domain of baseline units above the baseline to fit three times better than the constant interval model. Our downstep rule has the further advantage that the predicted value of the downstepped accents is asymptotic to the baseline. Thus, the model automatically avoids generating downstepped values below the bottom of the speaker's range. In the constant interval model, the predicted values are asymptotic to 0 Hz. Additional assumptions must be introduced to describe what happens when the predicted values fall below the baseline and are not physically realizable.

Given the promising results of this pilot experiment, a full scale experiment with more careful controls seems justified. In the projected experiment, the estimate of the baseline for the stepping contours will be arrived at by optimization, as in Chapter 3, rather than by trial and error. Since the stepping intonation patterns differ from those examined in Chapter 3 in not providing an estimate of the baseline which is independent from the coefficients of the fitted model, additional contours will be collected for this purpose.

It would be interesting to know whether downstep and downdrift in African tone languages exhibit the same phonetic properties that were found here for English. Meyers' (1976) study of Hausa found that the step size as measured in Hz decreases through the phrase. "At the beginning of an utterance, within a given word, the distance between a high tone and a following low tone may be as great as 20 herz, while towards the end of the utterance the distance between a high and a low tone, even within the same word, may be only 2-3 herz" (85-86). This report obviously rules out the possibility that the step size is constant in Hz, and it also suggests that the step size cannot be a constant musical interval: If the step of 2 Hz is a fall from 100 Hz to 98 Hz, then the step of 20 Hz would have to be a fall from 1000 Hz to 980 Hz in order to represent the same musical interval. This is an implausibly large range: for comparison, the total range of tonal values in Silverstein's study of Hausa was 97 Hz to 186 Hz.

Painter's (1979) study of Gwa is also addressed to the question how step size is related to serial position. However, egregious errors in data processing make the results less worthwhile than they might have been. The primary error was in averaging together all measurements for the n^{th} high tone in a stepping sequence, regardless of the total number of steps in the sequence in which the measurement was taken. Since the sequences were not all of the same length, this means that the number of measurements per mean value presented decreases as n increases. Table V presents a hypothetical set of measurements which shows one reason why this procedure is misleading: in this table, the measurements for each sequence are descending and all steps are the same size but these equalities are completely lost in the averaged data.

Some situation of this sort must have arisen in Painter's data, because the highest FØ value reported in the results for terraced sequences is for the 9th tone, for which there is only one measurement.

Painter made an additional error in collapsing sequences of H !H (H downstepped H) with nonterraced H sequences rather then with sequences in which H downdrifted after L. The comparison one is interested in is between stepped sequences and sequences with declination but no stepping. By grouping the downstepped H sequences with the plain H sequences as "non-terraced spans", Painter makes it impossible to make this comparison. He points out this problem himself, but does not suggest any purpose for which averages based on his grouping of the data would be useful.

4.5 Upstep

The discussion of the phrase accent and the boundary tone in Chapter 2 pointed out that the boundary tone is upstepped after H^-. The magnitude of the difference between L% and H% after H^- appears to be comparable to that after L^-; but L%, whose value is 0 after L^-, has the value of H^- after H^-, and H% is accordingly higher. Since there are only

Table V: A Hypothetical Set of Data Showing Why Painter's Averaging Procedure is Inadvisable

	H Tones			
Sentence	First	Second	Third	Fourth
1)	300	285	270	255
2)	250	235	220	
3)	230	215		
Means	260	245	245	255

two tones in the system, there is no third option of a boundary tone on or near the baseline after a H^- phrase accent. Figure 43, repeated from Chapter 2, illustrates this observation.

As in the case of downstep, partial similarities between upstepped and non-upstepped boundary tones suggest that upstep be formulated as a readjustment rule. A preliminary statement of this value is given in Section 1 as rule 9). When T is L%, the old value is 0 and so the new value is $/H^-/$. When T is H%, the old value can be anywhere in a fairly large range, given that /H%/ varies as an expression of how non-final the utterance is. As a result, the upstepped /H%/ is also quite variable.

The fact that upstep, unlike downstep, has an addend follows from the qualitative features of the patterns observed. Upstep readjusts a tone whose value is 0 to have the non-zero value of the preceding tone. It would be impossible for a pure ratio rule to do this, since $k \cdot 0 = 0$. Under upstep, no tonal values are readjusted to a value less than that of the preceding tone. Under downstep, a tone whose value is 1/k times the value of the preceding tone is readjusted to match the preceding tone and tones with lower or higher values come out lower or higher than the preceding tone. The downstepped value of a hypothetical H tone with value 0 would be 0.

Because downstep generates an exponential curve asymptotic to the baseline, downstepped sequences can in principle be of indefinite length. Clearly, the functional form of our upstep rule incorporates no comparable prediction. Given that our system for scaling F0 values is referred to the bottom of the speaker's range rather than the top, and that the upstep function has no asymptote, nothing in principle prevents the upstep rule from generating values higher than the top of the speaker's range. What prevents this in practice is the fact that one's normal speaking range is nearer the bottom than the top of one's maximum range, and that the context for upstep is not met iteratively. The existence of tags which undergo upstep makes it possible to get utterances with two successive upsteps, like that in Figure 44. However, we do not seem to find sentences with more than one tag. For example, both 16) and 17) are possible:

16) He said he was sorry, didn't he?
H* L^- H^- H%

17) He said he was sorry, Benjamin.
H* L^- H^- H%

18) , in contrast, seems rather odd:

18) He said he was sorry, didn't he, Benjamin?
H* L^- H^- H^- H%

A much more natural version collapses "didn't he, Benjamin" into a single tag:

19) He said he was sorry, didn't he Benjamin?
H* L^- H^- H%

If this general observation is correct, it means that the structural description for upstep cannot be met an indefinitely large number of times, as that for downstep can.

Whether any languages have iterative upstep is an interesting question. Clements (1980) and Huang (1979) cite a number of languages which are reported to have upstep, but it appears that in many of these the upstep is not iterative. Huang reports that upstep frequently occurs in anticipation of downstep. In such cases, a rising and falling rather than a monotonically rising pattern would be generated, and so problems of exceeding the speaker's range would not arise. Discussion in Clements (1979) suggests that some languages which are reported to have upstep display what we would call reinitialization of tonal values. This is the change in pitch range which can occur at an intonation phrase boundary. There are two sources for such a change: one is baseline resetting, with its consequences for the scaling of tones above the baseline. The other is the expressive use of pitch range. For instance, the speaker may select a reduced pitch range for a parenthetical remark, and resume a larger pitch range when he picks up the main topic again. After a phrase with a series of downsteps, reinitialization to something like the initial pitch range is typical. As our terminology suggests, reinitialization arises from a fresh choice of values for tones, rather than from a rule relating the value of tones on the right to the values of their predecessors. Tonal rules proper, such as downstep and upstep, appear to apply only within the phrase. Given that phonological rules in general apply across weaker boundaries and become blocked by stronger ones, it would be surprising to find an upstep rule which applied only across intonation phrase boundaries.

If iterative upstep does exist, further instrumental investigation will be needed to determine whether it can be accommodated within the framework described here. The framework does not automatically rule out iterative upstep, since a straightforward generalization of the functional form for downstep can generate both downstep and upstep. A possible general recursive formula for subsequent upstepped or downstepped tones is:

20) $/T_{n+1}/ = (1-a)\ /T_n/ + at \qquad (0 < a < 1)$

(This is a reexpression of $/T_{n+1}/ = a/T_n/ + b$ which brings out the asymptotic value, t.) The form this takes as a readjustment rule if recursion is carried out by independent rules such as rule 2) is:

21) $/T/ = (1-a)\ /T/ + at$

The rule generates downstepping sequences if $/T_1/ > t$, and upstepping sequences if $/T_1/ < t$. Our downstep rule corresponds to the case where $t = 0$ and the addend is therefore 0; perhaps all downstep rules have $t = 0$. In the case of upstep, t is the ceiling of the pitch range for the phrase as a whole. t need not correspond to the absolute top of the range; in fact, given our earlier observation that the speaker's typical speaking range is in the lower part of the maximum range, one might expect that values of t lower than the absolute maximum would be selected. The curve sketched in Figure 45 shows an upstepped sequence generated by this formula on the assumption that t is 3 baseline units above the baseline. Given these observations, we cannot agree with Huang's suggestion that the existence of monotonically upstepping patterns which do not tax the top of the speaker's range would necessarily be evidence for preplanning. If such patterns show asymptotic behaviour, they can be generated without advance planning for the allocation of pitch range over the length of the phrase. This would mean that we do not need a phonological representation which permits such preplanning. However, the existence of iterative upstep would require loosening of one of our assumptions; it requires reference to a ceiling whereas all previous rules have referred only to a floor, the baseline. On the assumption that the ceiling is defined in baseline units above the baseline, its tilt would not be subject to linguistic variation. One would want to investigate, however, whether its level was varied.

A comparison between rule 9) and the general formula 21) shows that 9) is not an instance of 21). The case where the underlying value of T is 0 shows that a_t would have to be $/H^-/$; but then the case where it is greater than 0 shows that 1-a would have to be 1, hence a would have to be zero. This is an inconsistency. However, our data on upstep in questions is sparse enough to permit an alternative formulation which is an instance of 21). Taking $a_t = /H^-/$ and assuming t = 3, this would be:

22) In H^- T: $/T/ = (1-\frac{/H^-/}{3})\ /T/ + /H^-/$

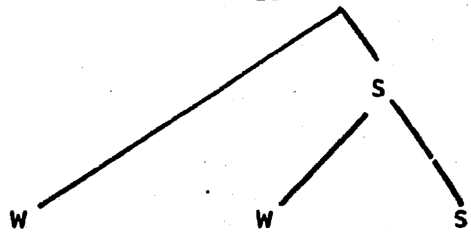
$0 < a < 1$ on the reasonable assumption that $/H^-/ < t$. This rule, unlike 9), predicts that the difference between L% and H% is reduced when they are upstepped. More detailed investigation is needed to determine whether this is correct. A second consequence of 22) is a dilemma: either the rule predicts that $/H\%/ - /H^-/$ becomes smaller, the greater $/H^-/$, or the theory must permit t to increase with $/H^-/$.

4.6 Does an Upstepping Pitch Accent Exist in English?

In the past sections, we have discussed a downstep rule which applied iteratively to two-tone pitch accents, and an upstep rule which applies to boundary tones. We have not seen in English a situation which is a logical possibility, that of a pitch accent which upsteps iteratively, creating an ascending staircase in the FØ domain. There is reason to examine this possibility carefully, since Palmer (1922) discusses a "scandent head" which might correspond to it, and this suggestion is taken up in Crystal (1969), where the type E rising head "comprises a rising series of stressed syllables, with or without intervening unstressed syllables, each stressed syllable being higher than, or occasionally at the same pitch as the preceding pitch-prominent syllable". (p. 230). However, the system we have outlined already generates a number of FØ patterns which rise up to the nuclear accent. These patterns involve rising prominence or L H tone sequences. We conclude that there is no motivation for an additional pitch accent which engenders upstep.

Chapter 2 gave examples of FØ contours with two H* accents, in which the nuclear stress on the second meant that the second peak was higher than the first. While altering prominence configurations in general are preferred for phrases with more pitch accents, it is also possible to find instances of monotonically ascending prominence in particular contexts. With such a prominence configuration, a series of H tones is then mapped into a rising FØ contour. Figure 46 is an example of such a contour. The metrical structure for the sentence in this figure is:

23)



It's indexed by the keywords in the abstract

The speaker's decision to emphasize "keywords" results in it being more prominent than "indexed". "Abstract" is in turn more prominent than "keywords". The rising FØ contour shown for 23) does not actually have H* accents, but rather the H*+H^{-} accent which will be introduced in Chapter 5. The H*+H^{-} scales in the same way as the H*, but generates FØ plateaus instead of dipping between the accented syllables.[5]

A second major source of rising FØ contours is tonal representations with a L* or L*+H^{-} early in the phrase. A contour with the accents L* H* H* is shown in Figure 47. This tonal sequence generates a three level rise even if the first accent is on a more prominent syllable than the second. The contour may be compared to the L* H* intonation pattern for the same sentence shown in Figure 29. Figures 26 and 27 showed contours in which L*+H^{-} H* resulted in a rising configuration of accents. Of course, rising contours similar to those reviewed so far can also be generated with a H+L or L+H accent in nuclear position, provided the preceding context is not one which would downstep the H tone. One such contour is shown in Figure 48.

The type of intonation which most resembles iterative upstepping was shown in Figure 44. In this contour, the existence of a second phrase accent in a tag created the conditions for two successive upsteps. However, the upstep in such contours applies to the second phrase accent and the boundary tone rather than to pitch accents. In the last section, we noted that the number of upsteps in such contours stops at two, in contrast to the large number of downsteps which could be found in phrases with downstepping pitch accents.

All the rising F0 contours we have seen can be analyzed as the above examples were, without positing an upstepping pitch accent. The lack of an iteratively upstepping accent is not accidental, but appears to be related to the relationship of phonetic value and prominence in English. One of the main types of evidence for the existence of downstep was provided by F0 contours like Figure 2, where it is clear that the nuclear pitch accent has greater prominence but lower phonetic value than the prenuclear pitch accent. This is clear, because we know independently that the nuclear stress is both the stongest stress and the last stress carrying a pitch accent. The comparable evidence for upstep would be provided by an F0 contour in which an accent on the right clearly had both lesser prominence and higher F0 than the accent on its left. Unequivocal cases of this sort are not to be found however, because of the left-right assymmetry of the nuclear stress rule and post nuclear deaccenting. Any F0 contour which had the requisite F0 relations could be analyzed as involving greater prominence on the right, where the higher phonetic value occurred.

4.7 Ambiguities

A number of the distinctions in tonal pattern introduced in Chapter 2 collapse when compressed onto little segmental material. For example, we suggested that L^*+H^- H^- and L^* H^- are distinct when there is enough material after the nuclear stress for the normal placement of the H^- phrase accent to be later than that of the H^- tone of the pitch accent. Ordinarily, the nuclear stress falls on the last content word of the phrase and these two patterns would not be distinct. The topic of this section is a different set of ambiguities which arise from the rules pertaining to the downstep accents. These ambiguities are intrinsic to the intonation system rather than coincidental, in the sense that they arise from the tonal representations and the tone mapping rules without regard to the segmental material.

The relationship of downstep and prominence is one possible source of ambiguity. Consider the FØ contour in Figure 49 which has a w w s metrical structure with a higher peak on the first w than on the second. As we pointed out in Chapter 2, the phonology does not determine a prominence relationship between the two w's in such a case; either one may be more prominent, depending on the semantic importance of the material it dominates. Thus, the FØ contour in Figure 49 could arise because the first w is more prominent than the second. Or, it could arise even with the first less prominent if the second accent were downstepped. These two analyses differ in their predictions for how the prominence of the nuclear accent would be mapped into a phonetic value. Given our principles for carrying over tonal value, the nuclear H* would, under the first analysis, be equal in prominence to the first accent if it had the same phonetic value. Under the second analysis, it is separated from the first accent by one downstep, and is equal to H^*_1 if its value is $/\frac{H^*_1}{k}/$. Clearly, however, there is a range of values for H^*_3

which permit either analysis, with the analyses differing in the interpretation of the prominence of H^*_3.

A related case is shown in Figure 50. Here there are two peaks with the second slightly higher than the first. The obvious interpretation of such a contour is as H* H*, with the nuclear H* slightly more prominent than the prenuclear accent which is subordinated to it. A second possible interpretation is H^*+L^- H*. This interpretation is possible because the downstep rule does not actually require that the downstepped tone be lower than the preceding H tone: it only shifts how a given prominence relation if reflected in the FØ contour. If the downstepped tone were considerably more prominent than the preceding H, it could come out higher even after downstep.

The existence of these last two ambiguities is possible rather than certain, because of our incomplete understanding of how prominence is expressed phonetically. It is possible that some estimate of prominence based on amplitude, vowel duration, or force of articulation of consonants can be brought to bear in analyzing the FØ contour. If this were the case, the listener would be able to pick one analysis of the ambiguous FØ contour. On the other hand, it may be that these other phonetic reflexes of prosody are varied independently from FØ, and for somewhat independent reasons. If so, they would not help the listner disambiguate the FØ contours in Figures 49 and 50.

Rule 10), which prevented the L^- in H^*+L^- from appearing on the surface, is another source of ambiguity. Since downstep affects only H tones, this rule neutralizes H^*+L^- with H* before a L phrase accent or a L* pitch accent. Chapter 5 will introduce a similar case: the H*+H^- accent, which is introduced in that chapter, is neutralized with H* before any tone with a lower or equal phonetic value. This situation arises because H* and H*+H^- are distinct only when the environment for spreading the H^- in H*+H^- is met and this is when the next tone is higher. One consequence is that the FØ configuration in Figure 48 is three ways ambiguous, having the analysis H*+H^- H* H* in addition to the analyses just mentioned. Also, contours which have been described as H* L^- have the additional analyses H*+L^- L^- and H*+H^- L^-.

Does these ambiguities in analysis actually correspond to a three way ambiguity in meaning? To answer this question, we need a better account than we have now of intonational meaning and its relation to the phonological decomposition of the intonation contour. If meaning of the intonation contour is related compositionally to the meanings of the pitch accents, phrase accents, and boundary tones, then we would expect that the contours in question would be truly three ways ambiguous. The situation would in this case correspond to the situation that arises when dental flapping applies across a word boundary neutralizing /t/ and /d/. The resulting phonetic form can be analyzed as involving a word ending in underlying /t/ or underlying /d/; the listener makes his choice on the basis of what yields a meaning most consistent with the context, and in some contexts may be unable to choose. If on the other hand, intonation contours are lexicalized sequences of pitch accents, phrase accent, and boundary tone, then one would expect that the contours in question would not be ambiguous. Rather, the unmarked analysis of the contour would presumably be picked as the lexical representation. We would judge that the unmarked analysis would be the one with the single tone pitch accent, H* L. The situation in this case would be similar to that which arises in German because of the rule devoicing word final obstruents. Obstruents which are final in stems that do not accept suffixes always undergo the rule, and have become lexicalized as unvoiced.

We feel it is unlikely that all phrasal contours have lexicalized meanings. In contours with a mixture of different pitch accents, it often seems that each pitch accent is a comment on the particular material it is associated with. The discussion of Ladd's stylized contours in Chapter 2 above suggests furthermore that the final boundary tone has an identifiable meaning, which contributes compositionally to the meaning of the intonation for the phrase. However, it seems possible that the sequences of nuclear pitch accent and phrase accent have become lexicalized. Nor is the position that the meaning of intonation phrases is in general compositional necessarily at odds with the position in Liberman and Sag (1974), Liberman (1975) and Sag and Liberman (1975) that particular tonal specifications for phrases are tunes with specialized meanings. In our framework, such tunes would represent the idiom chunks of the intonation system. Two candidates for such a treatment, the "surprise/redundancy contour" and the "contradiction contour" were illustrated in Figures 29 and 40 respectively.

4.8 The Representation of Pitch Range

In our discussion of tonal implementation, we have assumed that the value of each tone is computed as a function of its phonological type and of the types and values of immediately preceding tones. Pitch range does not exist in the system except as it is represented in the value of particular tones. It is worth thinking about two alternative formulations in which pitch range is represented more explicitly. The first alternative, proposed in Clements (1979), is that the pitch range at each tonal location is reified in the description. In his description, a value for both (or all) possible tones is computed at each location, and then the value for the actually occurring tone is instantiated as an FØ target. We assume that prominence effects would be handled by computing the range on the basis of the prominence for each tonal location. Given that the system states relations between pitch ranges rather than tonal values, it can in effect refer to the value the non-occurring tone at some location would have had, in computing the pitch range for other locations. A second alternative is that rules like downstep expand or compress the graph paper on which tonal values are computed. This account would have a unit of range, say the "Amana", which would be initialized at x baseline units above the baseline but whose definition in baseline units above the baseline could be changed by rules like downstep. Rules instantiating tones in Amanas would still be needed, but could be somewhat simpler than our rules. For example, in the case of $H^*{+}L^-$ H^*_i H^*_{i+1}, $/H^*_{i+1}/$ would no longer be determined as a function of $/H^*_i/$. Instead, it could be mapped onto Amanas on the basis of its prominence alone, and its lowered FØ value relative to $H^*{+}L^-$ would arise because the number of baseline units above the baseline constituating one Amana was reduced after $H^*{+}L^-$. A modification of the theory might have two graph papers, one for H tones and one for L tones. We do not have in hand the facts to decide among these theories, but it seems worthwhile to discuss what facts would tend to support which theory. One case which was not covered by our partial set of rules for English tonal implementation H* after L*. In order to formulate rules for these cases, one would like to know whether or not /H*/ is the same after L* as after H*, for the same prominence configuration. It may be difficult to determine the answer to this question, because equating prominence across sentences with different intonation patterns is problematic. By locking onto particular usages of different intonation patterns, subjects in an experiment might associate different degress of emphasis with different patterns. For example, a subject who decided that H* H* was a "reading pattern", while L* H* was a "surprised pattern", would be apt to produce instances of L* H* with more overall emphasis. The opposite might be true for a subject who had in mind the "polite greeting" usage of L* H*, and the "explicit explanation" usage of H* H*. Supposing that such problems can be circumvented, the results would help to decide whether range should be represented explicitly in the theory. Under our original proposal, there is no reason to expect that /H*/ would behave in the same way after H* as after L*. It would be possible to write rules which produced the same output for both cases, but the identical behaviour would be completely coincidental. For the two alternative proposals, the result that /H*/ behaved the same after H* and L* would be a readily captured generalization, perhaps even the simplest possible outcome. The generalization would be captured under Clements' approach by a rule which computed the top of the range (or /H/) at location i+1 as a function of the pitch range at location i. Taking into account what the tone at location i actually was, as a downstep rule does, would only be a complication. Under the Amana theory, the result would arise because the rule for implementing H* in Amanas is insensitive to tonal context. Making H* come out differently after L* than after H* would require making the rule context sensitive, or adding a rule similar to downstep which redefined the Amana.

On the other side of the coin, adding an explicit representation of pitch range to the tonal implementation system significantly increases its power. In the absence of additional restrictions, such a proposal makes it possible to write rules changing the value of the tone which does <u>not</u> occur at a given location. The effects of such a change would shows up only later, on the first subsequent tone of the type changed. For example, consider an intonation pattern involving the sequence H^*+L^- L^* H^*, and suppose that the language has a downstep rule lowering the top of the pitch range after H+L, but not the bottom. This rule would lower the value of H at the location of L*; since the occurring tone is L, the effect of downstep would not show up at once, but only on the H* which follows L*. To block downstep in H^*+L^- L^-+H^*, as English actually does, would require a slight complication of the rule. To generate downstep in H^*+L^- L^* H^* would be impossible in our theory, without giving up the claim that tonal implementation rules cannot refer further back than the last pitch accent. In principle under this approach, the downstepped H value could be passed under the table for any number of L's separating the H*+L$^-$ from the H*.[6] Such regularities could not be handled under our approach without the introduction of the starred constants used in Chomsky and Halle (1968).

The version of the Amana theory with a single graph paper for both tones provides only very limited means for passing tonal values under the table. For example, in the case of H*+L$^-$ L* H*, the Amana could be redefined after H*+L$^-$, with lowering of H* as a consequence. However, the compression of the overall range would also lower L*, so that some effect for downstep would be seen at once. The only way for H* to be lowered without an effect on L* would be if L* was on the baseline. (The modified version with separate graph papers for the different tones of course has the power of Clements' approach).

However, the Amana theory provides extra power of a different sort. It appears that even with rules which redefine the graph paper as described, we still need context sensitive rules stating where on the graph paper a particular tone is realized. This is so because the interpretation of L varies with context in ways that do not depend on prominence, with results which are not propagated rightward. /L*/ in H$^-$+L* is greater for greater prominence, whereas the L* accent is lower for greater prominence. /L*/ in H$^-$+L* had no effect on the value of the following tones; when /T/ in H$^-$+L* T is computed, /H$^-$/ is still within the window and all rules computing /T/ refered to /H$^-$/ rather than to /L*/. With this addition to the theory, any local feature of the FØ pattern which could be described by a rule redefining the Amana could also be described by a context sensitive tonal implementation rule. The expected difference is that the effect of the rule redefining the Amana would show up on all subsequent tones, while those of the tonal implementation rule would show up only if overtly propagated by subsequent rules.

In our account, there are only restricted circumstances in which a context sensitive tonal implementation can fail to have consequences for tones to its right. Specifically, this can happen only if the tone evaluated is one tone in a bitonal accent or the phrase accent, sot that some other tone still remains in the window to serve as the basis for tonal evaluation on the next iteration. In the Amana theory, changes in the graph paper replace rules relating tonal values to preceding ones as the basic mechanism for carrying rightward the effects of rules like downstep. The effects of tonal implementation rules do not in general propagate rightward (they propagate only if they happen to feed a context sensitive tonal implementation rule). In this formulation, it does not appear possible to capture the restriction on non-propagation which fell naturally out of our theory. Instead, the Amana theory essentially has a diacritic for marking whether or not a given context sensitive rule has consequences which propagate rightward; the rules which redefine the Amana have effects to the right, while the context sensitive tone implementation rules do not. This leaves us with a system of description which is probably too rich.

The Amana theory could be revised to eliminate this dual system of rules by eliminating context sensitive tone implementation rules and handling all contextual variation in the rules defining the Amana. We would judge that a revision in this direction which was thorough enough to be successful would make the Amana theory the same as Clements' account. There would no longer be a graph paper, but rather a pitch range defined by H at the top and L at the bottom. The rules for implementing tones would have become as trivial as the tonal instantiation in Clements' account.

4.9 The Hierarchical Representation of Downstep

The previous sections have developed a picture of downstep in which it is one of a package of rules which map tonal representations into FØ contours. These rules apply iteratively left to right, and this iterative application results in a descending series of values when the underlying representation meets the structural description for downstep more than once. An alternative formulation of downstep, suggested in Clements (1980) and Huang (1979), attributes the descending form of the contour to a hierarchical representation similar to the metrical tree, rather than to iterative application of phonetic rules. Specifically, the proposal is that the surface string of tones is organized into domains for which a pitch range is defined, referred to as "terracing spans", are gathered into a right branching tree, with sister nodes labelled *h* and *l* in the case of downstep.[7] An example is shown in Figure 51. This representation must of course be supplemented by phonetic rules, which generate the FØ values corresponding to the tones. Specifically, the phonetic rules must spell our what lowering of pitch range results from the *h l* relationship, and how much lower L is than H when in the same domain. An explicit stipulation is needed to prevent H at the start of a new terracing span from being downstepped below the level for L in its sister domain on the left.

In this section, we discuss how the hierarchical representation of downstep would work out in the description of English intonation. One false prediction for English makes at least some modifications necessary. In general, the hierarchical representation appears to be unmotivated for English, since rules mapping tones into FØ values can generate the observed patterns without making reference to a hierarchical representation. Since the hierarchical representation is not an alternative to such tone mapping rules, but rather must be supplemented by a somewhat recast set of rules expressing the same regularities, this means that the hierarchical representation is superfluous for the description of English. This suggests to us that it is also superfluous in the description of African tone languages, which appear to present simpler phenomena than English does. We will also suggest that the hierarchical representation is consistent with a greater variety of different phonetic implementations of downstep than our phonetic rules could generate. If it turns out that the phonetic form of downstepped sequences in African tone languages share the properties of the English relatives, then the hierarchical representation is too rich.

We will discuss first how the facts of English do not support the division into the terracing spans on which downstep is defined. Then we will turn to the hierarchical organization and its implementation.

In our discussion of English, we have seen three levels which contribute to the phonetic value of a tone. The speaker makes a choice of pitch range for each intonation phrase. For example, he might raise his pitch in a cutsy style of speech, or lower it in an aside. This choice is reflected in the tonal value selected to intialize the FØ pattern for the phrase. Within the phrase, the effects of prominence on tonal values are defined for each accent. Lastly, tonal values are defined at each tonal position; for instance, in $H^-+L^*\ L^-$, $/L^*/ \neq /L^-/$ even though both are under the umbrella of the same prominence. To relate these observations to the Clements/Huang proposal, we might say that a pitch range is defined in English for every tonal location by the value for H and the value for L at that location. The values for these ranges reflect the division of the utterance into phrases, the metrical organization within the phrase and the tonal context.

Now consider what domains would be defined for English by the occurrence of downstep and upstep. As we have seen, English downstep is triggered by particular pitch accents, L^*+H^-, L^-+H^*, H^*+L^-, and H^-+L^*. If each new terracing span is defined by the occurrence of a downstepped tone, then a whole utterance can comprise one span if it does not include any of these accents in a context for downstep. In a chain of H+L accents, a new terracing span would begin with the H of each accent, so that each span would correspond to the group of tones governed by one prominence specification. In a chain of L+H accents, a new terracing span would still begin with the H of each accent, so in this case it would include the second tone governed by one prominence specification and the first tone governed by the next. In the case of H+L H^-, the H^- phrase accent would start a new terracing span, while for all other nuclear pitch accents the phrase accent would not start a new span. The boundary tone would start a new span after H^-, where it is upstepped, but not after L^-.

Therefore, the domains needed for a hierarchical representation of terracing in English do not coincide with the domains needed for any other purpose. It follows that the tree structure constructed on these domains does not coincide with the two metrical structures which have played a part so far, the metrical structure of the text which controls text-tune alignment and prominence, and the rather impoverished metrical structure of the tune which also plays a part in alignment.

Given that the downstep domains do not coincide with the tonal organization otherwise needed, the only way to motivate them is to claim that they are needed to account for the facts of downstep or upstep. This turns out not to be possible. Obviously, the domains are not needed to account for the value of first tone in each domain (the stepped tone); any description of the structural conditions under which a new domain is started could as well serve as the structural description for a phonetic rule which actually stepped the first tone. To motivate the domains, it is necessary to find a process which applied within a domain but not across the boundary between two such domains, or else a process which refers to the domain taken as a whole.[8] The Clements/Huang proposal is that downstep refers to the domain as a whole, and that the phonetic values of H and L are then carried through within a domain but not across the boundary between two. In English, as we have seen, rule 2) relates the phonetic value of a H tone to that of a previous H tone both in downstep contexts and in non-downstep contexts. This formulation was motivated by the need to describe the interaction of prominence and downstep in determining the value of downstepped tones.

Thus, the domain for rule 2) appears to be the intonation phrase, and it does not refer to the terracing spans. The situation for L is the opposite; the value of L is not carried through even when prominence is not varied and downstep does not intervene; in $H^- + L^* \ L^- \ L\%$, the three L tones all have different values. The rule for L% is non-relational, and so can provide no evidence for organization into domains. The rules for L and L^- apply only within a downstep domain, but this need not be stipulated since their structural description is never met crossing the boundary between two such domains. Thus, tonal values are accounted for without any rules of the sort which would motivate the terracing spans.

In the Clements/Huang proposal, the descending series of terraces in a downstepping contour results from the hierarchical organization of the domain. As both point out, the hierarchical organization constrains the relation between one terrace and the next, but does not actually determine their phonetic differences. For this, the hierarchical representation needs to be supplemented by rules which spell out the actual phonetic values of the tones. Clements (1980) presents a proposal for the character of these mapping rules, which takes the following form for a system with total downstep.

24a) Each tone level is numbered 1, 2, ..., $\underline{n}$ starting from the highest.

b) An increment of 1 is added to each tone for each l dominating it in the tree.

In order to compare this to our proposal, we propose the following restatement, under which the value for each tone comes out one lower than under Clements' proposal.

25a) Each tone level is numbered 0, 1, .. n starting from the highest.

b) An increment of 1 is added to each tone for each l dominating it in the tree.

These numerical values still do not represent FØ values. On the assumption that the phonetic facts are the same as for the case of total downstep we examined in English, the following mapping rule seems appropriate:

26) $/n/ = V(k^n) \quad 0 < k < 1$

Here, $V = V(k^0)$ is the value of the highest tone. Given the speaker's freedom to vary his overall range, it can be taken to be arbitrary. k is the downstep factor. Values are in baseline units above the baseline.

Rules 24) through 26) describe a system with total downdrift, total downstep, or both, depending on how the tones are organized into the domains on which the tree structure is defined. Figure 52 shows how the rules work out for a simple two-tone system with total downdrift and no downstep. Figure 53 shows the result for Kikuyu, which, according to Clement's report, has no downdrift of H after surface L, but total downstep of H after a floating L which does not appear on the surface.

In the present framework, the situation in Figure 52 would be described by the following rules:

27a) $/T_1/ = V$

b) $/T_{i+1}/ = k/T_i/$ for $H_i\ L_{i+1}$

$= /T_i/$ otherwise

The more complicated situation in Figure 53 would be described as follows.

(A vertical line indicates an associated tone; a raised hyphen, a floating tone, and a feature analysis of H as [+ high] and L as [- high] is assumed).

28a) $/T_1/ = V$

b) $/L_{i+1}/ = k/H_i/$

c) $/H_{i+1}/ = \frac{1}{k} /\overset{|}{L}_i/$

d) $/H_{i+1}/ = /L^{-}_i/$

d) $/[\alpha\, high_{i+1}]/ = /[\alpha\, high]_i/$

The features of these rules which we take to be universal are that the value of each tone is computed in relation to the value of tones immediately to the left,[9] that the computation is carried out as a ratio of baseline units above the baseline; and that $/L_{i+1}/ \leq /H_i/$. The things which have to be stipulated under either treatment of Figure 52 are: the value of V, the value of k, and the fact that H is lowered all the way to the level of a preceding L^{-} rather than partially. This last fact is expressed in part b) of rule 27), and for the hierarchical representation it is expressed by the selection of convention 25) over alternative conventions for partial downstep which we will state shortly. These stipulations also apply to the case of Kikuyu. In addition, both accounts of Kikuyu distinguish the treatment of H after floating L from the case of H after an associated L. In 28), the distinction is made by c) and d), while in the hierarchical formulation it is made by the rule for setting up the terracing spans.

There is one thing which the hierarchical formulation must stipulate which we take to arise from general convention. This is the fact that the downstepping sequence has an exponential form, in baseline units above the baseline. In our framework, this follows automatically from the claims that the tone mapping rules are ratio rules, and that they are local. The locality constraint on its own restricts the system to recursively specifiable functions. By contrast the hierarchical account is intrinsically non-local. In Clements' formulation, the n's provide a way to keep track of how far from the beginning a particular step is. Speaking more generally, the point of using trees in metrics is to encode non-local rules, and carrying the trees over into the tonal domain entails carrying over this capability. Thus, there is nothing to prevent the sequence of steps taking the form $V(\frac{k}{n+1})^n$, for example, instead of $V(k^n)$ as above. In fact, the description is consistent with the n's being implemented by any monotonically decreasing series of values; the fact that the steps have the same phonological status in no way implies that they are phonetically the same. The occurrence of non-local regularities in the phonetic implementation of downstep would be a strong argument for the hierarchical theory. If non-local regularities are not found, we would conclude that the full power of the hierarchical representation is not needed.[10]

There are two features which are stipulated in our account but arise through general conventions in the hierarchical account. The first is that successive like tones have the same value.[11] This was expressed in 28e). The fact that e) is stated explicitly amounts to a claim that it could have been otherwise, and we believe this claim is correct. As we just noted, successive L tones in English can differ in value without either terracing or a prominence difference. We suspect that similar facts will come to light through more careful instrumental analysis of African tone languages. For example, Meyers (1976) reports an utterance final L tone is lower than an immediately preceding L tone in Hausa. It is well known that the effects of consonants on FØ can become regularized and exaggerated by the phonetic rules, with the result that successive like tones are given different phonetic values; if the conditioning environment is lost, this phonetic difference can become a phonological difference.

Now, perhaps someone might take the view that the convention in the hierarchical theory that like tones in the same domain have the same value is not meant to deal with this level of phonetic detail, but rather to express the fact that like tones count as the same at some more abstract level. In our view, this fact is adequately expressed by the fact that like tones are phonologically the same tone. Claiming they are the same phonetically means they are the same in some well defined quantitative descriptive framework. We have not found any motivation for a level of description between the underlying tonal description and the quantitative description.

The second feature which is stipulated in our account but not in the hierarchical account is that the coefficients in 28b) and 28c) are reciprocally related. Nothing in our system prevents the coefficient in 28c) from being less than 1/k; if it were, the result would be total downstep of H after a floating L (from 28d)), plus partial downstep after an associated L.

As we have observed, English has partial downstep in H L+H and total downstep in H*+L¯H. The existence of a language which exhibits partial downstep in one context and total downstep in another appears to be a problem for both Clements' and Huang's proposals, though for different reasons. Huang proposes interpretative conventions which exclude this possibility: stepping arises only at the boundary between two terracing spans, the implementation of stepping is insensitive to the tonal makeup of the domain, and the implementation of contrasts within a span is invariant. Thus it would not be possible to have greater downstep when a terracing span ends in +L than when it ends in L+, nor is it possible to have the same downstep in both cases but a different interpretation of the L H difference within the span.

The problem for Clements' proposal arises because he handles partial downstep by an implementation system which is completely different from the implementation system for total downstep. In his words, the pitch interpretation for partial downstep ..." can be defined in terms of a 'comparator' which examines a particular set of pairs of tones to determine which of the two, in each case, belongs to the higher register. Specifically, only like tones that are non-adjacent and separated only by un-like tones (i.e., tones not identical to them) are compared. Thus, the comparator will examine each pair of High tones separated by one or more L tones to determine which of the two is higher; a similar procedure may be applied to each such pair of L tones, if necessary. For example, the 'comparator' will not be able to make a direct comparison between, for instance, the first and last High tones in (Clements' example) 12), the claim being that such a comparison is linguistically irrelevant (i.e., never used for linguistic purposes). With this preliminary clarification, the pitch interpretation principle can be stated as follows:

29 Pitch interpretation (partial downstep)

a) Given two non-adjacent like tones T_j, T_k separated in the tone group only by tones not identical to them,

i) T_j is higher in pitch than T_k if the (unique pair of) sister nodes dominating T_j and T_k are labelled (*h*, *l*) respectively.

ii) Otherwise (i.e., if the nodes are labelled (*h*, *l*), T_k is higher in pitch than T_j.

b) Neighboring tones do not cross levels ...."[12]

It would be surprising if a language used both of the two interpretative conventions, especially if different conventions were applied in closely related contexts, as in English. It would be yet more surprising if the relations among the terraced levels came out the same when these two different conventions were applied, as appears to be the case for English.

Clements' pitch interpretation convention for partial downstep seems far closer to the proposal made here than the convention for total downstep. As far as we can see, 29) does not make crucial use of the hierarchical structure at all. The comparator refers to the hierarchical labelling and to local properties of the tonal sequence. However, the rules which set up the hierarchical structure and labelled it also referred only to local properties of the tonal sequence. Thus reference to the hierarchical structure in 29) could be replaced with reference to the local tonal features on which it was based. For example, Clements discusses a dialect of Zulu documented in Cope (1970), where H downsteps after a floating L before a "final tonal morpheme" (generally, the last two syllables of a word) and upsteps elsewhere. He proposes to explain these facts by a dialect specific rule for labelling the right branching tree into which the terracing spans are organized. Specifically, the labelling is (l,h) for the pair of nodes in which the left hand node immediately dominates a floating L before at least three tone-bearing units within the word. In all other cases, (when the floating L is less than three tone-bearing units before the end of the word, or when H follows an non-floating L) one finds downstep rather than upstep, and the labelling is thus (h,l).

While these facts are somewhat complex, it is still the case that local properties of the tonal sequence determine whether downstep or upstep occurs. Thus it would be possible to account for the facts with dialect particular rules for interpreting the tonal sequence, without the intermediate level of the hierarchical representation. A theory which reified pitch range, as suggested in Clements (1979), would say something like this: Lower the pitch range wherever H follows an associated L, or a floating L no more than two tone-bearing units before the end of the word. For all other cases of H after L, raise the pitch range. In a theory in which pitch range was not reified, the same general idea could be recast in terms of rules for evaluating H.

4.10 Downstep and the Layered Theory of Intonation

One of the themes of our description of intonation is that differences in overall configuration among intonation patterns may be traced to tonal differences and the local phonetic rules which implement them. FØ contours with a steeper downslope than declination would engender have been analyzed in terms of downstepping pitch accents; the extra high pitch at the end of the typical yes/no question was attributed to an upstep rule affecting boundary tones. One ramification of this approach is that declination can be taken to be quite invariant. A second is that the same theoretical primitives, the tones, can be used to describe both the FØ correlates of stress (the pitch accents), and the FØ correlates of phrasal intonation type (the phrase accent and the boundary tone). This has made it possible to be precise about how these two aspects of intonation conspire to determine the FØ contour. Third, the approach has been very successful in giving an account of FØ contours which start off one way and continue a different way.

A number of authors have suggested a contrasting approach, in which the FØ pattern is built up by superposition of layers. The bottom layer represents the phrasal intonation type, and pitch accents are added onto it. For example, Bing (1979) characterized the yes/no question patterns discussed in Section 5 as involving a basically rising and expanding pitch range, in contrast to declarative contours which exhibit declination. In Lea (1973), declarative intonation is characterized by adding FØ contours marking stress to a phrasal FØ contour, which is a rise-plateau-fall. Another version of this approach may be found in

Thorsen (1978, 1979a, b, c and d, 1980). This is the only work we know of that presents experimental data which appears to require this approach over the approach taken here. Our aim in this section is to show how Thorsen's data may be accommodated in the framework proposed here.

Thorsen studied the intonation of completed declarative sentences, non-final declarative clauses, and questions in Danish. According to Thorsen's data, Danish has only one type of pitch accent, which appears to correspond to L^*+H^- in our framework; the stressed syllable is low, and then there is a rapid rise on following unstressed material followed by a gradual fall to the next low. When there is more than one unstressed syllable after the last stressed syllable in the sentence, one can make out a L% in questions and declarative sentences alike; the FØ rises from the stressed syllable and then falls again to the end of the sentence. If there are one or no unstressed syllables, the L% is less evident. It is unclear whether there is a phrase accent. Contrastive emphasis results in an increase in the value of the H^- on the emphasized stress group. Nearby H^-'s are lowered. There is greater lowering to the right than to the left of the H^-, and greater lowering on the adjacent stress groups than on more distant ones. It would not appear to be correct to say that contrastive emphasis causes other pitch accents to be deleted. In Thorsen's schematized representation of an intonation pattern containing three pitch accents with contrastive stress on the first one, both L and H in the third accent are still clearly visible. A similar result is reported for Swedish in Bruce (1977).

Questions in Danish are distinguished from declaratives by the overall trend of the pitch accents rather than by local features of the FØ contour. The contrast is shown in Figure 54, which is taken from Thorsen (1979c). Here, the heavy circles are the L*s and the light circles are the FØ of unstressed syllables. The line connecting the heavy circles is what Thorsen views as the phrasal contour. In the completed declaratives, the line falls steeply; in questions, it falls slightly if at all; in the non-final declarative clauses, its slope is in between the slopes for the other two cases.

As Thorsen points out, this summary of the facts suggests that the difference among these three sentence types is not marked by the FØ configuration at the end of the sentence but rather by relations throughout the sentence. This contrasts with our conclusion that the comparable distinctions are made in English by the phrase accent and boundary tone. She tested her hypothesis by a series of perception experiments using segmentally homophonous sentences read with the intonation patterns of interest. These experiments are reported in Thorsen (1980). The first two experiments established the ability of listeners to recognize the intonational categories of interest. In the first experiment, subjects successfully distinguished among interrogative, non-final declarative, and completed declarative utterances. In the second experiment, subjects had two choices, declarative versus non-declarative, instead of the three choices of the first experiment. It was found that under these conditions, the interrogative and non-final declaratives were grouped together in a single category of non-declarative contours. Spearman rank order correlations between the responses in the first experiment and a large number of different measures of the FØ contour were computed. Only two measures divided the stimuli into the same categories that the listeners did: the slope fitted through the three L*s and the FØ contour on the last L*. Obviously, these two are highly correlated. Two subsequent experiments established that either of them separately was sufficient to enable the listners to distinguish the three categories. In one experiment, a series of stimuli truncated from the end were constructed. Subjects were able to make the three way distinction under all truncations except the most severe one, which left only the first L^*+H^-. This supports the hypothesis that the contours differ in slope, since the slope can be computed by the listener from any sequence containing two like tones even if the end of the phrase is not included. The fact that the initial L^*+H^- was not sufficient for recognition of the contour rules out the hypothesis that overall range rather than slope was the basis of the subjects' judgements. A second experiment involved a series of stimuli truncated from the beginning rather than from the end. This experiment established that the last L^*+H^- was a sufficient basis to make the three way discrimination. In fact, under the slope hypothesis, the pitch level of the last L^*+H^- would differ in the three intonation patterns.

We believe that Thorsen's results can be accommodated within the present framework by supposing that Danish has a downstep rule in declaratives which is suspended in questions. There are good precedents for such an analysis; Igbo and Hausa are also known to suspend downstep rules in questions (Clements 1980). It appears that in Danish, however, there are degrees of downstep which correspond to the degree to which the utterance is non-final. This situation can be described by allowing the downstep coefficient k to vary between its minimum value and 1 as a reflex of the relevant semantic continuum.

Under this account, we would expect interrogative sentences to exhibit declination even though they are not downstepped. In fact, it appears from Figure 54 that H tones later in the phrase are somewhat lower than H's earlier in the phrase. Because the output of downstep is exponential in baseline units above the baseline, the account predicts that the rises from L to H should grow smaller through the course of a declarative sentence, even when plotted in semitones. Figure 54 shows that this is the case. A further advantage of this account is that it reduces the variation among the different types of phrasal intonation layers in Thorsen's theory to variation of a single parameter, k. Thorsen suggests that different phrasal contours may be described as lines with different slopes. However, without a phonological account of the phrasal contour, there is no systematic reason to exclude any other functional form one might think of. Lea (1973) for example, proposes that the phrasal contour for neutral declarative intonation may be described as a rise-plateau-fall.

A possible difficulty with this account of Thorsen's results can be resolved in an instructive fashion. In the experiment in which subjects categorized stimuli with the ends truncated, it was found that hearing just the first L^*+H^- was an insufficient basis for discrimination. When the subjects heard just the initial L^*+H^- L^*, however, their performance was quite good. Now, our discussion so far has treated downstep as a rule which lowers H after L. If this rule applied in declarative sentences but not in questions, we would expect that the level of the first H would provide the relevant information to the listeners, and that the next L would provide no additional information. This problem arises because unempirical assumptions have crept in. A sequence with partial downstep, like that schematized in Figure 54, only supports the conclusion that the interval for L H is smaller than the interval for H L. It does not provide any information about which interval is basic, and is therefore neutral between an interpretation in which H assimilates to a preceding L and one in which L dissimilates from a preceding H. It is only if the language suspends downstep is some context that we can see which interval is varied. For Danish, it is clear that it is the H L interval which is varied by different degrees of downstep, while the L H interval remains constant. This says that the L^*+H^- is implemented in a constant way, while the relation between one pitch accent and the next varies. Thus, listeners are only able to judge what phrasal type they have heard when they hear evidence about how the second pitch accent relates to the first.

Footnotes to Chapter 4

1) H would be lower than a preceding L if n were greater than k. Languages with downstep seem to avoid this situation. Current accounts which note this fact block lowering of H below L by stipulation; one hopes that an explanation based on the general character of paradigmatic distinctions will eventually be worked out.

2) Specifically, H* in L^{-}+H* is ordinarily found at the end of the accented syllable, as the H* accent is. H* from either accent can occur earlier in the syllable when crowded by other tones, such as a following L^{-} phrase accent.

3) In our account of downstep, the rule readjusts the phonetic value for a H tone corresponding to a given stress level. Given our claim that tone mapping rules only look back as far as the last pitch accent, it is possible to show that an alternative account, under which downstep affects the phonetic value indirectly by readjusting the underlying prominence, is not feasible. If P_n represents the underlying prominence for the n^{th} step, and $P_n(DS)$, the prominence value as readjusted by the alternative downstep rule, then we have the following relations:

i) $P_n(DS) = k^n P_n$ (where k is the downstep coefficient as above)

And, therefore:

ii) $P_{n-1}(DS) = k^{n-1} P_{n-1}$

Under the hypothesis, these prominence values would be mapped into F∅ values using the same rules as for the nondownstepped H* accents. The relation between $P_n(DS)$ and $P_{n-1}(DS)$ is then:

iii) $\frac{P_n(DS)}{P_{n-1}(DS)} = \frac{kP_n}{P_{n-1}}$

Hence:

iv) $P_n(DS) = \frac{k\ P_{n-1}(DS)\ P_n}{P_{n-1}}$

Computing $P_n(DS)$ as a function of P_{n-1} is unacceptable; P_{n-1} is the underlying prominence value, and the derived value $P_{n-1}(DS)$ has already been determined, and so a rule computing $P_n(DS)$ as a function of P_{n-1} would be global. P_{n-1} can be eliminated from the equation using the relationship:

v) $P_{n-1} = \frac{P_{n-1}(DS)}{k^{n-1}}$

However, in this case, the rule for computing $P_n(DS)$ has to refer to n. This means that the rule would no longer be a local recursive rule, in contradiction to our claim about the character of tone mapping rules.

In general, we have taken the position that tones are predicated on prominence relations but do not change them. Allowing tonal rules to alter prominence relations would open the door to many rules which appear not to be found. For example, this move would make it possible to write a rule by which H* accents downgraded the prominence for any following accent which was not a H*. The regularities resulting from such a rule could not be described in the present framework, in which the selection of accent type is completely independent from the prominence representation.

4) The value of H* on "really increased slightly with sentence length, and this increase results in slightly higher values for subsequent tones, and slightly larger step size. The fact that the longer sentences started higher is not surprising, in view of the results in Sorensen and Cooper (1980), Enkvist and Nördstrom (1978), and Lehiste (1975) which were discussed in Chapter 4, Section 3. As we noted there, it appears that larger semantic units are in general marked by a higher FØ onset.

5) The pattern in Figure 46 apparently corresponds to Crystal's type E rising head, which is schematized with unstressed syllables on the same level as preceding stressed syllables. The lack of concrete examples in Crystal's account makes it impossible to check whether the sentences he would view as exhibiting a type E head may be plausibly analyzed as involving ascending prominence.

6) Clements and Ford (1979) and Clements (1980) introduce additional assumptions which would prevent downstep of H at the location of L* in H^*+L^- L* H*. However, as far as we can see, these assumptions do not completely prevent passing tonal values under the table; examples similar to this one, though more complicated, can be constructed for hypothetical languages with more than two tones. In Clements (1980), the mechanism for passing on tonal values is spelled out as a hierarchical structure on pitch range. Other properties of this representation are discussed in Section 9.

7) In Huang (1979), upstep is handled by positing an additional tree structure, while in Clements (1980), it is treated by reversing the

node labelling to h, ℓ. Clements (1980) also uses the h, ℓ relationship to account for the feature decomposition of systems with more than two tones. In a four tone system, for example, there is a split into a h register with a ℓ register, and each of these registers is then split into h and ℓ. A three tone system arises when only one of the two registers has split. The English tone system has only two tones, so this theory will not play a part in the discussion. For the sake of clarity, we will continue to use H and L for the two tones, which would correspond to terminal h and ℓ nodes in Clements' account.

8) A third possibility, a process which refers to the designated terminal element of a terracing span, does not come into question here since a span can contain several like tones and therefore does not have a unique designated terminal element.

9) We are not in a position to say how far back tone mapping rules may refer in a language in which tones are not organized into pitch accents.

10) The same observations can, of course, be made about the left to right iterative accounts of downstep proposed in Schacter and Fromkin (1968), Fromkin (1972), Peters (1973), and Meyers (1976). These descriptions also generate as an intermediate level of representation a series of integers which would support nonlocal phonetic rules.

11) Huang (1979) claims that successive like tones have the same value, and Clements (1980) makes this claim for languages with total downstep. The interpretative conventions he gives for languages with

partial downstep have enough latitude that successive like tones would not necessarily have the same phonetic value on the surface. However, he assumes elsewhere in the paper that they would.

12) As Clements notes, in some special cases, provision b) of 29) is superseded.

Chapter 5

TONE SPREADING

5.1 Introduction

In the preceding chapters, a number of examples involving tone spreading came up. In Chapter 2, we noted that an L^- or H^- before H% spreads to the right, with the result that the FØ contour does not show a gradual rise from the phrase accent to the boundary tone, but instead a plateau followed by a sudden rise. Two figures from Chapter 2 which illustrate this phenomenon are repeated as Figures 1 and 2 here. Figures 3 and 4 show two examples of another case of tone spreading. Here, H^- in L^*+H^- spreads to the right when followed by H*. The discussion of rising intonation patterns in Chapter 4 mentioned a pitch accent which is one of the main topics of this chapter, the H^*+H^-. H^*+H^- is responsible for patterns in which the FØ on unaccented syllables maintains the high level of a preceding accented syllable, instead of dipping as between H* accents. An example is shown in Figure 5. Sag and Lieberman (1975) discuss the use of this pattern in questions; it is also a common pattern for statements.

The aim of this chapter is to give a unified treatment of tone spreading in English, by considering systematically where it does and does not occur. In the cases we have seen so far, the tone which spreads is a T^-. In fact, spreading of T* or T% does not have to be assumed in any of the figures in the preceding chapters. The hypothesis that only

T^- is eligible for tone spreading will drive the investigation here, and will turn out to be justified by its success in describing the additional facts that will be presented.

Section 2 considers cases of tone spreading to the right. The main conclusion is that the tone spreading rule must be stated in terms of the relative phonetic values of a T^- and the following tone; the result of spreading is that T^- perseverates in time. Specifically, the rule is:

1) T^-_i spreads towards T_{i+1} if $/T_{i+1}/ \geq /T^-_i/$ [1]

It is impossible to formulate the rule in terms of tonal values rather than phonetic values because the contributions of prominence and upstep to the phonetic value of tones plays a part in determining when spreading occurs. Section 3 shows how rule 1) explains the apparently idiosyncratic set of contexts in which H^*+H^- contrasts with H^*. These accents contrast only in contexts where H^- would spread, therefore, only when the next tone is equal or higher

Section 4 investigates the possibility of leftward tone spreading in English. If leftward tone spreading occurs, it is restricted to the case where the tone preceding T^- has the same phonetic value:

2) T^-_i spreads towards T_{i-1} if $/T_{i-1}/ = /T^-_i/$

The result that leftward spreading is more restricted than rightward spreading is not surprising, in view of Hyman and Schuh's (1974) report that the perseveration of tones is universally less marked than anticipation. We are not completely certain that English does have leftward tone spreading; we will suggest that it may be possible to explain away all of the cases in which it appears to have taken place.

5.2 Rightward Spreading

Two situations arise in which spreading to the right is a possibility. These are the case of the phrase accent before the boundary tone, and the case of T^- in an accent of the form T^*+T^-. In this section, we will discuss first the case of the phrase accent, and then turn to the two cases of T^*+T^- that have been introduced so far, namely L^*+H^- and H^*+L^-. The H^*+H^- accent is taken up in the next section.

It has already been established that L^- and H^- spread to the right before H%. Because of the upstep rule, the H% is in both of these cases higher than the phrase accent, and so the spreading is apparent in the timing of the final rise. In the two cases which lack this final rise, L^- L% and H^- L%, it is less apparent whether the spreading rule has applied. As we have seen in Chapter 1, interpolation between L and any other tone is monotonic. Since L% in H^- L% is upstepped to the level of H^-, such interpolation would generate a flat F0 contour for this tonal sequence, even if tone spreading did not apply. The expected outcome for L^- L% is similarly quite flat. Consideration of two phenomena, however, leads us to conclude that the phrase accent does not spread in L^- L% but does in H^- L%.

One of the observations about the H* L^- L% contour made in Maeda (1976) was that endpoint of the fall was "below the baseline". The baseline, as he defined it, was a line fit by eye through the low points in the FØ contour, and the observation is that the L% lies below this line. In our model, the L% is on the baseline and the observation is that other low points, including the L^- phrase accent, typically lie above the baseline.[2] In designing the intonation synthesis program described in Pierrehumbert (1979a), it was found essential to incorporate this observation as it affects the relation of the L% to a preceding L^-. If the fall from a nuclear H* went straight to the baseline and tracked the baseline until the end of the phrase, the FØ sounded too low too soon. This problem was corrected only by raising the level of the corner in the contour, the L^-, relative to the level of the endpoint, the L%. The FØ between the L^- and L% as a result fall faster than the baseline did over the rest of the phrase. A similar conclusion about the form of the H* L^- L% is reached in Ashby (1978), although he states it in different theoretical terms. Thus, it appears that the L^- phrase accent is significantly above the baseline in H* L^- L%, as we saw in Chapter 3 that it was in H* L^- H%. Given this conclusion, what would the outcome of spreading L^- before L% be? One would expect to find an FØ plateau somewhat above the baseline, with a sudden drop to the lower valued L% at the end. In fact, the fall from L^- to L% is gradual, suggesting that tone spreading does not occur. Figure 6 summarizes the contrast between FO contours for L^- H% and L^- L% which is predicted by these observations: in L^- H%, a value somewhat above the baseline is maintained until the rise for H% begins, while in L^- L% a fall from this value to the baseline begins immediately at the phrase accent. This prediction has been informally confirmed. For JBP, a value of 160 Hz is typical for L^- in an utterance with peak values around 300 Hz. Table III of the last chapter gives the median value for L% in utterance final position as 137 Hz for JBP. In contours involving L^- H%, the F0 never falls below about 160 Hz, while in contours involving L^- L%, there is a gradual fall after the nuclear accent amounting to about 20 Hz. This observation holds when the nuclear accent is L* as well as when it is H*. One consequence is that phrasal patterns ending in L* L^- L% in fact have a noticeable fall at the end. One such F0 contour is shown in Figure 7. Similar tendencies were found for MYL. MYL has a smaller pitch range, and the size of the fall which is found in L^- L% but lacking for L^- H% is about 10 Hz.

Let us now turn to the question of whether H^- spreads before L%. We have only been able to find one phenomenon which bears on this question. This is the phenomenon of echo accents: Accentable syllables past the nuclear accent often carry a miniature replica of the nuclear accent. That is, in H* L^- contours, one may see small peaks on accentable syllables following the H* nuclear accent; in L* H^- contours, one seen small dips. Echo accents can be seen in Figures 8 and 9 where they are marked by parenthesized tones. Figure 10 illustrates an outcome when they are only minimally present.

Due to the scale of the echo accents, it is difficult to separate them phonetically from segmental effects. For this reason, it has not been possible to answer many questions on which a complete phonological account of this phenomenon would depend. For example, in the case of an L*+H^- H^- question, it is unclear whether the L*+H^- is echoed, or merely the L*. However, it seems that the F∅ excursion on the post-nuclear accentable syllables systematically mirrors the nuclear accent, but is produced at very low prominence. If this is correct, then it would be appropriate to handle the phenomenon by a rule copying tones from the nucleus to sufficiently strong syllables on the right.

The relation of the echo accents to tone spreading is brought out in an example like this:

3) The Uruguayan bulldozer drivers' union
L* H^- L%

Here, the intonation pattern is L* H^- L%, an appropriate pattern for a nonfinal element in a list or for a rhetorical question. Because of the focus on "Uruguayan", "Ur" has the L* nuclear accent. The word stresses of "bulldozer", "drivers", and "union" are eligible to receive echo accents. At the level where these have echo accents, the representation is something like 4): echo accents are parenthesized.

4) The Uruguayan bulldozer drivers' union
L* H^- (L*) (L*) (L*) L%

Now, if the H^- did not spread to the right, the F∅ contour sketched in Figure 11 would result, given our interpolation rule for L's. The observed contour does not have this form, but rather the one shown in Figure 12, where the F∅ between the echo accents reverts towards the level of the H^-. This is a subtle point to the eye but completely obvious to the ear. In order to generate such a pattern, H^- must propagate to the right so as to come out interleaved with the echo L*s. In spite of uncertainties about how this interleaving is accomplished, it seems reasonable to conclude that H^- spreads to the right before L%,

The F0 between two echo H*s in an H* L^- L% similarly reverts towards a lower level. However, F0 between H*s in general reverts towards a lower level, so spreading of the L^- is not necessary to explain the form of the F0 contour.

So, the phrase accent spreads to the right in the cases of H^- H%, H^- L%, and L^- H%, but not in the case of L^- L%. This distribution is captured by Rule 1), since in the first three cases the boundary tone is as high as the phrase accent or higher, while in the last case, the boundary tone is lower.

Of the pitch accents introduced so far, two, the L*+H^- and the H*+L^-, have a floating tone to the right of the starred tone which could in principle be subject to Rule 1). The F0 contours in Figures 3 and 4 show that the rule works correctly for the case of L*+H^- H*. In both contours, spreading of H^- is triggered by the presence of a nuclear H* with greater phonetic value to the right. In Figure 3, the F0 plateau resulting from spreading is somewhat obscured by the large F0 obtrusion from the [k] in "quintillion." This is not a problem in 4, where the reader must, however, trust to the ear and intentions of the author regarding the underlying accentuation. Numerous F0 contours in Chapter 2 show that H^- in L*+H^- does not spread before an L. Two such contours are Figures 2 and 13. In all of these cases, Rule 1) fails to apply because the value of the H^- in $L*+H^-$ is greater than that of the following L.

In the case of $H*+L^-$, L^- trivially fails to spread. As we argued in Chapter 4, L^- is not mapped into a target value at all. As far as the spreading rule is concerned, the two adjacent targets are the two H* targets, and since these correspond to starred tones, spreading is not applicable.

At this point, we already have some evidence for the phonetic formulation of rightward spreading over a formulation in terms of tonal types. Spreading occurs in sequences of the form H^- H%, L^- H%, $L*+H^-$ H, and H^- L%, but not in L^- L% and $L*+H^-$ L. The major obstacle to a tonal formulation is the fact that spreading occurs in H^- L% but $L*+H^-$ L. The phonetic formulation differentiates these cases because the L% is upstepped to the phonetic value of a preceding H^-, whereas L after $L*+H^-$ is lower than H^-. This difference would not be available for a tonal formulation, and so the structural description for a tonal formulation would have to inelegantly duplicate that for upstep. Further obstacles to a tonal formulation of spreading will arise in the next section, where the interaction of spreading in $H*+H^-$ with prominence is discussed.

5.3 <u>The $H*+H^-$ Accent</u>

Figure 5 introduced a type of intonation in which the high F0 is sustained on unaccented syllables, instead of dipping as between H* accents. This section proposes that such patterns arise from $H*+H^-$

accents. In the present theory, the only mechanism for generating a sustained high FØ value is tone spreading, and we have argued that the only tones eligible for spreading are floating tones. In view of the fact that the FØ is sustained in the middle of the intonation phrase, the floating tone must be the floating tone of a compound pitch accent. Thus, the only choices for describing such a plateau are H*+H$^-$ H*, with H$^-$ spreading to the right under Rule 1), or H* H$^-$+H*, with H$^-$ spreading to the left. Figure 5 is neutral between these two logically possible analyses, but an examination of other FØ contours suggests that H*+H$^-$ be added to the inventory. In Figure 14, the accent on "raingear" is more prominent than that on "bring" and, therefore, the H tone it carries is higher than the H on "bring." Now, it is the value on "bring" which is carried across the FØ plateau. This would be the expected outcome for the sequence H*+H$^-$ H*, since prominence is assigned per pitch accent and it is plausible that the two H tones in an H*+H$^-$ would have the same phonetic value. The analysis H* H$^-$+H* would require the less plausible assumption that H$^-$ takes on the prominence of the preceding accent rather than that of the accent it belongs to. A second contour which illustrates the need for the H*+H$^-$ accent is shown in Figure 15. Here, "med" in "intermediate" must have an H*+L$^-$ accent: The level of the nuclear accent is downstepped and the FØ falls over "diate." If the tonal transcription for "intermediate levels" were H* H$^-$+L*, the FØ on "diate" would remain higher, as in Figure 16. Now, in this theory, pitch accents may have an unstarred tone on the left or the right of the starred tone, but not both. Given that the accent on "intermediate" uses up its unstarred tone in generating downstep to the right, it cannot have an additional unstarred tone responsible for the FØ plateau on the left. Therefore, the H^- which produces this plateau must belong to a H^*+H^- accent on "many."

One of the important predictions of Rule 1) is that prominence relations, as reflected in phonetic value, can affect the applicability of tone spreading. This point has particular significance for the H^*+H^- accent, since this accent contrasts with H* only in environments where spreading can occur. Why is this so? Measurements of the L^*+H^- accent before L suggested that the target corresponding to the H^- is located as soon as possible after the L*. In the case of H^*+H^-, the speaker does not execute an FØ change between the H* and the H^-; thus, "as soon as possible" would in this case mean at the very location of the H*. Under these assumptions, the H^*+H^- in a nonspreading environment would be realized as a single peak, just as an H* is. Although one could conceive of a phonetics in which H^*+H^- and H* could be distinguished in non-spreading contexts by, say, breadth of peak, the assumptions made here are confirmed by our failure to find any such contrast in our own corpus or in the literature.

The consequence of this state of affairs is that the distinctive attributes of the H^*+H^- accent are seen only in level or rising configurations. Three typical contours with H^*+H^- accents are shown in Figures 17, 14, and 18. In Figure 17, the H^*+H^- accents on "took" and "advantage" have the same prominence as the H* on "Amanda", and so there is an FØ plateau extending all the way from the first stressed syllable to the nuclear stress.[3] In Figure 14, "raingear" is somewhat emphasized, with the result that the nuclear H* has greater prominence than the two prenuclear H^*+H^- accents, which share the same prominence. The FØ contour has a plateau from the first stressed syllable up to just before the nuclear stress, where there is a sudden rise. In Figure 18, "keywords" is more prominent than "indexed," and "abstract" is in turn more prominent than "keywords." So, the tonal transcription H^*+H^- H^*+H^- H* in this case results in an ascending staircase.

None of the FØ contours just discussed have the alternating prominence configuration which we have said was typical for a series of H* accents. Such a typical H* contour is shown in Figure 19. A hypothetical contour combining the alternating configuration of Figure 19 with the sustained FØ values of Figures 17 and 18 is sketched in Figure 20. Rule 1) predicts that such a contour would be impossible, because H^- in the H^*+H^- on "remember" should fail to spread before the lower valued accent on "bring." It is our impression that the intonation in Figure 20 does not in fact exist. Similarly, we would claim that the contour in Figure 21 is also impossible. Here, an inadmissible spreading of H^- in H^*+H^- before L^*+H^- has been attempted. The existing contour which comes closest to these two hypothetical contours is shown in Figure 22. The transcription for this contour is:

5) And remember to bring along your raingear.

H^*+H^- H^-+L^* H* L^- L%

In this intonation pattern, "bring" has an H^-+L* accent whose H^- enables the H^- of the preceding H*+H^- to spread. Since H^-+L* has its unstarred tone on the left rather than the right, there is no way of generating the sudden rise at this point in the contour shown in Figure 21, nor the plateau of Figure 20. Instead, there is a gradual rise from L* to the nuclear H*. The nuclear H* is downstepped after H^-+L*, but is sufficiently prominent that it still comes out higher than the L*.

Given the failure of H^- to spread before a lower valued H, the FØ contour in Figure 19 is additionally ambiguous. In Chapter 4, we pointed out that such a contour represent either H* H* H* L^- L% or H*+L^- H* H* L^- L%. Now we see that H*+H^- H* H* L^- L% is also a possibility, since H*+H^- H* and H* H* are neutralized when the prominence pattern is descending.

H*+H^- is also neutralized with H* in nuclear position. When the phrase accent is L^-, neutralization occurs because the condition for spreading is not met. When the phrase accent is H^-, we only find examples in which the H* value is sustained; there are none where we see dipping between H* and H^-, with H^- then spreading rightwards toward the boundary tone. This suggests either that H^- spreads leftward under equality with H*, or that the phrase accent is placed earlier when it matches the nuclear accent than in the cases of L* H^- and H* L^- examined in Chapter 2. Under either account, H*+H^- and H* are also neutralized before H^-.

5.4 Leftward Spreading

If leftward tone spreading exists in English, it is very restricted. Here, we will first review the facts for the phrase accent, and show that if it does spread left, it spreads only under phonetic equality to a preceding tone. Then we will consider pitch accents of the form T^-+T*. It will turn out that most and possibly all candidates for leftward spreading from a pitch accent can be explained without it. The hypothesis that English lacks leftward spreading is attractive, because it makes the lack of H^-+H* systematic rather than accidental.

In the discussion of rightward spreading of the phrase accent, there were four cases to consider. Fourteen cases are relevant to the question of whether the phrase accent spreads to the left, that is, all possible combinations of the seven pitch accents and the two phrase accents. Table I summarizes these cases and our observations about them.

The first eight entries in Table I indicate that leftward spreading is not allowable when the tone preceding T^- does not have the same phonetic value. Rule 2) of section 1 would generate the correct result for the remaining cases, where /T^-/ is the same as that of the preceding tone. However, in four of these cases, the correct result can clearly be generated without positing leftward spreading. In the cases of L*+H^- H^- and H*+H^- H^-, rightward spreading of the unstarred tone in the pitch accent independently gives the right outcome. For L* L^- and H^-+L* H^-, the monotonic interpolation which is observed generally between L and either L or H results in a flat F0 contour. This leaves

Table I

Pitch Accent	Phrase Accent	Phonetic Relation of Phrase Accent to Preceding Target	Correct Output Derived With Spreading	Correct Output Derived Without Spreading
$L*$	H^-	>	no	yes
$H*$	L^-	<	no	yes
$L*+H^-$	L^-	<	no	yes
L^-+H*	L^-	<	no	yes
$H*+L^-$	L^-	<	no	yes
$H*+L^-$	H^-	<	no	yes
H^-+L*	L^-	<	no	yes
$H*+H^-$	L^-	<	no	yes
$H*$	H^-	=	yes	maybe
$L*+H^-$	H^-	=	yes	yes
L^-+H*	H^-	=	yes	maybe
H^-+L*	H^-	=	yes	yes
$H*+H^-$	H^-	=	yes	yes
$L*$	L^-	=	yes	yes

the cases of H* H^- and L^-+H* H^-. We would expect dipping between H* and H^- in these two cases if there were no leftward spreading, provided that H^- was located as far from the starred nuclear tone as it is in H* L^- and L* H^-. The lack of dipping seems to imply that H^- has spread leftward. However, given that the resulting FO contour is flat, it is exactly not possible to see where the H^- was originally located. Thus, the correct result could also be generated if T^- is located earlier after a tone of the same value than after a differently valued tone, and spreads only rightward.

H^-+L* and L^-+H* are the two pitch accents with an unstarred tone which might in principle spread leftwards. Our discussion of the phrase accents suggests that L^-+H* will not provide evidence for leftward spreading: after H, L^- would fail to spread because of the phonetic inequality, whereas after L, monotonic interpolation independently generates a flat contour. Thus the issue comes down to the behaviour of the FO contour around H^-+L*. Figures 23 and 24 show, respectively, a contour in which H^- in H^-+L* is higher than a preceding tone, and one in which it is on the same level. In the first, it comes out as clear peak in the contour, while in the second, it marks the end of a plateau. Figure 24 could be generated from H* H^-+L* via leftward spreading. However, it could also arise from H*+H^- H^-+L* through rightward spreading of the first H^-. In order to argue for leftward spreading, it would be necessary to show that a contour with dipping between H* and H^-+L* is impossible. Unfortunately we do not know whether or not such contours are found.

In the last section, we showed that the H^*+H^- exists, but did not show that H^-+H^* is missing. The discussion in this section indicates that the H^-+H^* could be at best very marginal in the language, because it could be distinctive in only one context. Provided that leftward spreading existed, it would be distinctive only after a L^-+H^* with the same prominence, and would be neutralized with other analyses in all other situations. Two possible outcomes for this context are sketched in Figure 25. It seems to us that the contour with the plateau is quite odd compared to the one with dipping between the the H*s. The conclusion is that H^-+H^* does not exist as a distinctive type of pitch accent. The gap is systematic if there is no leftward spreading.

Footnotes for Chapter 5

1) In stating the rule this way, we do not mean to imply that the implementation system can generate and refer to arbitrarily fine differences in phonetic value. The conclusion that phonetic value is continuously variable is consistent with the assumption that comparison between different phonetic values is fuzzy. Under this assumption "$\leq$" means "not saliently lower than." How big a difference counts as salient is presumably determined by the precision of the production and perceptual systems, and might even vary with the speaker's carefulness of articulation or alertness.

2) Maeda does not examine intonation patterns involving L* accents. The theory outlined here predicts that L* accents could approach the baseline under sufficient emphasis.

3) Even though "advantage" has no FO inflection, it is possible to argue that it does have a pitch accent. The argument is based on the oddity of the contour sketched in Figure 17B, which would be the outcome for i):

i) He took advantage of Amanda

H* H* L^- L%

The reason this contour is odd is that it is hard to image a discourse context in which it would be appropriate to make "took" more prominent than "advantage," so that "took" got an accent and "advantage" did not. Since the relation between relative prominence and accentuation does not depend on type of accent, the normalness of the contour in Figure 12 leads to the inference that it has an accent on "advantage" in addition to the one on "took." Thus, not all pitch accents are implemented as FØ excursions.

Chapter 6

CONCLUSION

The preceding chapters have proposed a phonological representation for English intonation and a characterization of the rules which map it into a phonetic representation. The phonological representation has three components. First, there is a grammar which generates well-formed tunes for the intonation phrase. The tunes are structured strings of H and L tones. They consist of one or more pitch accents, which are either a single tone or a pair of tones on which a strength relation is defined, plus two extra tones which characterize the intonation at the end of the phrase, the phrase accent and the boundary tone. After a pause, there may also be a boundary tone at the beginning of the phrase. Expanding this grammar to provide the option of an extra phrase accent and possibly an extra boundary tone was proposed in order to account for the intonation of tags. The second component of the phonological representation is a metrical grid for the text of the phrase. The third is a set of rules which align the tune with the text, on the basis of the structure of the tune and the metrical representation for the text.

A quantitative representation of the intonation pattern is computed from the phonological representation by a package of local context-sensitive rules, which applies iteratively left to right. One class of rules evaluates each tone in baseline units above the baseline, a transform of the FØ domain. This transform is proposed because it gives the rules we have examined a simple mathematical form, and explains without additional assumptions the fact that downstepped patterns are asymptotic to the baseline. A second class of phonetic rules fills in the F0 contour between one target and the next. The most interesting of these rules was tone spreading, which spreads T^- when the next tone is equal or higher. One of the main themes of the thesis was that the phonetic implementation rules have interesting and language specific properties We argued that downstep and upstep should be accounted for by tone evaluation rules, and compared the English versions of these rules to versions required in African tone languages. Our tone spreading rule is another phonetic rule which is clearly language specific; we showed that it must be formulated in terms of phonetic values of tones rather than tone types, but it also makes use of a distinction between starred and unstarred tones which is not universal.

The framework just summarized was used to explicate a good number of English intonation patterns. At this point, we would like to go over what classes of cases motivated particular features of this framework.

A theory with two tones and context sensitive implementation rules resolves a number of problems with four tone theories, while still maintaining the advantages of describing intonation in terms of a sequence of tone levels. As Bolinger (1951) noted, positing four different levels leads to chronic ambiguity because it confounds tonal differences with differences in choice of pitch range. For example, the rendition of the H% L* H* L^- L% contour in Figure 1 could, in a four tone theory, count as an instance of LM L LM L L produced in a very large pitch range, as an instance of HM L HM L L produced in a large pitch range, or as an instance of H L H L L produced in a more moderate pitch range. This problem does not arise in a two tone theory because this contour has only the one transcription just given. Differences in overall range arise only from prominence differences. Ladd (1978a) discusses a set of cases in which a particular meaning difference can be associated with the presence or absence of a rise at the phrase boundary. As he points out, the cases with a rise do not count as a natural phonological class under a four tone theory. In our two tone theory, the contours with a rise are a natural class, because they all end in H%. As we pointed out in Chapter 2, this description of the relevant class seems to be an improvement on Ladd's since it extends more naturally to additional cases he did not take note of. A four tone theory in which each tone is assigned a fixed portion of the overall range also has difficulties with downstepped contours, which can exhibit more than four distinct levels. When context sensitive rules are introduced which can alter where in the range a given tone is implemented, two tones turn out to be sufficient to account for these contours.

The bitonal accents in our theory share an additional advantage of the approach in Bolinger (1958) over tone level theories in which each tone in the string is taken to be selected independently. A tone level theory of this character which restricts tones to metrically strong syllables fails to account for many intonation patterns, such as

L^* $H^-{+}L^*$ and $L^*{+}H^-$ L^-. On the other hand, if any syllable is allowed to carry a tone, the theory overgenerates. The observation to be captured is that metrically weak syllables may carry distinctive tone, but only on the strength of their proximity to a strong syllable (or of their occurrence in post nuclear position). This general observation is captured in Bolinger (1958) by describing pitch accents as F0 configurations. It is captured here by allowing the phonetic implementation to carry the unstarred tones in a bitonal accent off the syllable which is assigned the accent phonologically. We also have single tone accents, so that unlike Bolinger (1958), our theory does not require that a pitch accent induce tonal features on adjacent material.

A two-tone theory also offers some advantages over theories framed in terms of F0 changes. We noted a number of cases in which no F0 movement was associated with an accented syllable. These included the "contradiction contour" shown in Figure 2, which has no F0 movement on the nuclear stress, and the intonation pattern shown in Figure 3, which has no F0 movement on a word which can be argued to be accented. Analyzing the contours in this way made it possible to claim that the location of accents depends only on the metrical structure of the text, and not on the choice of accent type. In a theory like Clark (1978), in which pitch accents are defined as types of F0 movements, the rules for accent placement would be much more complex. (An F0 change theory which recognized "level" as a tonal primitive in addition to "rise" and "fall" could circumvent this problem by positing a level accent.)

There are also a number of cases which we feel are categorized more naturally in a tone level theory than in an FØ change theory. The theory proposed here recognizes the difference between H*+L^- H^- L% and H* L^- L%, illustrated in Figure 4, as a difference in type of intonation. In an FØ change theory, these are falls of two different sizes, just as the two examples of H* L^- L% in Figure 5 are, and thus count as the same intonation produced in different pitch ranges. A case which goes the other way is illustrated in Figure 6. In our view, these differ just as the contours in Figure 5 do; the nuclear accent is the same in both cases, but assignment of prominence has resulted in its being implemented in different pitch ranges. In an FØ change theory, on the other hand, the first contour would have to be characterized as rise-(level-)rise-fall, while the second would be rise-(level-)fall. These could not count as instances of the same intonation without positing an allomorphy rule. A third example is the relation of L* H^- H% questions to L*+H^- H^- H% questions discussed in Chapter 2. In our view, the L*+H^- adds just the same note of incredulity to a question that it can add in a declarative contour of the form L*+H^- L^- H%. An FØ change theory has no way of identifying the two nuclear accents in these cases, since one contour would be described as rise-(level-)rise, while the other is rise-fall-(level-) rise.

FØ change theories have recognized rising, falling, and sometimes level nuclear contours. When we examine the nuclear and post-nuclear intonation for contours with nuclear stress early in the phrase, we see that a theory must distinguish FØ movement near the nuclear stress from movement associated with the phrase boundary. For example, the L* H^- H% contour shown in Figure 7 would have to be described as rise-rise, or rise-level-rise. Once this decomposition is made, the FØ change theories offer no account of why there is no possibility of a fall in the phrase boundary slot. Given that the fundamental distinction is between rises and falls, with "level" playing a marginal role, we would expect either a two way distinction between rise and fall, or else a three way distinction between rise, level, and fall. The two tone theory, on the other hand, does have something to say about this gap. A three way distinction in treatment of the phrase boundary would require three different tone levels, whereas the theory has only two. The theory does not predict which of the three would be missing in a particular two-tone language, since languages may or may not upstep the boundary tone. Indeed we saw that Czech exhibits the pattern which is missing in English.

One of the issues raised in the thesis was to what extent intonation can be described using local specifications and rules. We claimed that the tonal correlates of the phrase taken as a whole are local: these are the phrase accent and the boundary tone, which are elements in the string of tones. We also claimed that the rules which implement tones phonetically are local. On the other hand, the interface between the tune and text could not be handled by local rules; the alignment of accents with the metrical structure is controlled by a nonlocal well-formedness condition. It is not possible to handle the alignment by a local rule applying left to right, because such a rule would have no way of distinguishing the nuclear stress of the phrase from a prenuclear local maximum. It is interesting that the nonlocal side of the intonation system arises exactly in its interaction with the metrical structure, which is well known for its nonlocal properties.

In this framework, features of the melody which are attributed to a nonlocal level of tonal representation in other accounts arise through the interaction of local specifications and rules. In particular, we found that generating downstepped contours using local rules correctly predicted that they would have an exponential form. A hierarchical representation of downstep was not adopted, because it neither supplanted nor supplemented phonetic rules relating tonal values. An approach in which pitch accents are superimposed on a phrasal intonation contour was also discussed. We suggested that the facts of Danish intonation, which make the strongest case we know of for such an approach, can also be accounted for in our framework using a downstep rule. We believe that the facts of English support a case against such an approach. Consider, for example, Figures 8 and 9, which exhibit downstep in only part of the phrasal contour. Such examples have a straightforward characterization in a theory in which downstep is a local rule applying to particular tonal configurations. Difficulties arise, however, if downstep is attributed instead to a separate layer of phrasal intonation. Under such an account, Figures 8 and 9 exhibit a complex phrasal contour;

8 might be a rise-fall, and 9 might be a level-fall. One consequence is that an account must be developed of what phrasal contours are possible, how they are aligned with the text, and how interpolation between one point of alignment and the next is carried out. A second is chronic ambiguity, which arises in much the same way as the chronic ambiguity in a four tone system. For example, an F∅ contour with two phonetically equal tones could be an instance of two phonologically equal tones riding on a flat phrasal contour, or it could represent a lower and a higher tone riding on a falling contour. The basic problem, as we pointed out in Chapter 3, is that recovering two independently varying layers from an F∅ contour is a mathematically underdetermined problem if the variation in either is too rich. One reason we pursued the hypothesis that declination is fixed was to avoid this problem.

A number of important problems have gone unanswered here. Our account of text-tune alignment was incomplete in several respects. One question is whether the metrical tree of the text can be dispensed with. Taking the metrical grid to be a device for interpreting the tree leaves two avenues for the pragmatics to influence the stress subordination: tree labelling, and options in constructing the grid. Given that the consequences in both cases are the same, it would clearly be desirable to posit only a single mechanism. A second question is how relative prominence, as it controls pitch range, is related to the metrical grid. Relative prominence, which is continuously valued, could be a more detailed specification of the grid. Or, it is possible that several different interpretations of the grid are constructed which separately control the different phonetic reflexes of prosody. We were also vague about whether the tune has a metrical grid of its own, and if so, how the phrase accent and boundary tone are represented in it so as to capture their alignment properties. These questions are important to an explanation of text-tune alignment in tags.

The evaluation of H tones was more thoroughly investigated in the thesis than the evaluation of L tones. More data on how L tones scale are clearly needed. A theoretical question which such data would help to clear up is whether the only representation of pitch range is the values of occurring tones, or whether the range should be reified in the description as proposed in Clements (1979). We are also lacking a theoretical explanation of a striking regularity in the rules for L we did propose: wherever possible, /L/ is computed in relation to the value of a neighboring H rather than a neighboring L.

Associated with our claim that tonal implementation rules were local was a hypothesis about what the window for such rules is in English: they have no right context and can refer only as far back as the previous pitch accent. This hypothesis will not serve as a language universal: we mentioned a case in Zulu in which right phonological context for a tonal implementation rule is needed, and also many languages do not have their tones organized into pitch accents. An important problem is what the universal constraints on the window for tonal implementation are.

Lastly, we have not said much about the respiratory and laryngeal control underlying FØ contours. There are three points on which questions about production are clearly related to our description. One is the timing of unstarred tones in bitonal accents. A second is the scaling in baseline units above the baseline, which could arise as a side effect of how respiratory and laryngeal control interact. The third is the character of the interpolation rules. Different accounts of interpolation could give rise to different tonal analyses of particular contours, and so a more thorough study of interpolation is important both to a phonetic and a phonological account of intonation.

REFERENCES

Abe, I., and T. Kanekiyo, eds. (1965). Forms of English. Harvard University, Cambridge, MA.

Armstrong, L., and I. Ward (1926). A Handbook of English Intonation. Teubner, Leipzig and Berlin.

Ashby, M. (1978). "A Study of Two English Nuclear Tones." Language and Speech 21 (4), 326-336.

Berman, A., and M. Szamosi (1972). "Observations on Sentential Stress." Language 48, 304-325.

Bing, Janet (1979). Aspects of English Prosody. Unpublished Ph.D dissertation, University of Massachusetts, Amherst.

Bolinger, D. (1951). "Intonation: Levels versus Configurations." Word 7(1), 199-210 (also in Abe and Kanekiyo).

-- (1957-1958). "Intonation and Grammar." Language Learning 8(1,2), 31-38 (also in Abe and Kanekiyo).

-- (1958). "A Theory of Pitch Accent in English." Word 14(2-3), 109-149 (also in Abe and Kanekiyo).

Bolinger, D. (1972). "Accent is Predictable (If You're a Mind-Reader)." Language 48, 633-644.

-- (1978). "Intonation Across Languages." In J.H. Greenberg et al., 471-524.

Boyce, S., and L. Menn (1979). "Peaks Vary, Endpoints Don't: Implications for Linguistic Theory." In Proceedings of the Fifth Annual Meeting of the Berkeley Linguistics Society.

Breckenridge, J., and M. Liberman (1977). "The Declination Effect in Perception." Bell Laboratories Technical Memorandum, Murray Hill, N.J.

Bresnan, J. (1971). "Sentence Stress and Syntactic Transformations." Language 47, 257-281.

-- (1972). "Stress and Syntax: A Reply." Language 48, 326-342.

Bruce, G. (1977). Swedish Word Accents in Sentence Perspective. Travaux de L'Institut de Linguistique de Lund, CWK Gleerup, Malmö.

-- and E. Gårding (1978). "A Prosodic Typology for Swedish Dialects." In E. Garding, G. Bruce, and R. Bannert (eds.), Nordic Prosody. Travaux de l'Institut de Linguistique de Lund 13, Department of Linguistics, Lund University, 219-228.

Carlson, L. (MS). Dialogue Games: An Approach to the Theory of Text. Academy of Finland.

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

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

Clements, G.N. (1979). "The Description of Terraced-Level Tone Languages." Language 55(3), 536-558.

-- (1980). "The Hierarchical Representation of Tone Features." To appear in Harvard Studies in Phonology, Vol. 2.

-- and K. Ford (1979). "Kikuyu Tone Shift and its Synchronic Consequences." Linguistic Inquiry 10, 179-210.

Cole, R.A. (ed.) (1980). Perception and Production of Fluent Speech. Lawrence Erlbaum Associates, Hillsdale.

Collier, R. (1975). "Physiological Correlates of Intonation Patterns." J. Acoust. Soc. Am. 58, 249-255.

-- and J. 't Hart (1971). "A Grammar of Pitch Movements in Dutch Intonation." IPO Annual Progress Report 6, 17-21.

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

Cope, A.T. (1970). "Zulu Tonal Morphology." Journal of African Languages 9(3), 111-152.

Crystal, D. (1969). Prosodic Systems and Intonation in English. Cambridge University Press, London.

Downing, B. (1970). Syntactic Structure and Phonological Phrasing in English. Unpublished Ph.D. dissertation, University of Texas.

Enkvist, N.E., and H. Nordstrom (1978). In B. Sigurd (ed.), Festschrift for Bertil Malmberg, Studia Linguistica, CWK Gleerup, Lund.

Fromkin, V. (1972). "Tone Features and Tone Rules." In Studies in African Linguistics 3(1).

Fry, D.B. (1958). "Experiments in the Perception of Stress." Language and Speech 1, 125-152.

Fukisaki, M., K. Hirose and K. Ohta (1979). "Acoustic Features of the Fundamental Frequency Contours of Declarative Sentences in Japanese." Research Institute of Logopedics and Phoniatrics Annual Bulletin No. 3, University of Tokyo, 163-172.

Goldsmith, J. (1976). Autosegmental Phonology. Ph.D. dissertation, Massachusetts Institute of Technology, published in 1979 by Garland Publishing Inc., New York.

Greenberg, J.H., C.A. Ferguson and E.A. Moravcsik (eds.) (1978). Universals of Human Language, Vol. 2: Phonology. Stanford University Press, Stanford.

Gunter, R. (1974). Sentences in Dialog. Hornbeam Press, Columbia, S.C.

Hadding-Koch, K., and M. Studdert-Kennedy (1964). "An Experimental Study of Some Intonation Contours." Phonetica 11, 175-185.

Halliday, M.A.K. (1967). Intonation and Grammar in British English. Mouton, The Hague.

Hayes, B. (1980). A Metrical Theory of Stress Rules. Unpublished Ph.D. dissertation, Massachusetts Institute of Technology.

Hirst, D.J. (1974). "Intonation and Context." Linguistics 141, 5-16.

Hirvonen, P. (1970). Finnish and English Communicative Intonation. Publications of the Phonetics Department of the University of Turku, No. 8.

Hodge, C., and H. Hause (1944). "Hausa Tone." Journal of the American Oriental Society 64(2).

Hombert, J.M. (1974). "Universals of Downdrift: Their Phonetic Basis and Significance for a Theory of Tone." Studies in African Linguistics, Suppl. 5.

-- J.J. Ohala, and W.G. Ewan (1979). "Phonetic Explanations for the Development of Tones." Language 55(1), 37-58.

House, A., and G. Fairbansk (1953). "The Influence of Consonant Environment upon the Secondary Acoustical Characteristics of Vowels." J. Acoust. Soc. Am. 25(1), 105-113.

Huang, C.T.J. (1979). The Metrical Structure of Terraced-Level Tones. Unpublixhed paper, Massachusetts Institute of Technology (abridged version under the same title in NELS 10, Cahiers Linguistiques d'Ottawa 9, 1980).

Hyman, L.M., and R.G. Schuh (1974). "Universals of Tone Rules: Evidence from West Africa." Linguistic Inquiry 5(1), 81-115.

Jackendoff, R. (1972). Semantic Interpretation of Generative Grammar. M.I.T. Press, Cambridge, MA.

Kingdon, R. (1958). The Groundword of English Intonatior. Longmans, Green, London.

Kiparsky, P. (1977). "The Rhythmic Structure of English Verse." Linguistic Inquiry 8(2):189-248.

Ladd, D.R. (1978). The Structure of Intonational Meaning. Ph.D. Dissertation, Cornell University.

-- (1978a). "Stylized Intonation." Language 54, 517-540.

Ladefoged, P., and D.E. Broadbent (1956). "Information Conveyed by Vowels." J. Acoust. Soc. Am. 29(1), 98-104.

Larson, J. (1971). "Downstep, Downdrift and Diacritics." Studies in African Linguistics, Suppl. 2, 171-181.

Lea, W. (1973). "Segmental and Suprasegmental Influences on Fundamental Frequency Contours." In Hyman, L. (ed.), Consonant Types and Tones. Southern California Occasional Papers in Linguistics No. 1.

Leben, W. (1976). "The Tones in English Intonation." Linguistic Analysis 2, 69-107.

Lehiste, I. (1970). Suprasegmentals. M.I.T. Press, Cambridge, MA.

-- (1975). "The Phonetic Structure of Paragraphs." In A. Cohen and S.G. Nooteboom (eds.), Structure and Process in Speech Production, Springer Verlag, Berlin.

Liberman, M. (1975). The Intonation System of English. Ph.D. Dissertation, Massachusetts Institute of Technology. Reproduced by the University of Indiana Linguistics Club, Bloomington.

-- and J. Pierrehumbert (1979). "A Metric for the Height of Certain Pitch Peaks in English." J. Acoust. Soc. Am. 66, Suppl. 1, S64.

-- and A. Prince (1977). "On Stress and Linguistic Rhythm." Linguistic Inquiry 8(2), 249-336.

-- and I. Sag (1974). "Prosodic Form and Discourse Function." Papers from the Tenth Regional Meeting of the Chicago Linguistic Society, 416-426.

Lieberman, P. (1967). Intonation, Perception, and Language. M.I.T. Press, Cambridge, MA.

Maddieson, I. (1978). "Universals of Tone." In Greenberg et al., 335-366.

Maeda, S. (1976). A Characterization of American English Intonation. Unpublished Ph.D. dissertation, Massachusetts Institute of Technology.

McCawley, J. (1970). "A Note on Tone in Tiv Conjugation." Studies in African Linguistics 1, 123-129.

Meyers, L. (1976). Aspects of Hausa Tone. Working Papers in Phonetics 32, University of California at Los Angeles.

Morton, J. and W. Jassem (1965). "Acoustic Correlates of Stress." Language and Speech 8, 159-181.

Mosteller, F. and J.W. Tukey (1977). Data Analysis and Regression: A Second Course in Statistics. Addison-Wesley, Reading, MA.

Nakagawa, S., and T. Sakai (1979). "A Recognition System of Connected Spoken Words Based on Word Boundary Detection." Studia Phonologica 13, 23-38.

Nakatani, L. and C. Aston (1978). "Acoustic and Linguistic Factors in Stress Perception." Bell Labs Technical Memorandum, Dept. of Speech and Communication Research (forthcoming in J. Acoust. Soc. Am.).

-- and J. Schaffer (1978). "Hearing 'Words' Without Words: Prosodic Cues for Word Perception." J. Acoust. Soc. Am. 63(1), 234-244.

Nash, R., and A. Mulac (1980). "The Intonation of Verifiability." In L.R. Waugh and C.H. van Schooneveld (eds.), 219-242.

O'Connor, J.D. and G.F. Arnold (1961). Intonation of Colloquial English. Longmans, London (new edition 1963).

Ohala, J. (1978). "Production of Tone." In V. Fromkin, (ed.) Tone A: Linguistic Survey. Academic Press.

Olive, J. (1974). "Speech Synthesis by Rule". In papers presented at the Speech Communication Seminar, Stockholm, August 1-3.

O'Shaughnessy, D. (1976). Modelling Fundamental Frequency, and its Relationship to Syntax, Semantics, and Phonetics. Unpublished Ph.D. dissertation, Massachusetts Institute of Technology.

Palmer, M.E. (1922). English Intonation, with Systematic Exercises. Heffer, Cambridge, England.

Painter, C. (1979). "The Pitch Range of Tonal Units in Gwa." Zeitschrift fur Phonetik Sprachwissenschaft und Kommunikations forschung 32, 272-281.

Peters, A. (1973). "A New Formalization of Downdrift." Studies in African Linguistics 4(2), 139-154.

Peterson, G.E., and H.L. Barney (1951). "Control Methods Used in A Study of Vowels." J. Acoust. Soc. Am. 24(2), 175-184.

Pierrehumbert, J. (1979a). "Intonation Synthesis Based on Metrical Grids." In J.J. Wolf and D.H. Klatt (eds.), Speech Communication Papers Presented at the 97th Meeting of the Acoustical Society of America. Acoustical Society of America, New York, 523-526.

-- (1979b). "The Perception of Fundamental Frequency Declination." J. Acoust. Soc. Am. 66(2), 363-369.

-- (1980). "Synthesizing Intonation." Forthcoming in J. Acoust. Soc. Am.

Pike, K.L. (1945). The Intonation of American English. University of Michigan, Ann Arbor.

Rando, E. (1980). "Intonation in Discourse." In L.R. Waugh and C.H. van Schooneveld, 243-278.

Sag, I. and M. Liberman (1975). "The Intonational Disambiguation of Indirect Speech Acts." In Papers from the Eleventh Regional Meeting of the Chicago Linguistic Society, 487-497.

Schachter, P. and V. Fromkin (1968). A Phonology of Akan: Akuapem, Asante and Fante. Working Papers in Phonetics No. 9. University of California, Los Angeles.

Schmerling, S. (1974). "A Re-examination of Normal Stress." Language 50, 66-73.

Selkirk, E. (1980). "The Role of Prosodic Categories in English Word Stress." Linguistics Inquiry 11(3), 563-606.

Silverstein, R. (1976). "A Strategy for Utterance Production in Hausa." Studies in African Linguistics, Suppl. 6, 233-242.

Sledd, J. (1956). "Superfixes and Intonation Patterns." Litera 3, 35-41.

Sorensen and Cooper (1980). "Syntactic Coding of Fundamental Frequency in Speech Production." In R.A. Cole (ed.), 399-440.

Sternberg, S., C.E. Wright, R.L. Knoll, and S. Monsell (1980). "Motor Programs in Rapid Speech: Additional Evidence." In R.A. Cole (ed.), 507-534.

Stewart, J.M. (1965). "The Typology of the Twi Tone System." Bulletin of the Institute of African Studies 1, 1-27. University of Ghana, Legon.

Streeter, L. (1978). "Acoustic Determinants of Phrase Boundary Perception." J. Acoust. Soc. Am. 64(6), 1582-1592.

J.H. Hart and A. Cohen (1973). "Intonation by Rule: A Perceptual Quest." J. of Phonetics 1(4), 309-327.

Thorsen, N. (1978). "An Acoustical Analysis of Danish Intonation." J. Phonet. 6, 151-175.

-- (1979a). "Lexical Stress, Emphasis for Contrast, and Sentence Intonation in Advanced Standard Copenhagen Danish." In Proceedings of the Ninth International Congress of Phonetic Sciences. Institute of Phonetics, Copenhagen, 417-423.

-- (1979b). "Interpreting Raw Fundamental Frequency Tracings of Danish." Phonetica 36, 57-78.

-- (1979c). "Intonation Contours in Declarative Sentences of Varying Length in ASC Danish." Annu. Rep. Inst. Phonet. Univ. Copenhagen 13, 1-7.

-- (1979d). "Lexical Stress, Emphasis for Contrast, and Sentence Intonation in Advanced Standard Copenhagen Danish." Annu. Rep. Inst. Phonet. Univ. Copenhagen 13, 58-85.

-- (1980). "A Study of the Perception of Sentence Intonation -- Evidence from Danish." J. Acoust. Soc. Am. 67(3), 1014-1030.

Trager, G.L., and H.L. Smith, Jr. (1951). An Outline of English Structure 41. Battenburg Press, Norman, Oklahoma.

Vaissiere, J. (1971). Contribution a la Synthese par Regles du Français. Unpublished doctoral dissertation, Universite des Langues et Lettres de Grenoble.

Vanderslice, R., and P. Ladefoged (1972). "Binary Suprasegmental Features and Transformational Word-Accentuation Rules." Language 48(4), 819-839.

Wales, R., and H. Toner (1979). "Intonation and Ambiguity." In W.E. Cooper and E.C. Walker (eds.), Sentence Processing: Psycholinguistic Studies Presented to Merrill Garrett. Halsted Press.

Waugh, L.R. and C.H. Van Schooneveld (eds.) (1980). The Melody of Language. University Park Press, Baltimore.

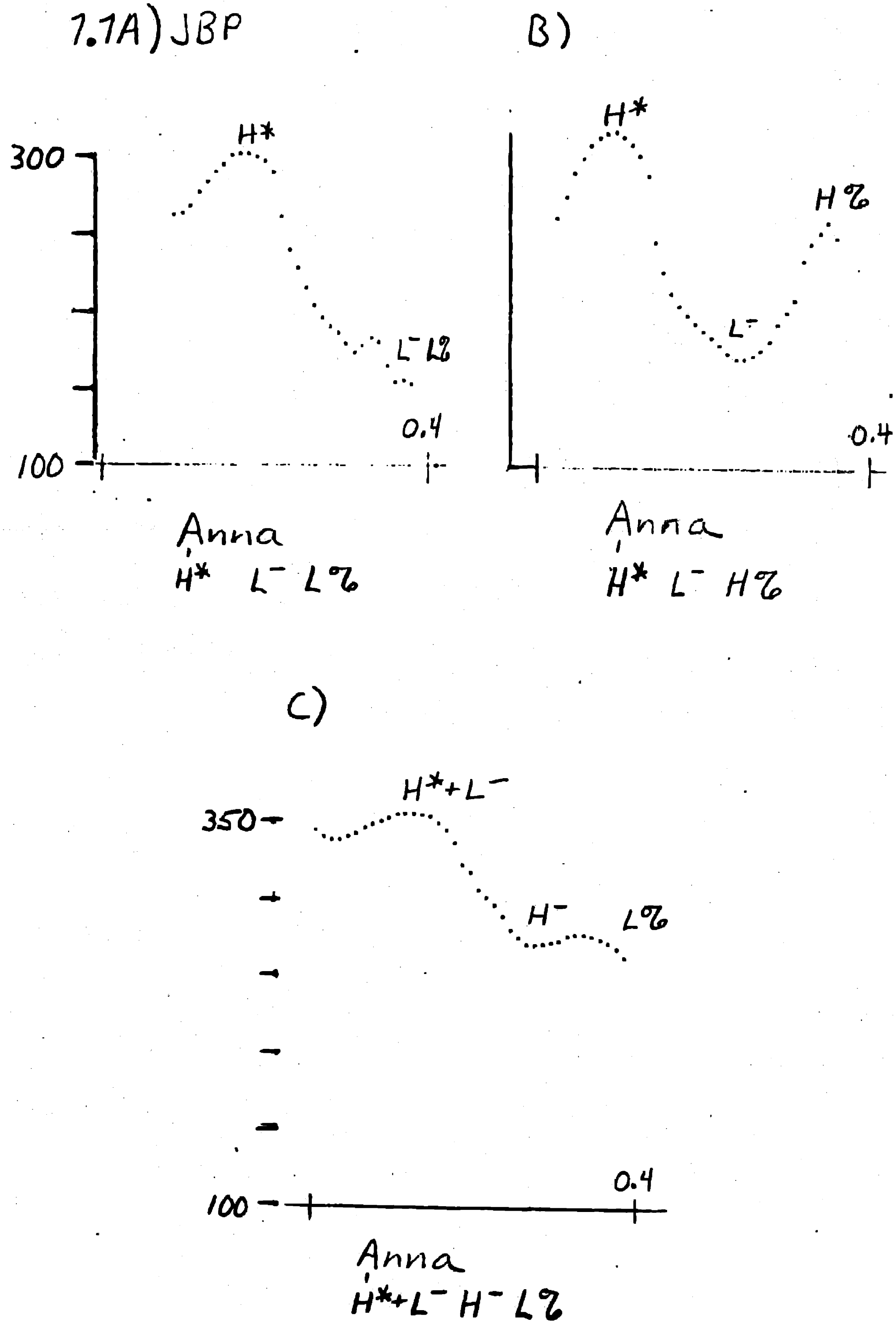
Yip, M. (1980). The Tonal Phonology of Chinese. Unpublished Ph.D. dissertation, Massachusetts Institute of Technology.

Figures

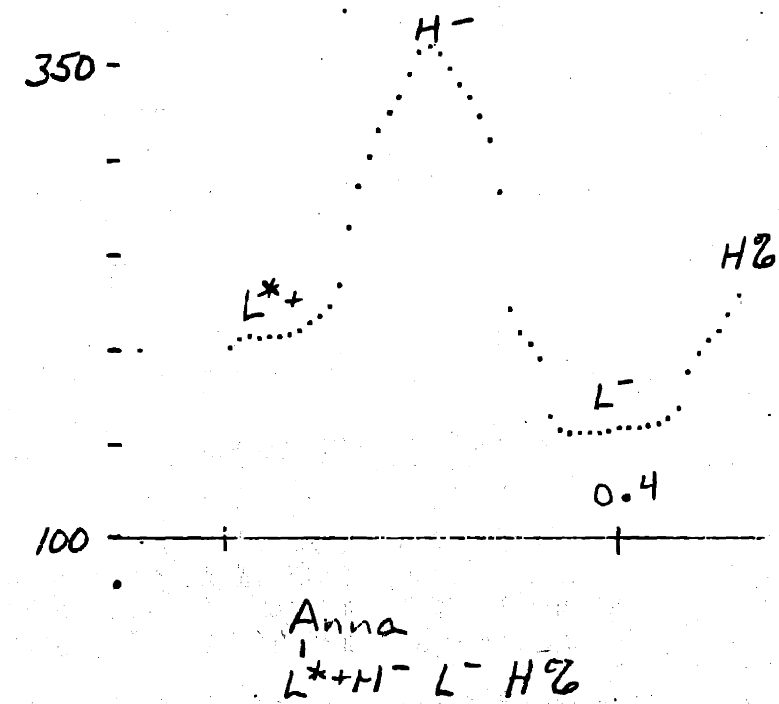
In the Figures, the vertical axis is fundamental frequency, in Hz. The horizontal axis is time, with ticks marking one second intervals unless otherwise indicated. To facilitate comparisons, the initials of the speaker for each FØ contour are given in the upper left hand corner. The line-up of the FØ contour with the text is indicated by two devices. A circled letter with an arrow points to the region in the FØ contour corresponding to the phoneme indicated. Also, a tonal transcription is given under the text and marked on the FØ contour. In some cases, additional transcriptions besides the one indicated would be possible; multiple transcriptions are given only when the ambiguity is important to the discussion.

Where low FØ values at the end of the phrase (L^- and L%) were produced with vocal fry, this is indicated by a scatter of points in the vicinity of the baseline for the speaker. In such cases, the pitch tracker fails to compute a continuous FØ contour, and can output values scattered over the whole range. The representation used was chosen to reflect the imporession on the ear made by such intonation patterns, and to make them visually distinct from contours ending in high FØ values.

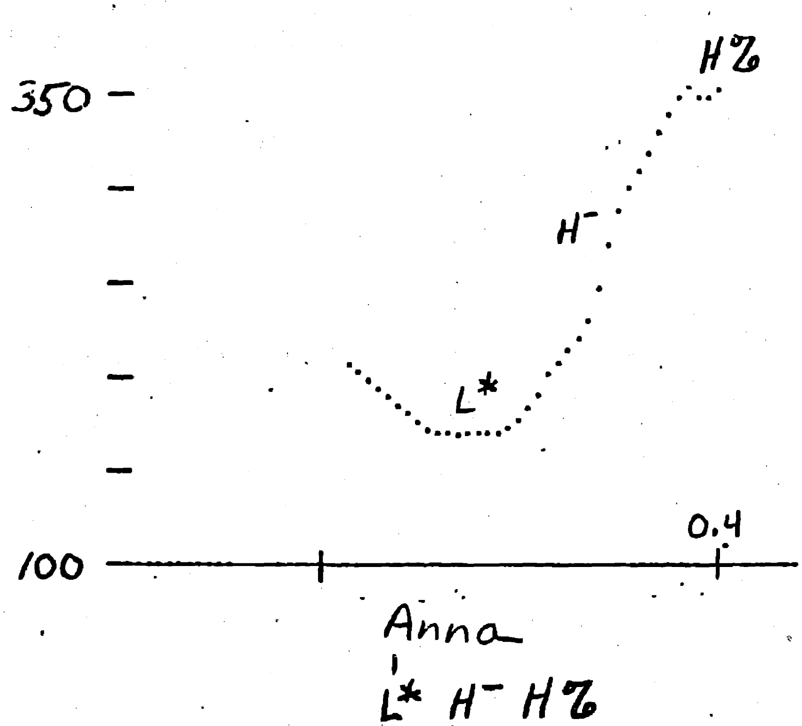
7.1A) JBP


300
H*
L⁻L%
0.4
100
Anna
H* L⁻ L%
B)
H*
H%
L⁻
0.4
Anna
H* L⁻ H%
C)
H*+L⁻
350
H⁻
L%
0.4
100
Anna
H*+L⁻ H⁻ L%

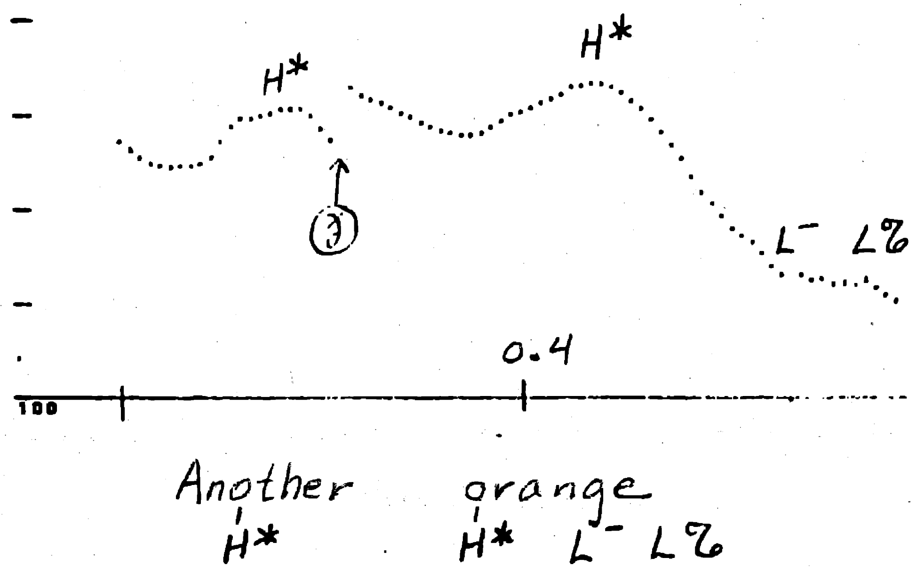
7.1 D)



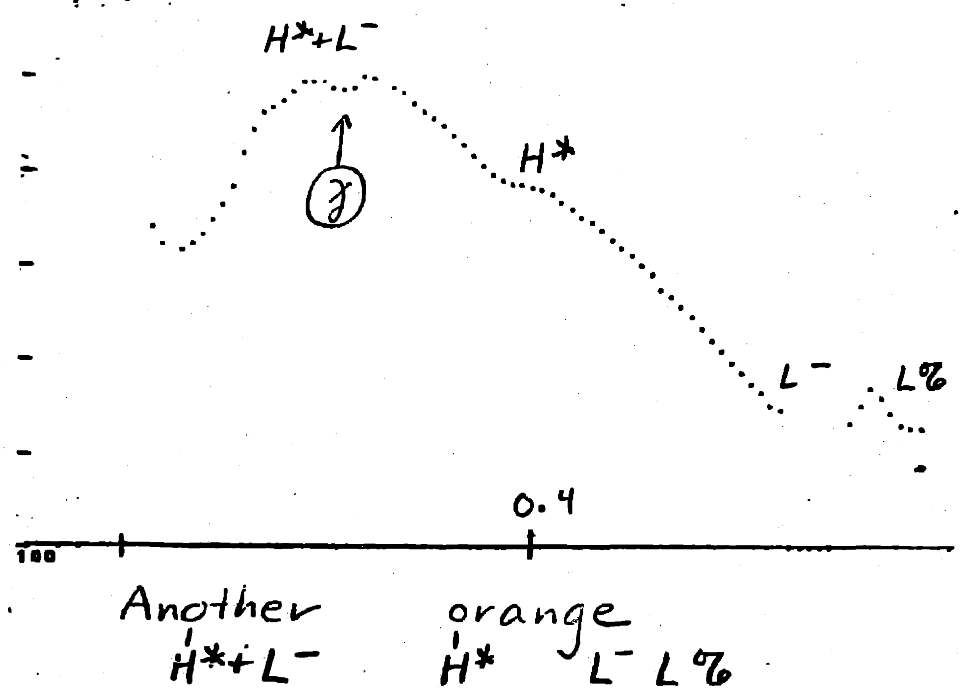
E)



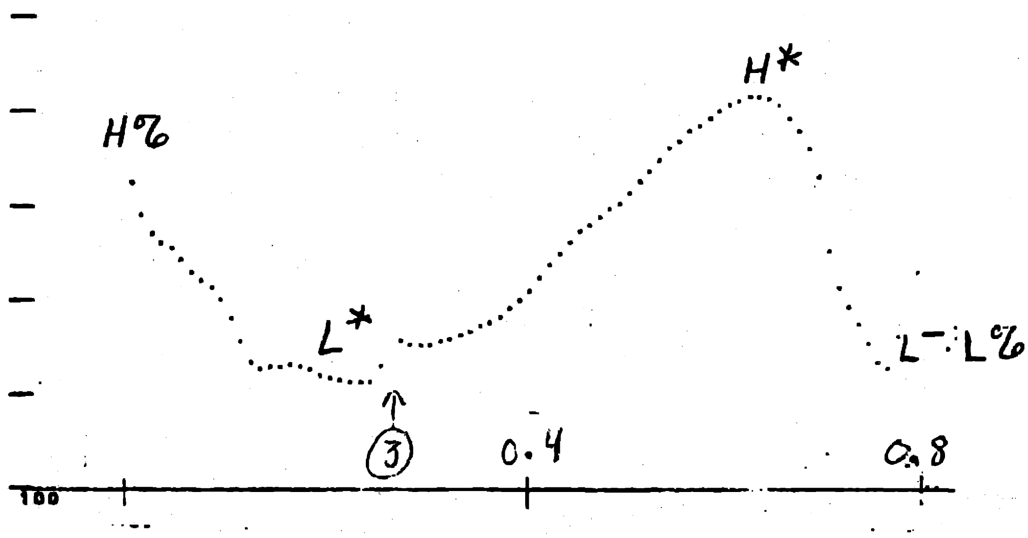
1.2A) JBP


H*
H*
ʒ
L⁻ L%
0.4
100
Another
H*
orange
H* L⁻ L%

B)


H*+L⁻
H*
ʒ
L⁻
L%
0.4
100
Another
H*+L⁻
orange
H* L⁻ L%

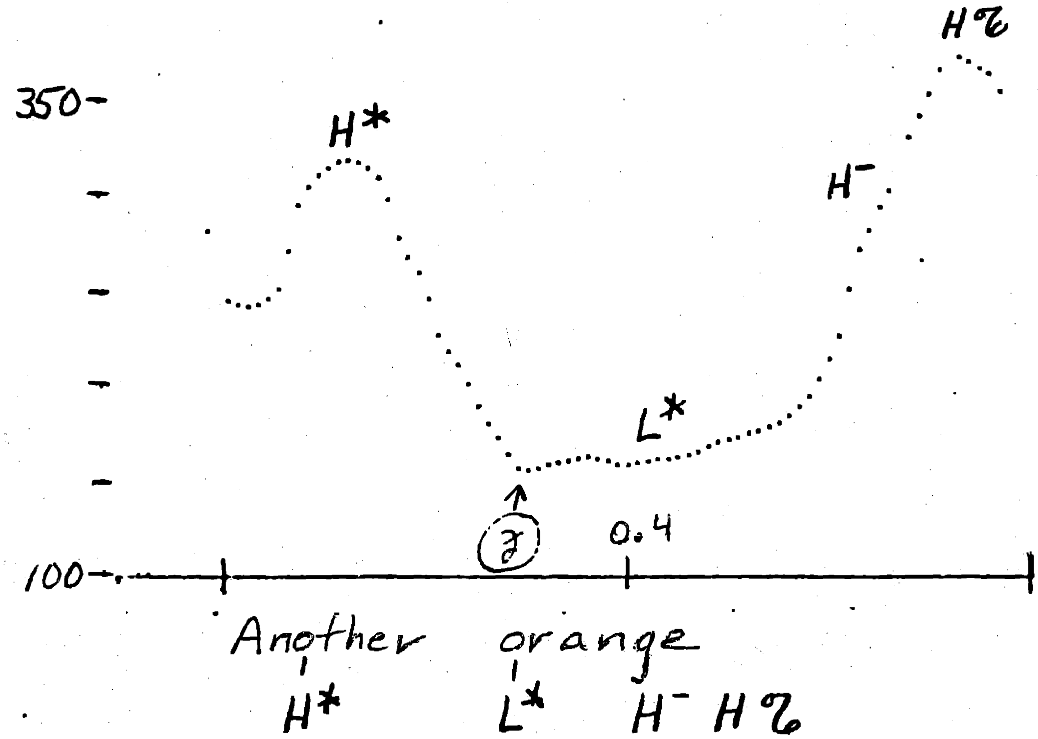
c)



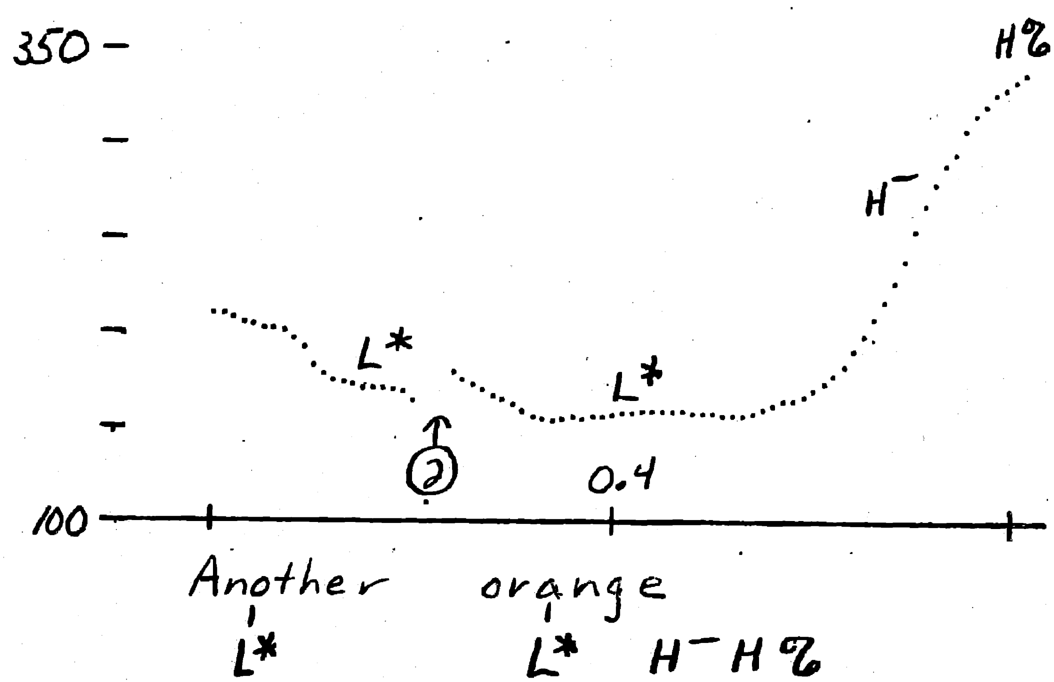
Another orange

H% L* H* L⁻ L%

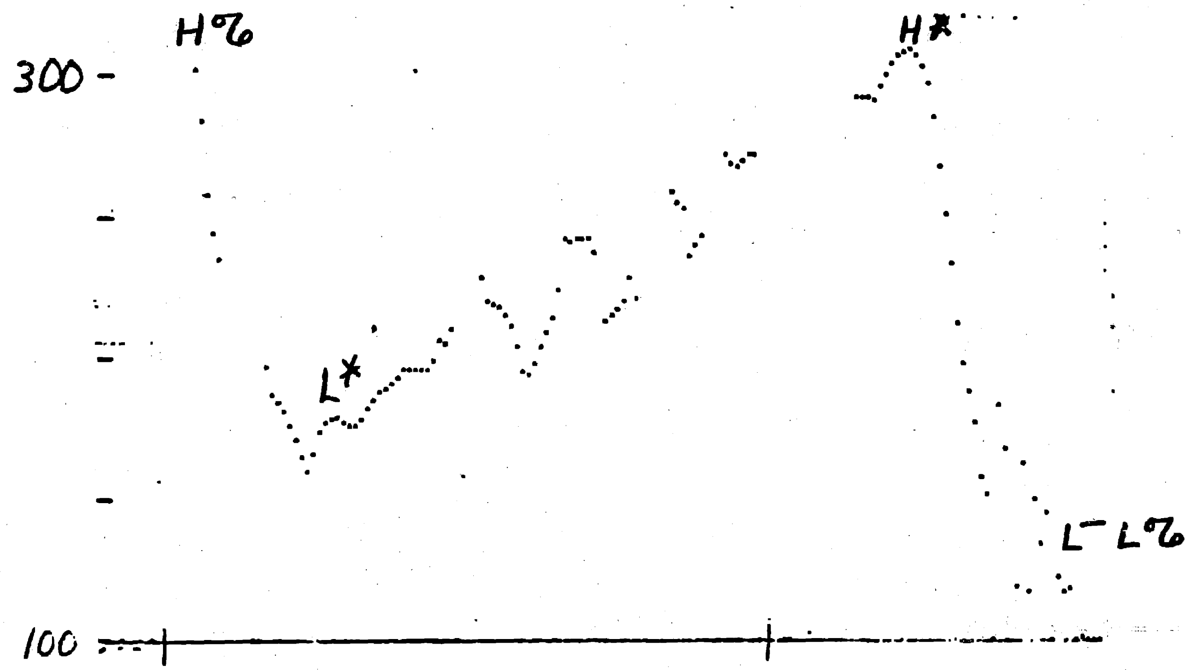
7.3A) JBP


H%
350
H*
H-
L*
0.4
100
Another orange
H* L* H- H%

B)


350
H%
H-
L*
L*
0.4
100
Another orange
L* L* H- H%

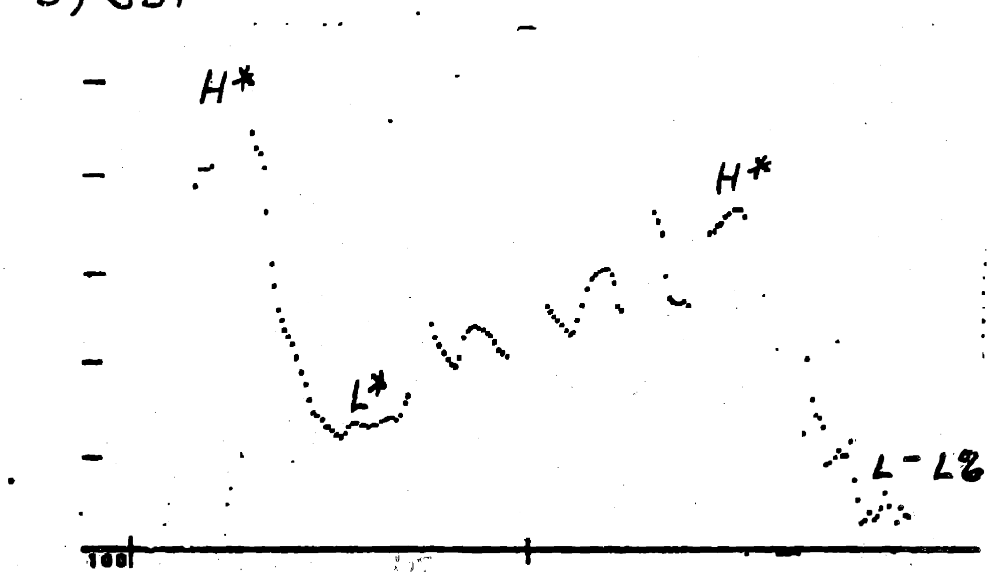
7.4A) JBP



It's really too good to be true

H% L* H* L⁻ L%

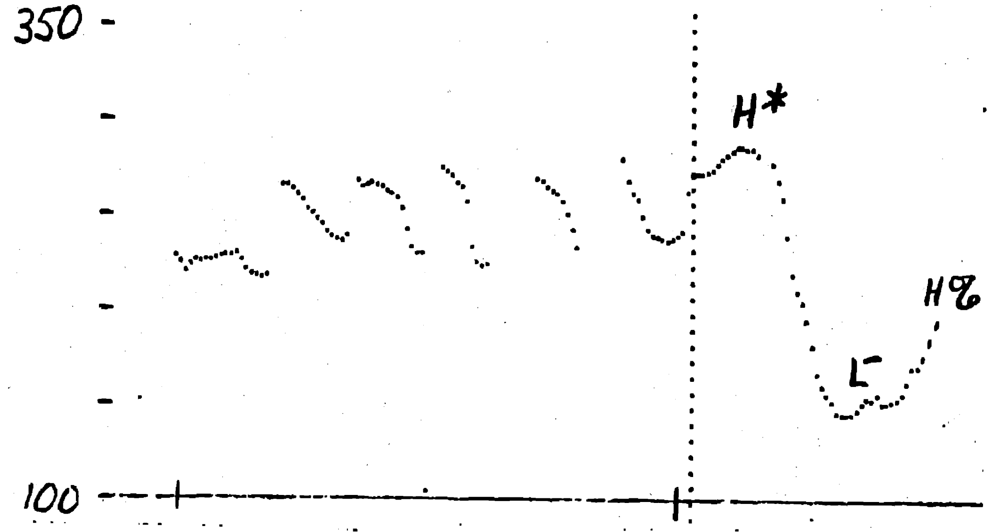
B) JBP



That's a remarkably clever suggestion

H% L* H* L⁻ L%

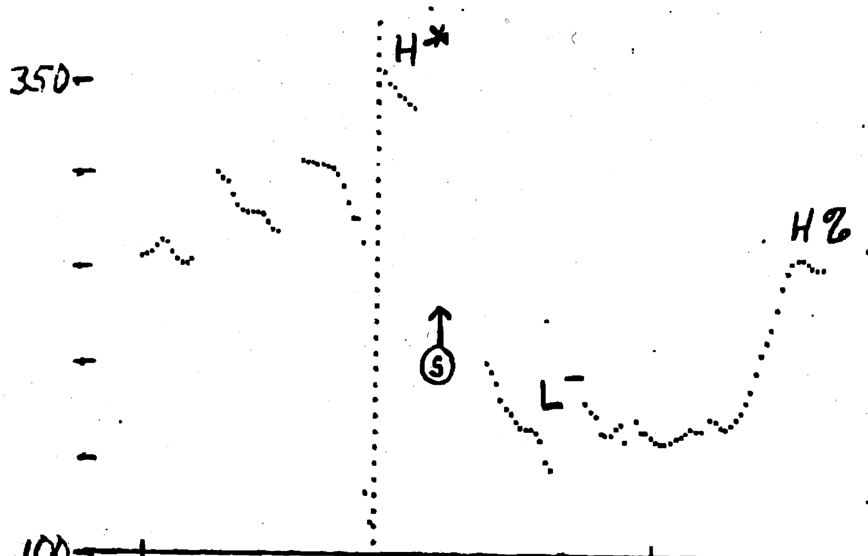
7.5 A)



Legumes are a good source of vitamins.

H* L⁻ H%

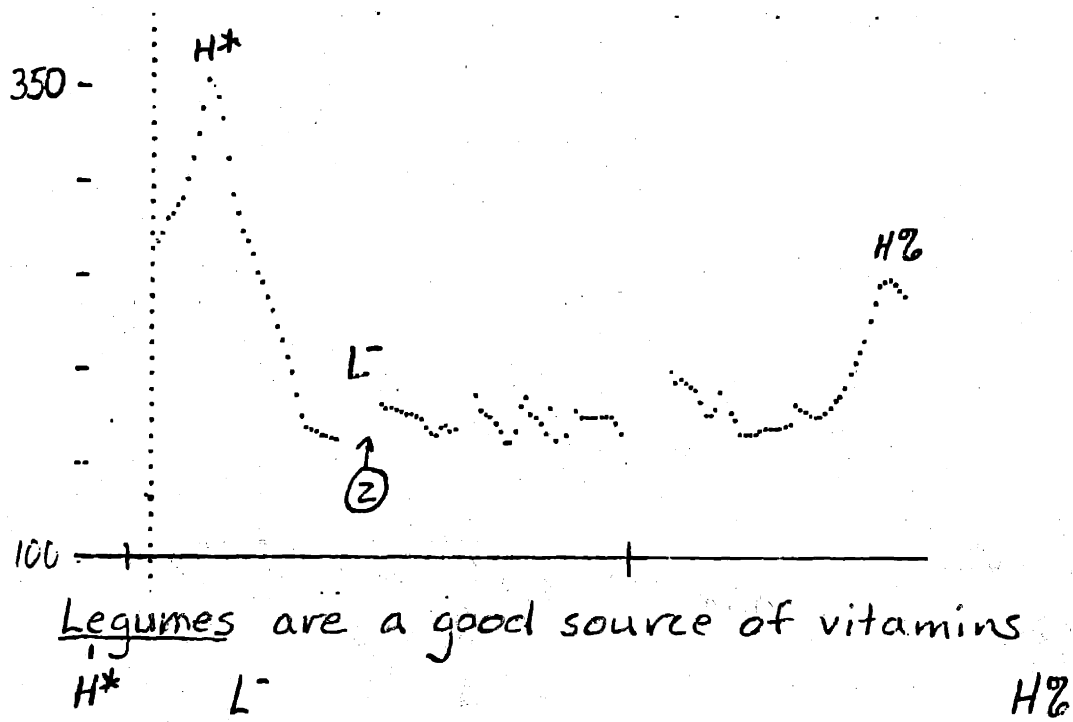
B)



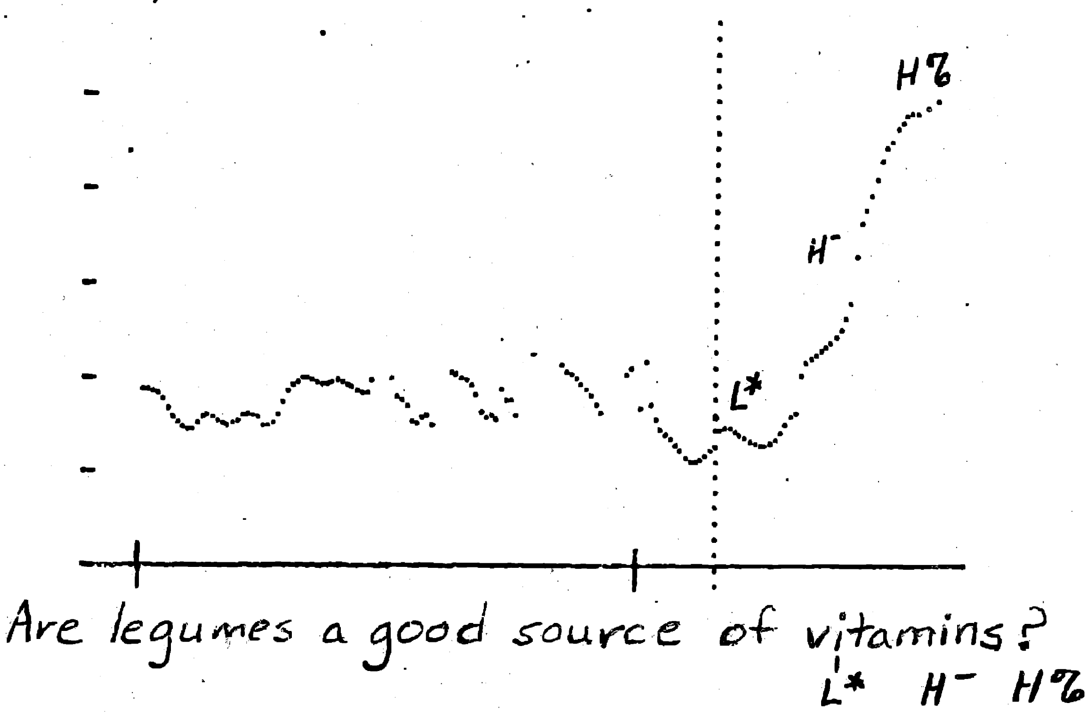
Legumes are a good source of vitamins.

H* L⁻ H%

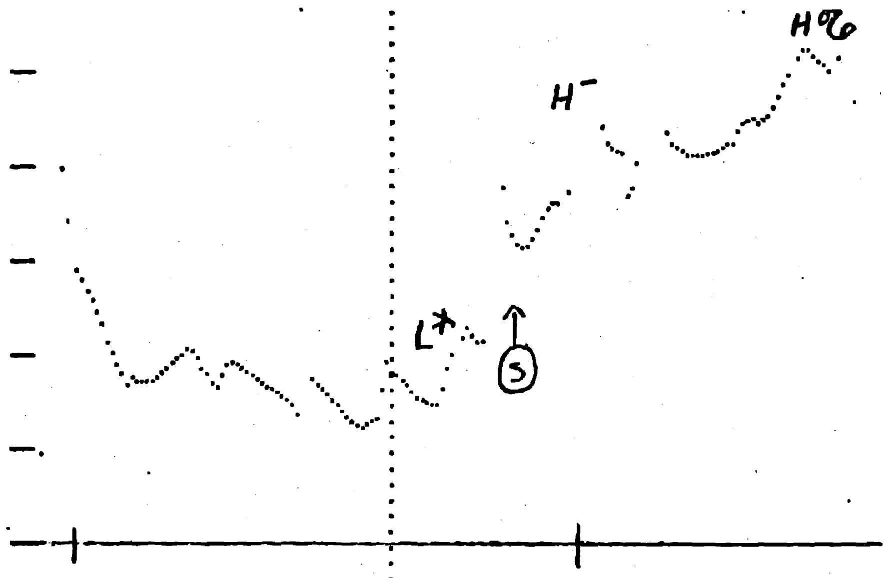
1.5C)


H*
350
L⁻
2
H%
100
Legumes are a good source of vitamins
H* L⁻ H%

1.6A)

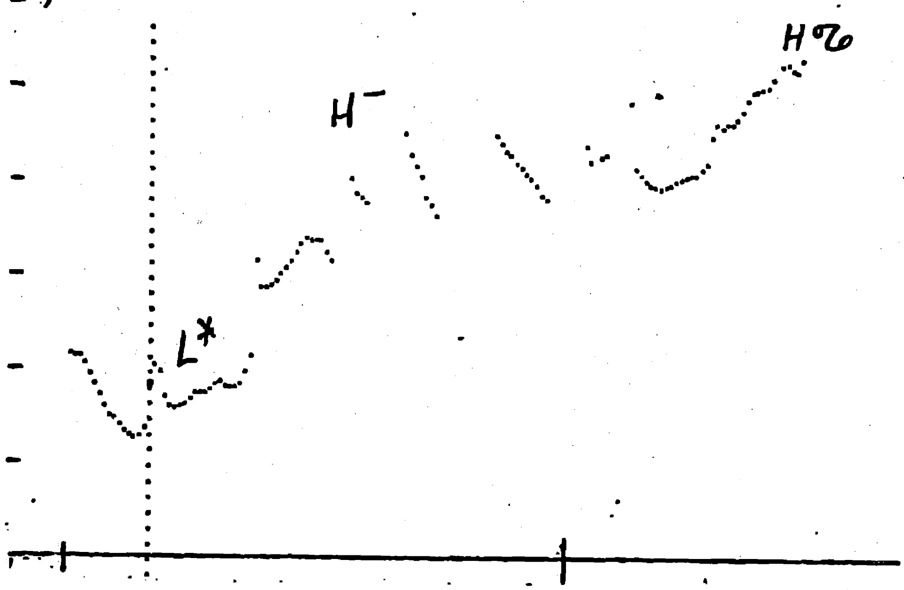

H%
H⁻
L*
Are legumes a good source of vitamins?
L* H⁻ H%

7.6B)



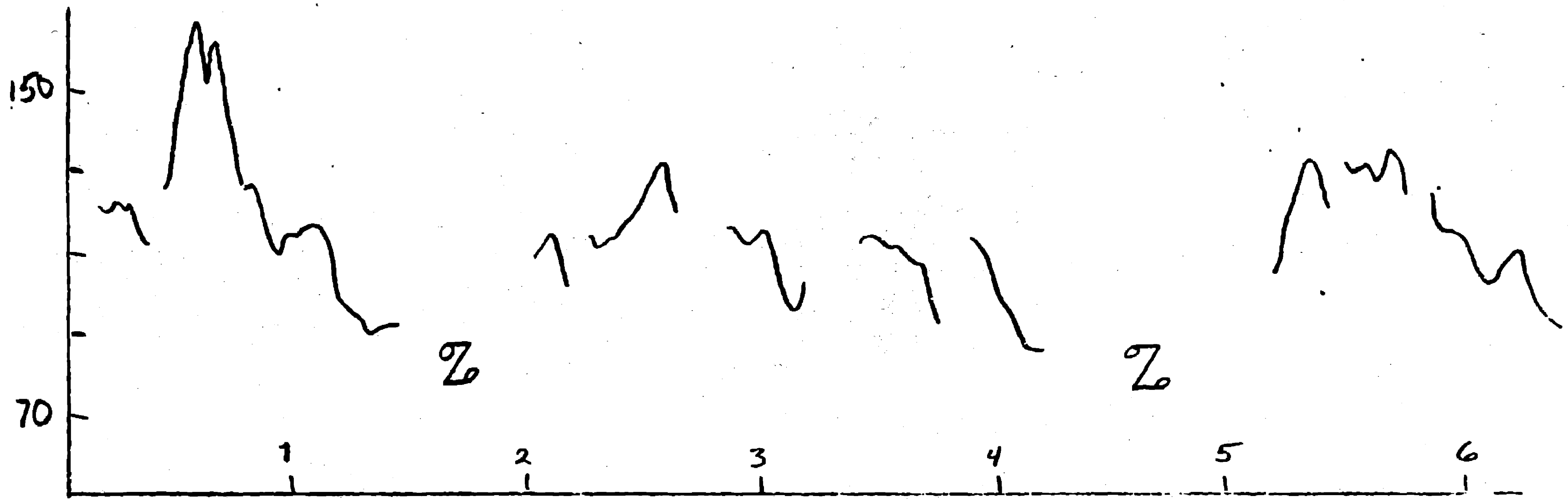
Are legumes a <u>good</u> source of vitamins?
L* H⁻ H%

C)



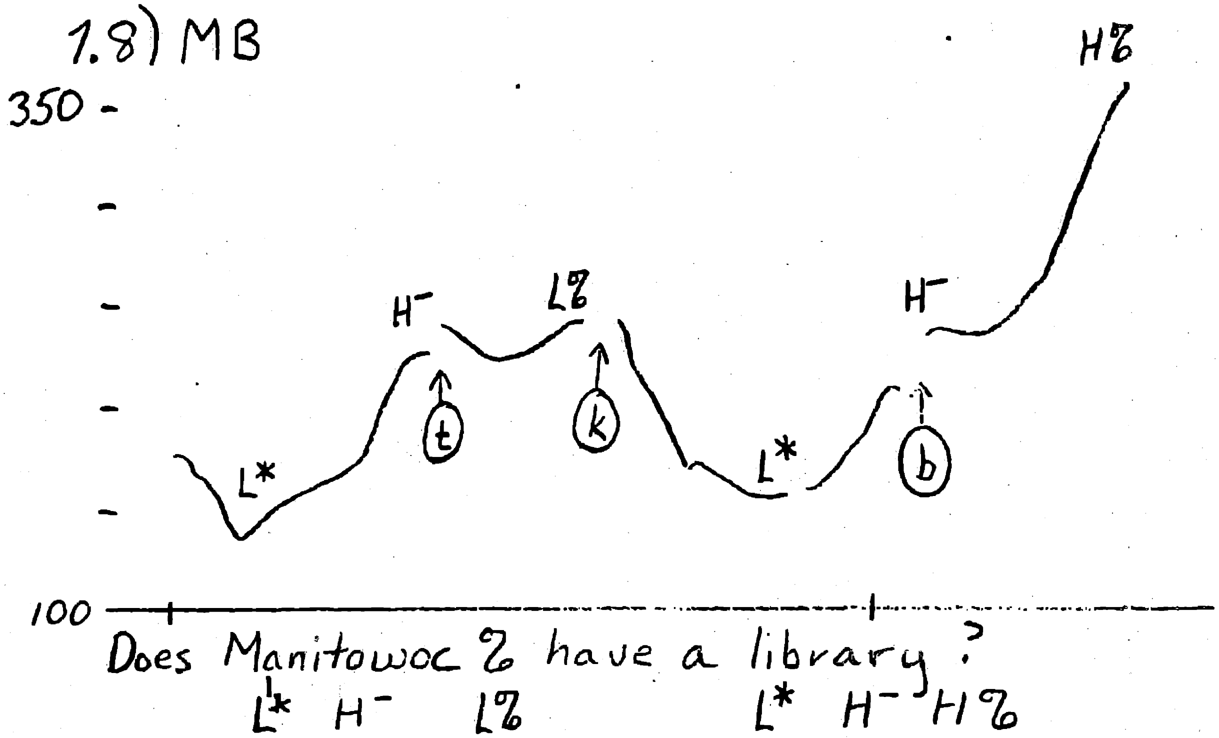
Are <u>legumes</u> a good source of vitamins?
L* H⁻ H%

1.7) JP (from Maeda 1976)

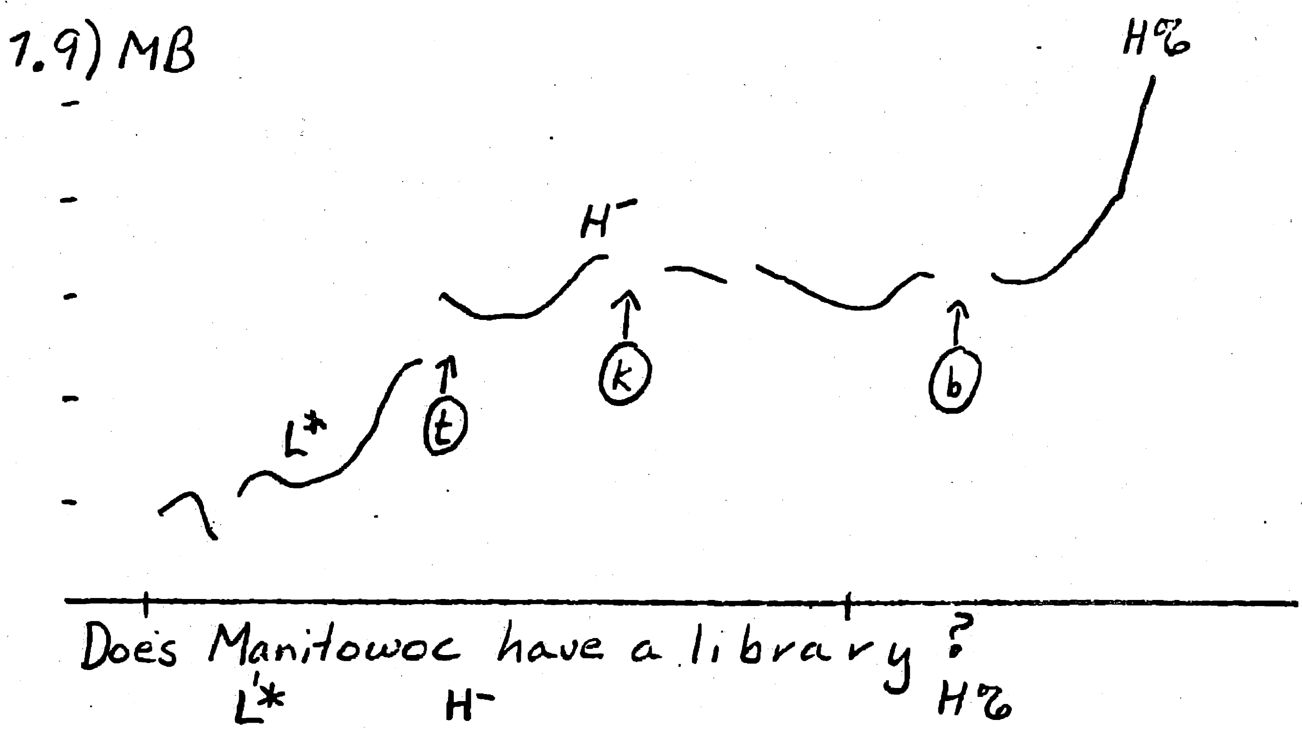


In the jungles of Asia % there is a large bird with brilliant colors % red feathers on the wings...

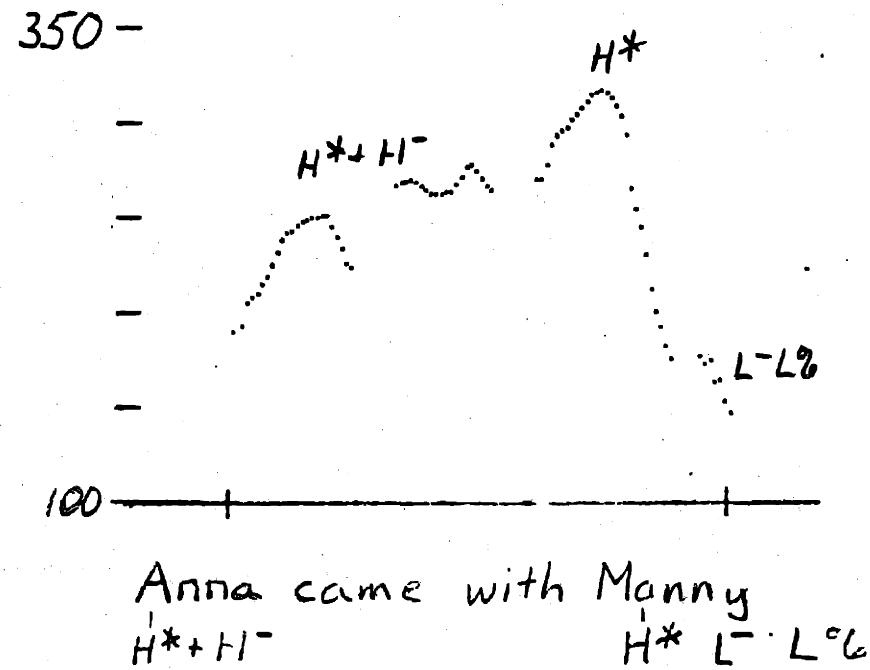
1.8) MB


H%
350
H⁻
L%
H⁻
t
k
L*
L*
b
100
Does Manitowoc % have a library?
L* H⁻ L% L* H⁻ H%

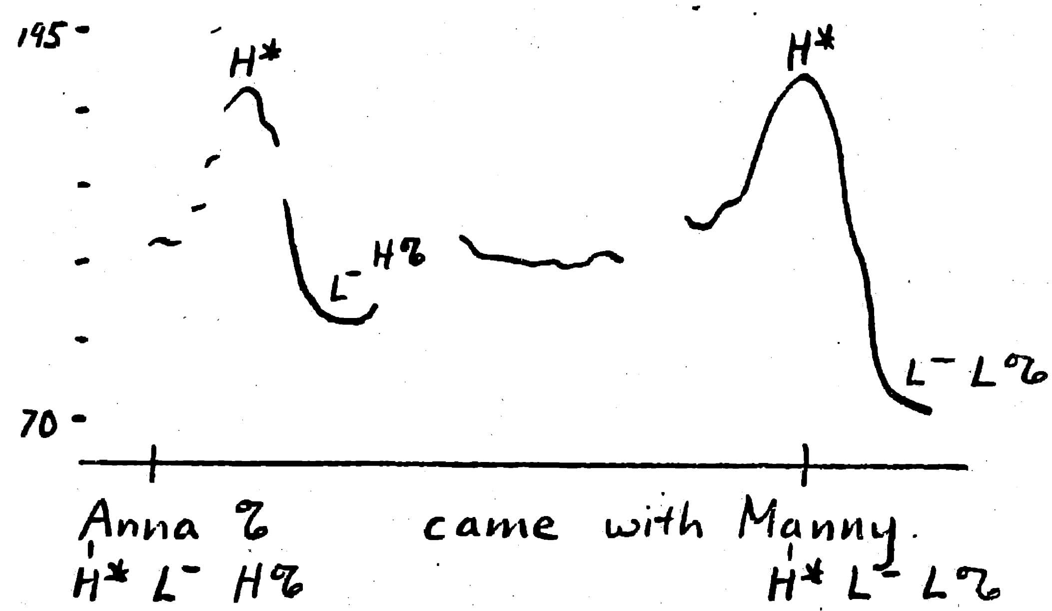
1.9) MB


H%
H⁻
k
b
L*
t
Does Manitowoc have a library?
L* H⁻ H%

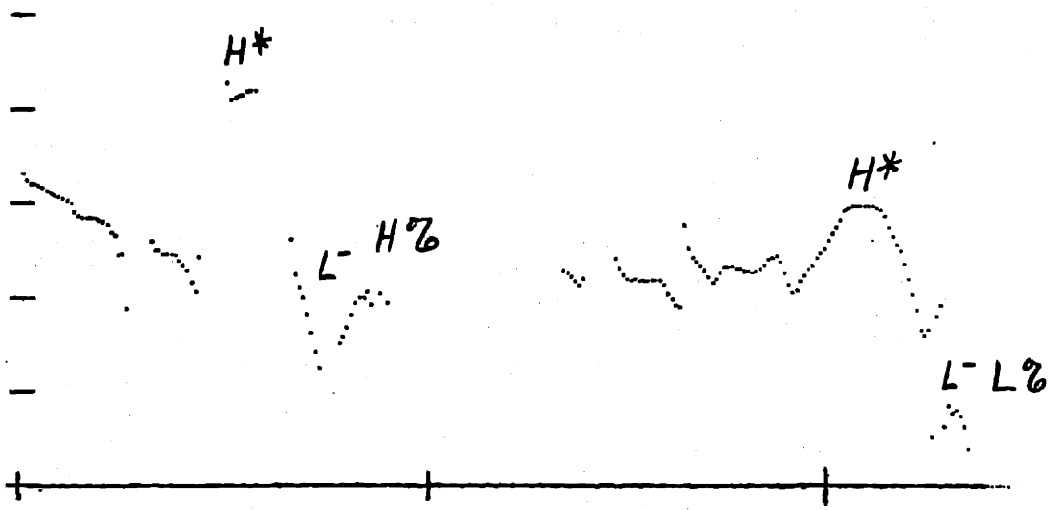
7.10) JBP



7.11) MYL



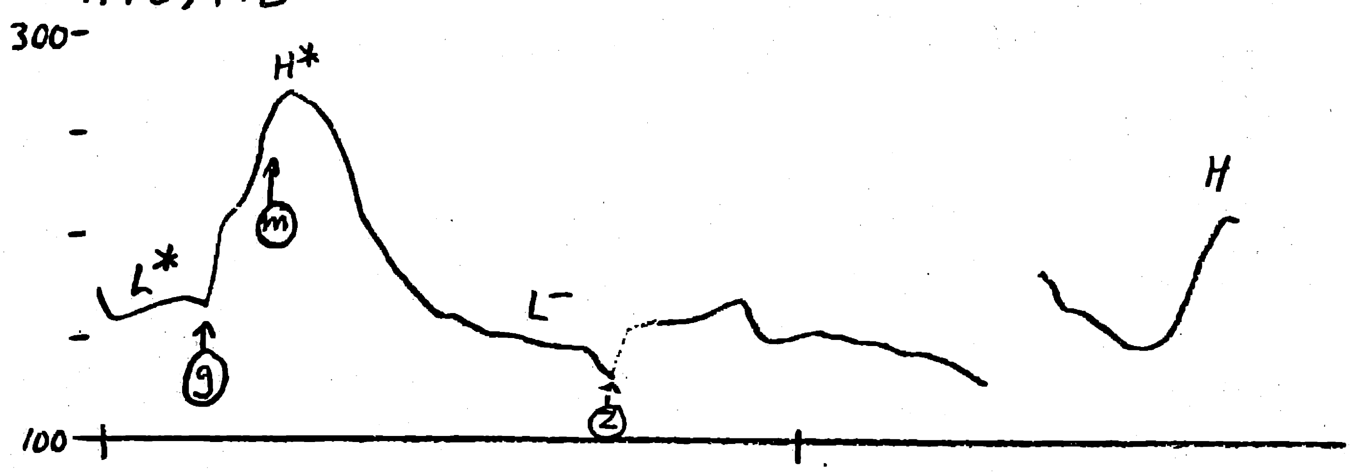
1.12) JBP



Three mathematicians % intend to rival Emma.

H* L⁻ L% H* L⁻ L%

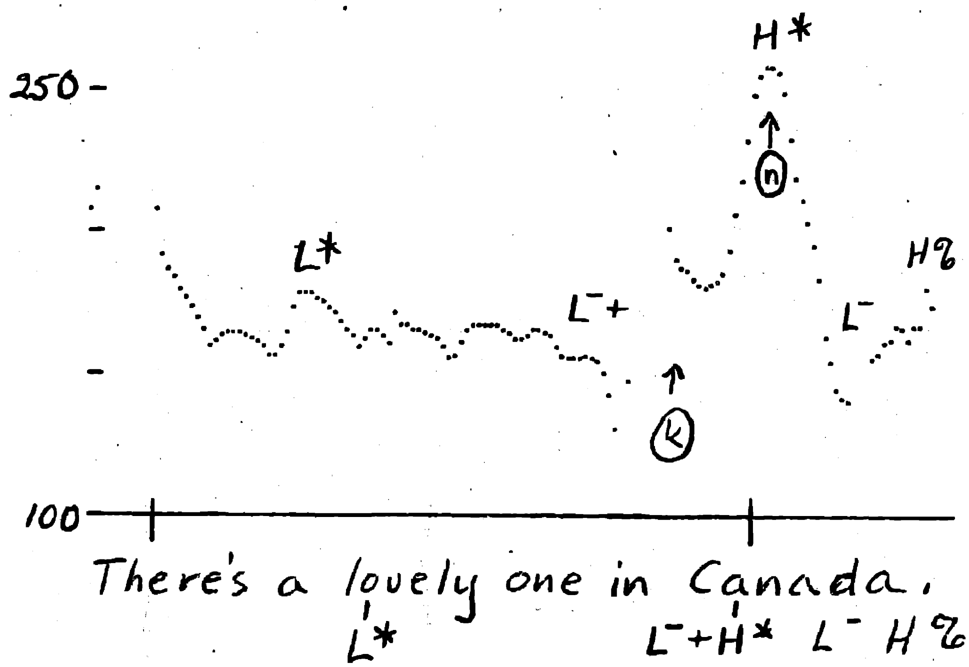
1.13) MB



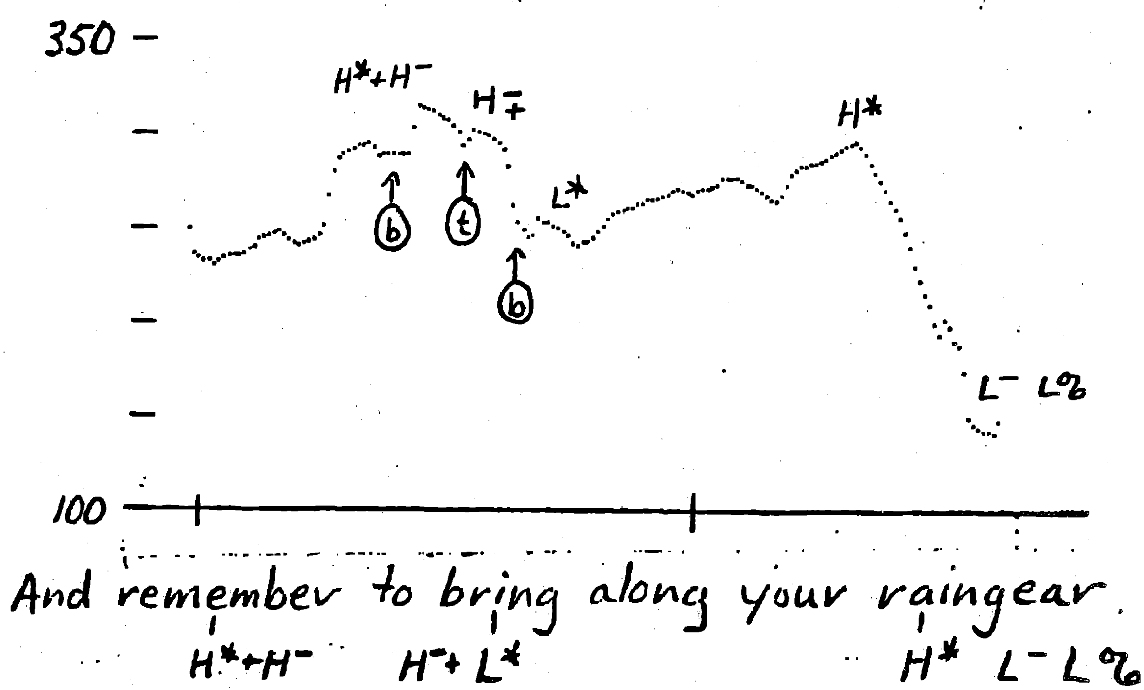
Rigamarole is monomorphemic.

L*+H⁻ L⁻ H%

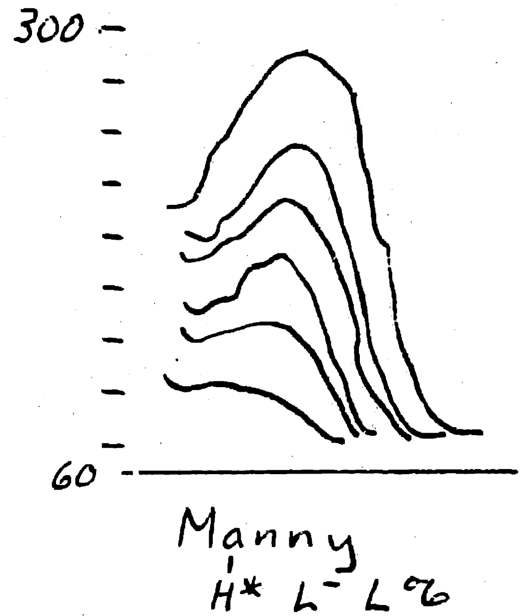
1.14) JBP



1.15) JBP

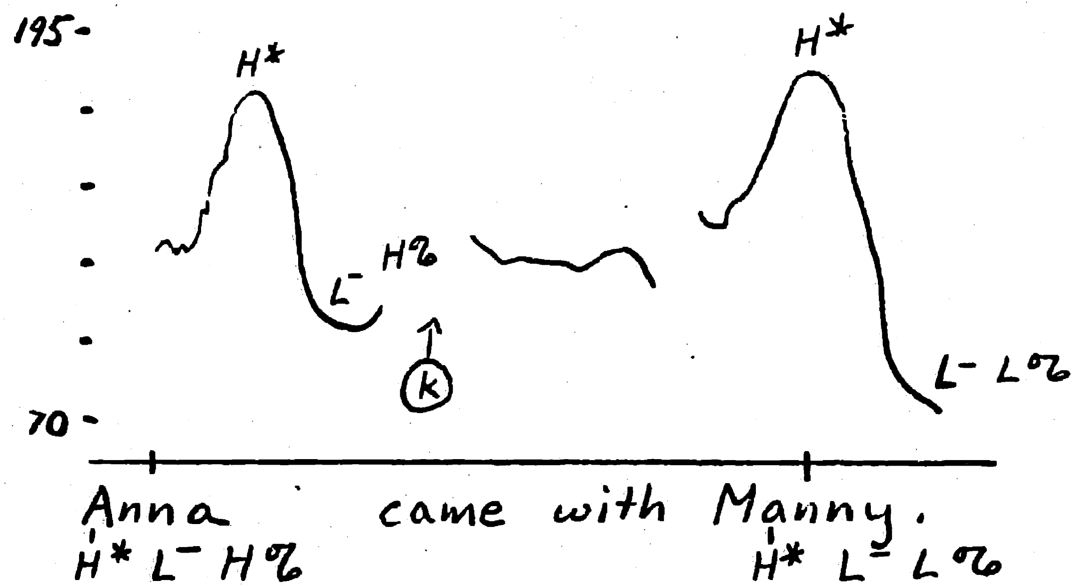


1.16 MYL)

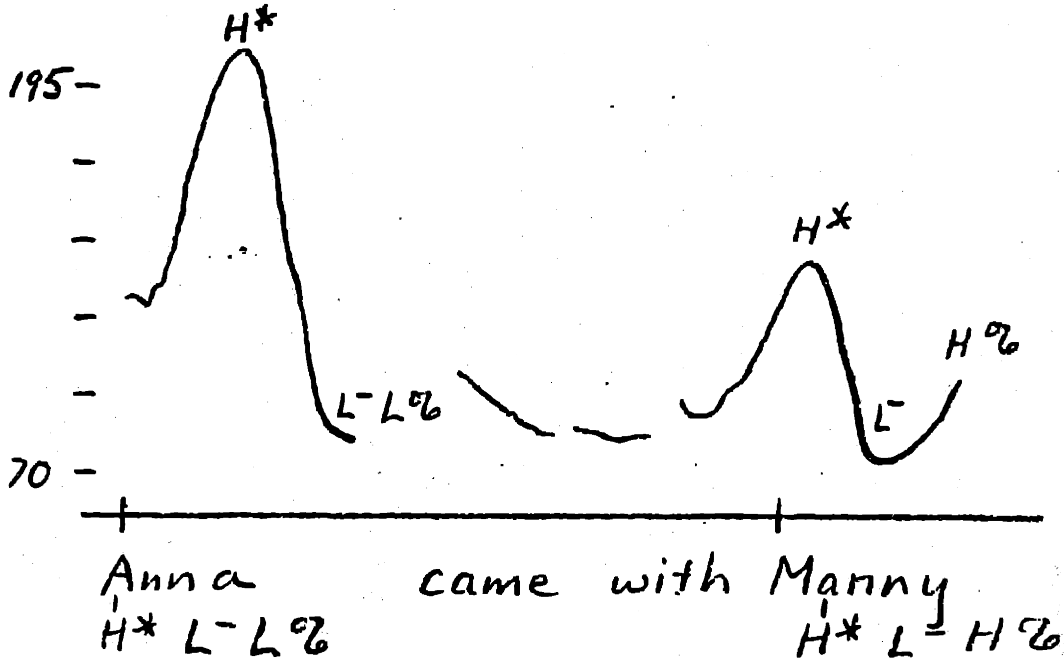


(in six different pitch ranges)

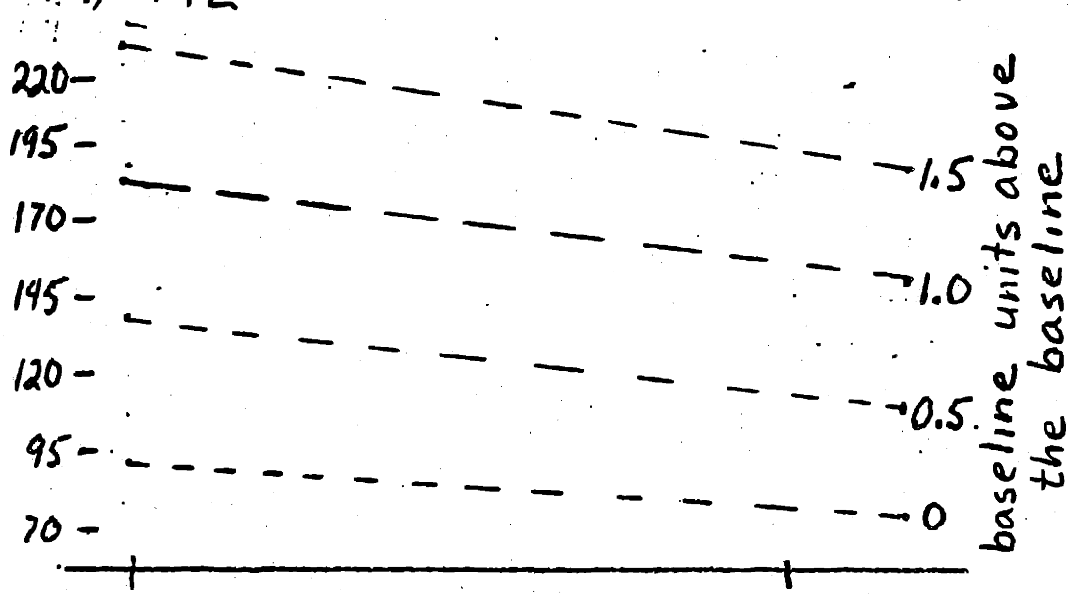
1.17) MYL



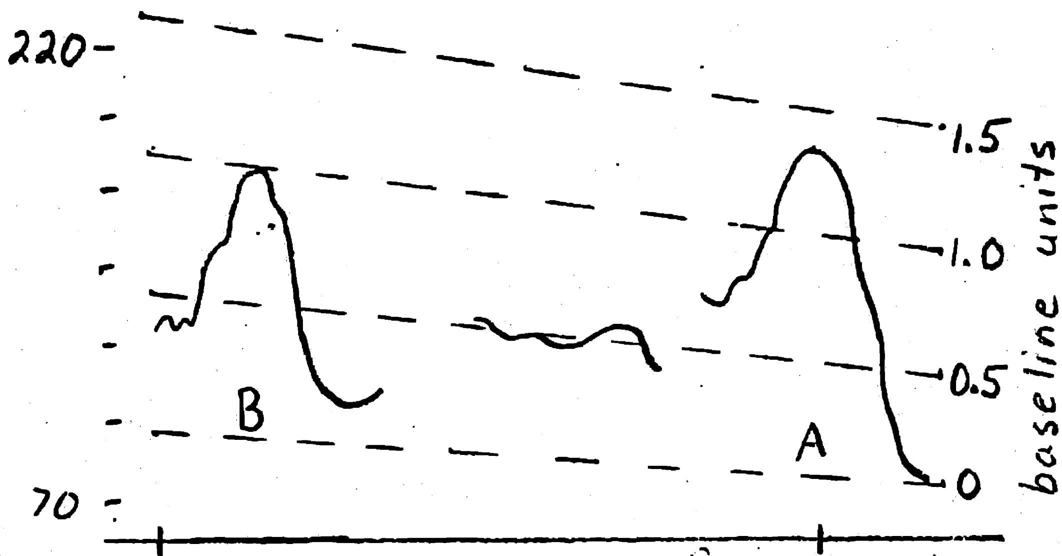
1.18) M4L



1.19) M4L



1.20 A) MYL

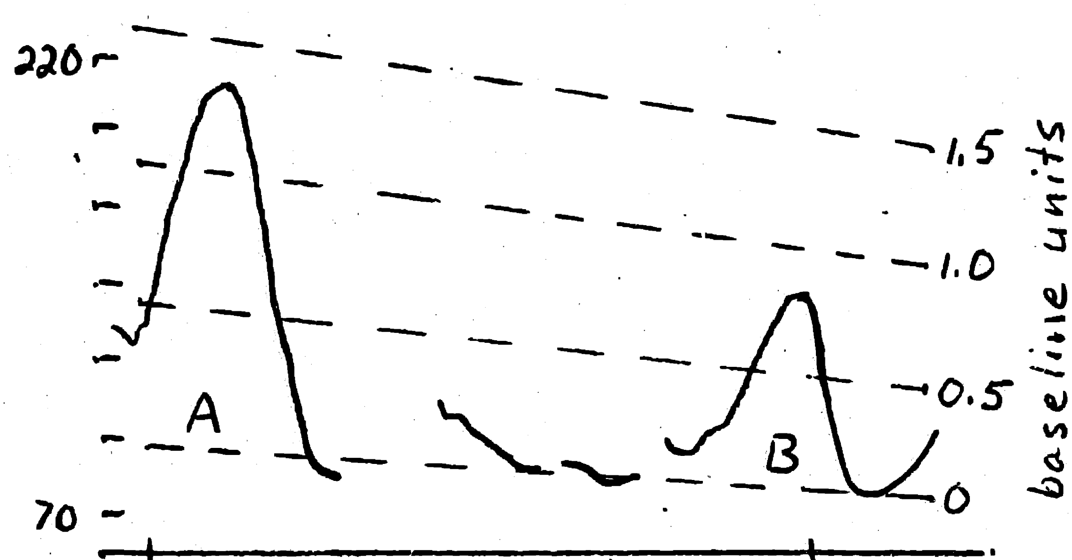


B = 180 Hz A = 185 Hz

⇒ 1 baseline unit ⇒ 1.45 baseline units

$\hat{A} = 1.45\ \hat{B}$

B)

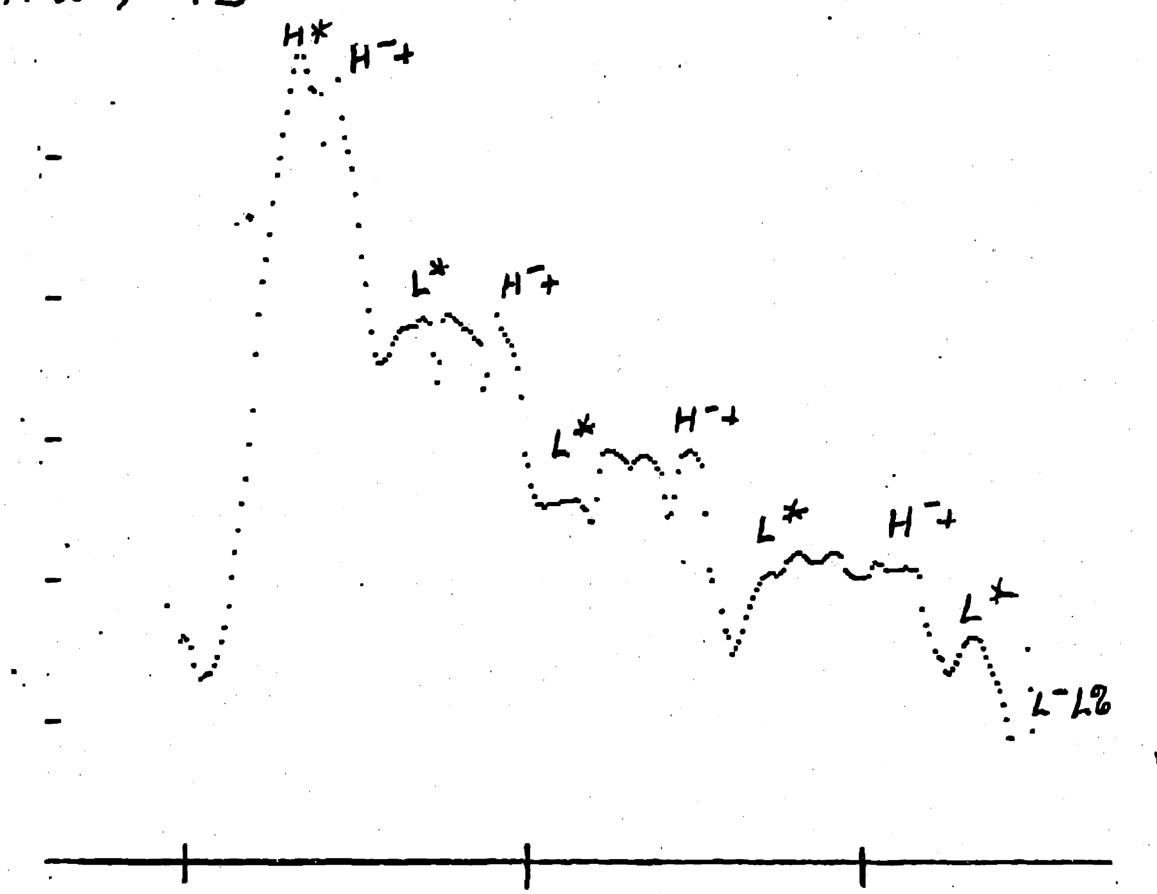


A = 210 Hz B = 141 Hz

⇒ 1.33 baseline units ⇒ 0.87 baseline units

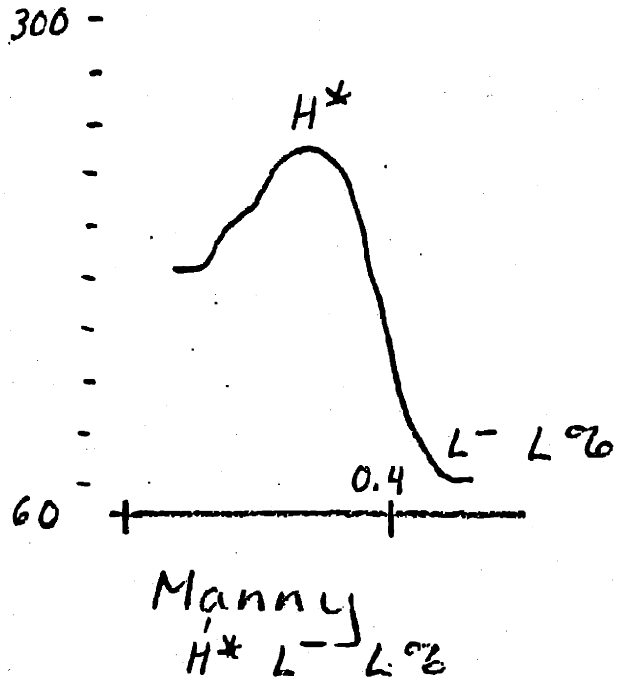
$\hat{A} = 1.53\ \hat{B}$

1.21) MB

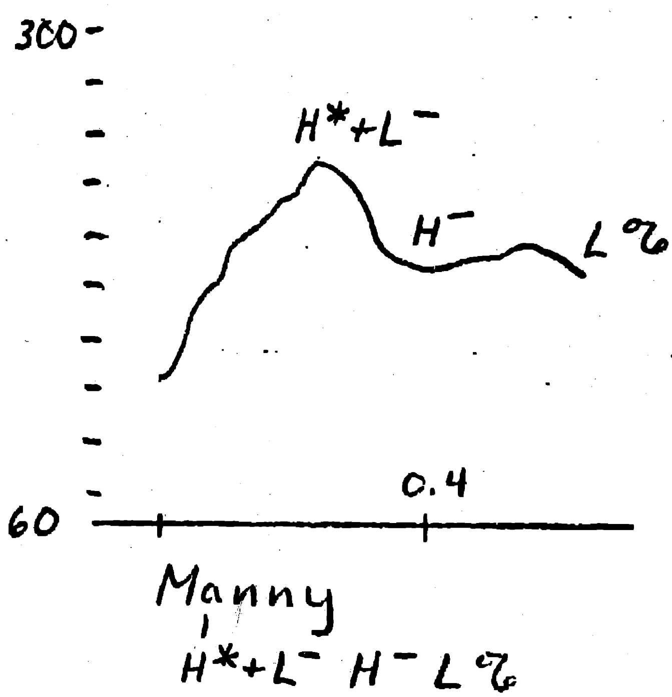


I really believe Ebenezer is a dealer in magnesium.
H* H⁻+L* H⁻+L* H⁻+L* H⁻+L* L⁻L%

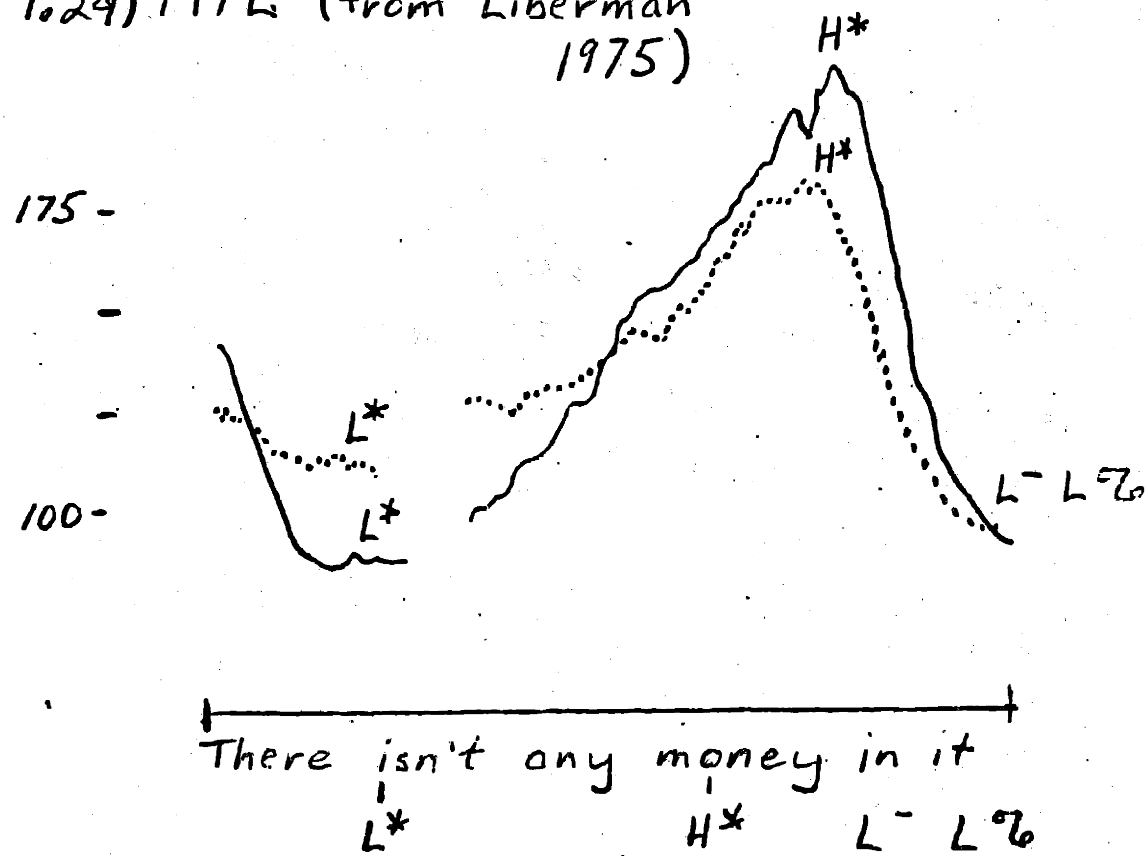
1.22) MYL



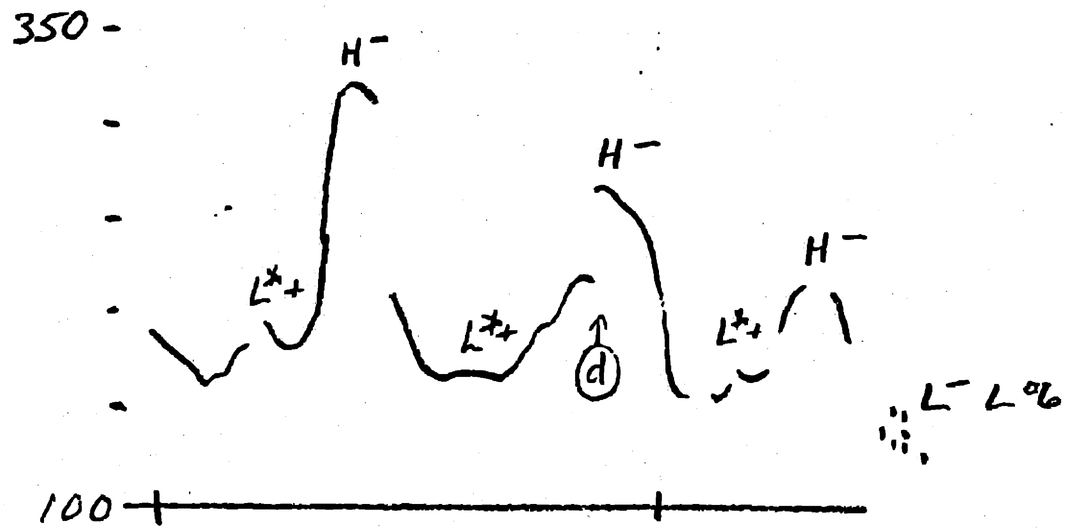
1.23) MYL



1.24) MYL (from Liberman 1975)



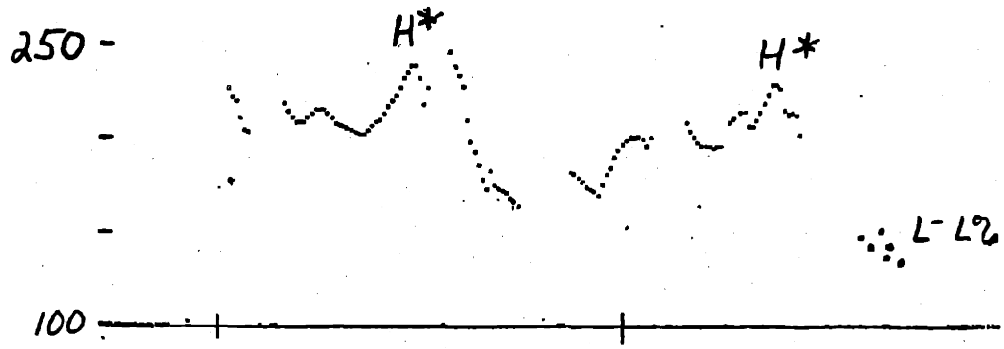
1.25) JBP



There are many intermediate levels

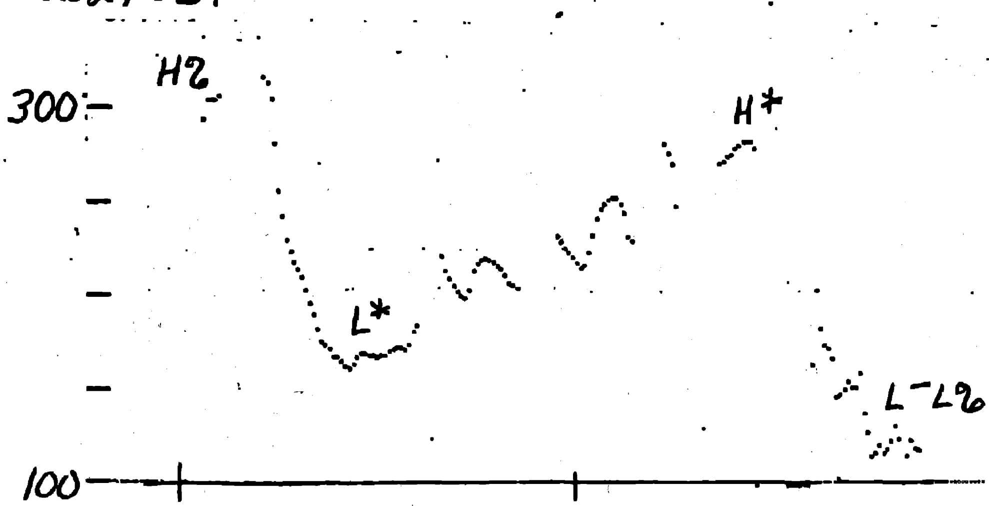
L*+H⁻ L*+H⁻ L*+H⁻ L⁻ L%

2.1) JBP



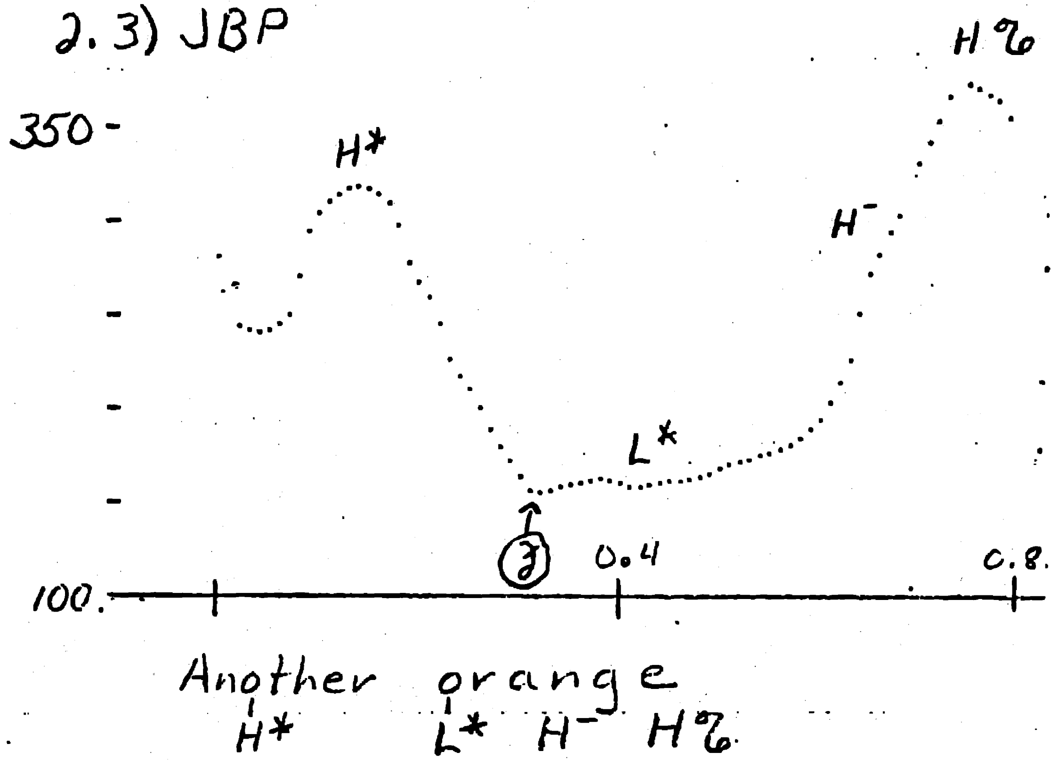
That's a remarkably clever suggestion.
H* H* L⁻L%

2.2) JBP

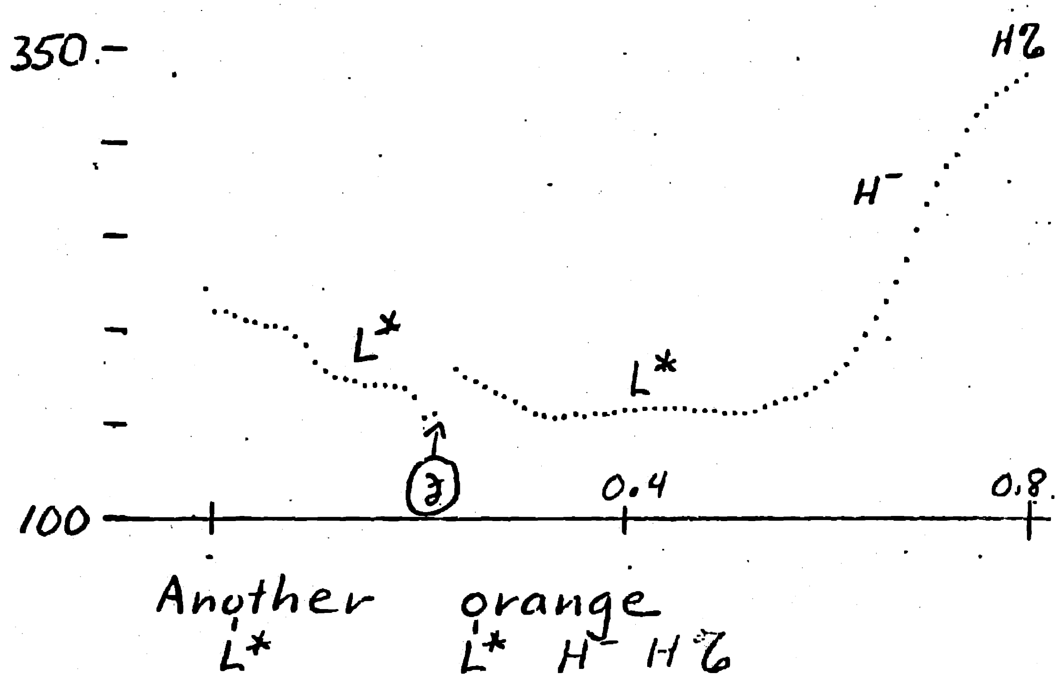


That's a remarkably clever suggestion.
H% L* H* L⁻L%

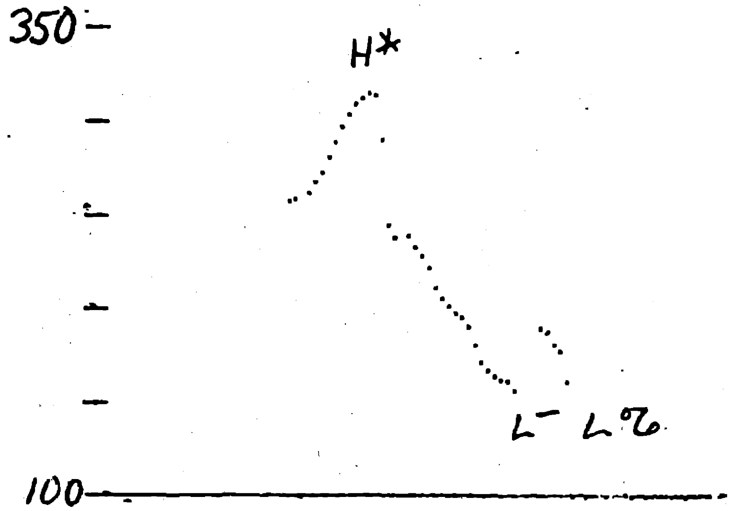
2.3) JBP


H%
350
H*
H⁻
L*
0.4
0.8
100
Another orange
H* L* H⁻ H%

2.4) JBP

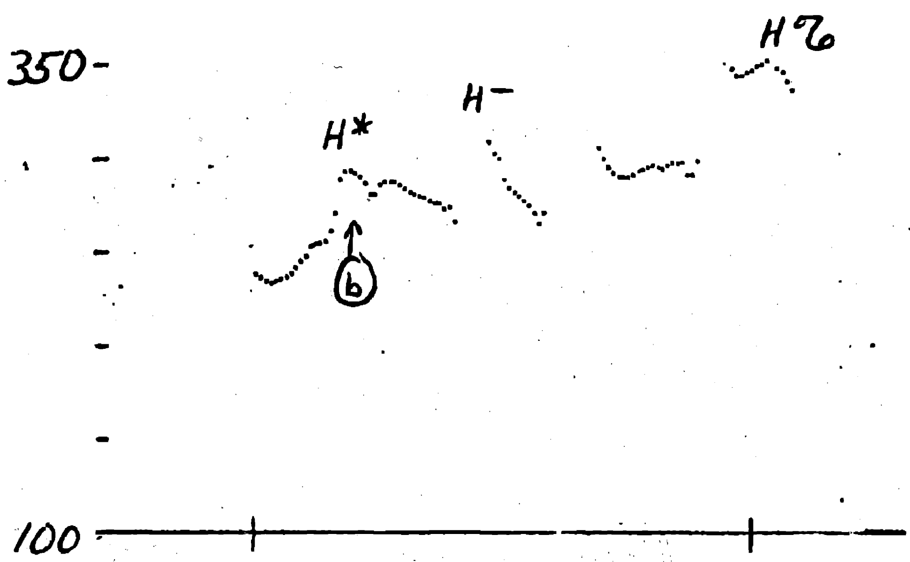

350
H%
H⁻
L*
L*
0.4
0.8
100
Another orange
L* L* H⁻ H%

2.5) JBP



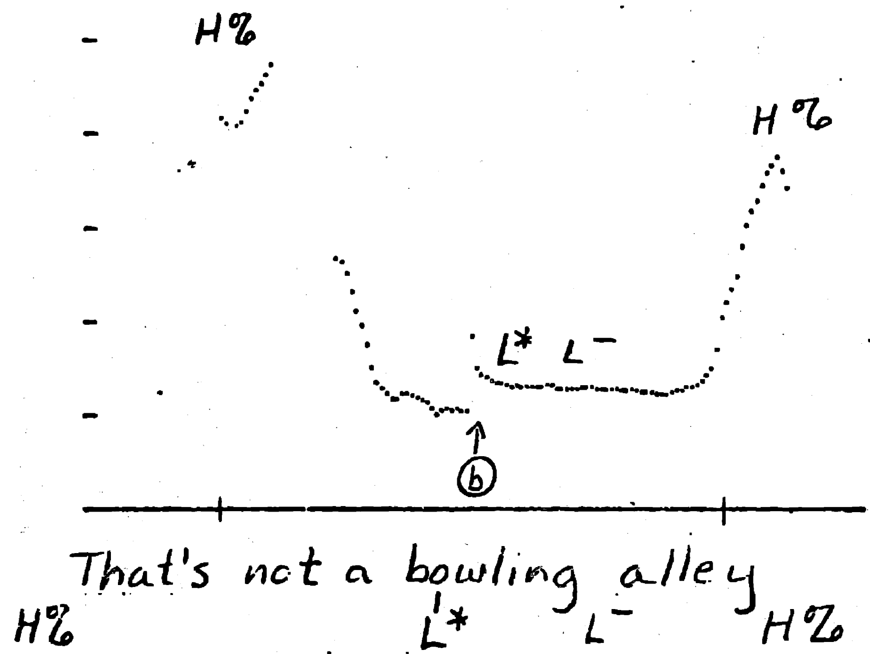
Abernathy
H* L⁻ L%

2.6) JBP

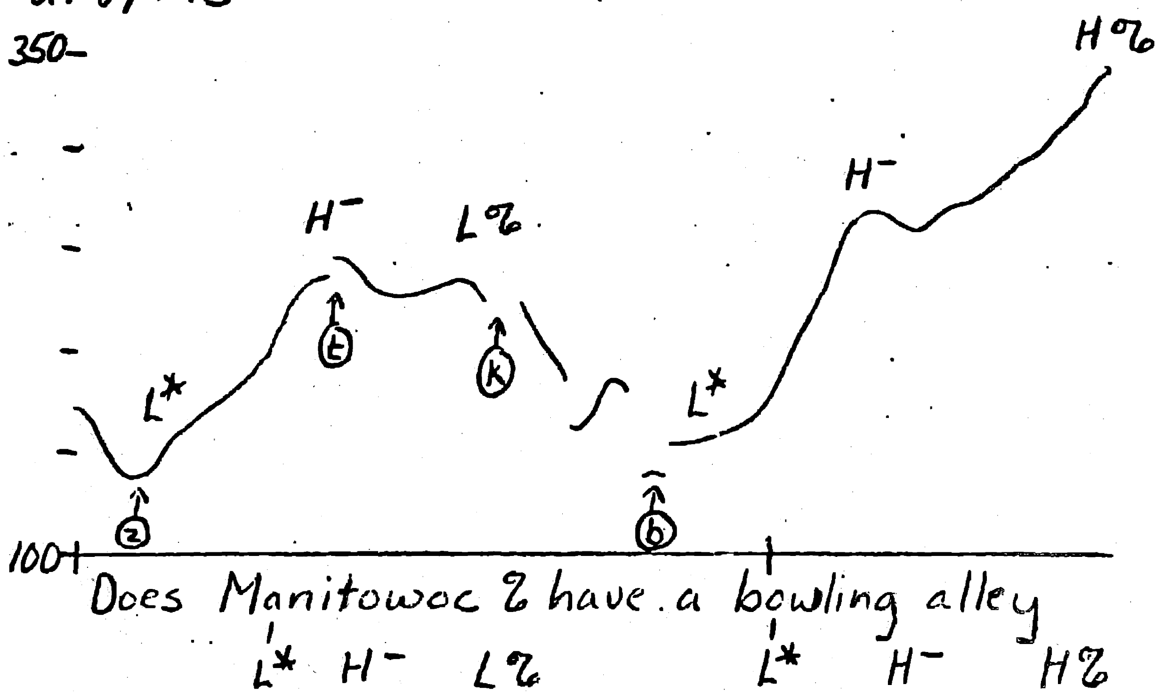


Abernathy is going to?
H* H⁻ H%

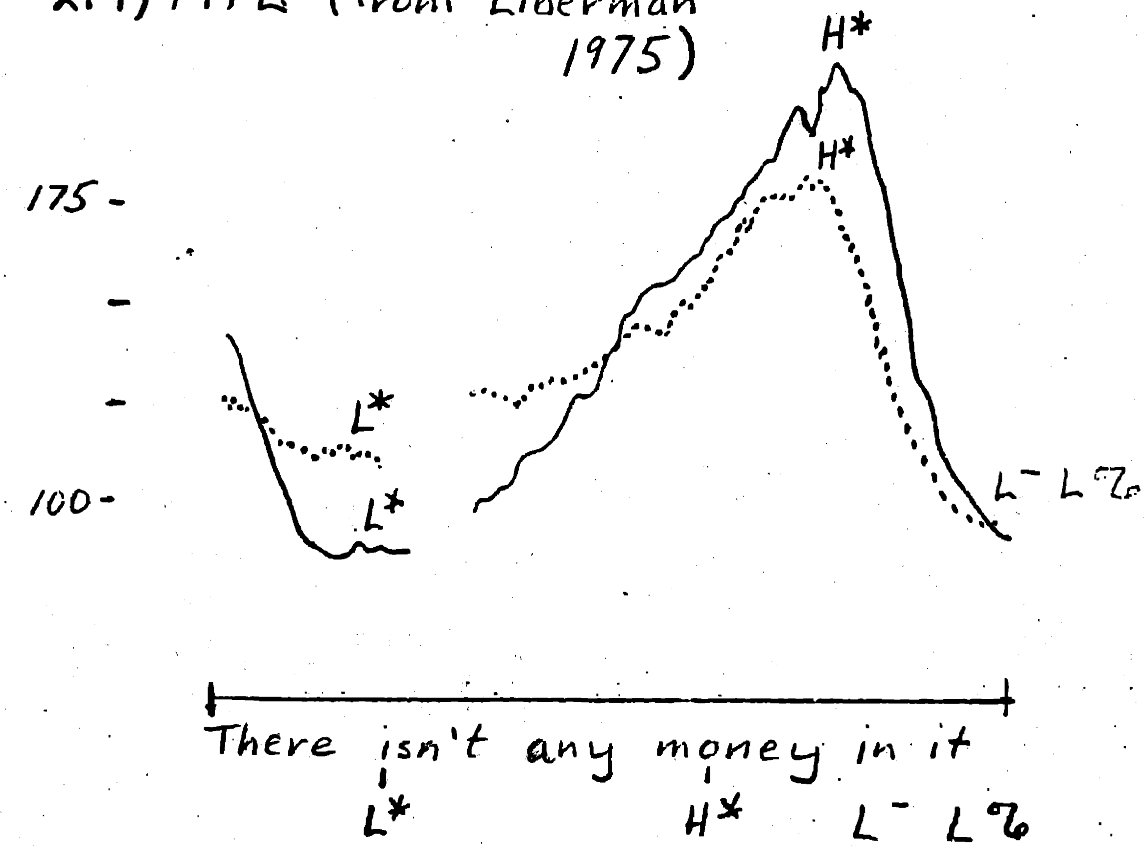
2.7) JBP


H%
H%
L* L⁻
b
That's not a bowling alley
H% L* L⁻ H%

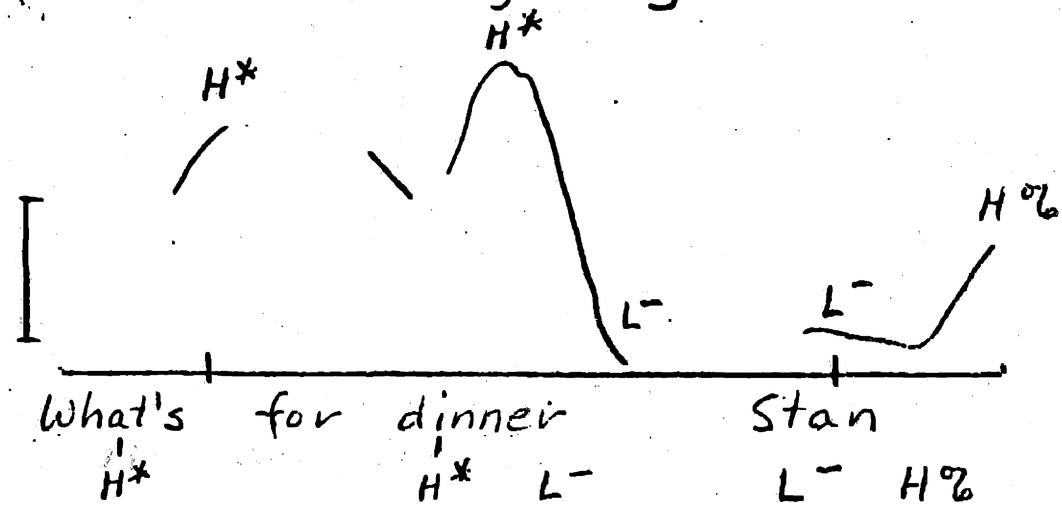
2.8) MB


350
H%
H⁻
H⁻
L%
L*
L*
t
k
z
b
100
Does Manitowoc % have a bowling alley
L* H⁻ L% L* H⁻ H%

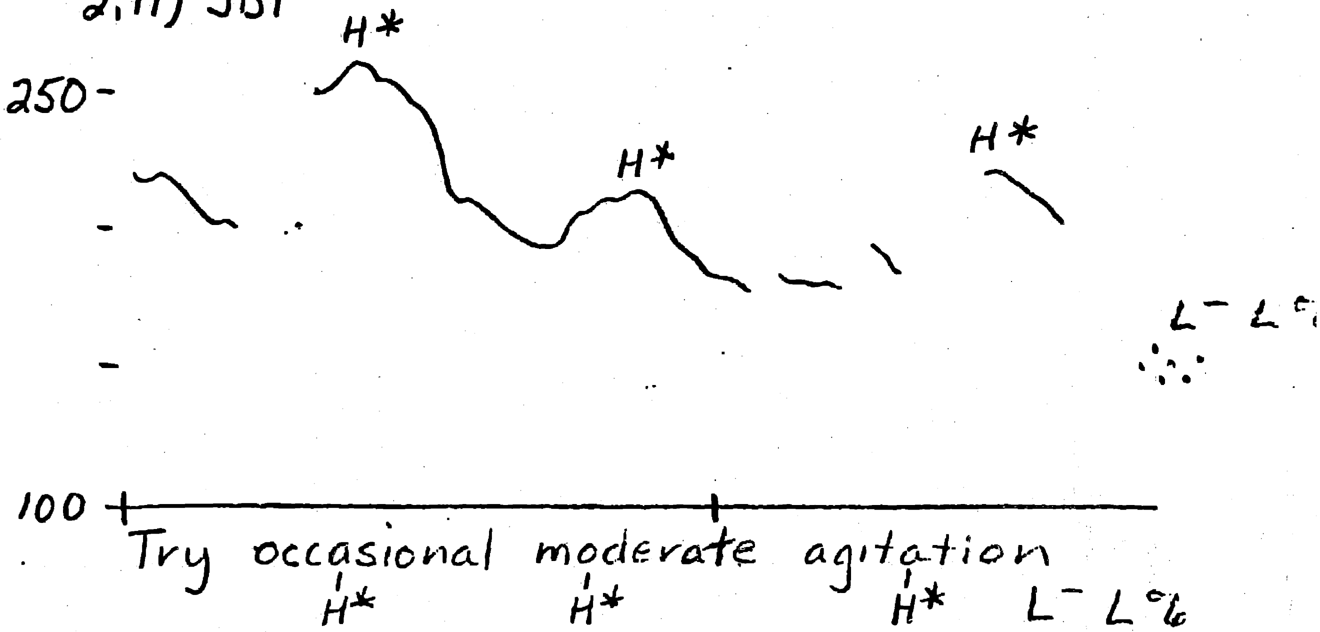
2.9) MYL (from Liberman 1975)



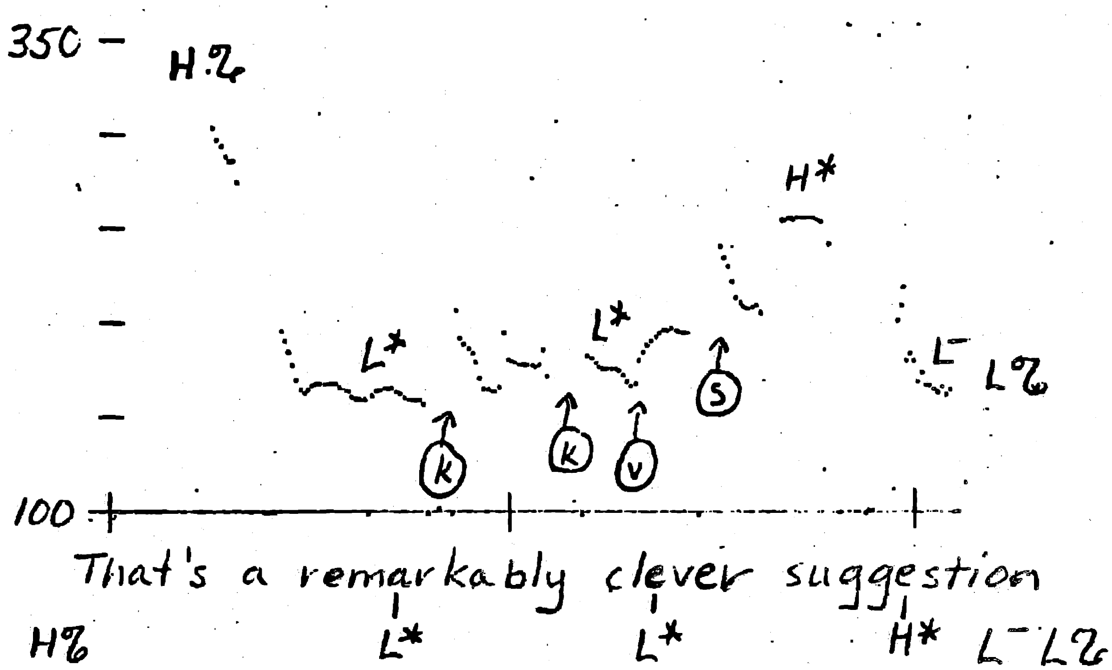
2.10) (from O'Shaughnessy 1976)



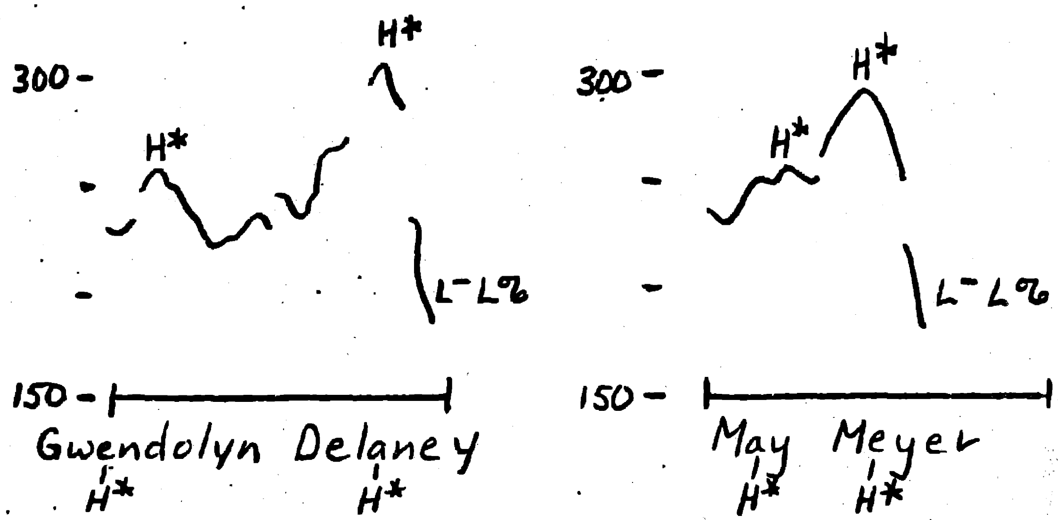
2.11) JBP



2.12) JBP

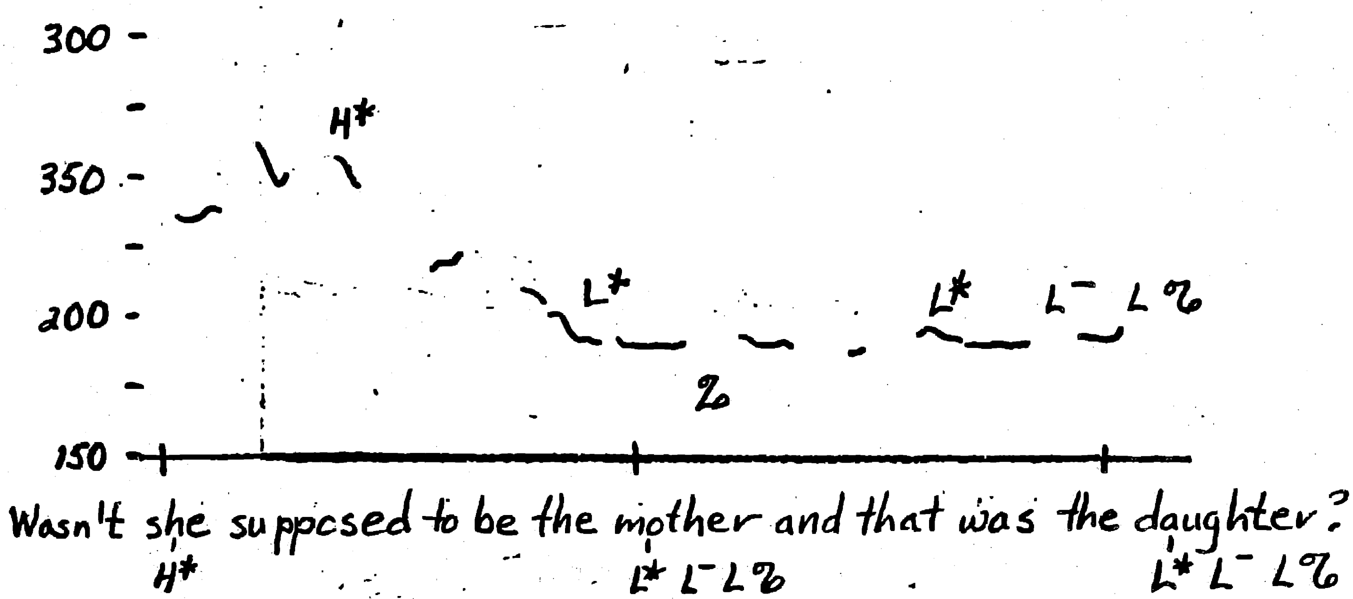


2.13 A) JBP B) JBP

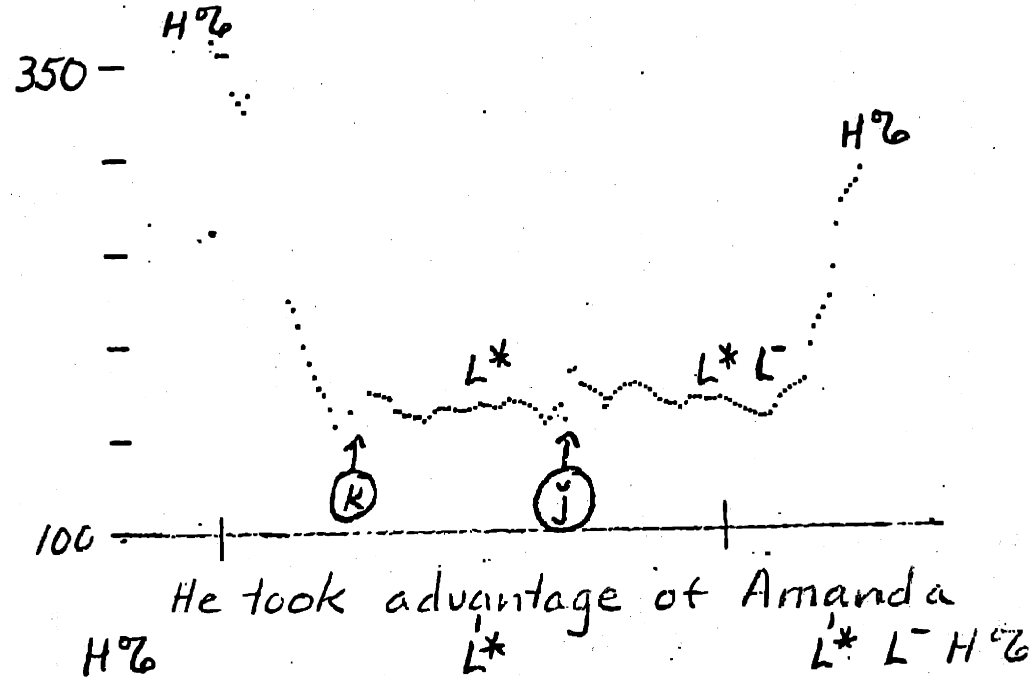


(excised from the context "I think you should discuss it with ____ .")

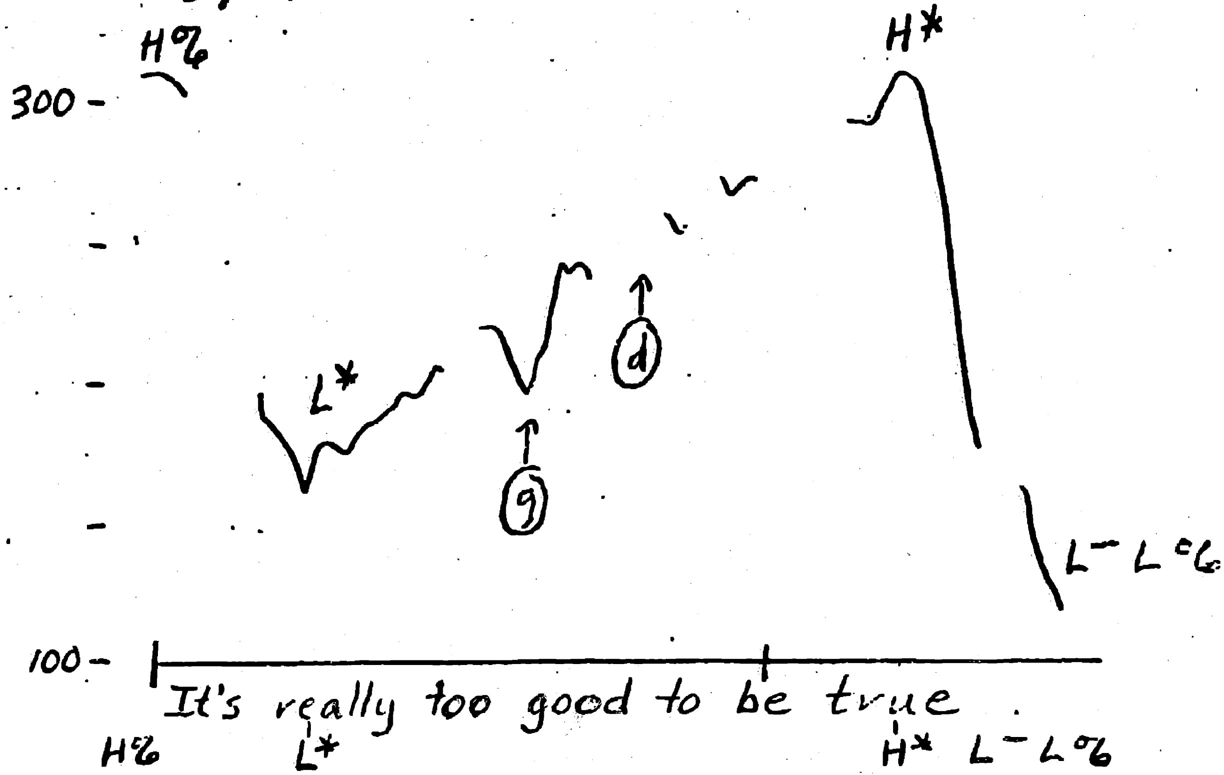
2.14)



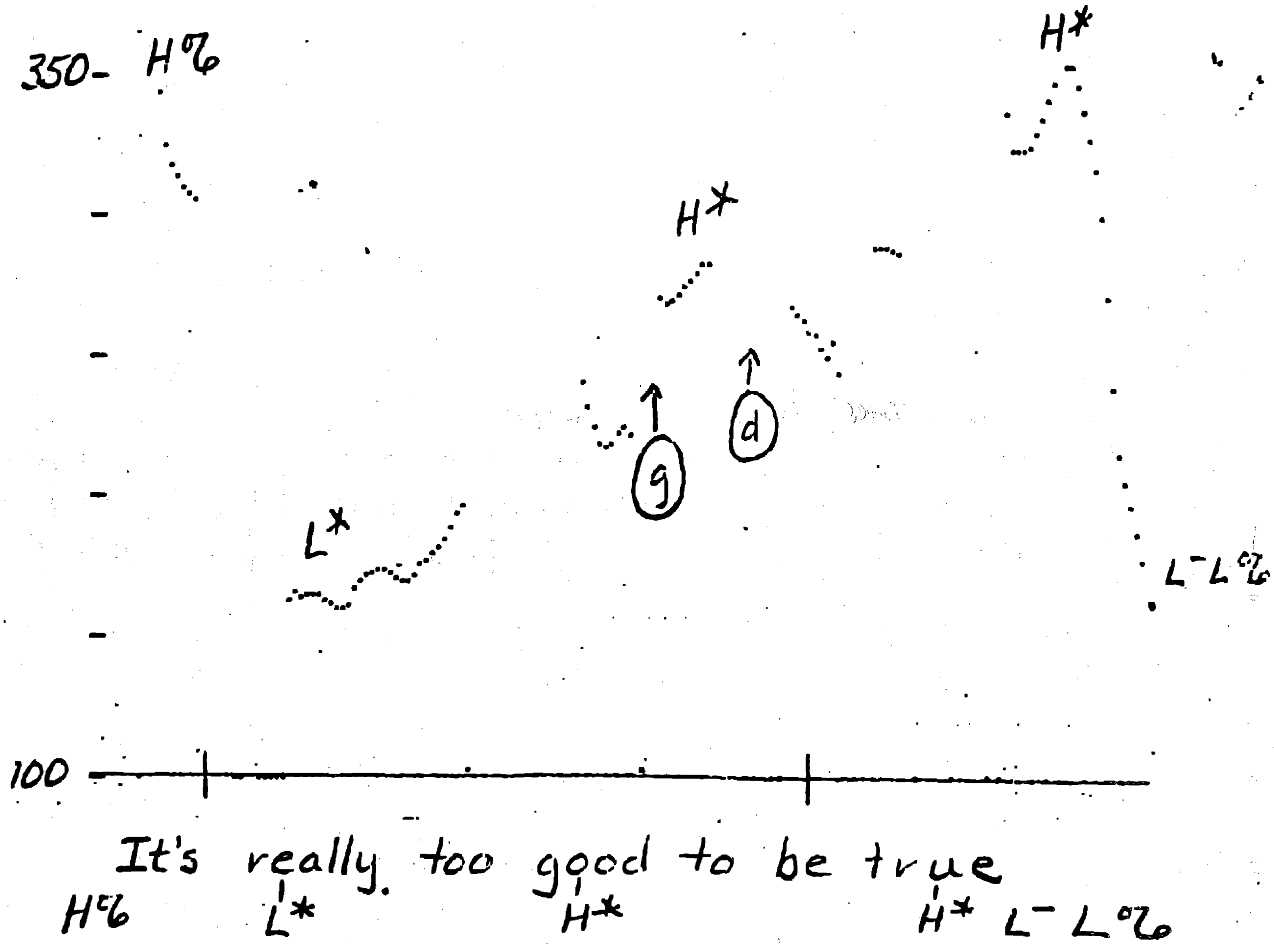
2.15) JBP



2.16) JBP

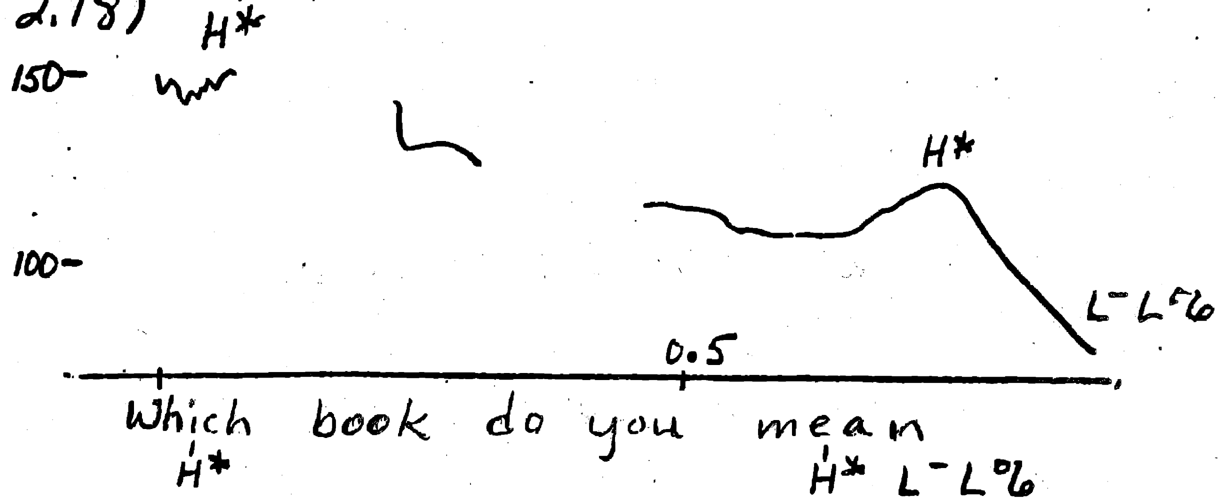


2.17) JBP



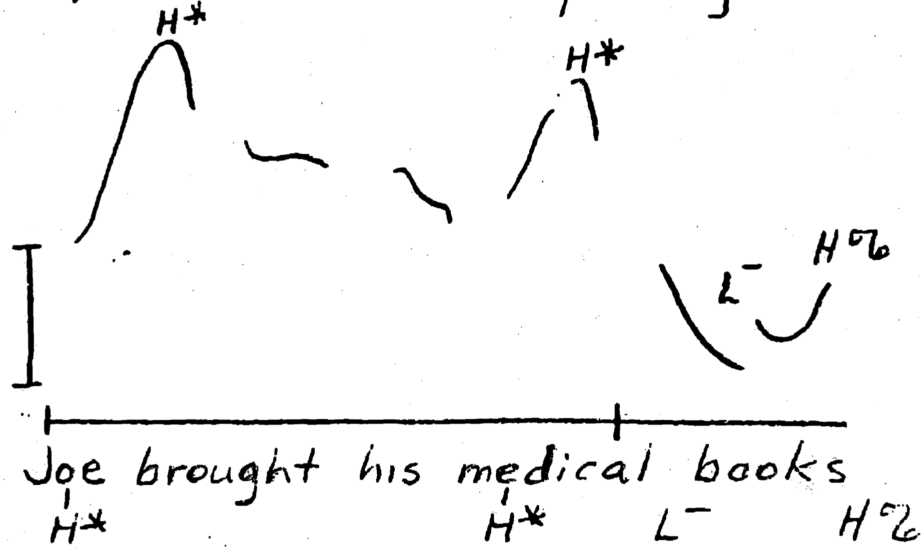
H%
H*
H*
L*
g
d
L⁻L%
350-
100
It's really too good to be true
H% L* H* H* L⁻ L%

2.18)

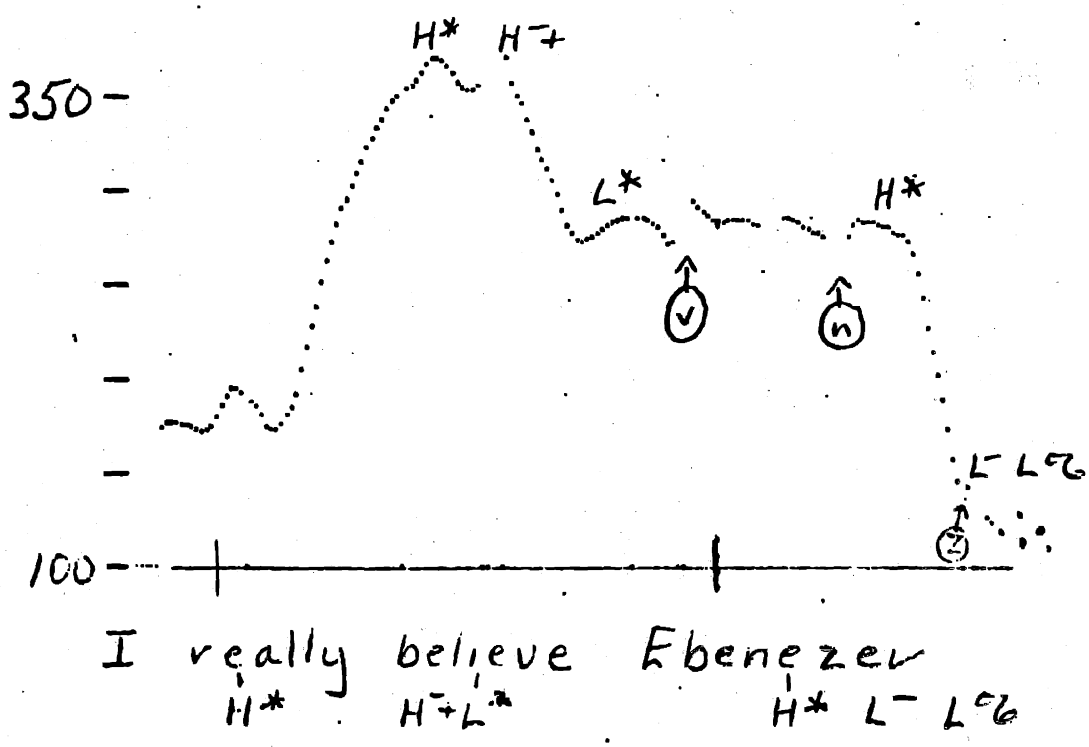


H*
150-
H*
100-
L⁻L%
0.5
Which book do you mean
H* H* L⁻ L%

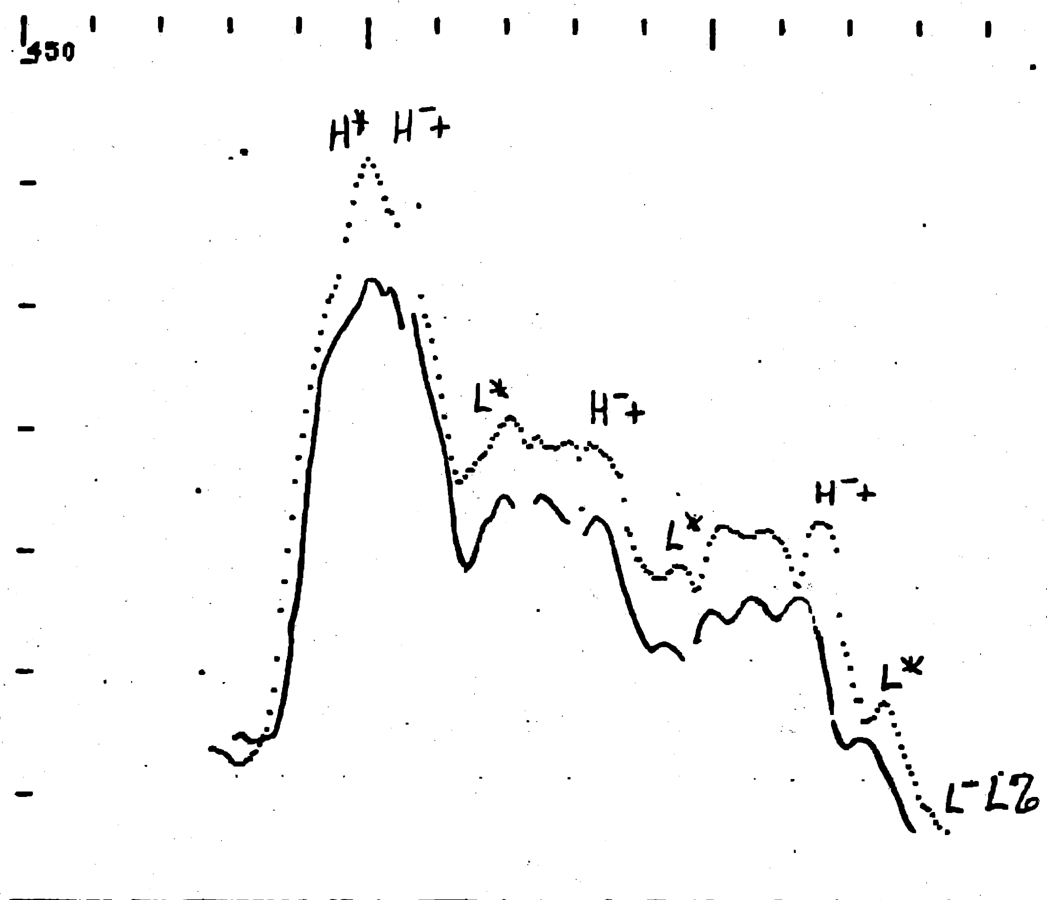
2.19) JA (from O'Shaughnessy 1976)



2.20) MB



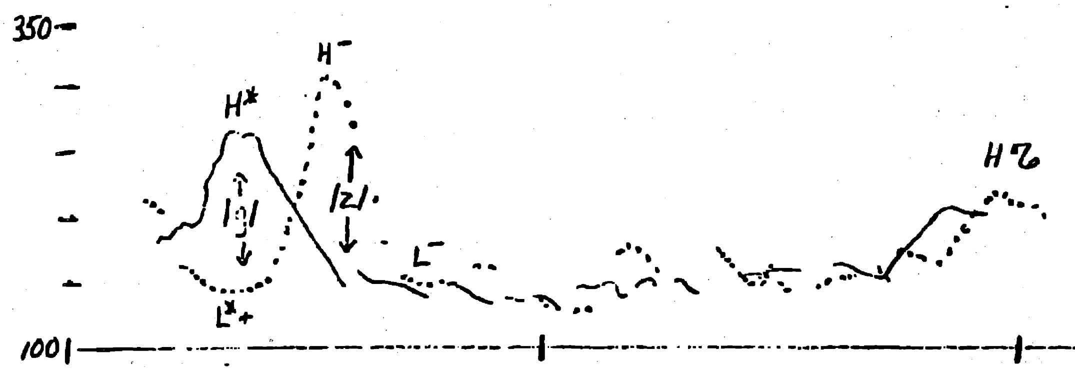
2.21) MB



Two F0 contours for:

I really believe Ebenezer was a dealer
H* H⁻+L* H⁻+L* H⁻+L* L⁻ L%

2.22) MB.

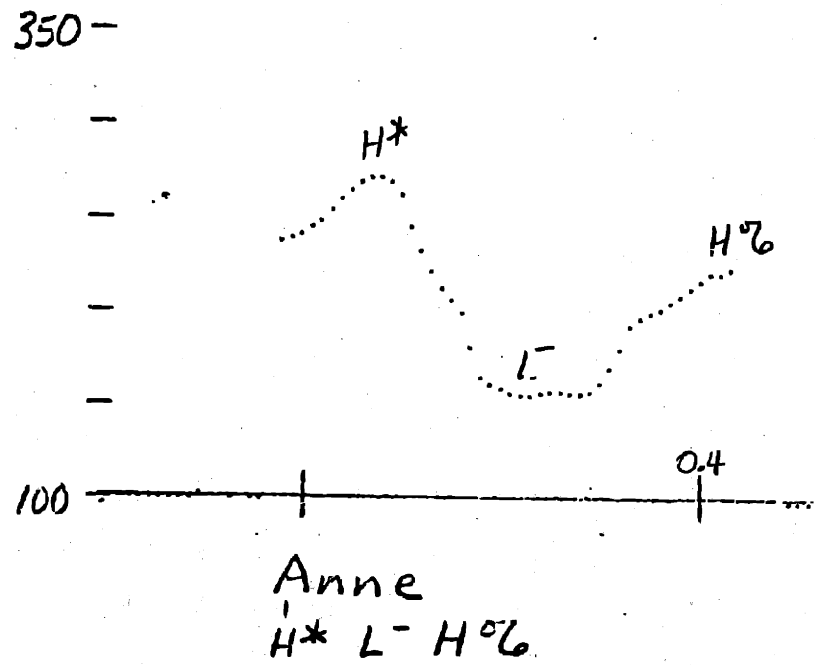


A) — Legumes are a good source of vitamins
H* L⁻ H%

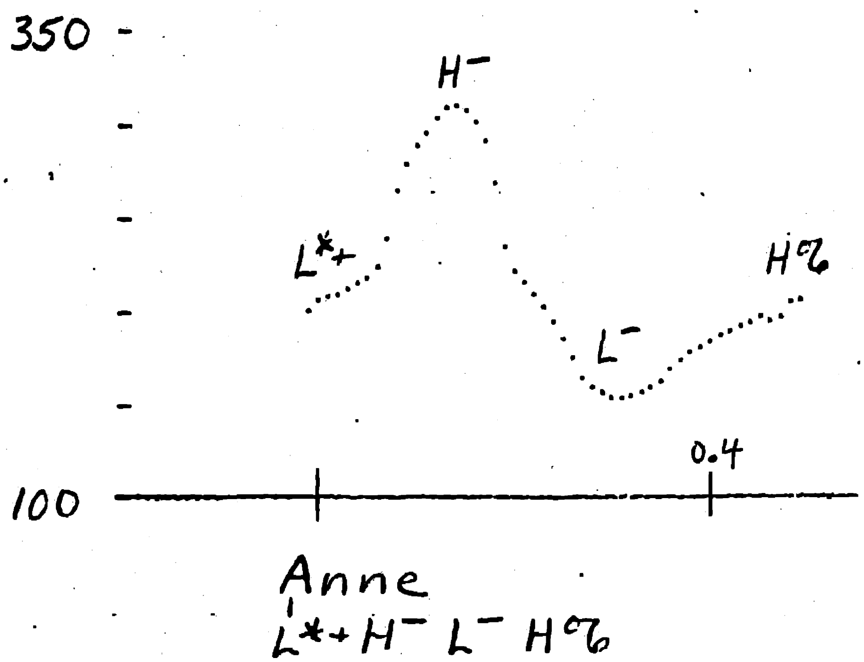
B) ···· Legumes are a good source of vitamins
L*+H⁻ L⁻ H%

(Lined up on /g/ and /z/)

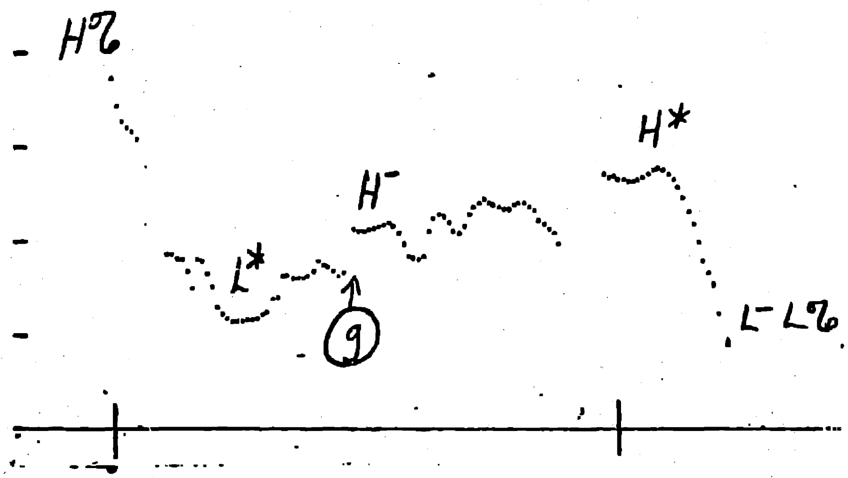
2.23 A) JBP



B) JBP



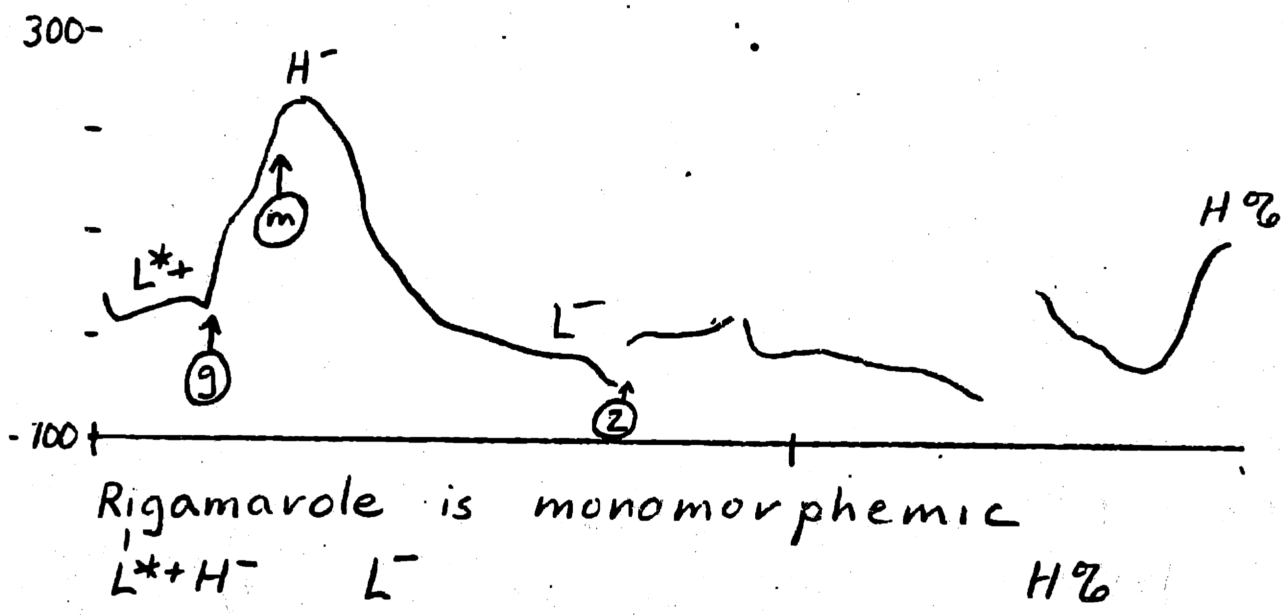
2.24) JBP



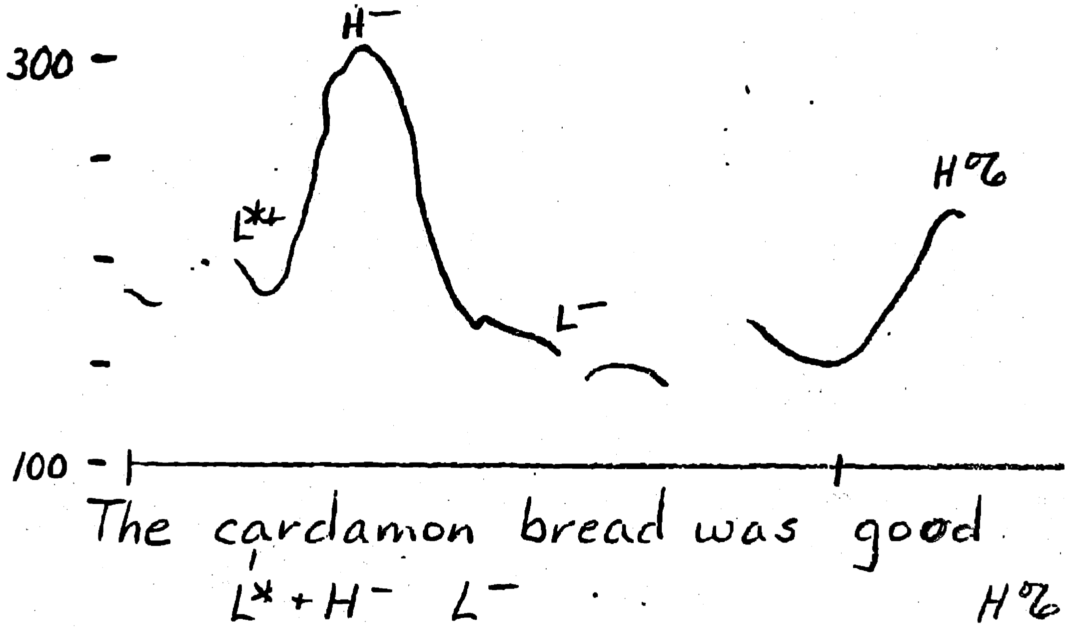
But Rob agreed to bring salad.

H% L*+H- H* L- L%

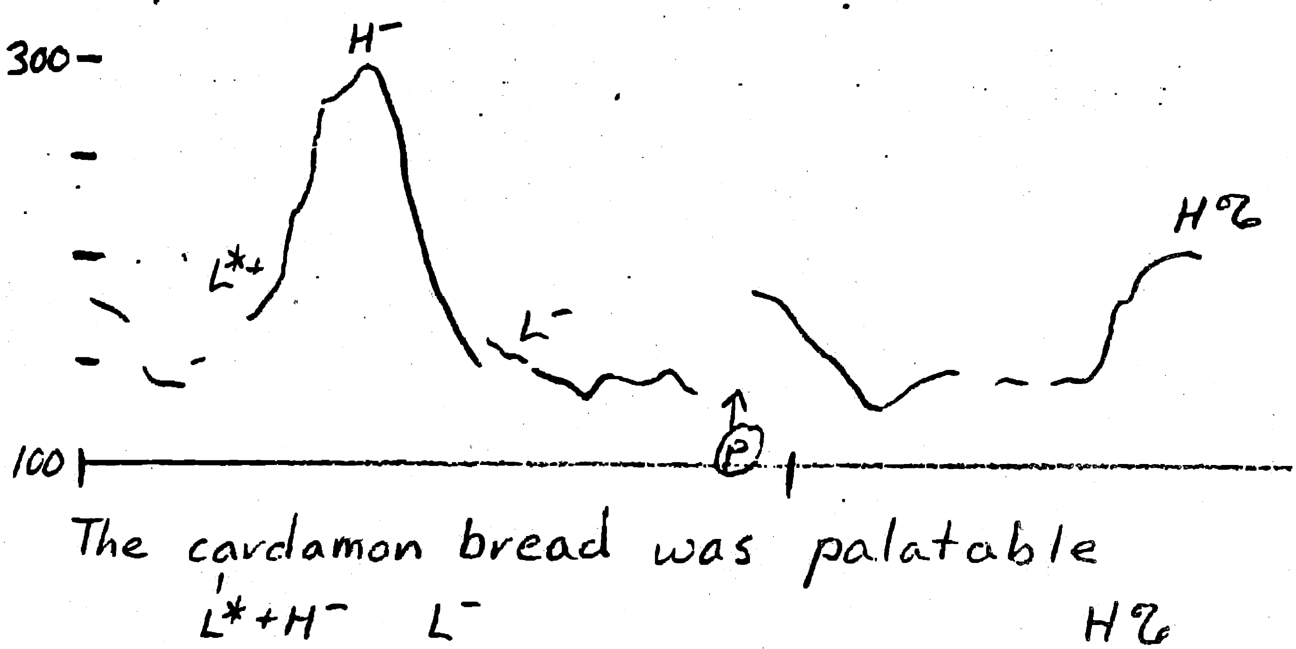
2.25) MB



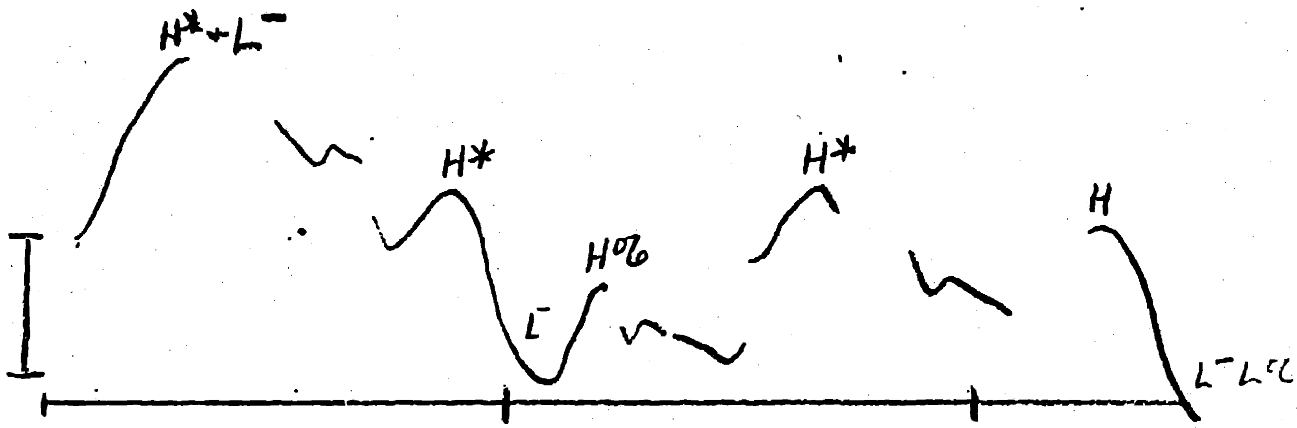
2.26 A) MB


H⁻
300
L*+
L⁻
H%
100
The cardamon bread was good
L* + H⁻ L⁻
H%

B) MB

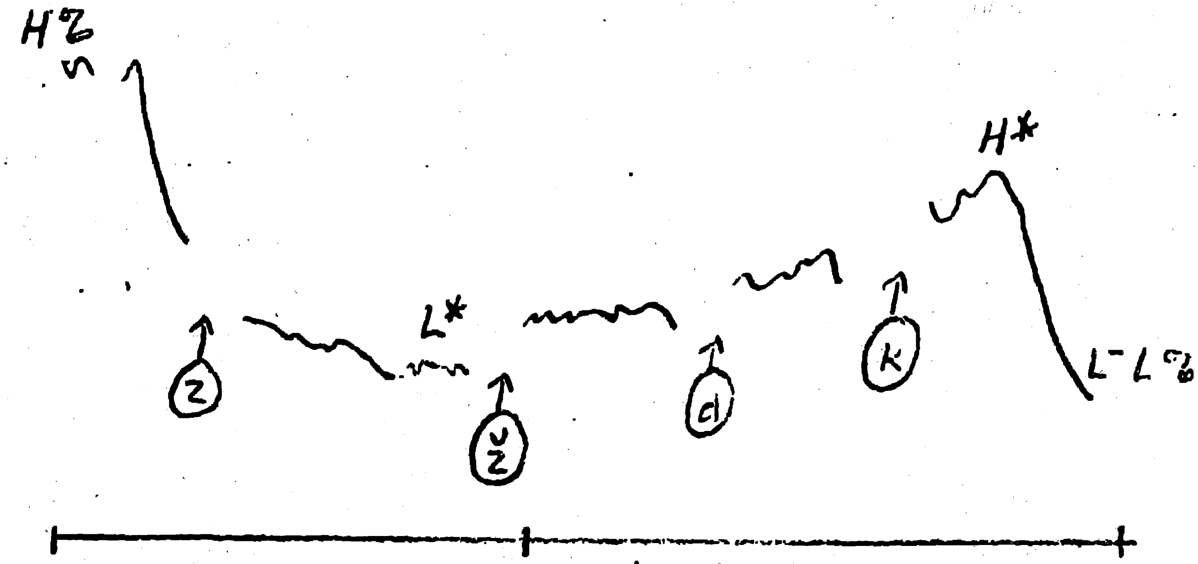

300
H⁻
L*+
L⁻
H%
P
100
The cardamon bread was palatable
L* + H⁻ L⁻
H%

2.26C) (from O'Shaughnessy 1976)



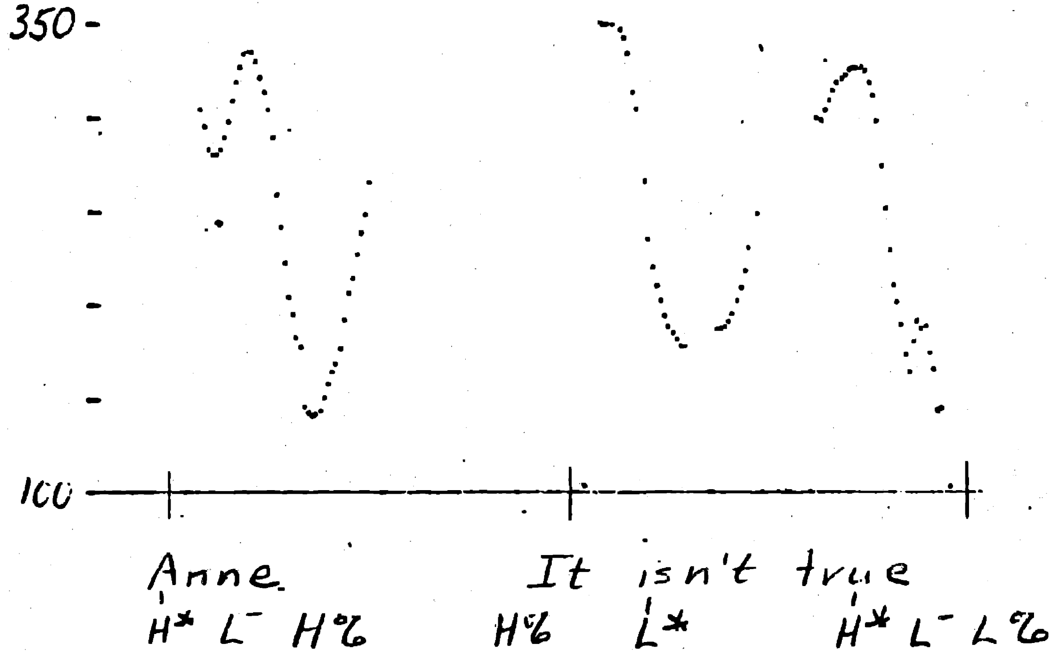
John studied Mary, and Bill studied Jane.
H*+L⁻ H* L⁻ H% H* H* L⁻ L%

2.27) MYL (from Liberman 1975)

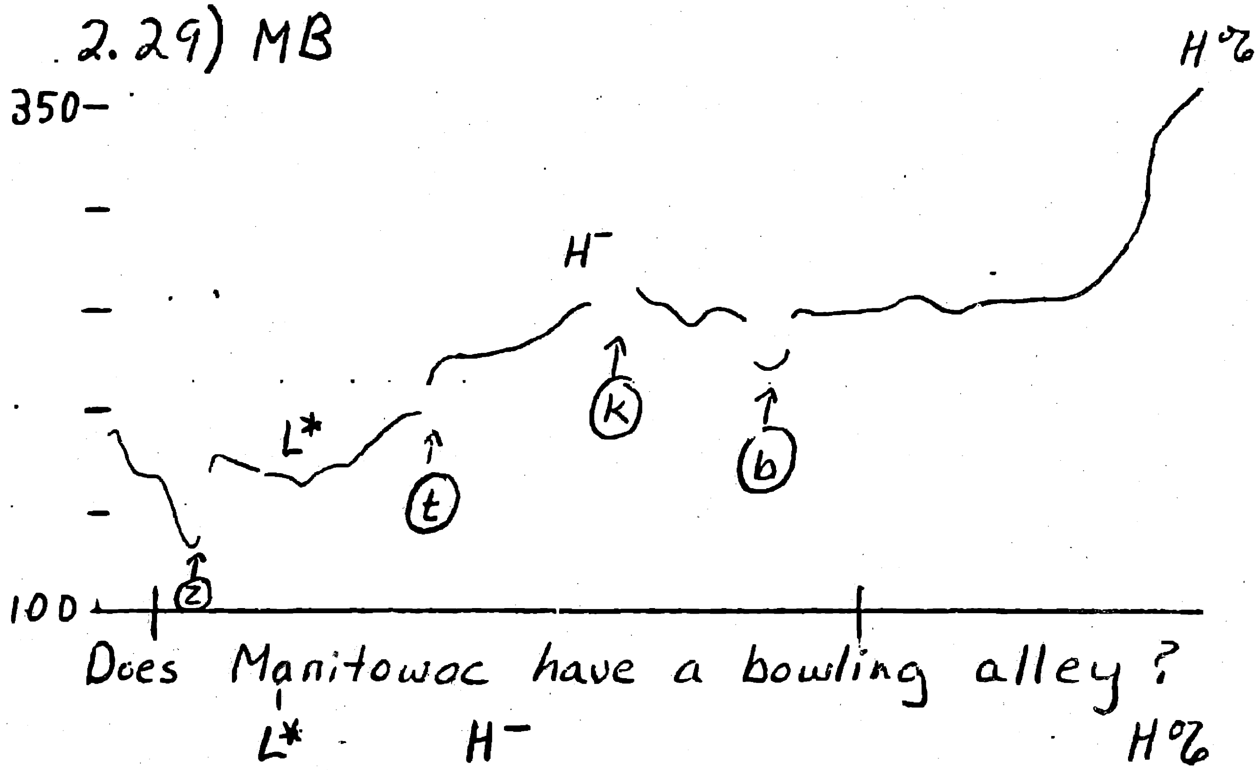


It was an unusually dark night
H% L* H* L⁻ L%

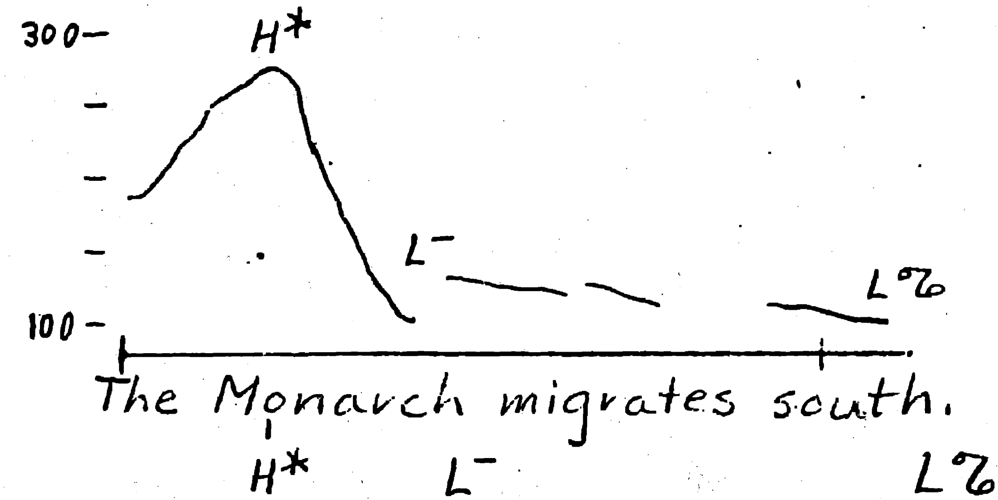
2.28) JBP



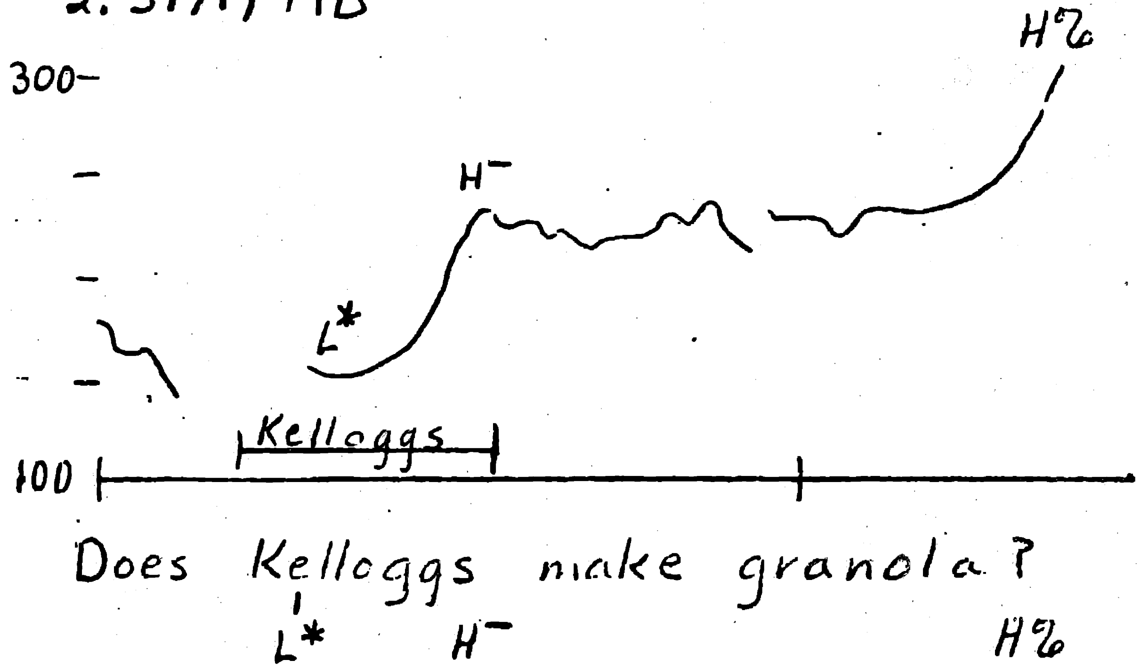
2.29) MB



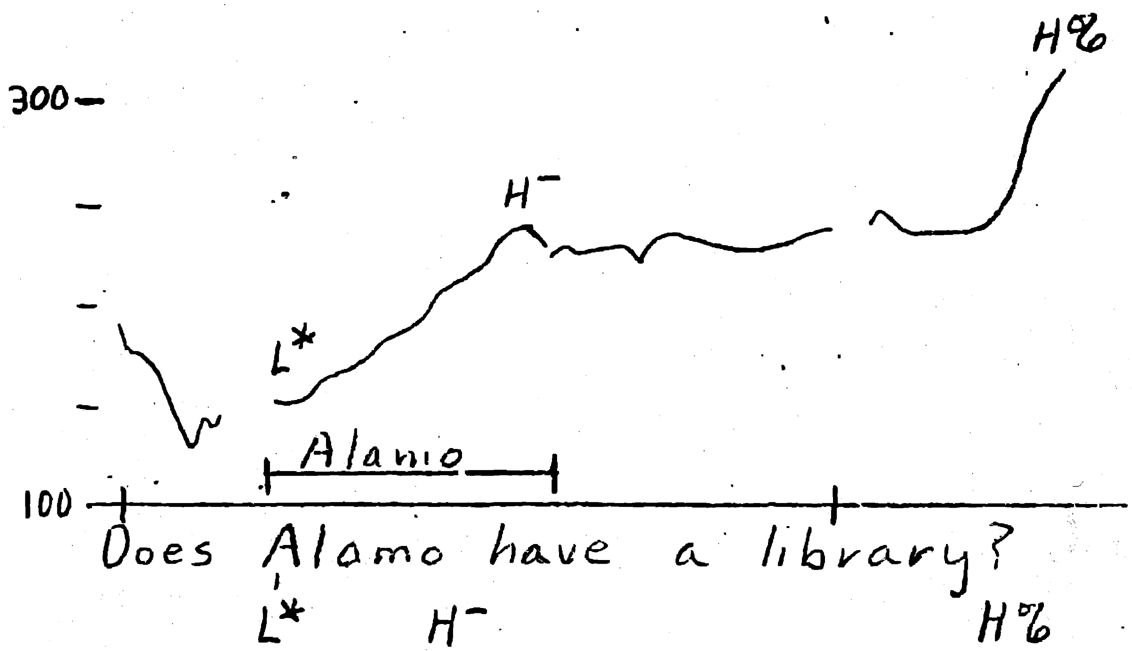
2.30) KXG



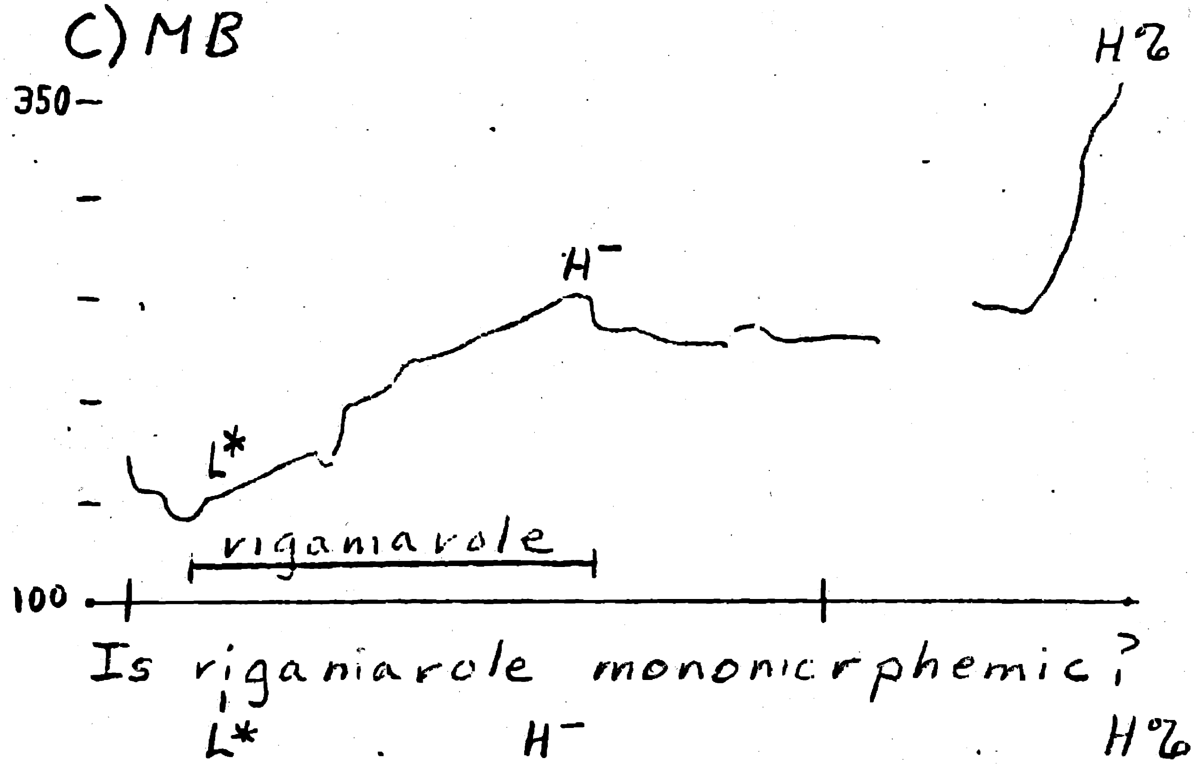
2.31A) MB



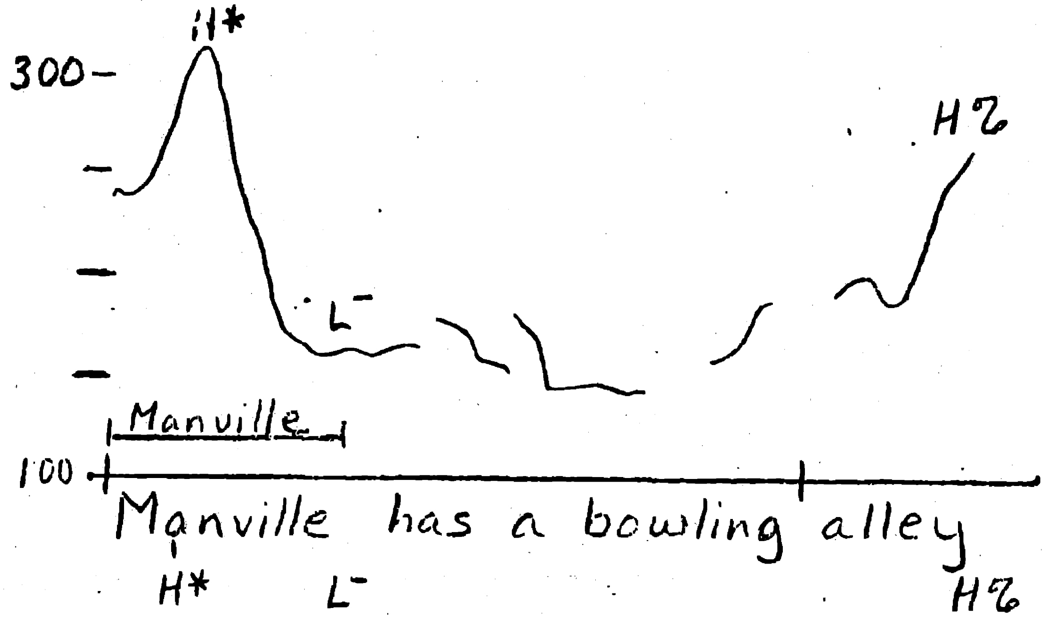
2, 31 B) MB


H%
300
H⁻
L*
Alamo
100
Does Alamo have a library?
L* H⁻ H%

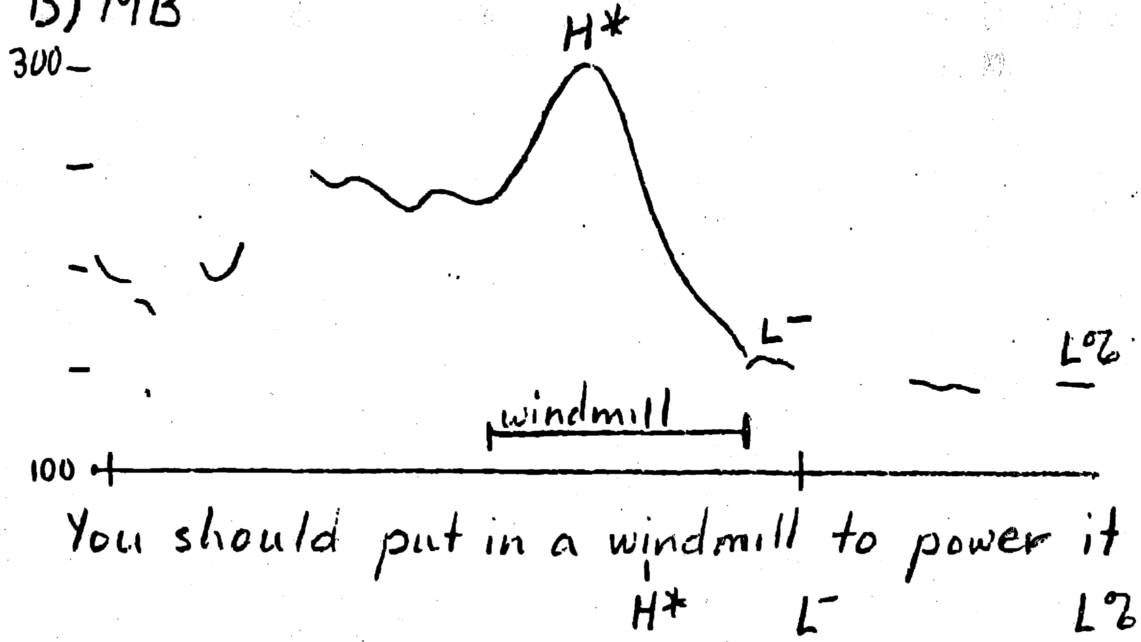
C) MB


H%
350
H⁻
L*
rigamarole
100
Is rigamarole monomorphemic?
L* H⁻ H%

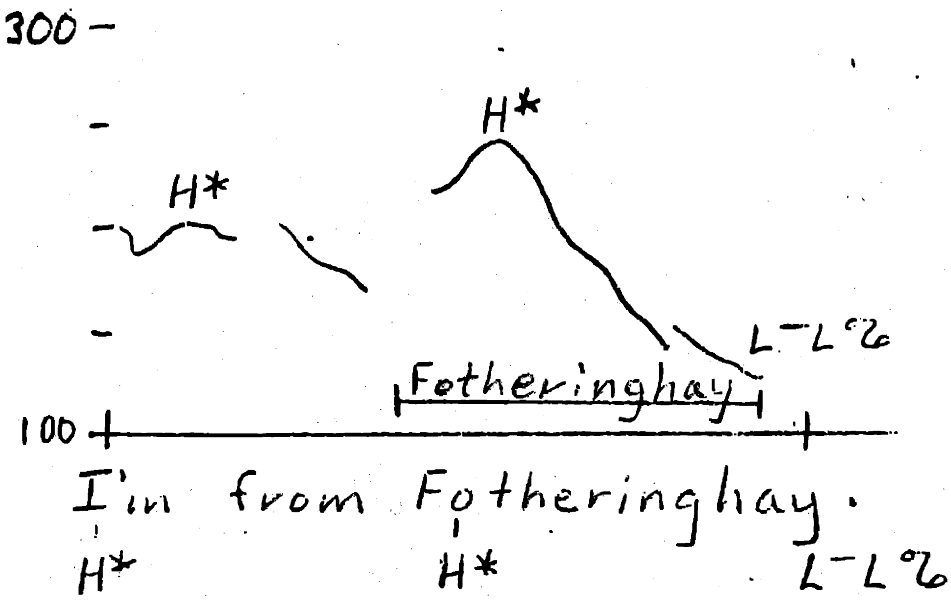
2.32A) MB


H*
300
H%
L-
Manville
100
Manville has a bowling alley
H*
L-
H%

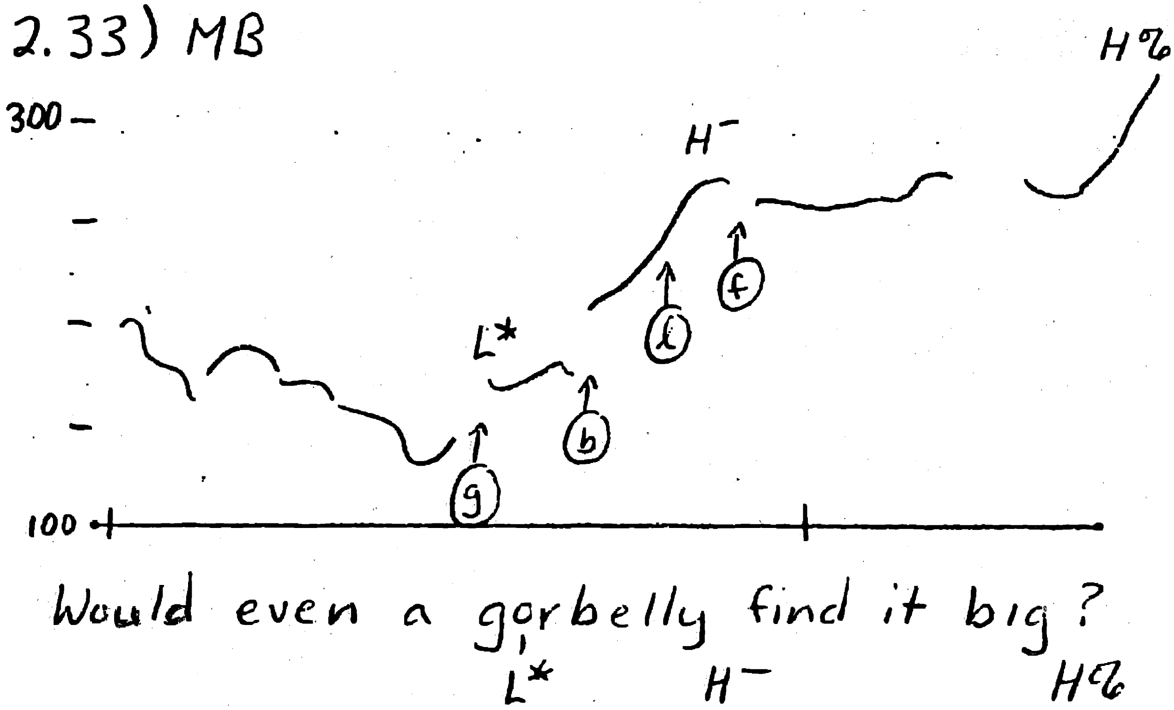
B) MB


300
H*
L-
L%
windmill
100
You should put in a windmill to power it
H*
L-
L%

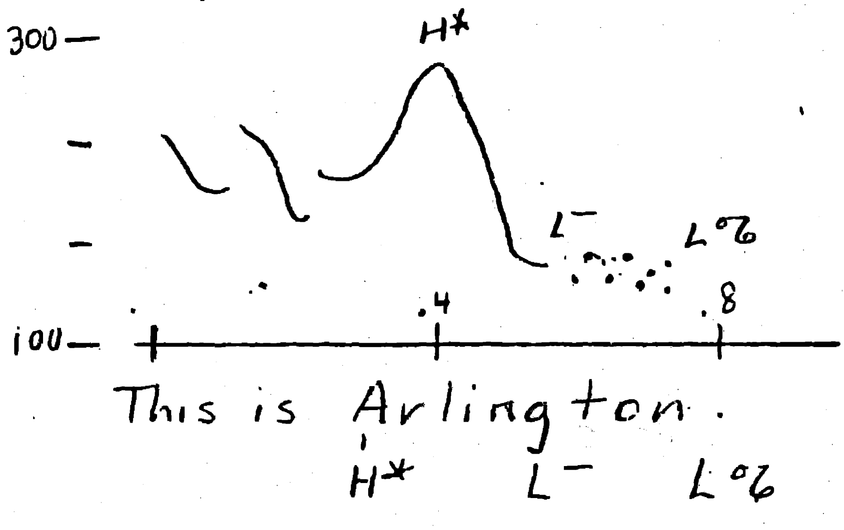
2.32c) MB



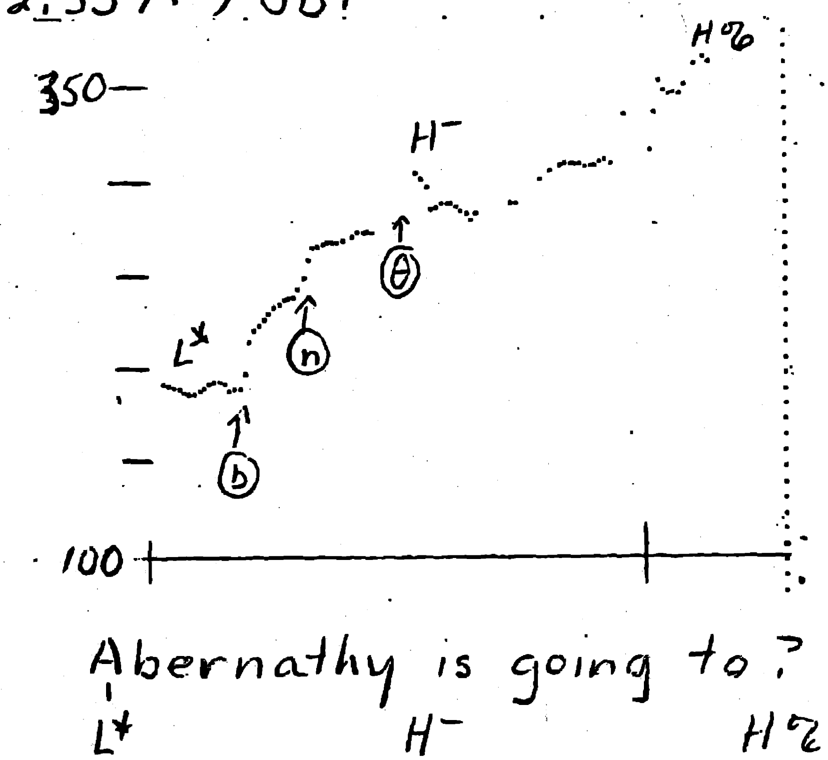
2.33) MB



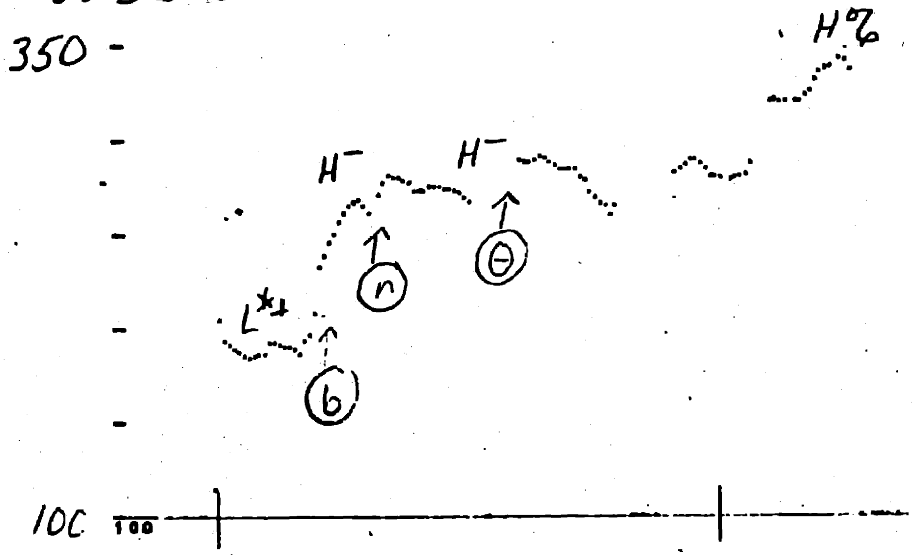
2.34) MB



2.35A) JBP



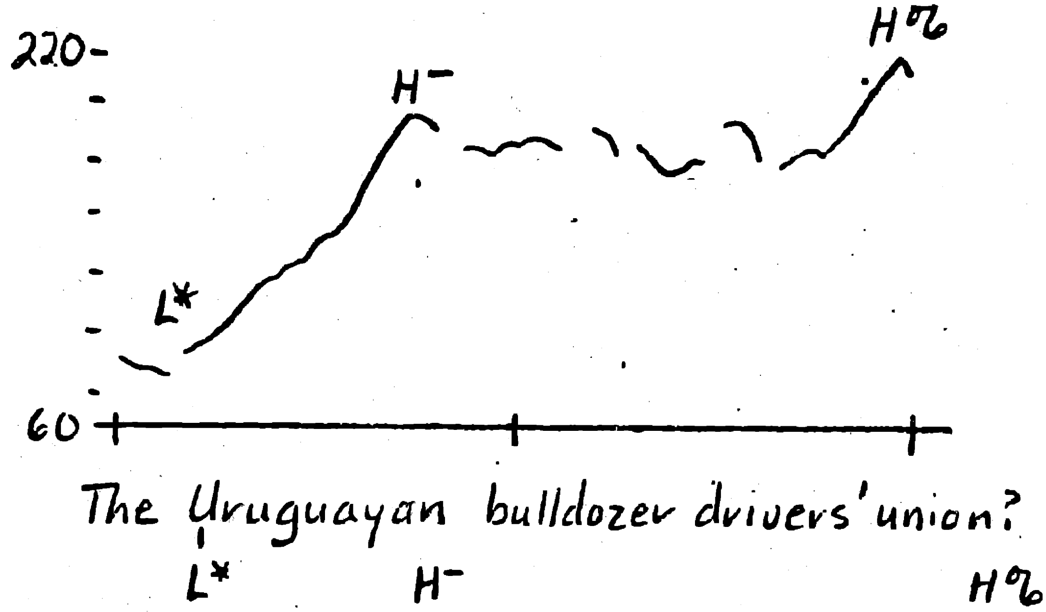
2.35 B) JBP



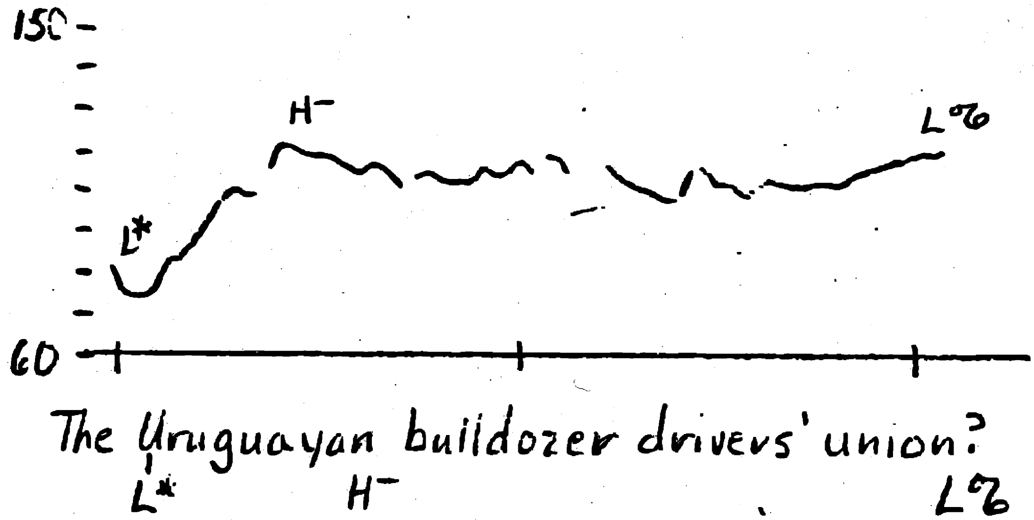
Abernathy is going to ?

L*+H- H- H%

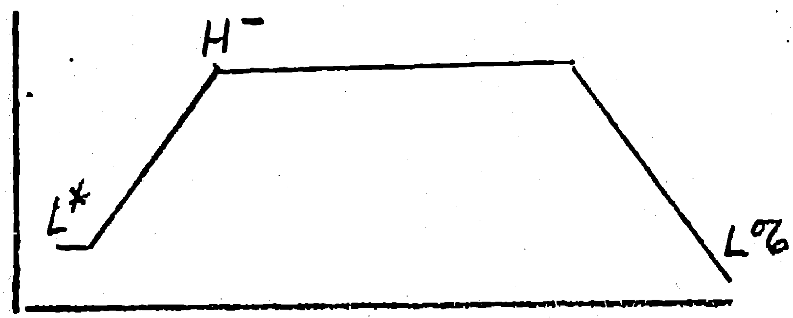
2.36A) MYL



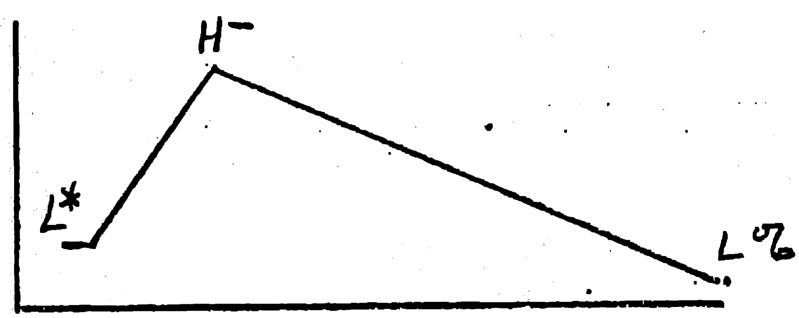
B) MYL



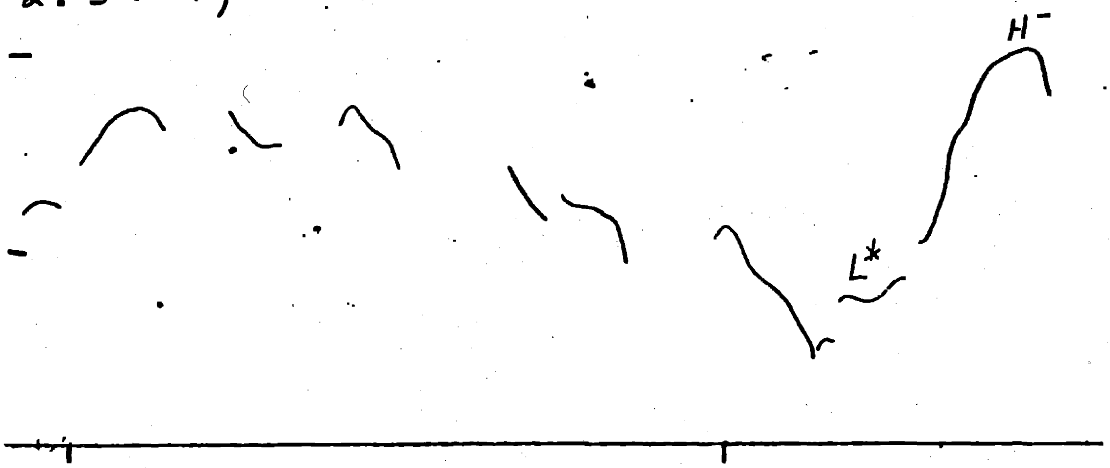
2.36C) impossible in English



D) also impossible



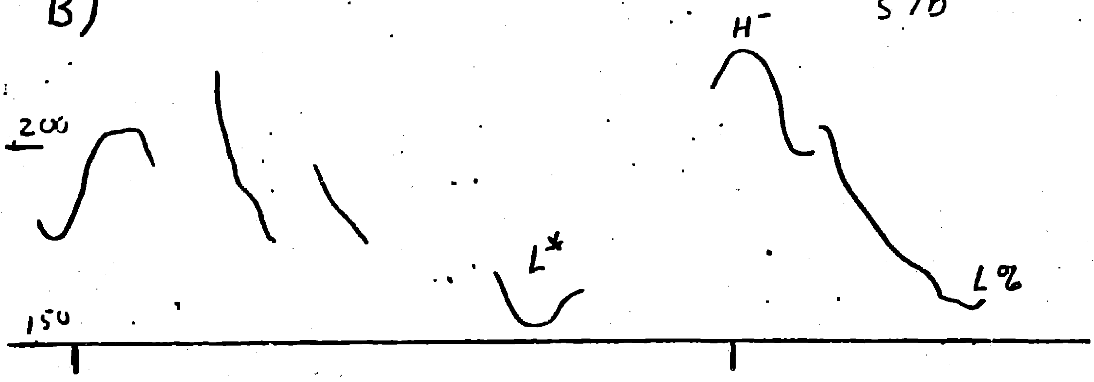
2.37A) 3c2



Karel koupil včerejši noviny ?
L* H⁻ (L%)

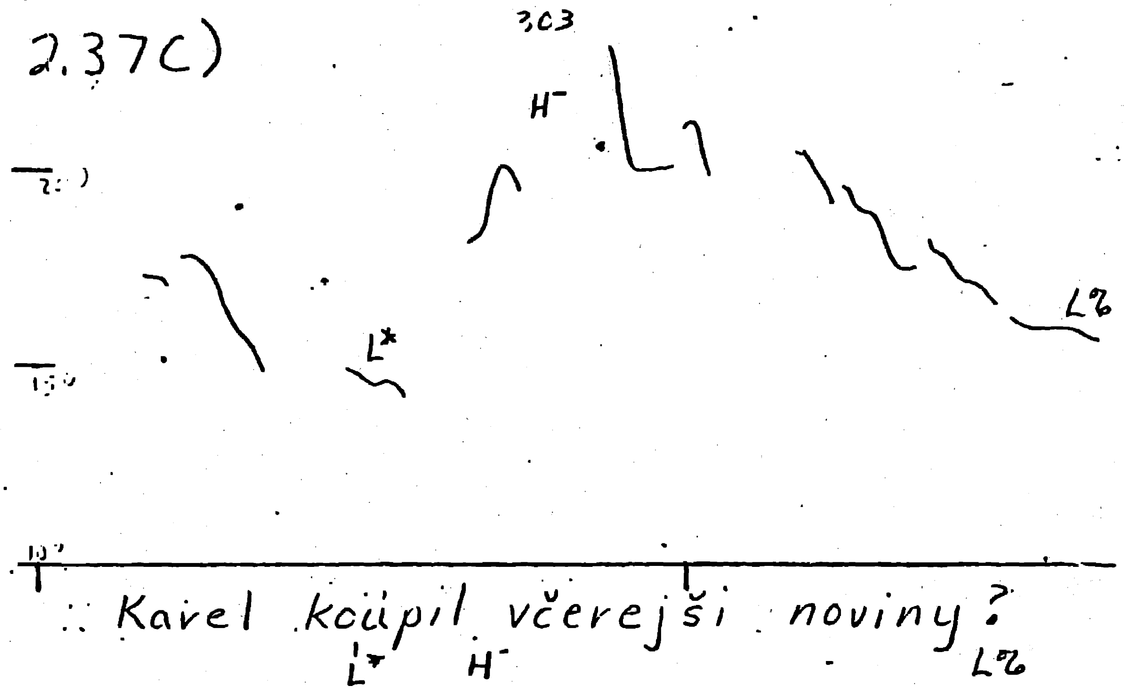
(Karel bought newspaper new)

B) 37b

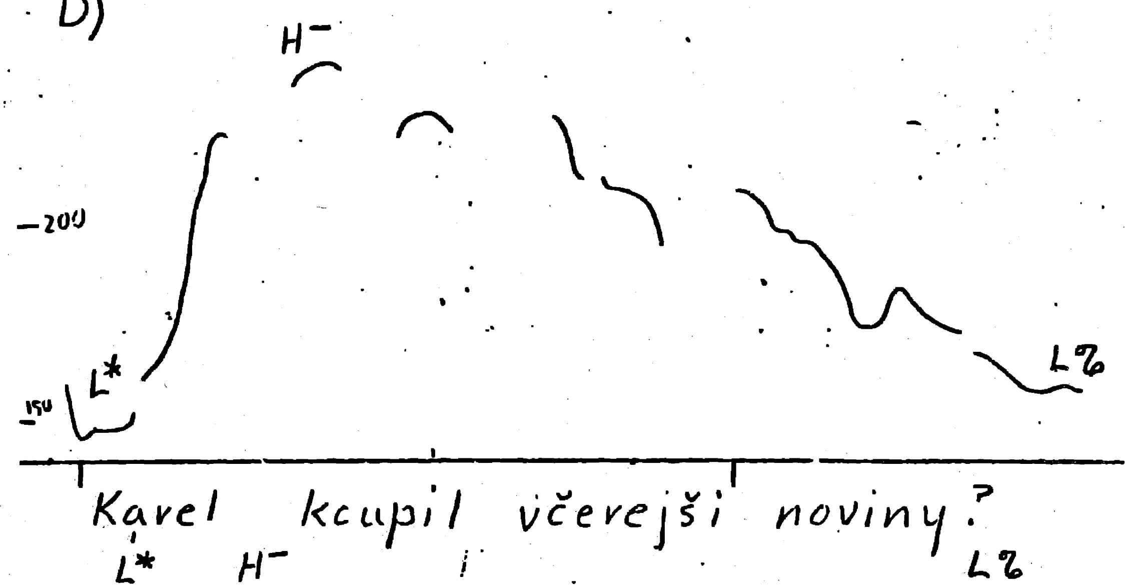


Karel koupil včerejši noviny ?
L* H⁻ L%

2.37C)

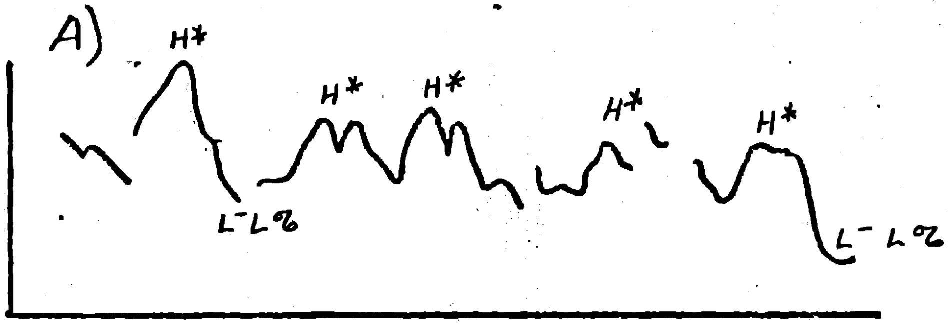

303
H⁻
200
150
L*
L%
100
Karel koupil včerejši noviny?
L* H⁻ L%

D)

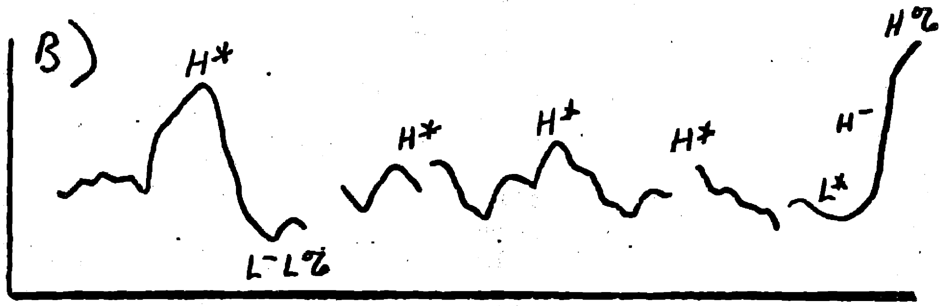

H⁻
200
150
L*
L%
Karel kcupil včerejši noviny?
L* H⁻ L%

2.38 A) MYL

A)

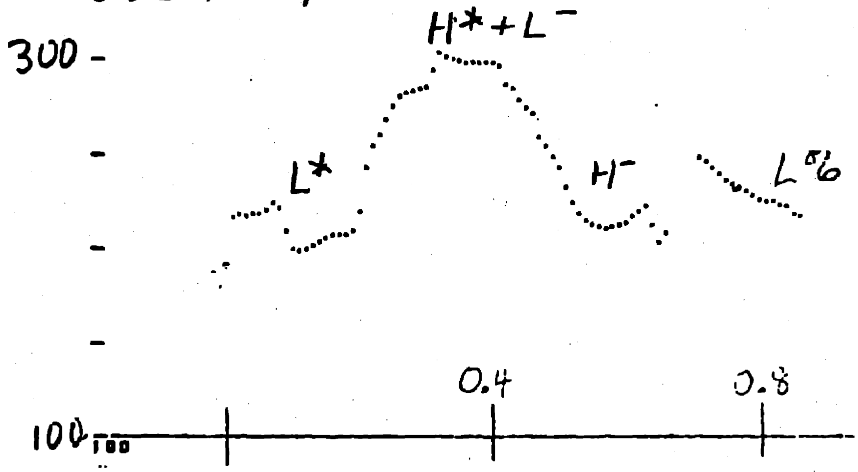


In November, the region's weather was unusually dry.
H* L⁻L% H* H* H* H* L⁻L%



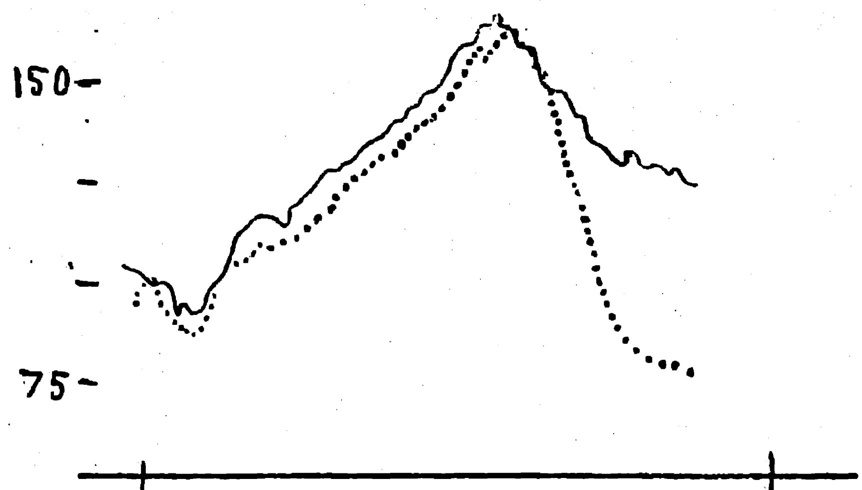
In November, was the region's weather unusually dry?
H* L⁻L% H* H* H* L* H⁻H%

2.39 A) JBP



Oh Abernathy
L* H*+L⁻ H⁻ L%

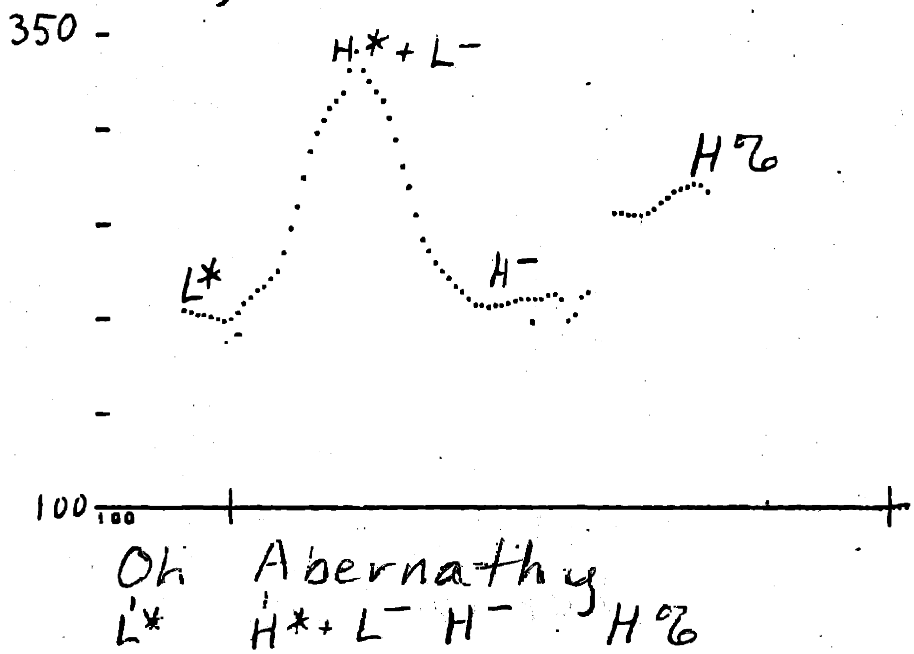
B) M4L (from Liberman 1975)



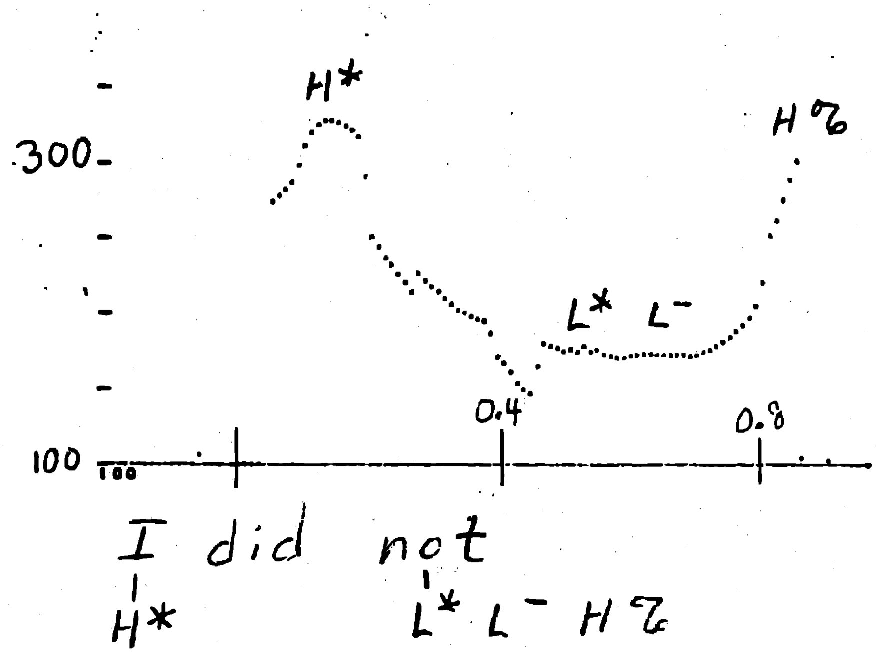
..... Cover your ears.
L* H* L⁻ L%

—— Cover your ears
L* H*+L⁻ H⁻ L%

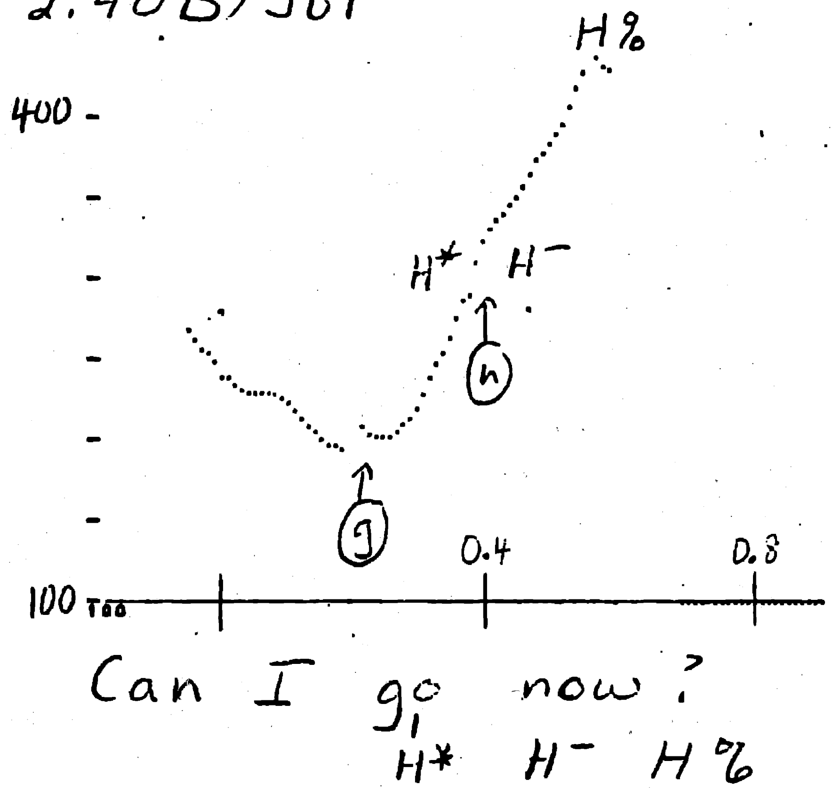
2.39 C) JBP



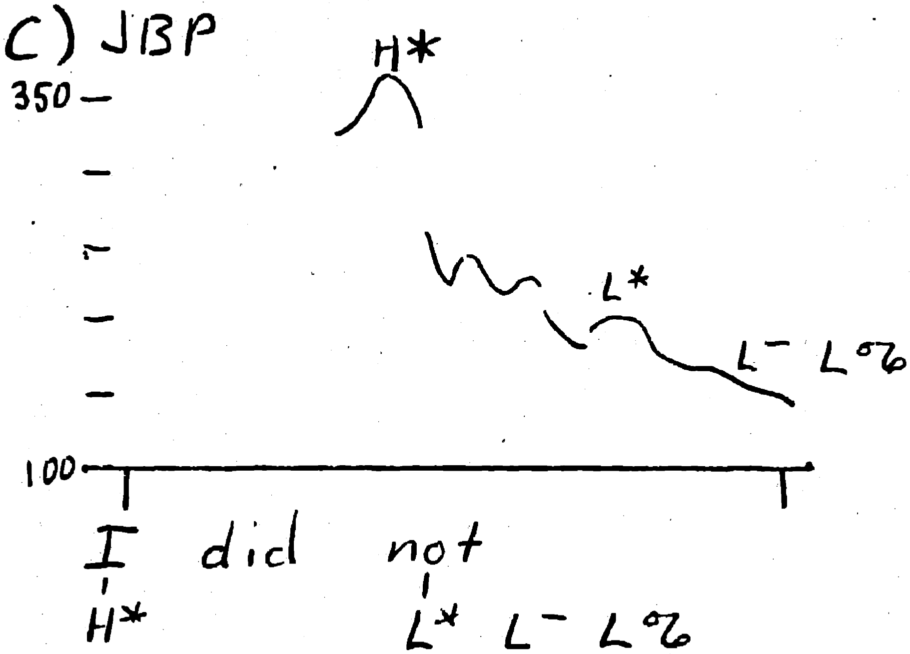
2.40 A) JBP



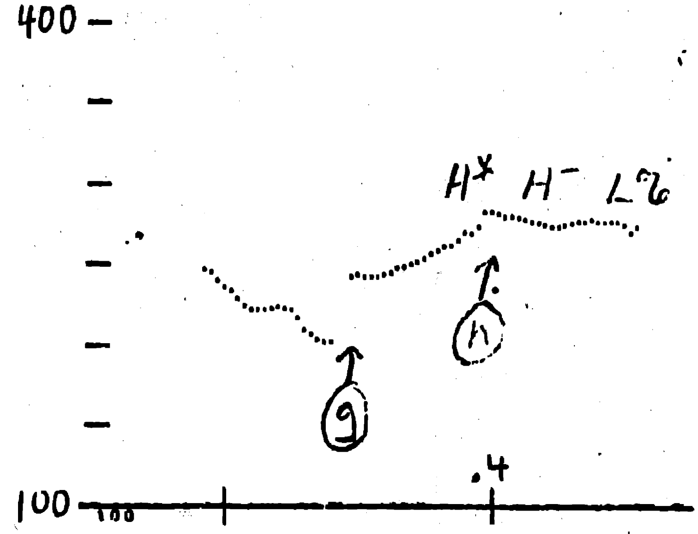
2.40 B) JBP



C) JBP

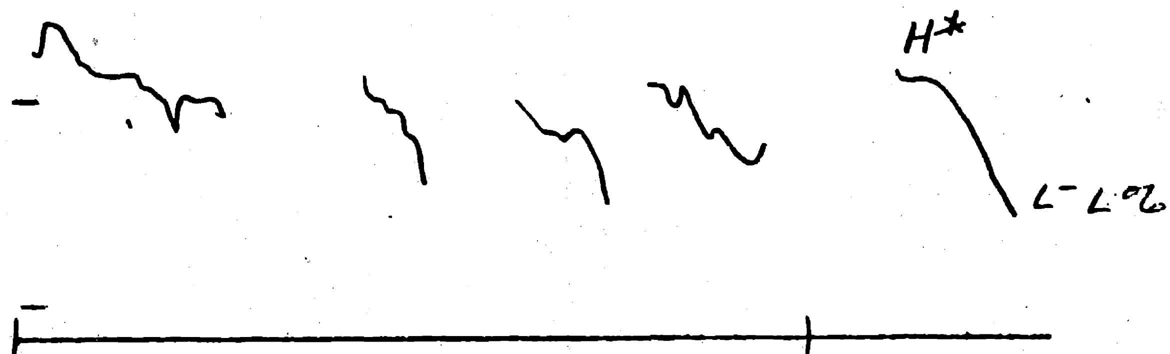


2.40D) JBP



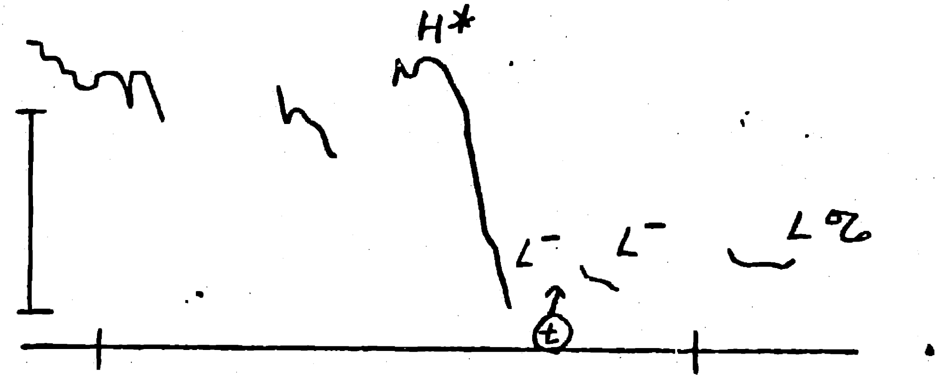
Can I go now?
H* H- L%

2.41) MYL (from Liberman 1975)



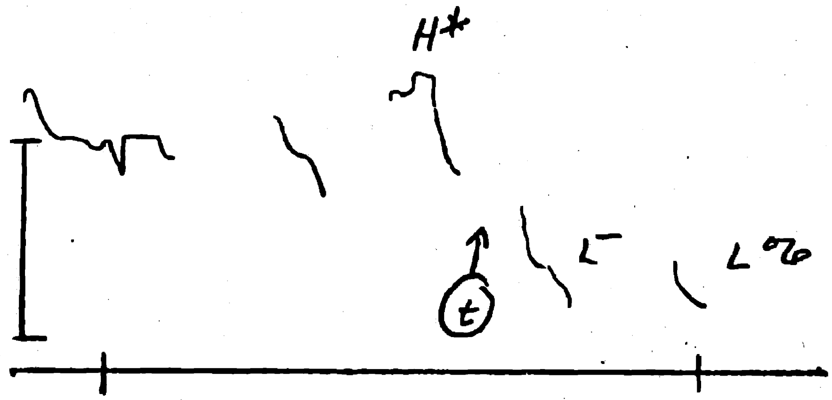
Sam struck out my friend.
H* L- L%

2.42) MYL (from Liberman 1975) 42



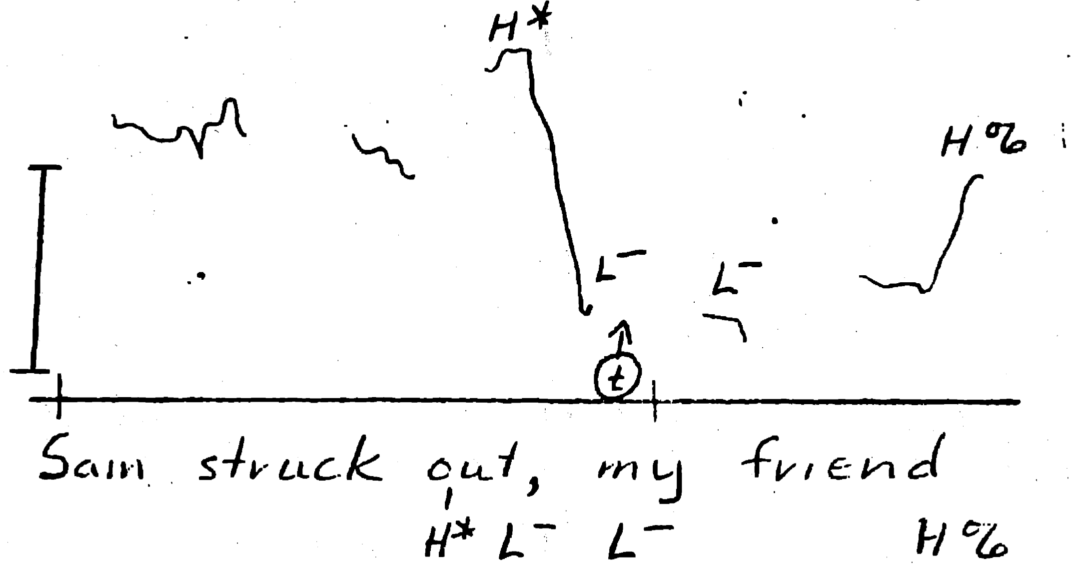
Sam struck out, my friend.
H* L⁻ L⁻ L%

2.43) MYL (from Liberman 1975)

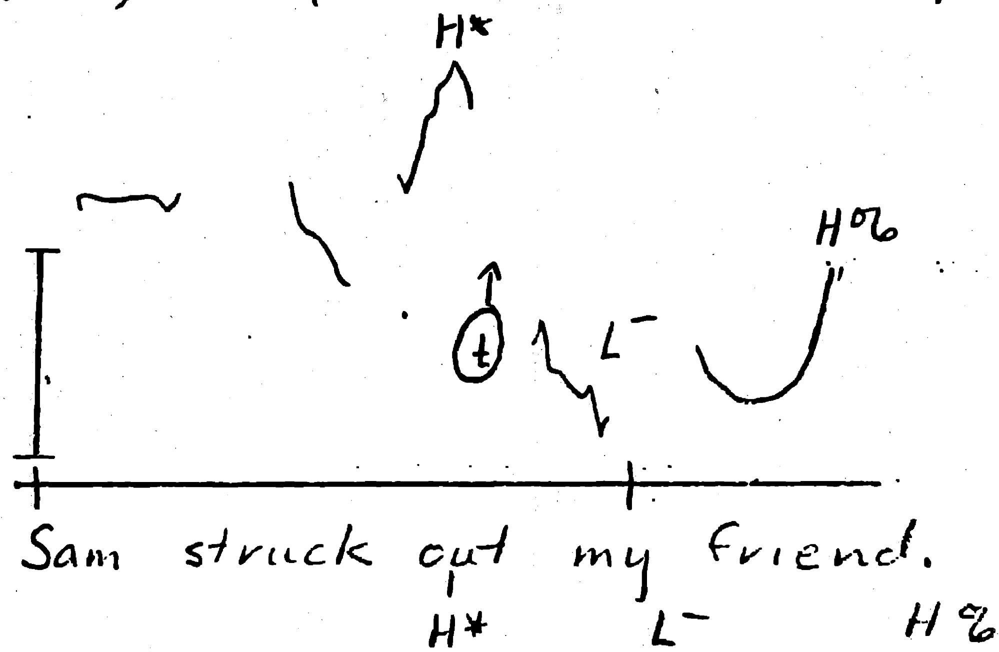


Sam struck out my friend
H* L⁻ L% 4,

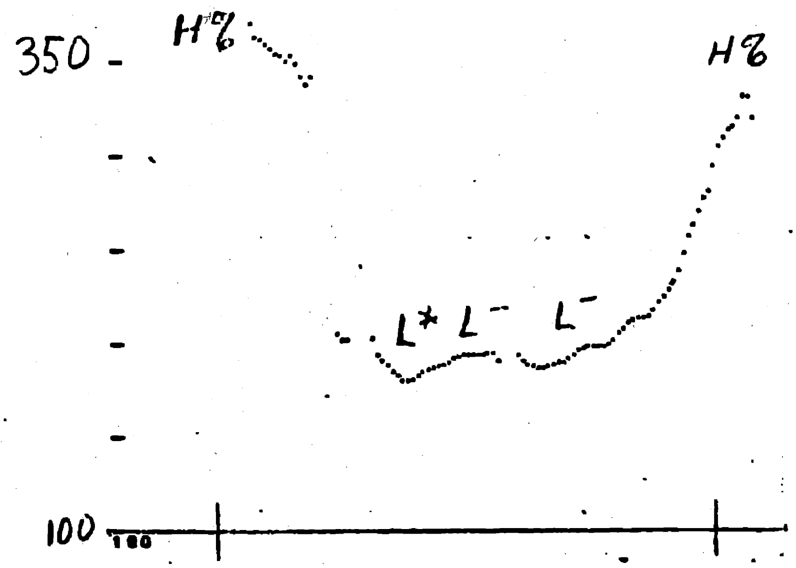
2.44) MYL (from Liberman 1975)



2.45) MYL (from Liberman 1975)

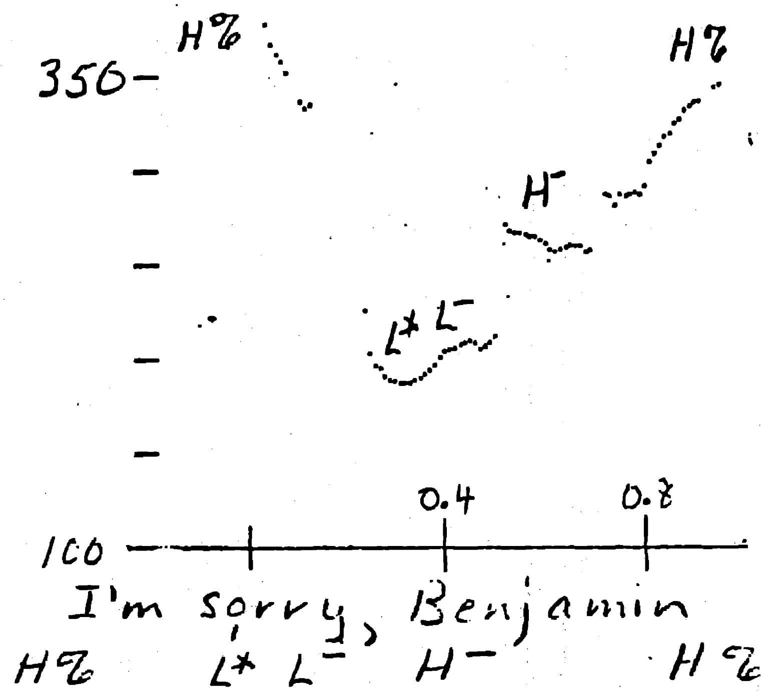


2.46) JBP

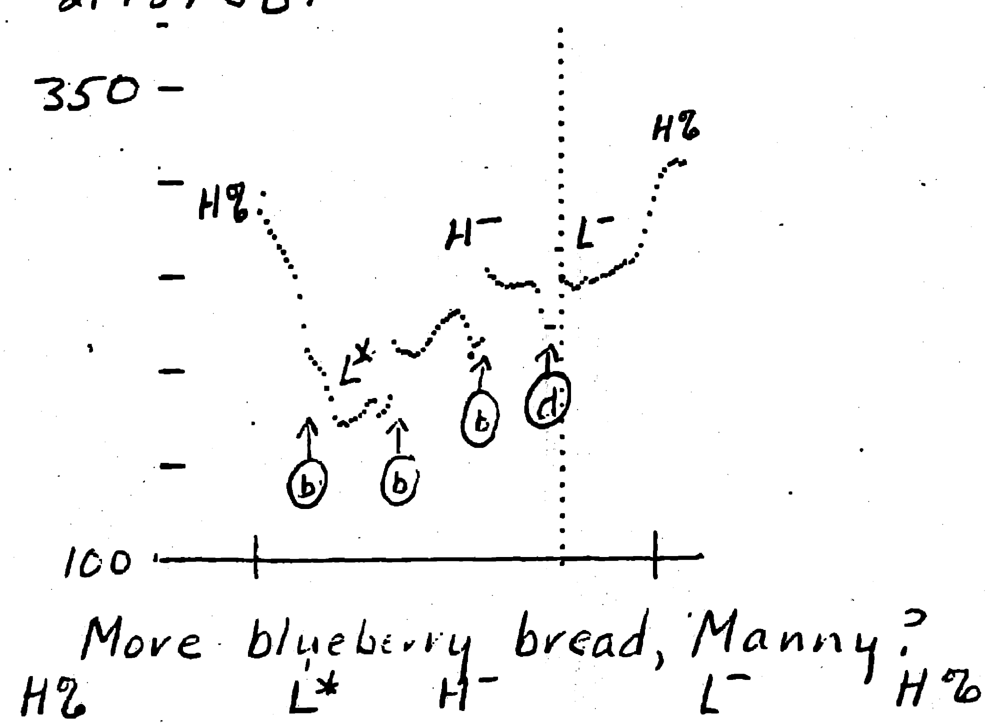


I'm sorry, Benjamin
H%. L* L⁻. L⁻ H%

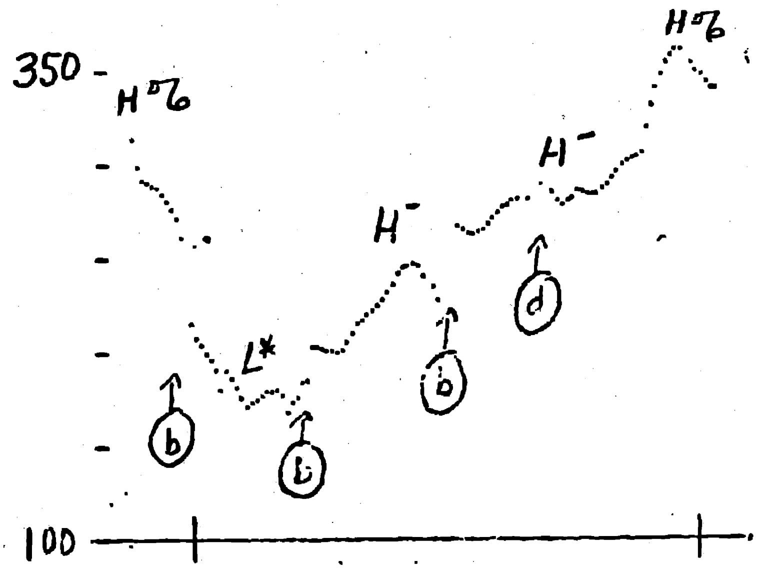
2.47) JBP



2.48) JBP

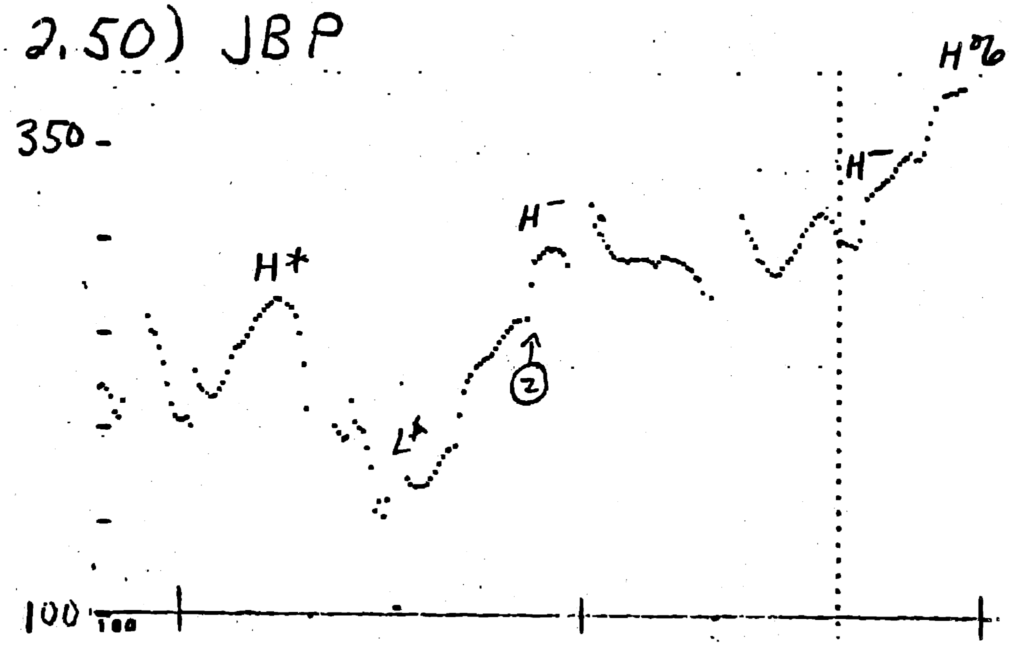


2.49) JBP



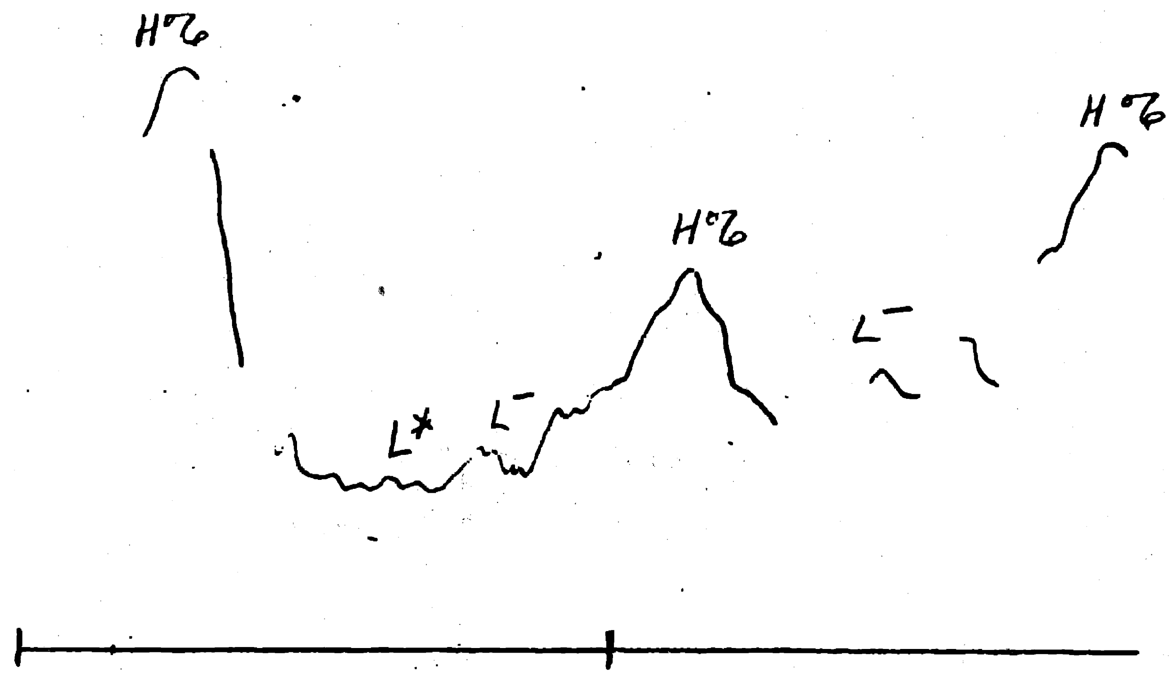
More blueberry bread, Manny?
H% L* H- H- H%

2.50) JBP



Is he a member of the bulldozer drivers' union, Manny?
H* L* H- H- H%

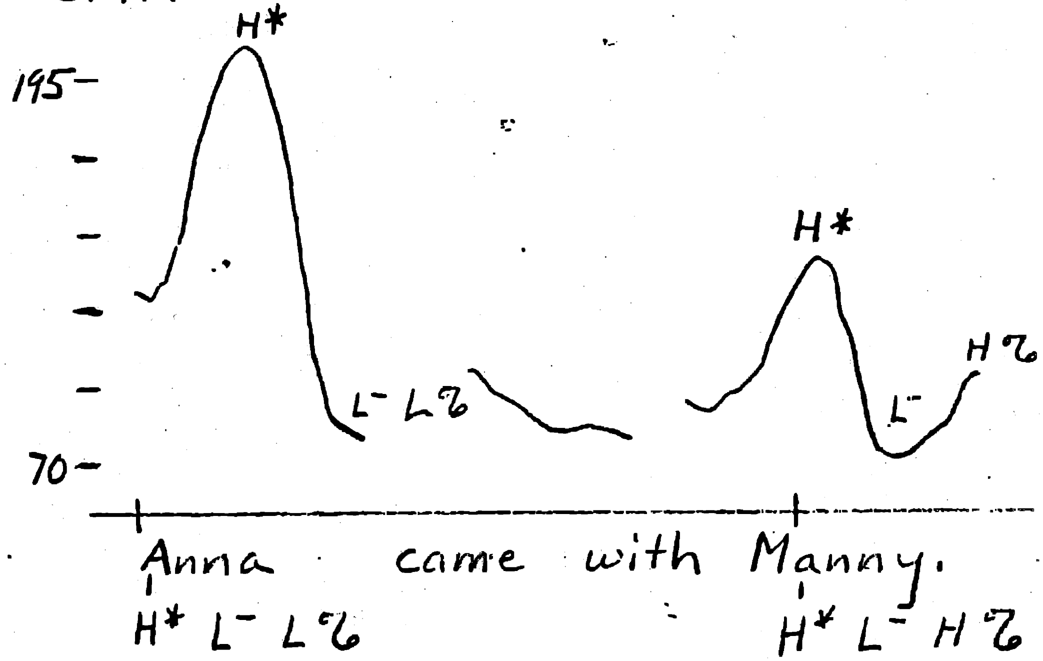
2.51) MYL (from Liberman 1975)



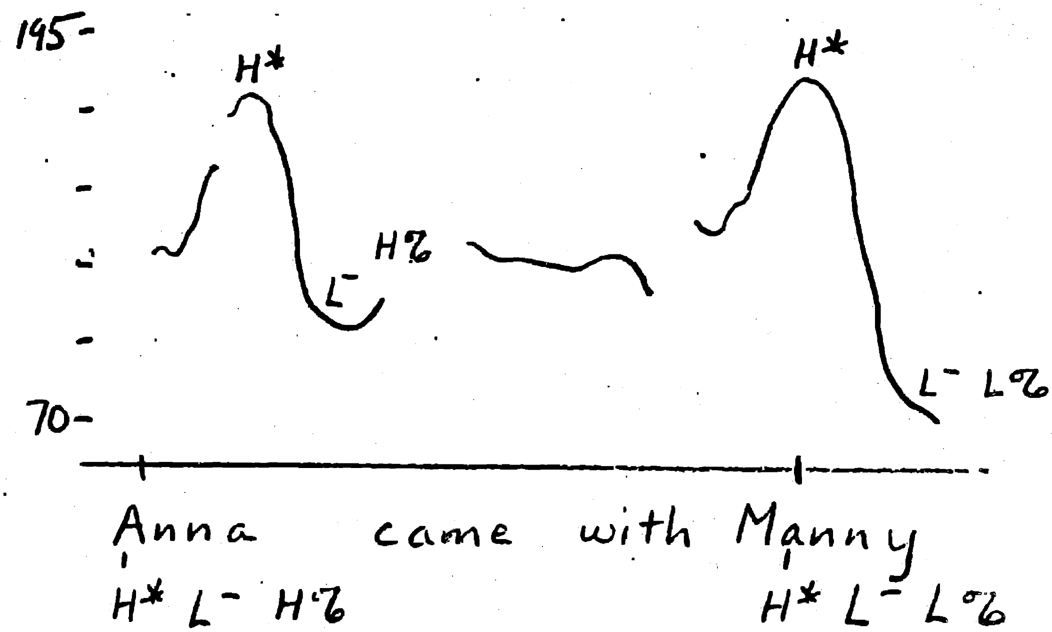
That isn't an elevator, you simpleton.

H% L* L⁻ H% L⁻ H%

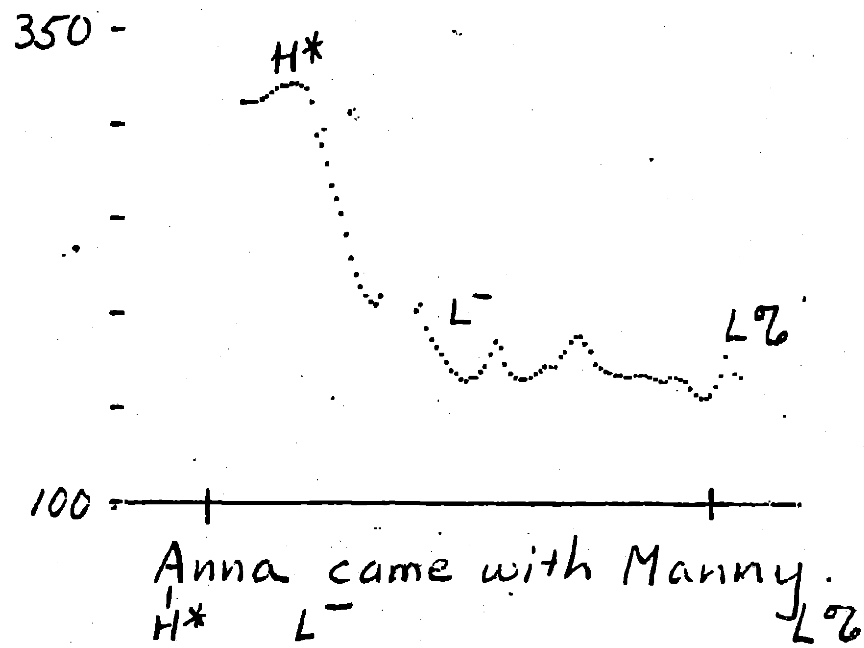
3.1) MYL


H*
195
H*
H%
L- L%
L-
70
Anna came with Manny.
H* L- L%
H* L- H%

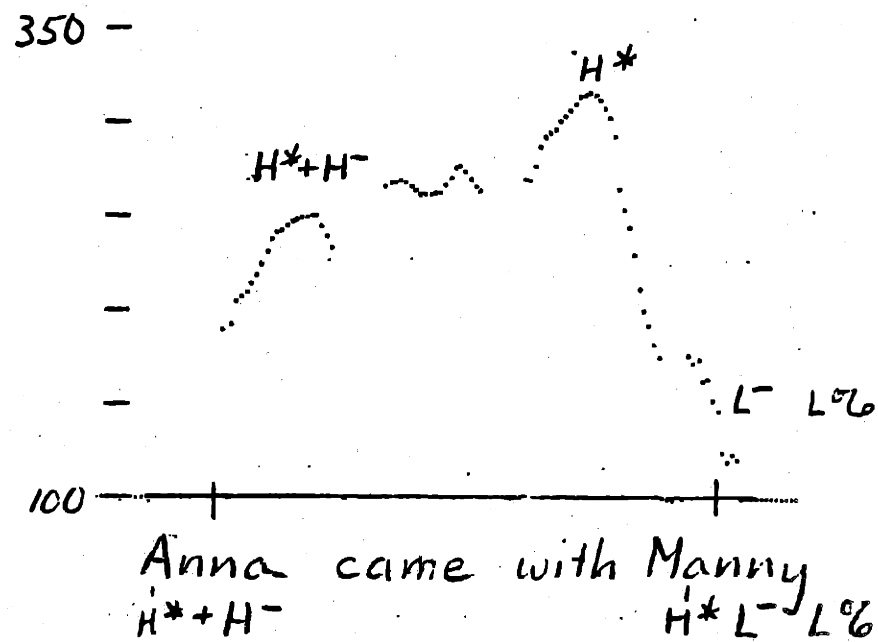
3.2) MYL


195
H*
H*
H%
L-
L- L%
70
Anna came with Manny
H* L- H%
H* L- L%

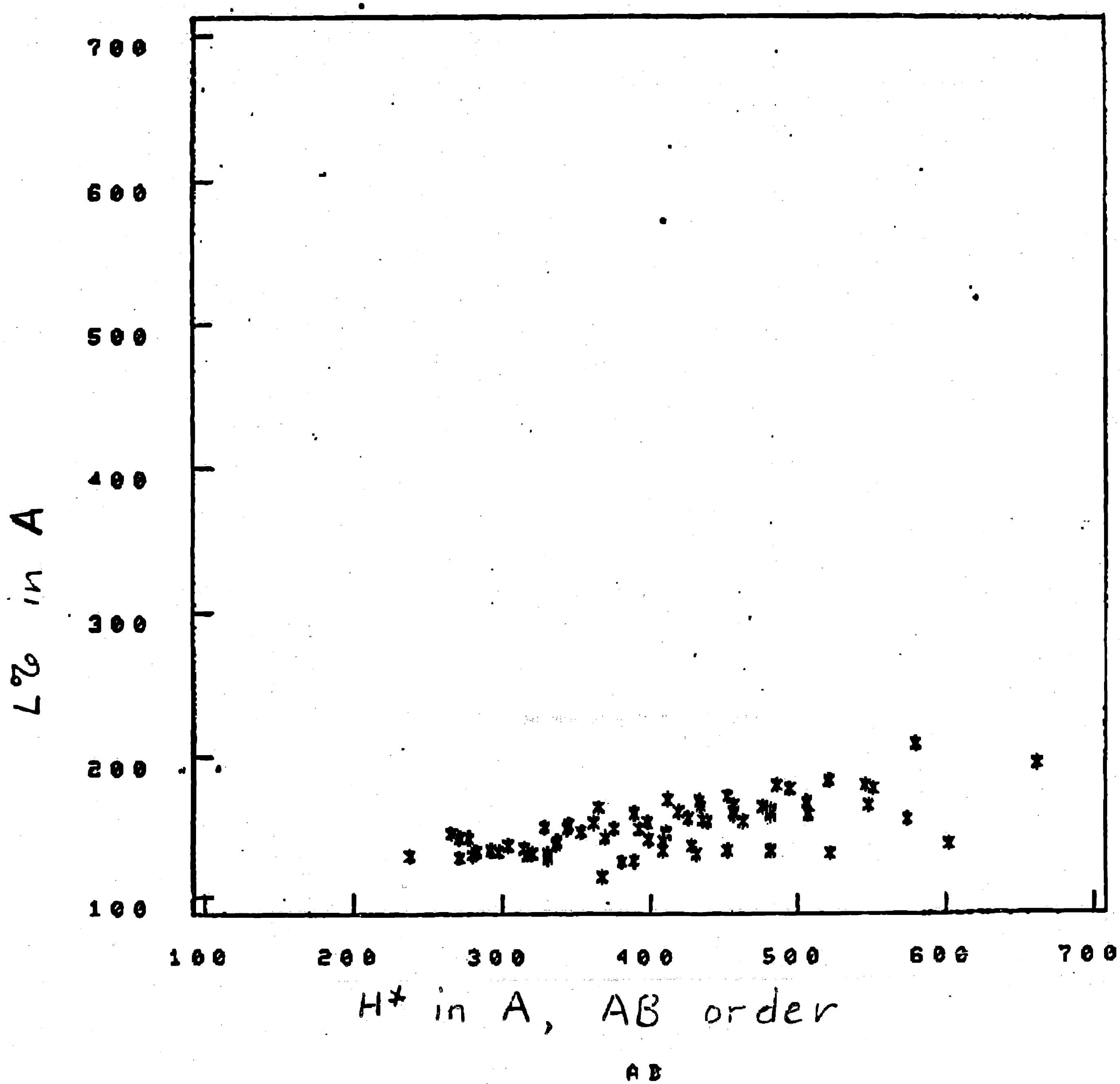
3.3) JBP


350
100
H*
L⁻
L%
Anna came with Manny.
H* L⁻ L%

3,4) JBP

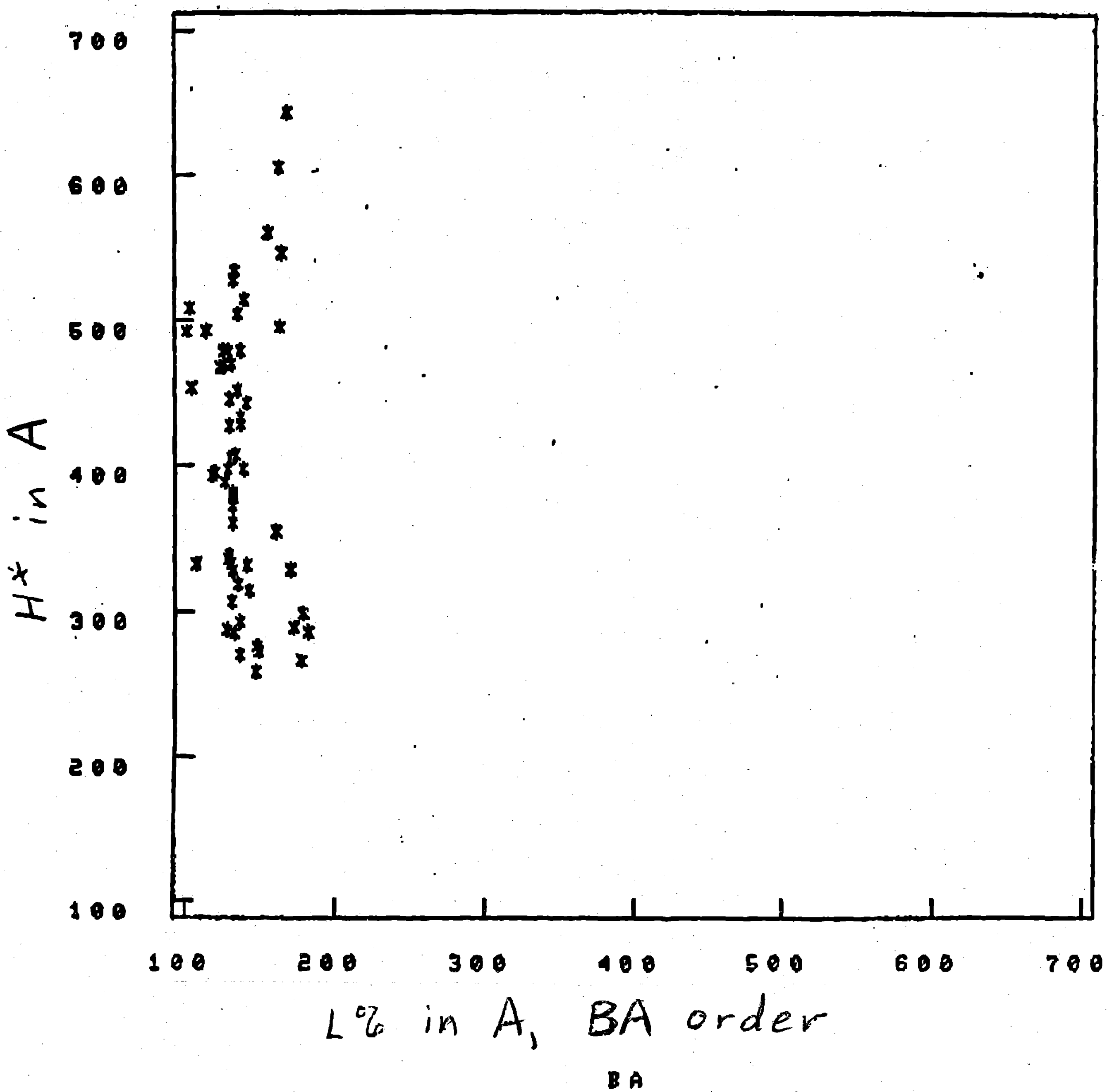

350
100
H*+H⁻
H*
L⁻ L%
Anna came with Manny
H*+H⁻ H* L⁻ L%

3.5) SUBJECT JBP

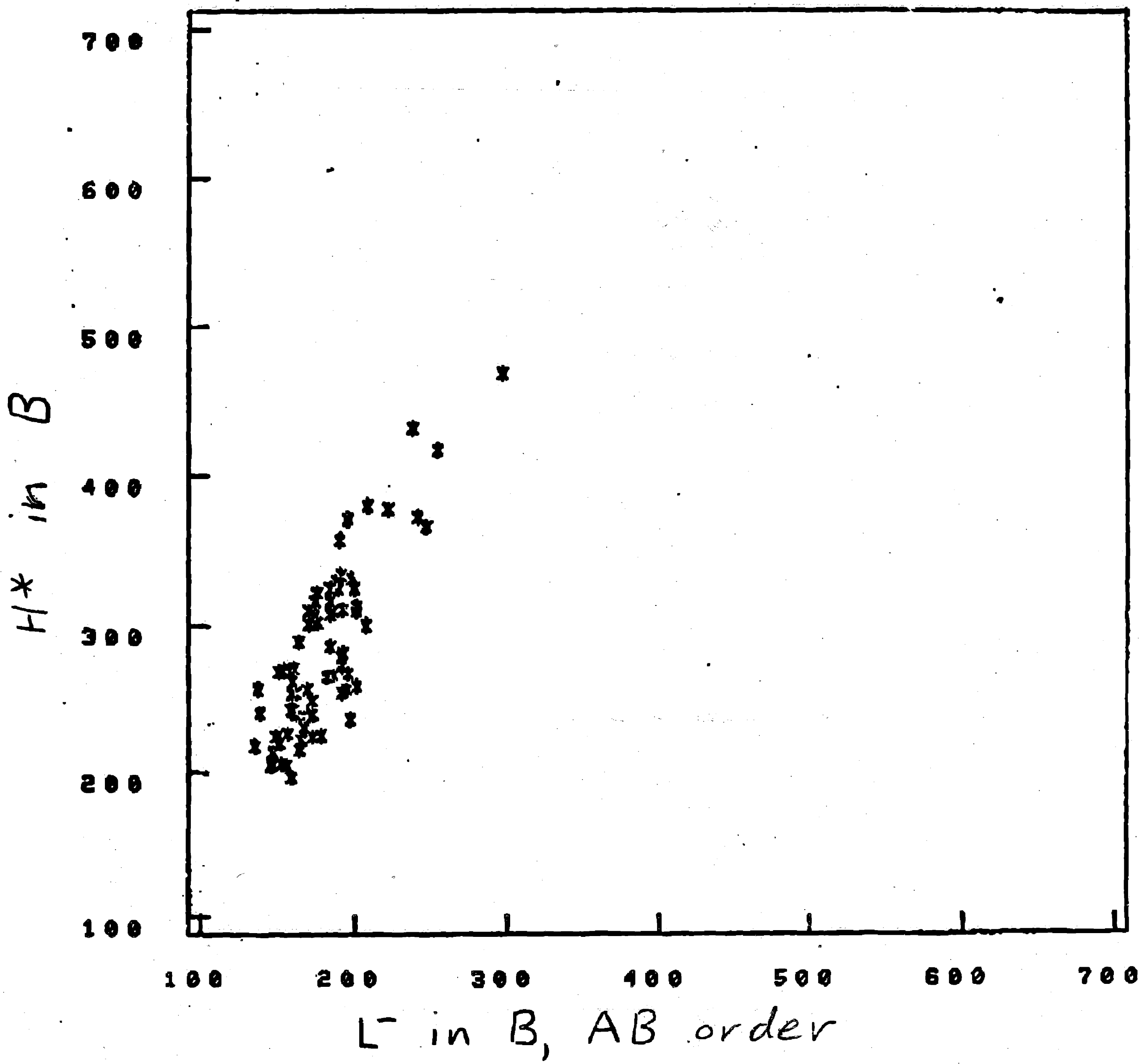


3.6)

SUBJECT JBP

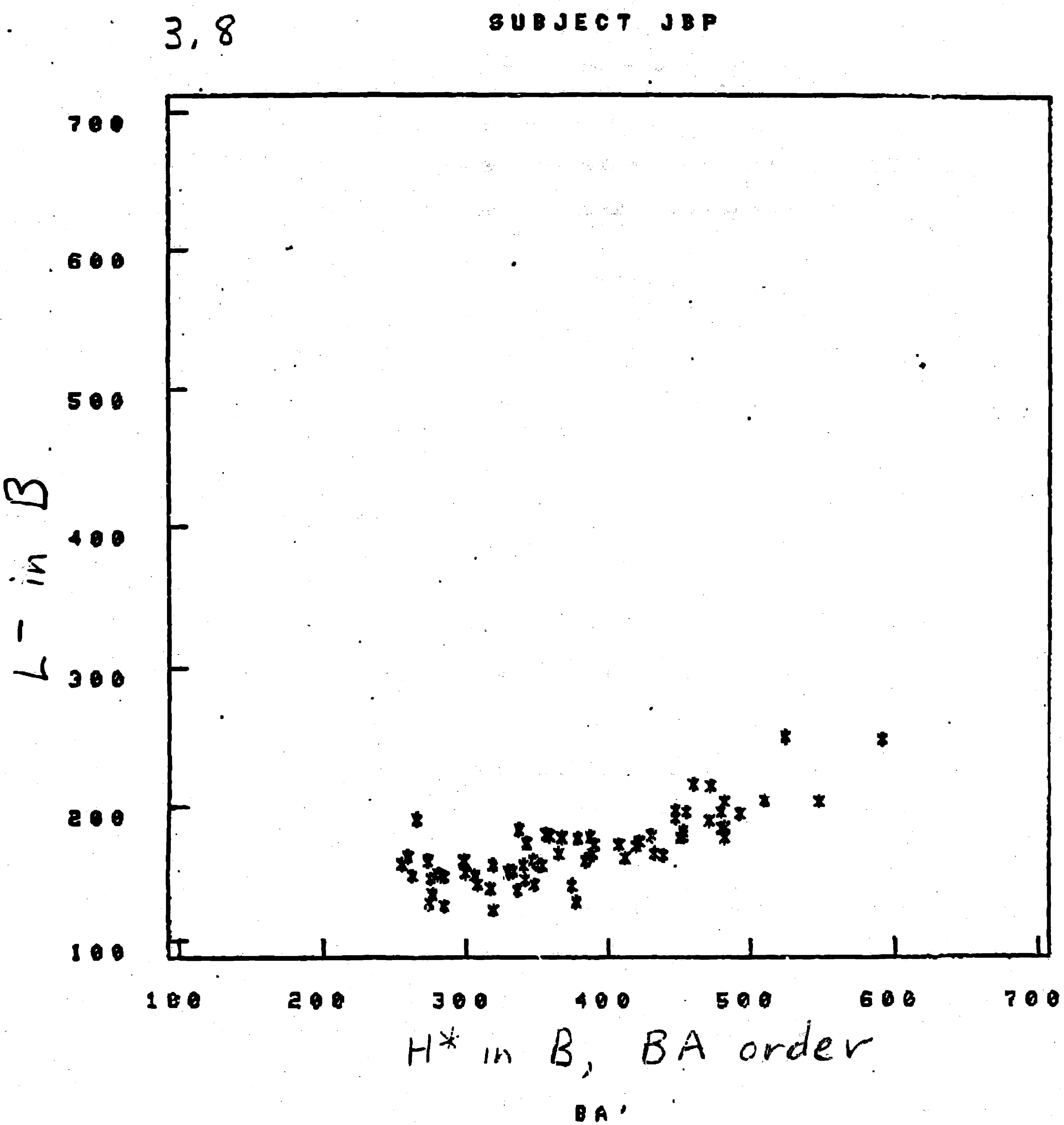


3.7) SUBJECT JBP

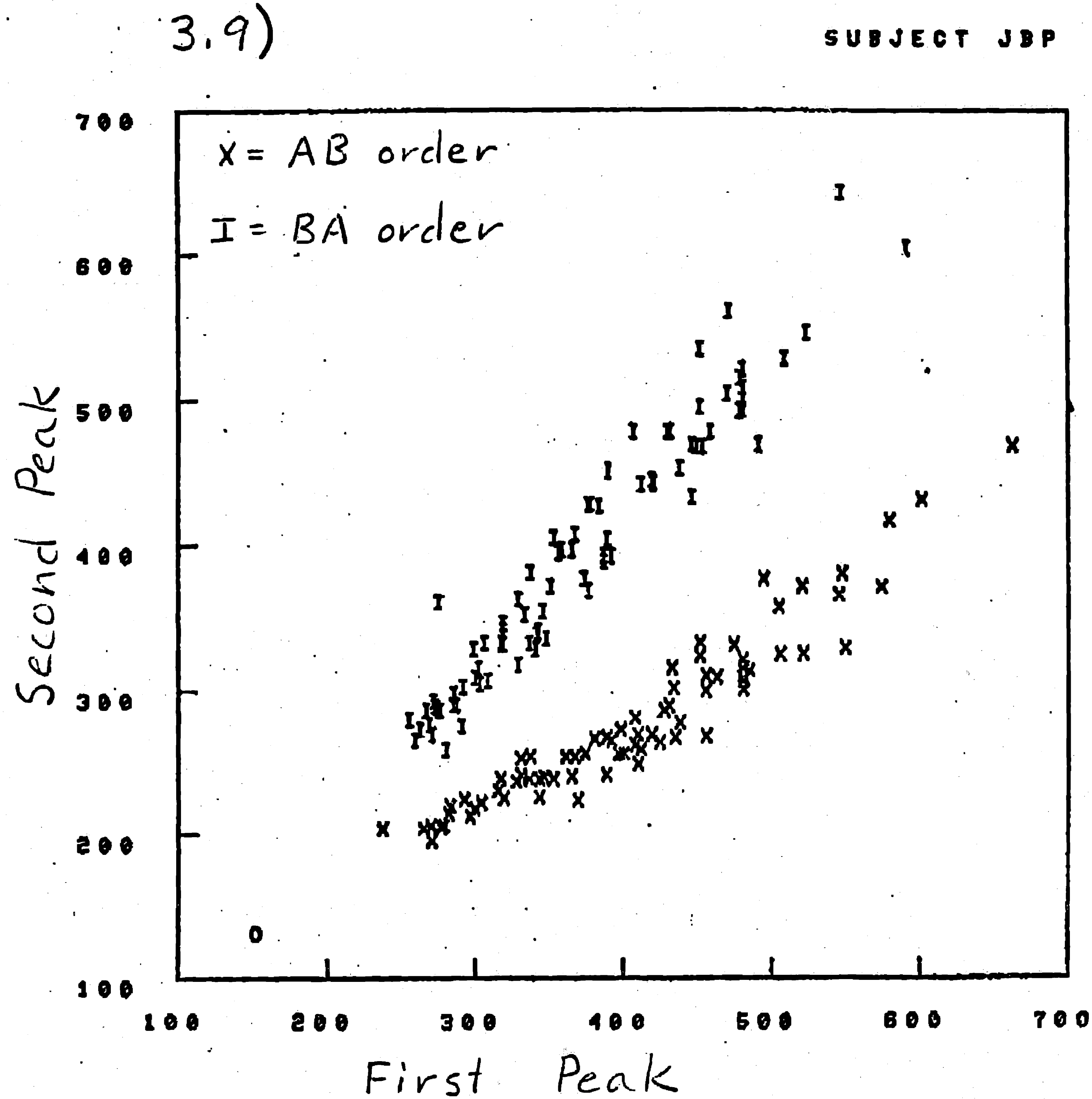


AB'

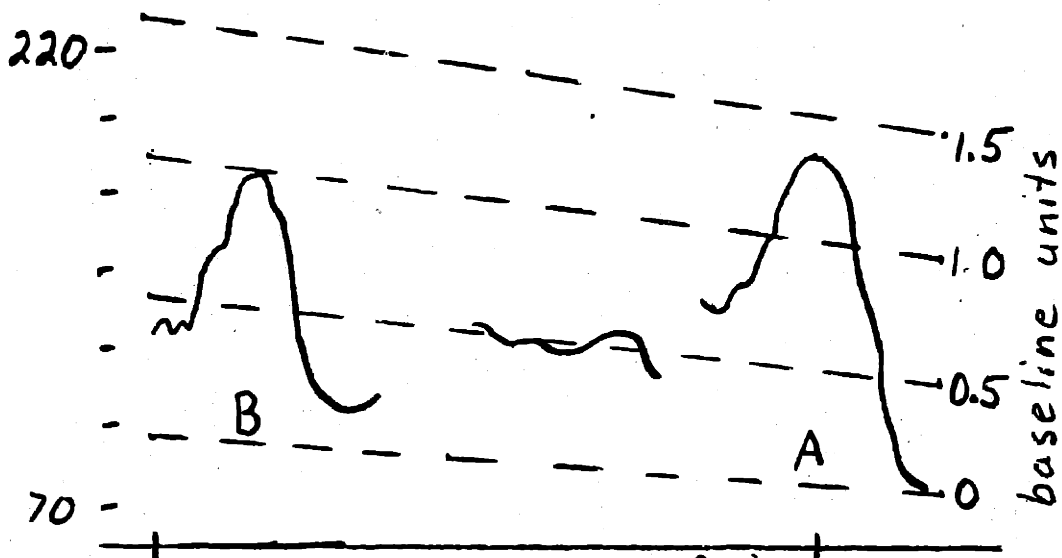
3, 8


SUBJECT JBP
700
600
500
400
300
200
100
L⁻ in B
100
200
300
400
500
600
700
H* in B, BA order
BA'

3.9)


SUBJECT JBP
X = AB order
I = BA order
Second Peak
First Peak
100
200
300
400
500
600
700
0

3.10 A) M4L

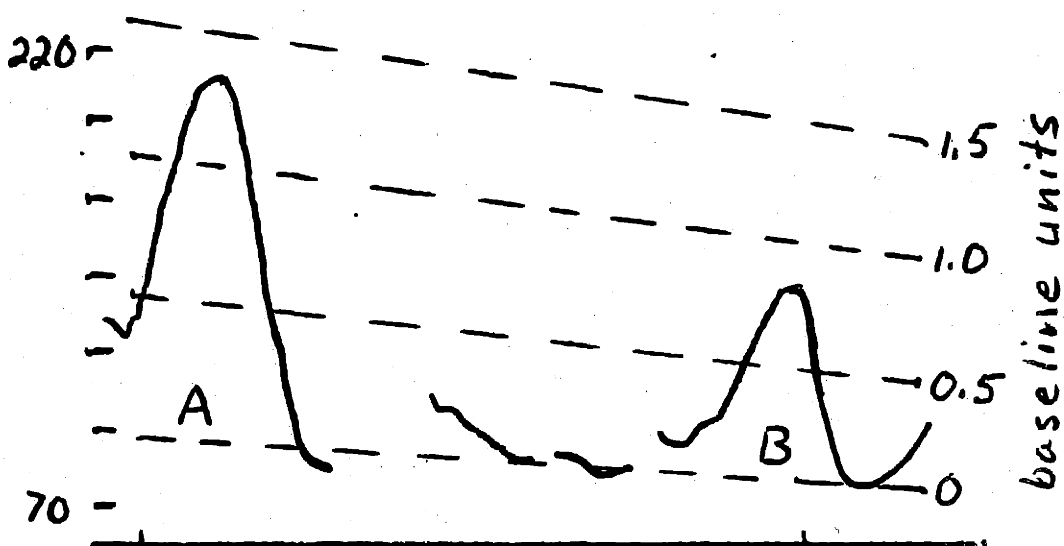


B = 180 Hz ⇒ 1 baseline unit

Â = 185 Hz ⇒ 1.45 baseline units

$\hat{A} = 1.45\ \hat{B}$

B)

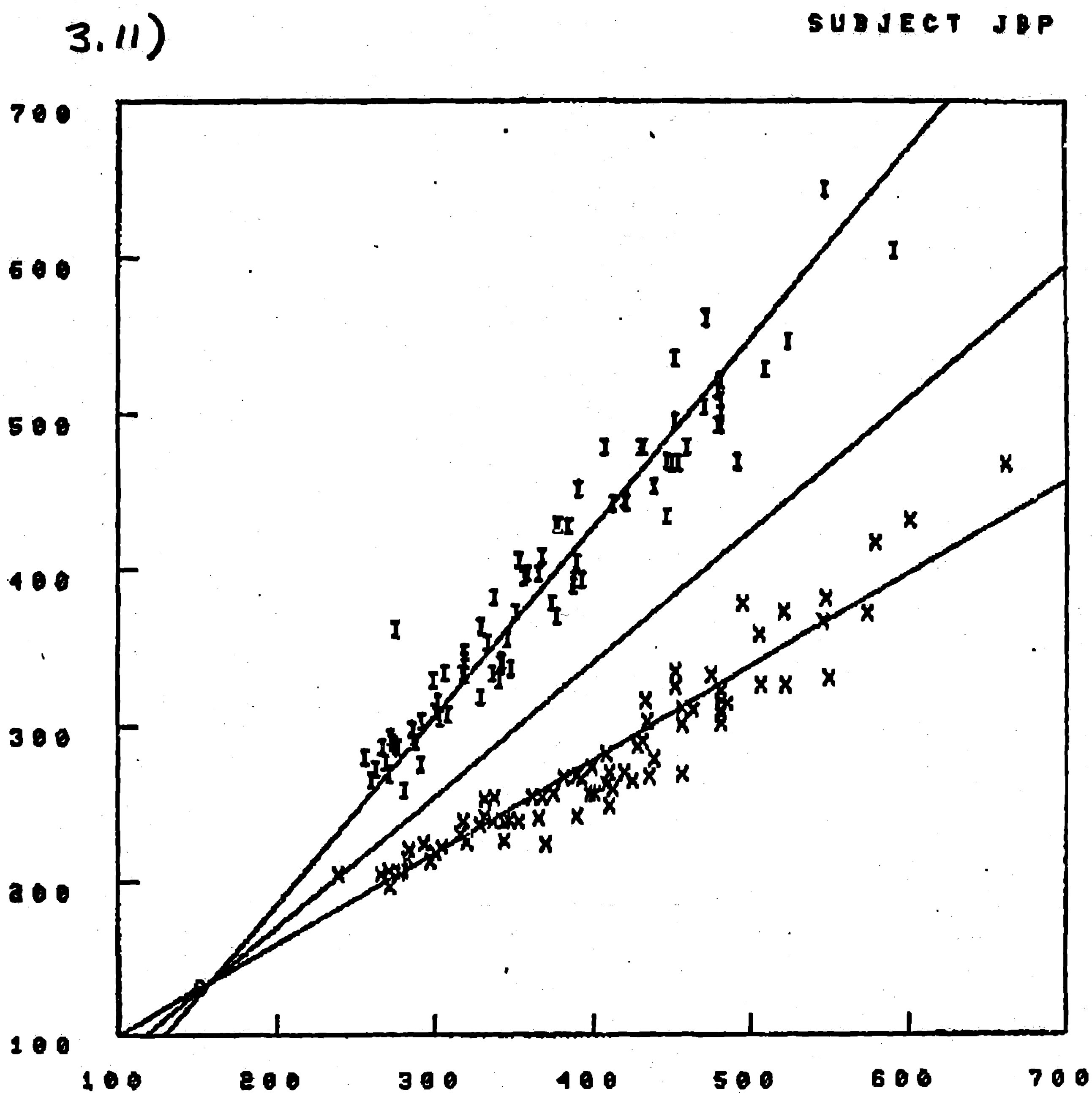


A = 210 Hz ⇒ 1.33 baseline units

B = 141 Hz ⇒ 0.87 baseline units

$\hat{A} = 1.53\ \hat{B}$

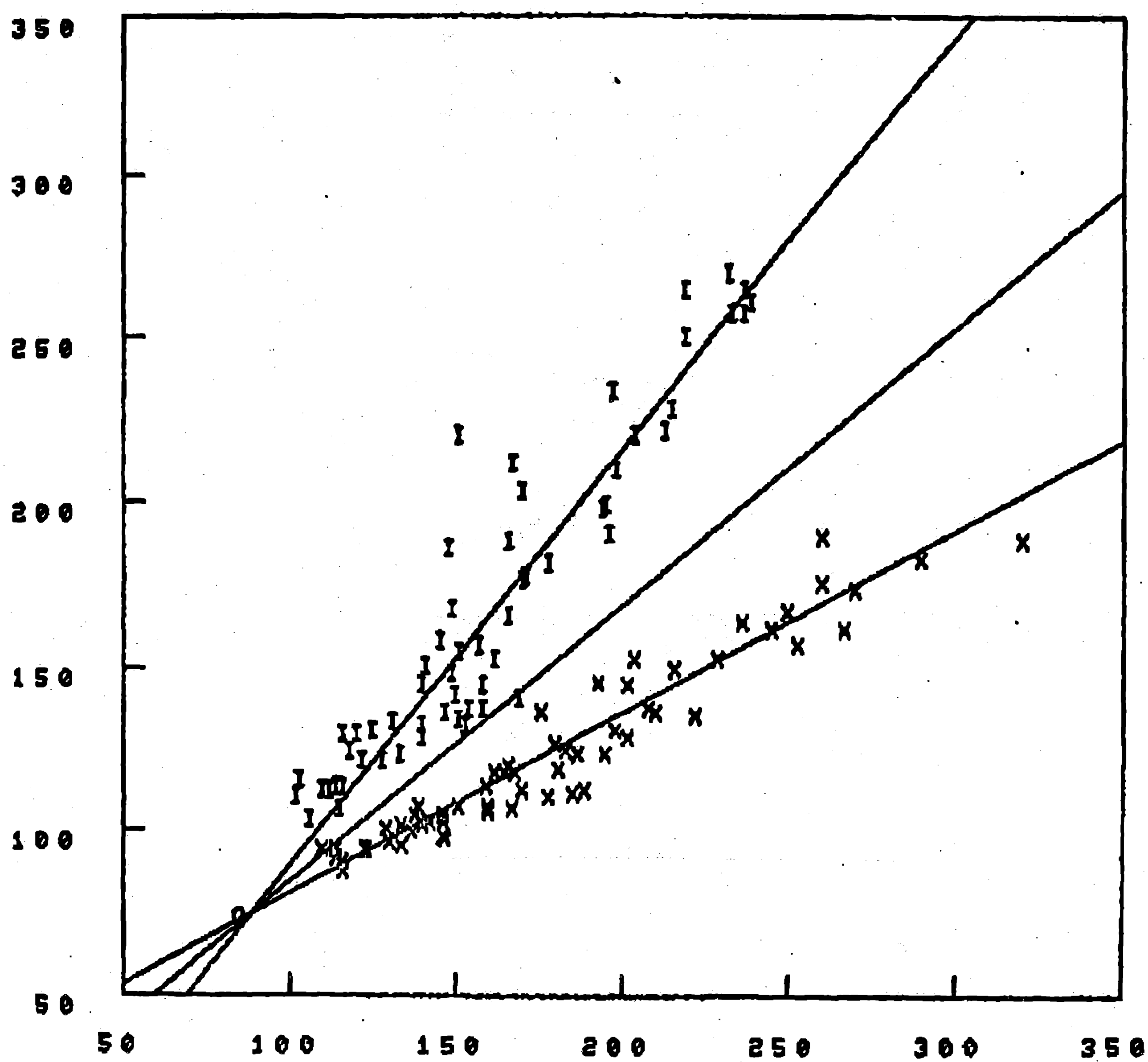
3.11)



SUBJECT JBP
700
600
500
400
300
200
100
100
200
300
400
500
600
700

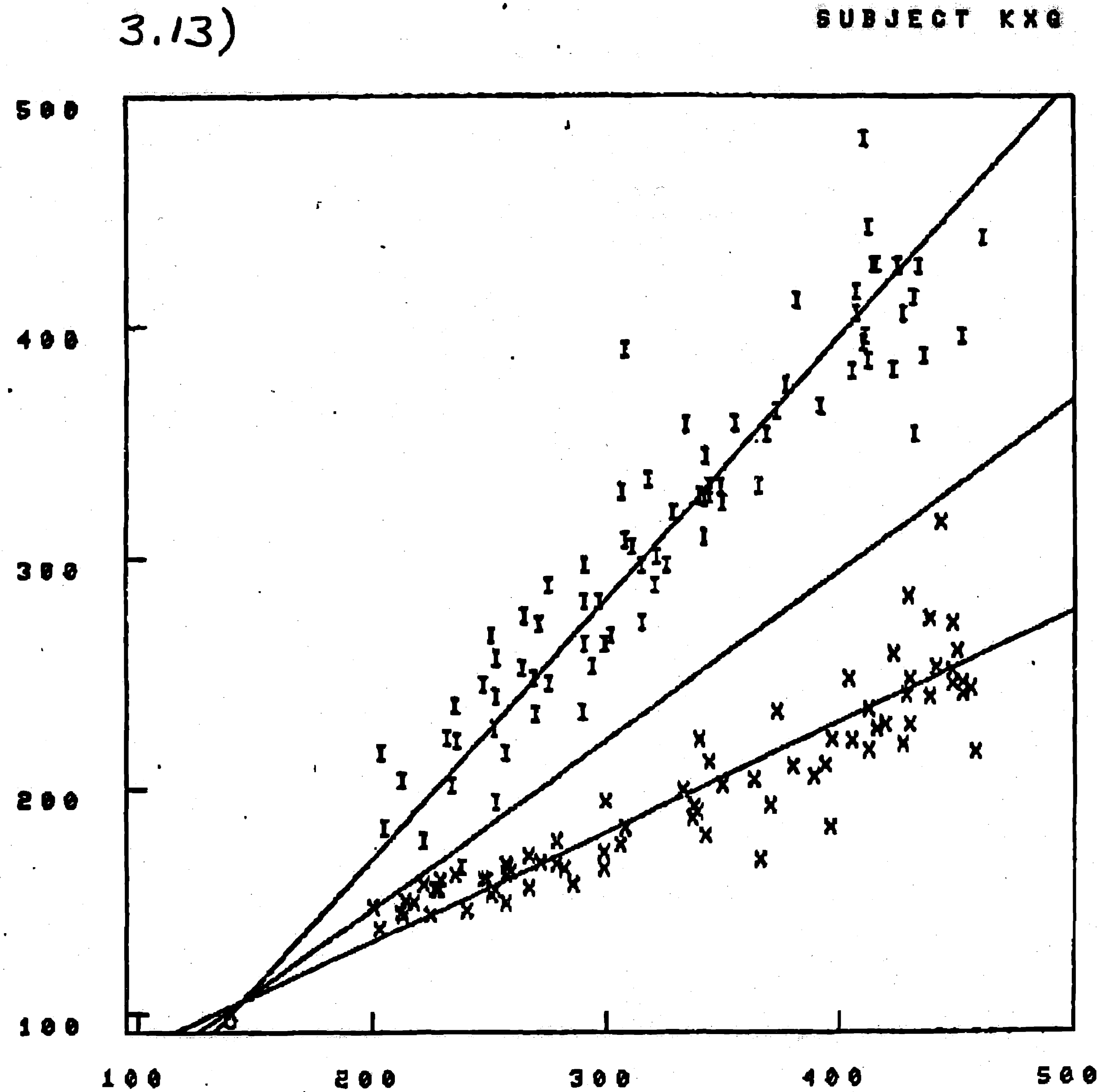
3,12)

SUBJECT MVL



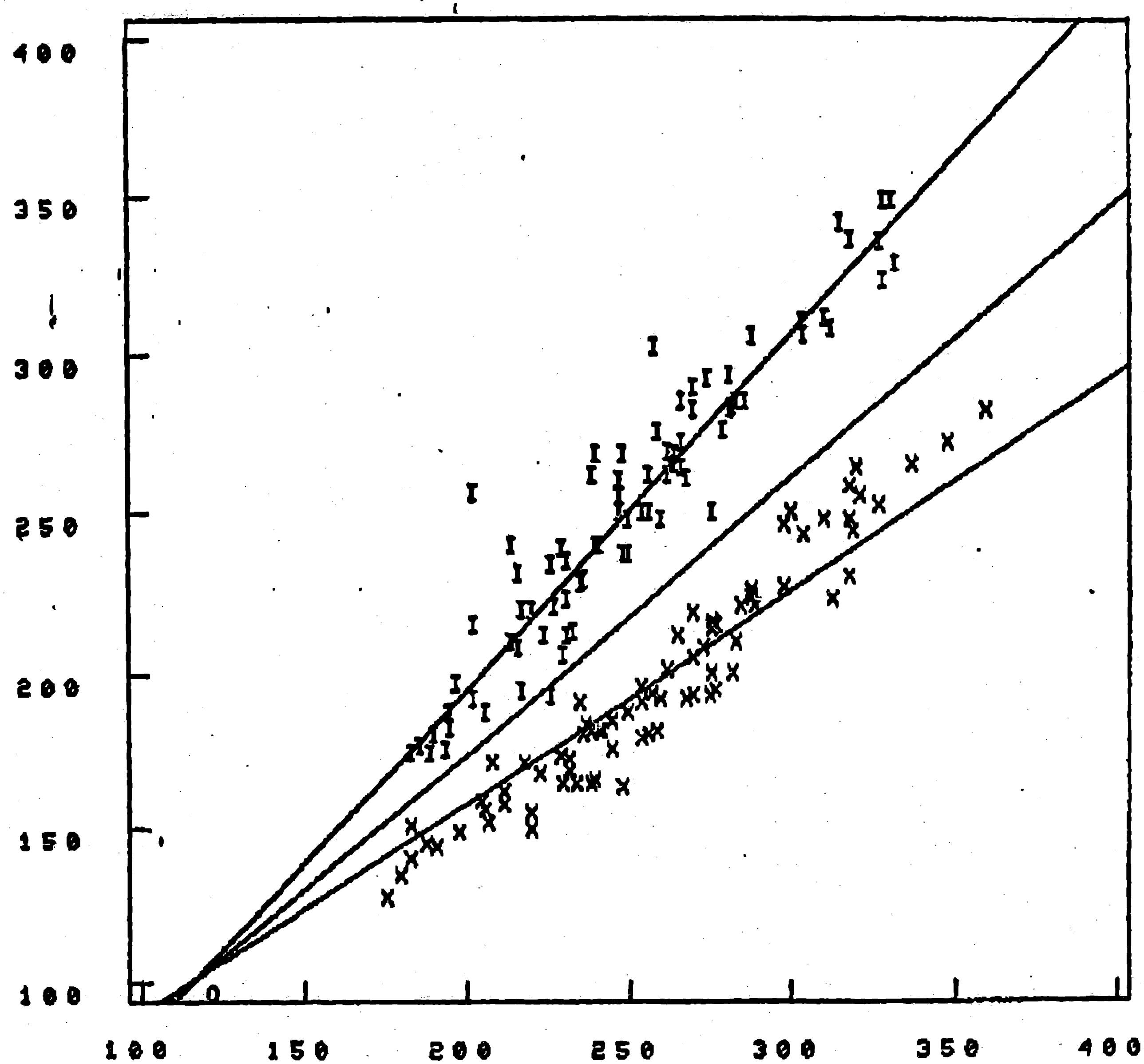
324

3.13)

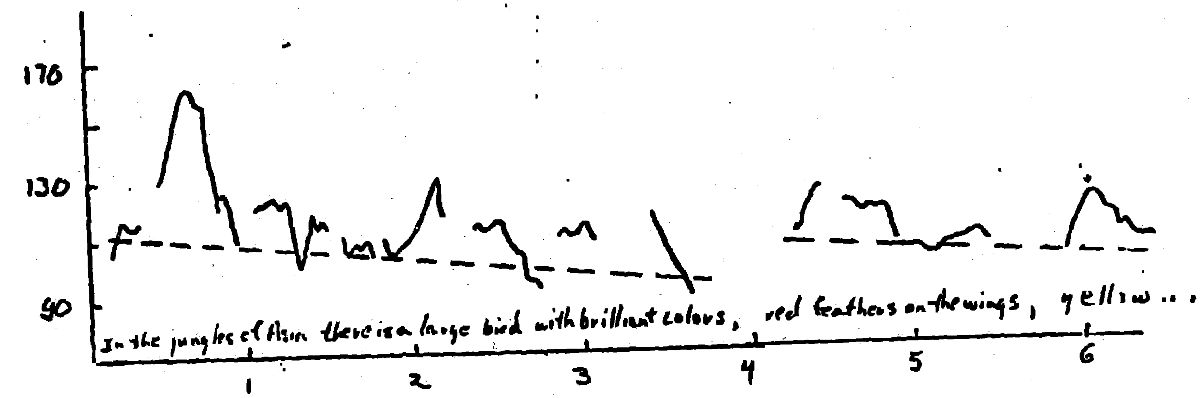


3.14)

SUBJECT DUS

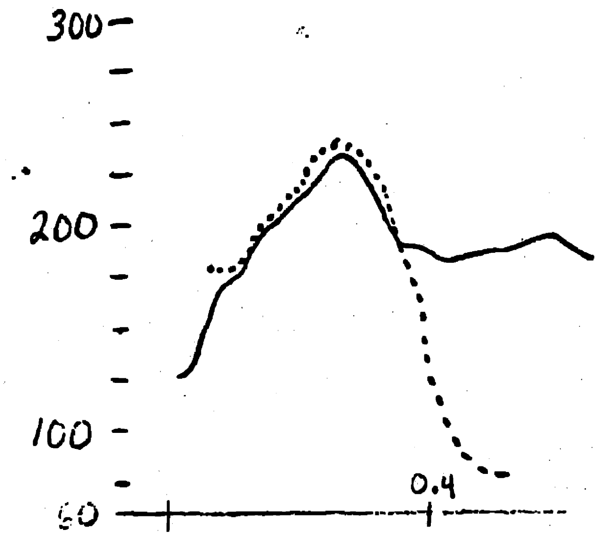


3.15) KN (from Maeda 1976)



An illustration of the baseline as fit by Maeda.

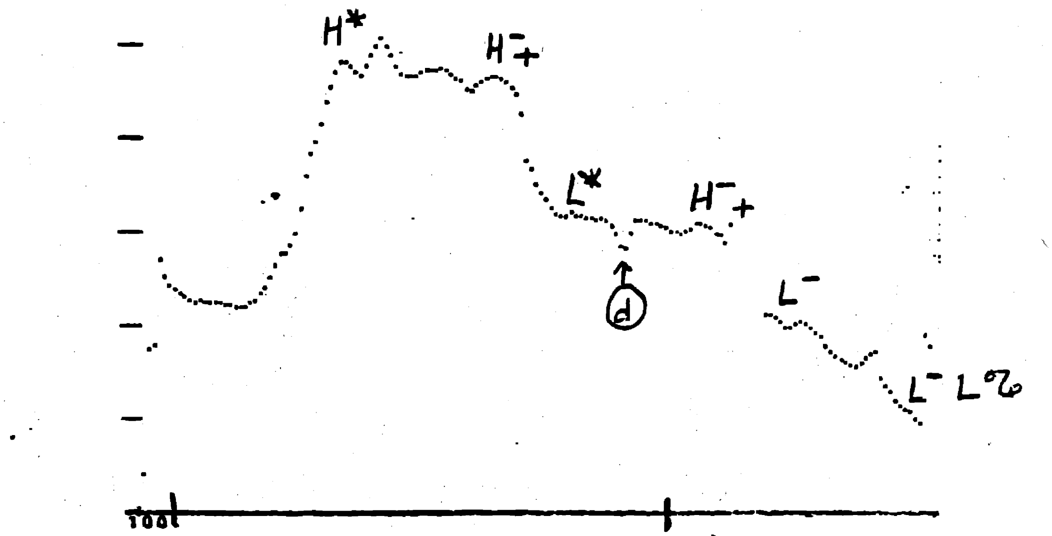
3.16) M4L



······ Manny
H* L^- L%

—— Manny
H*+L^- H^- L%

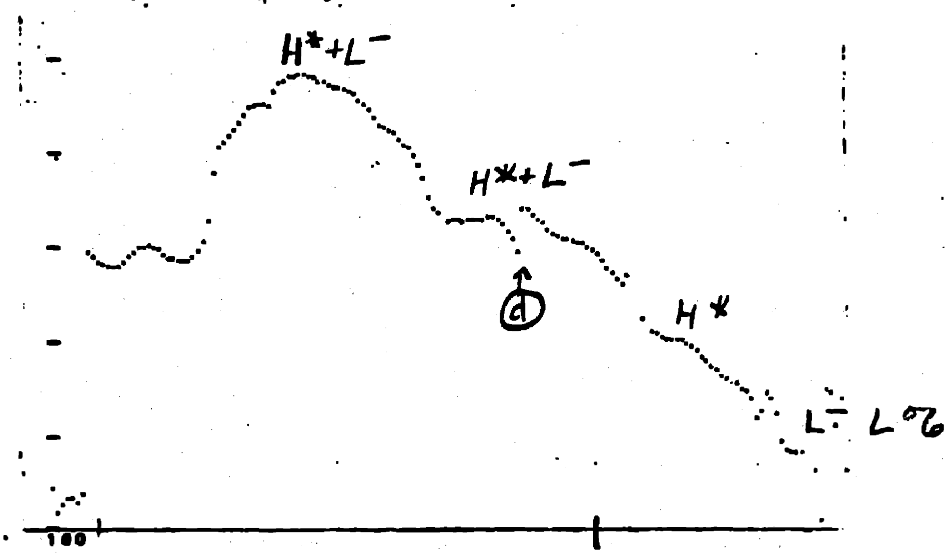
4.1) JBP



There are many intermediate levels

H^* $H^- + L^*$ $H^- + L^*$ L^- $L\%$

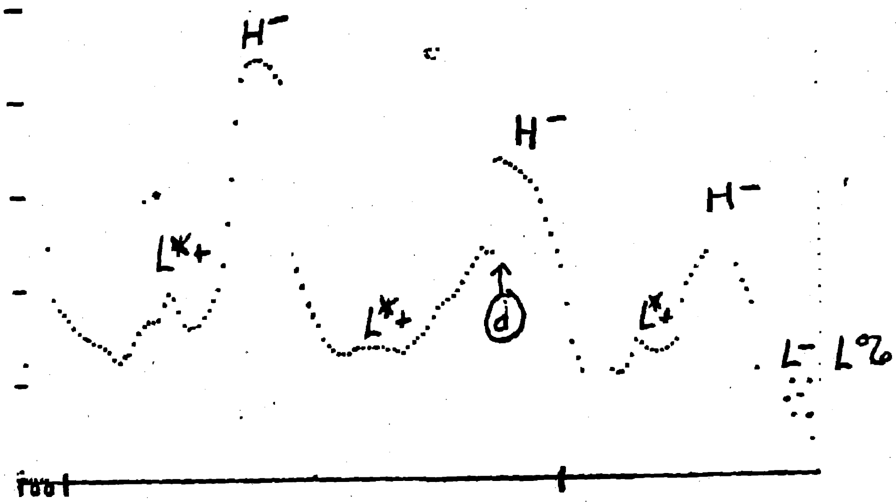
4.2) JBP



There are many intermediate levels.

$H^* + L^-$ $H^* + L^-$ H^* L^- $L\%$

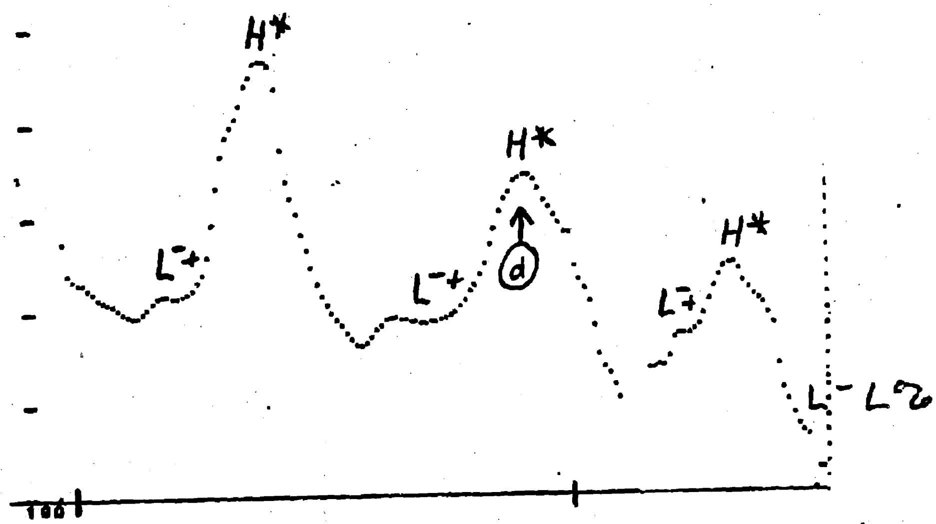
4.3) JBP



There are many intermediate levels.

L^*+H^- L^*+H^- $L^*+H^-\ L^-\ L\%$

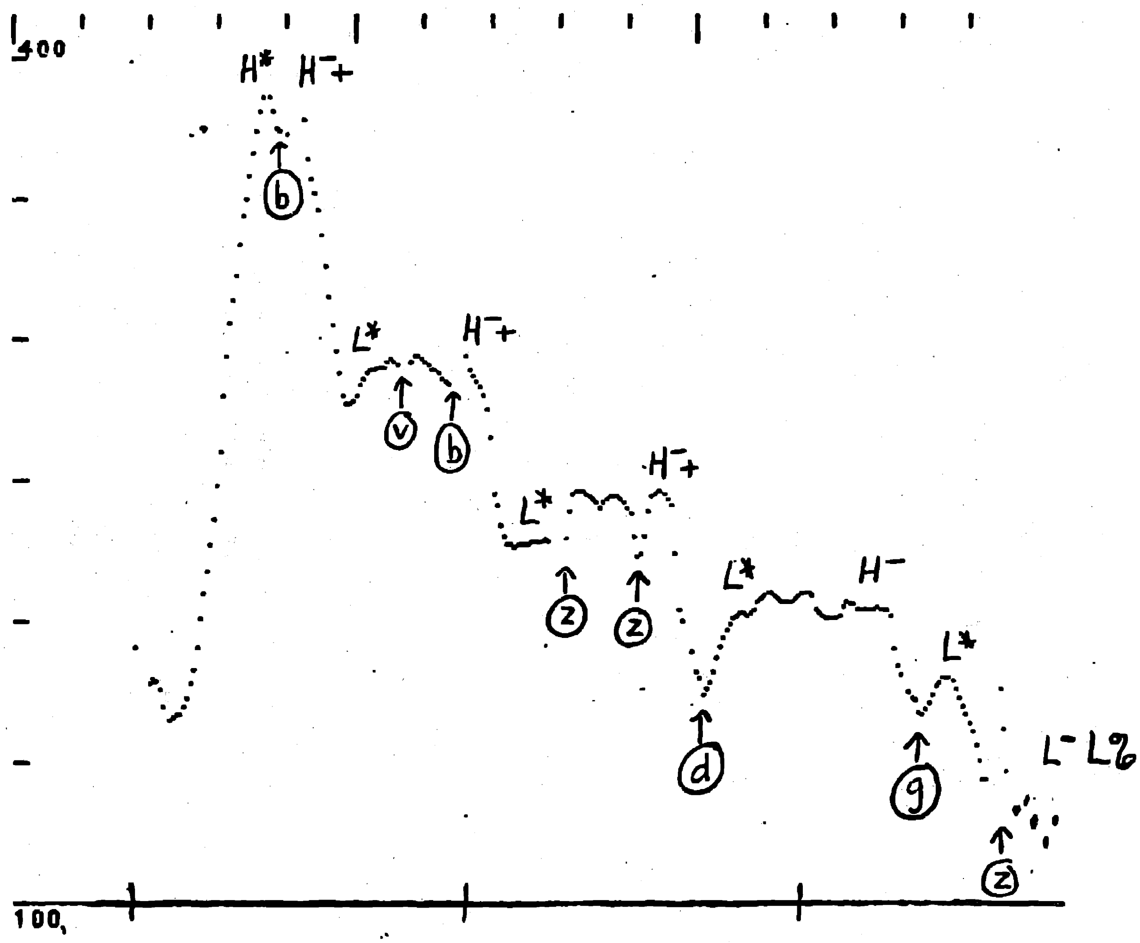
4.4) JBP.



There are many intermediate levels

L^-+H^* L^-+H^* $L^-+H^*\ L^-\ L\%$

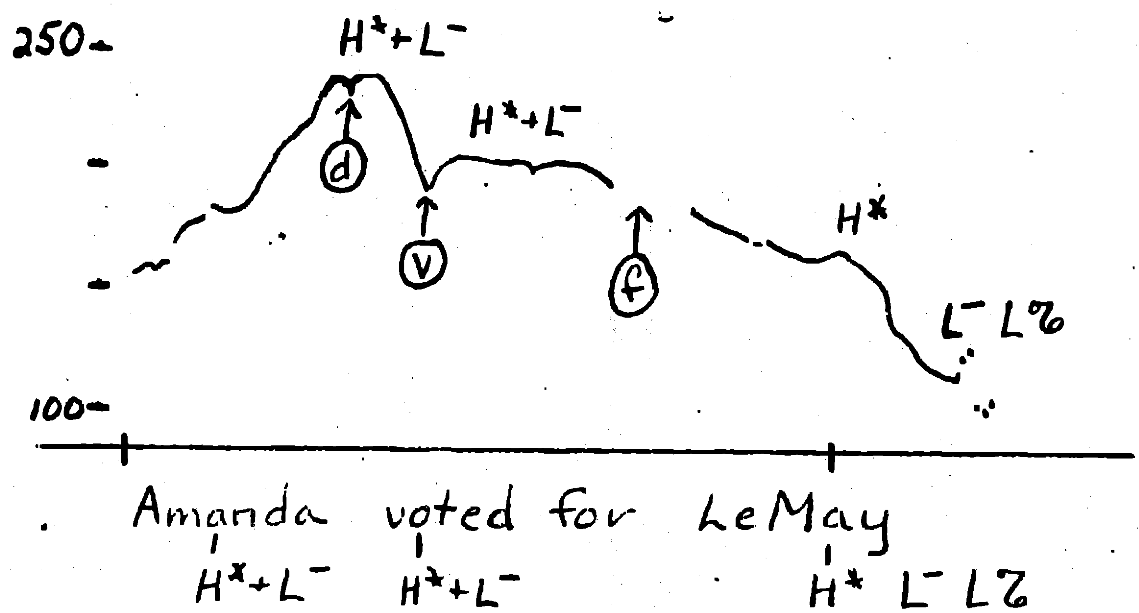
4.5) MB



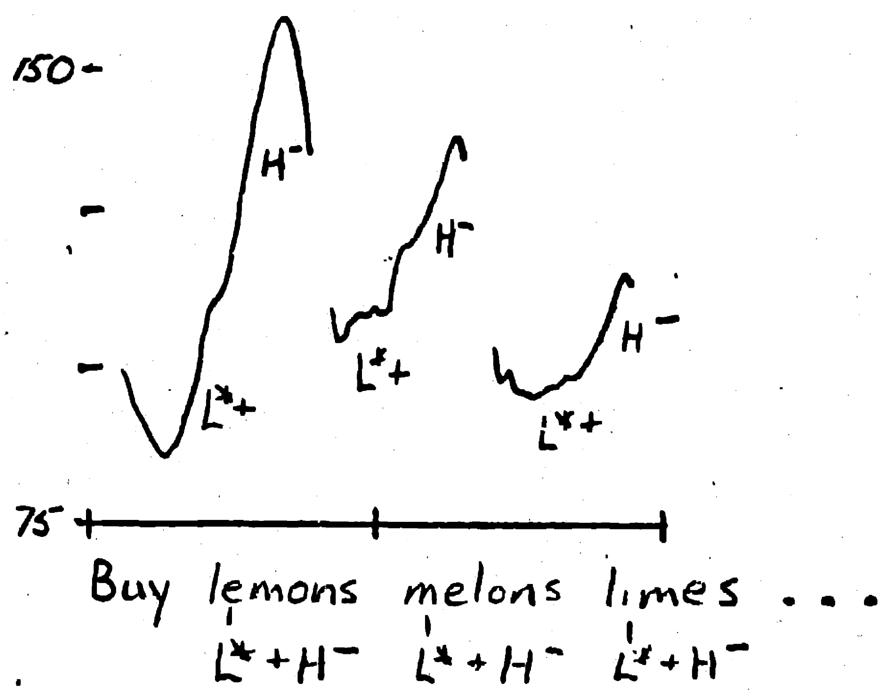
I really believe Ebenezer was a dealer in magnesium.

H*   H-+L*   H-+L*   H-+L*   H-+L*   L- L%

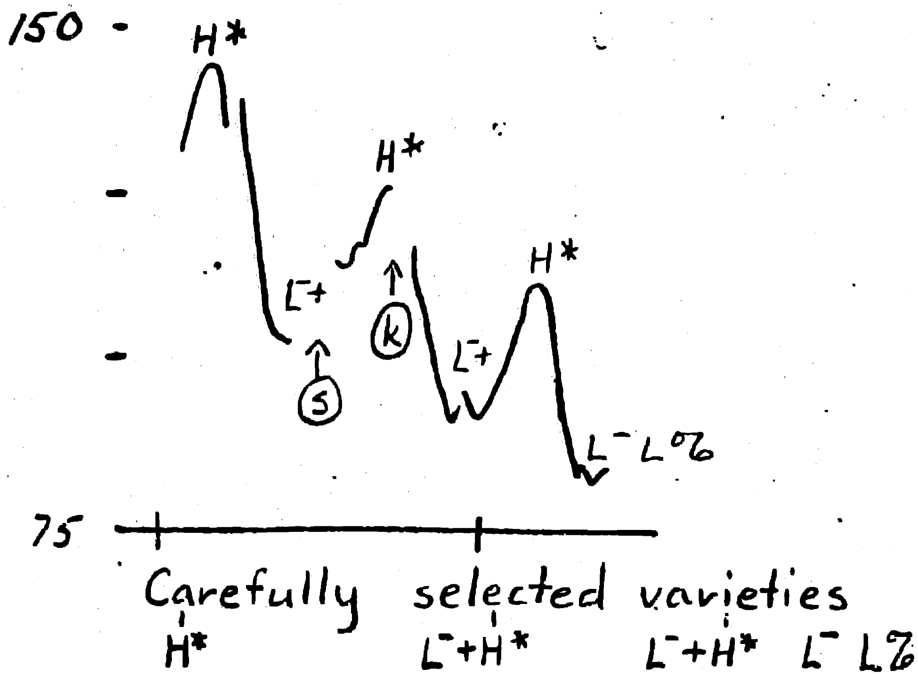
4.6) KXG


250
H*+L⁻
d
H*+L⁻
v
f
H*
L⁻ L%
100
Amanda voted for LeMay
H*+L⁻ H*+L⁻ H* L⁻ L%

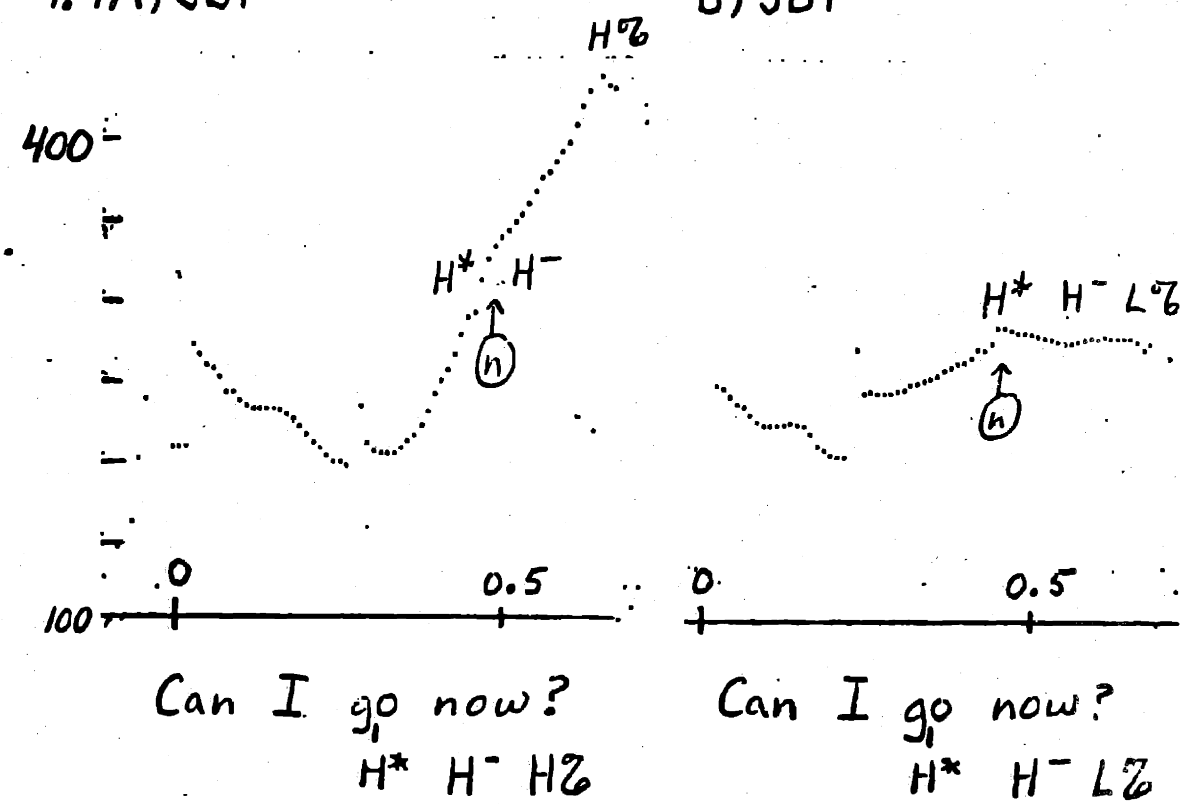
4.7) MYL


150
H⁻
H⁻
H⁻
L*+
L*+
L*+
75
Buy lemons melons limes . . .
L*+H⁻ L*+H⁻ L*+H⁻

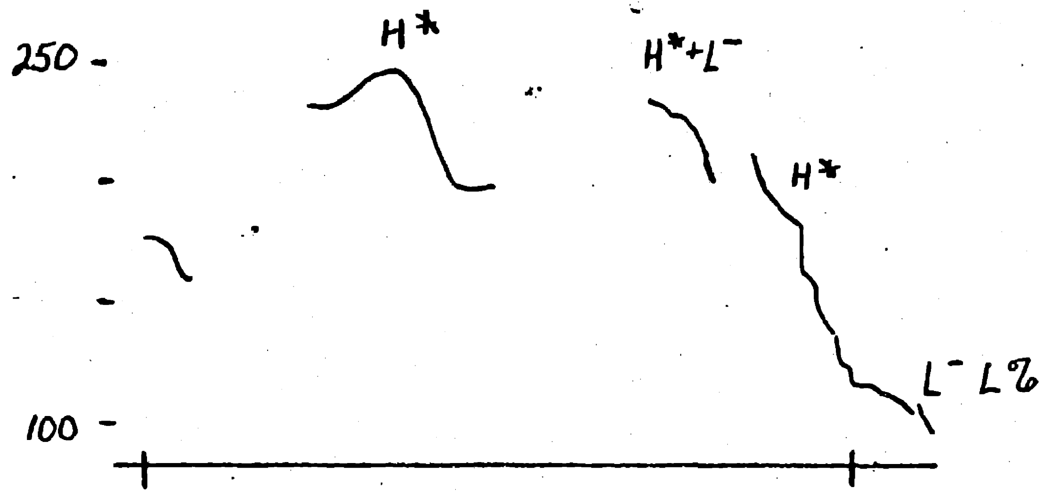
4.8) M4L



4.9A) JBP B) JBP



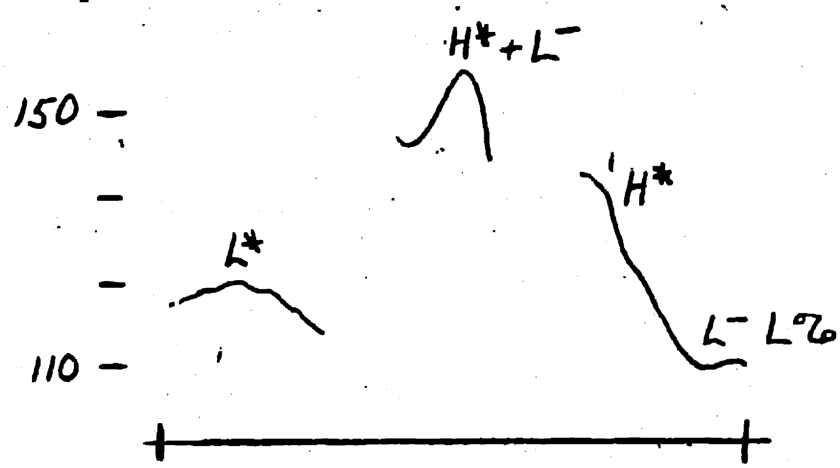
4.10) KXG



It's spelled with two dots
H* H*+L⁻ H* L⁻ L%

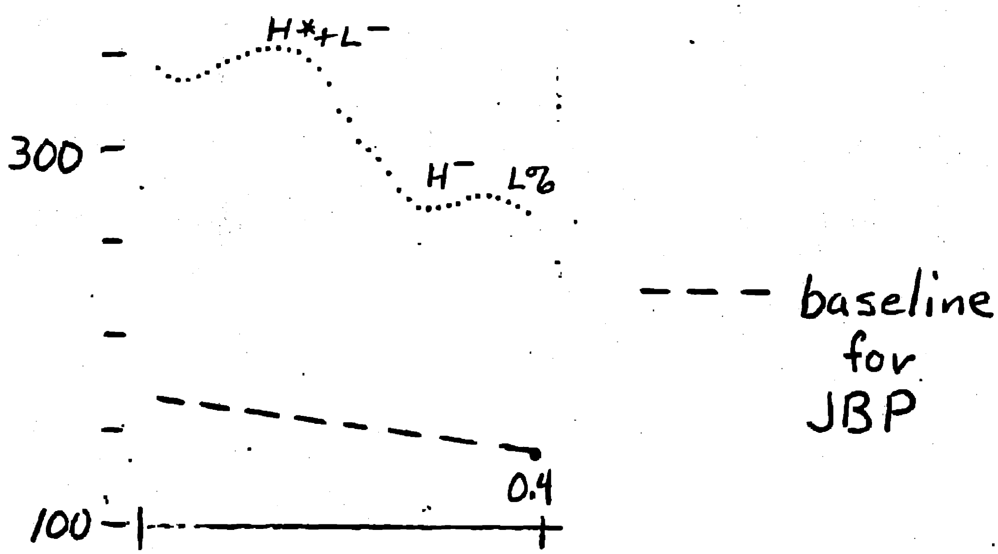
(To the ear, "dots" is clearly not deaccented.)

4.11) KS (taken from Maeda 1976)



Almost all farmers
L* H*+L⁻ H* L⁻ L%

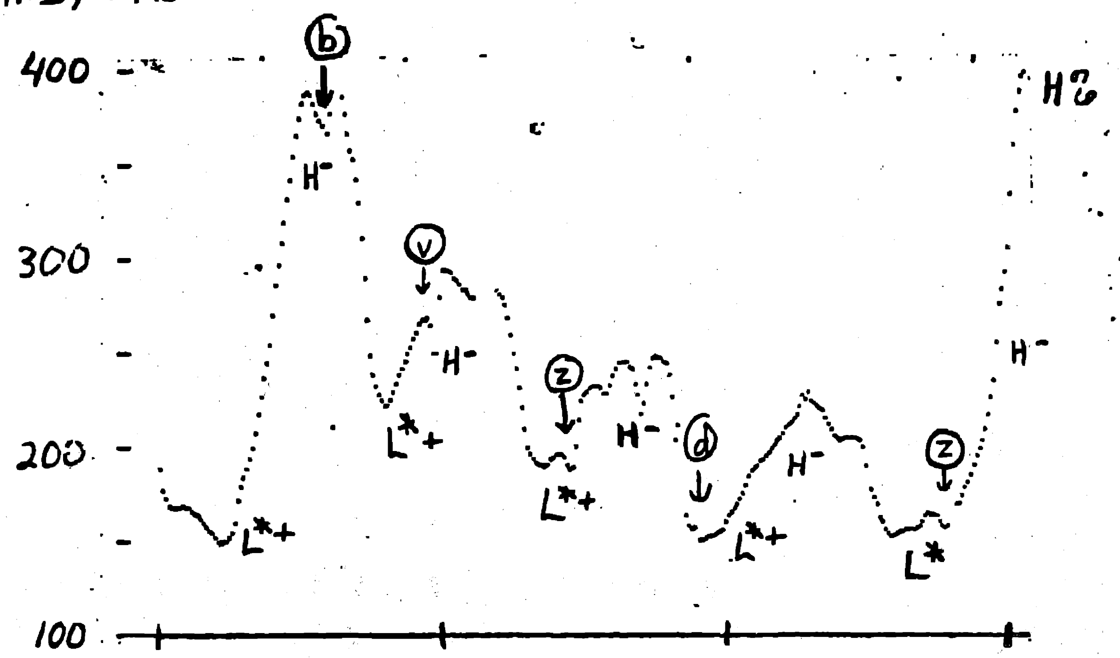
4.12) JBP



Anna!

H*+L⁻ H⁻ L%

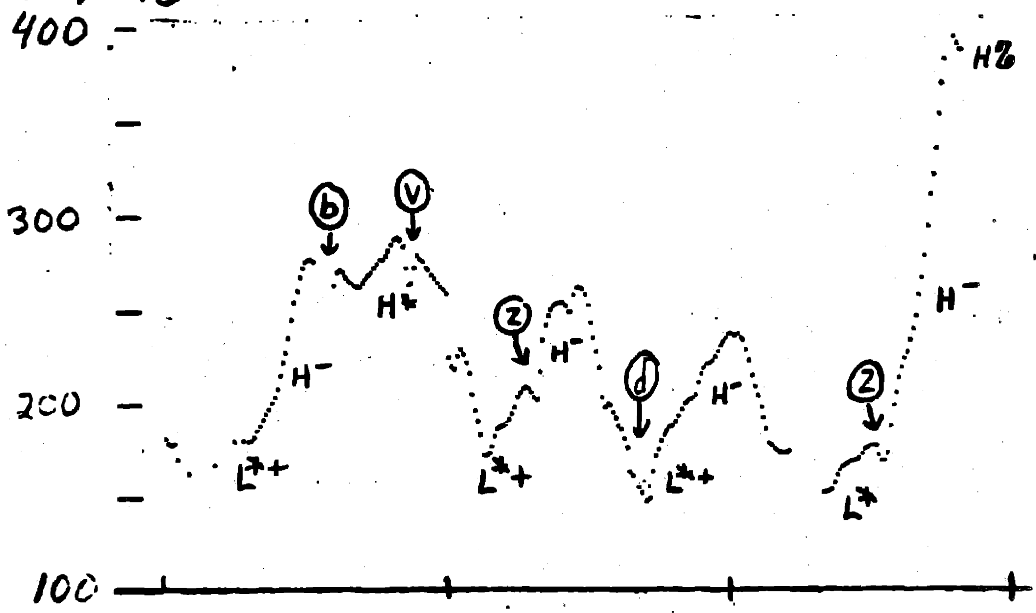
4.13) MB



Do you really believe Ebenezer was a dealer in magnesium?

L*+H⁻ L*+H⁻ L*+H⁻ L*+H⁻ L* H⁻ H%

4.14) MB

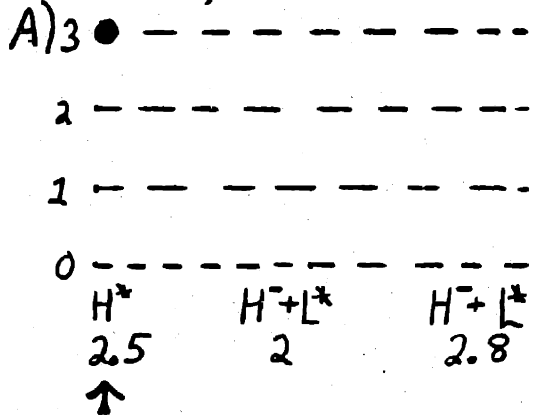


Do you really believe Ebenezer was a dealer in magnesium?

L*+H⁻ H* L*+H⁻ L*+H⁻ L* H⁻ H%

4.15) A step by step derivation for a downstepped sequence. Graph paper is in baseline units above the baseline. Numbers under accents are prominence values. ⟼ represents the current window, with ↑ at the current tone. $k = 0.6$

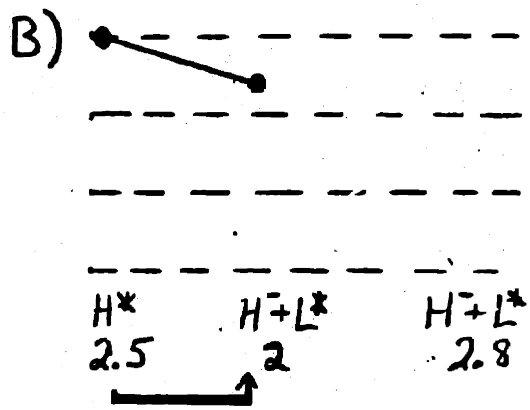
A)



Initialization:

$|H^*| = 3$

B)

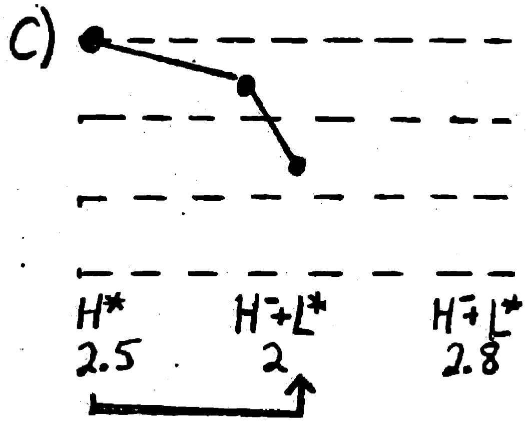


Increment

$$|H^-| = \frac{|H^*| \cdot \text{Prominence}(H^-)}{\text{Prominence}(H^*)} = \frac{3 \cdot 2}{2.5} = 2.4$$

Interpolation

C)

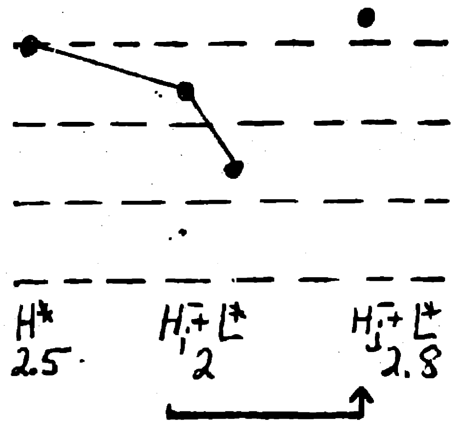


Increment

$$|L^*| = k\,|H^-| = .6 \cdot 2.4 = 1.4$$

Interpolation

D)

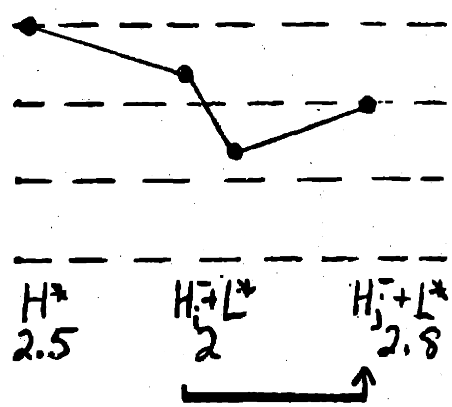


Increment

$$|H_j^-| = \frac{|H_i^-| \cdot Prominence\,(H_j^-)}{Prominence\,(H_i^-)}$$

$$= 2.4 \cdot 2.8 = 3.4$$

E)



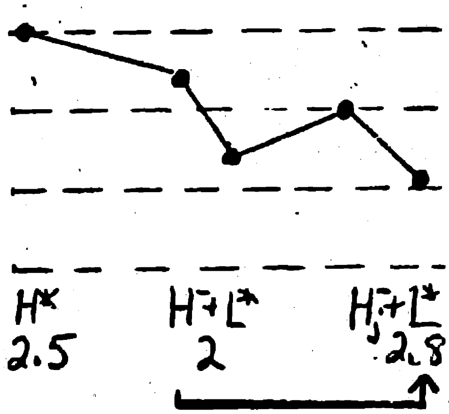
Downstep:

$$|H_j^-| = k\,|H_j^-|$$

$$= 0.6 \cdot 3.4 = 2$$

Interpolation

F)



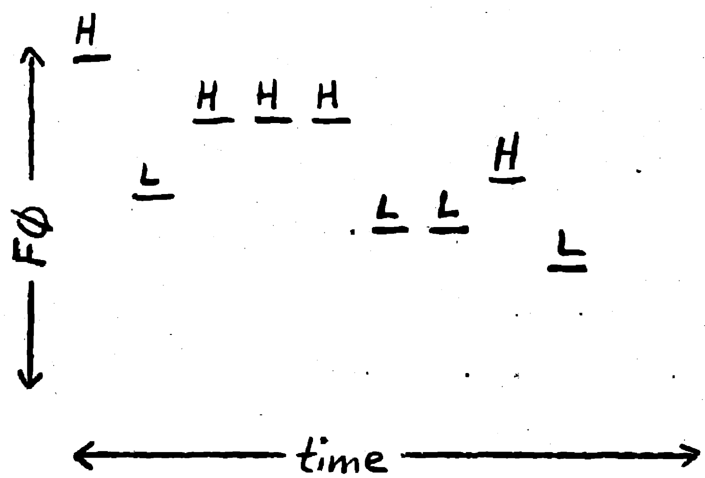
Increment

$$|L^*| = k\,|H_j^-|$$

$$= 0.6 \cdot 2 = 1.2$$

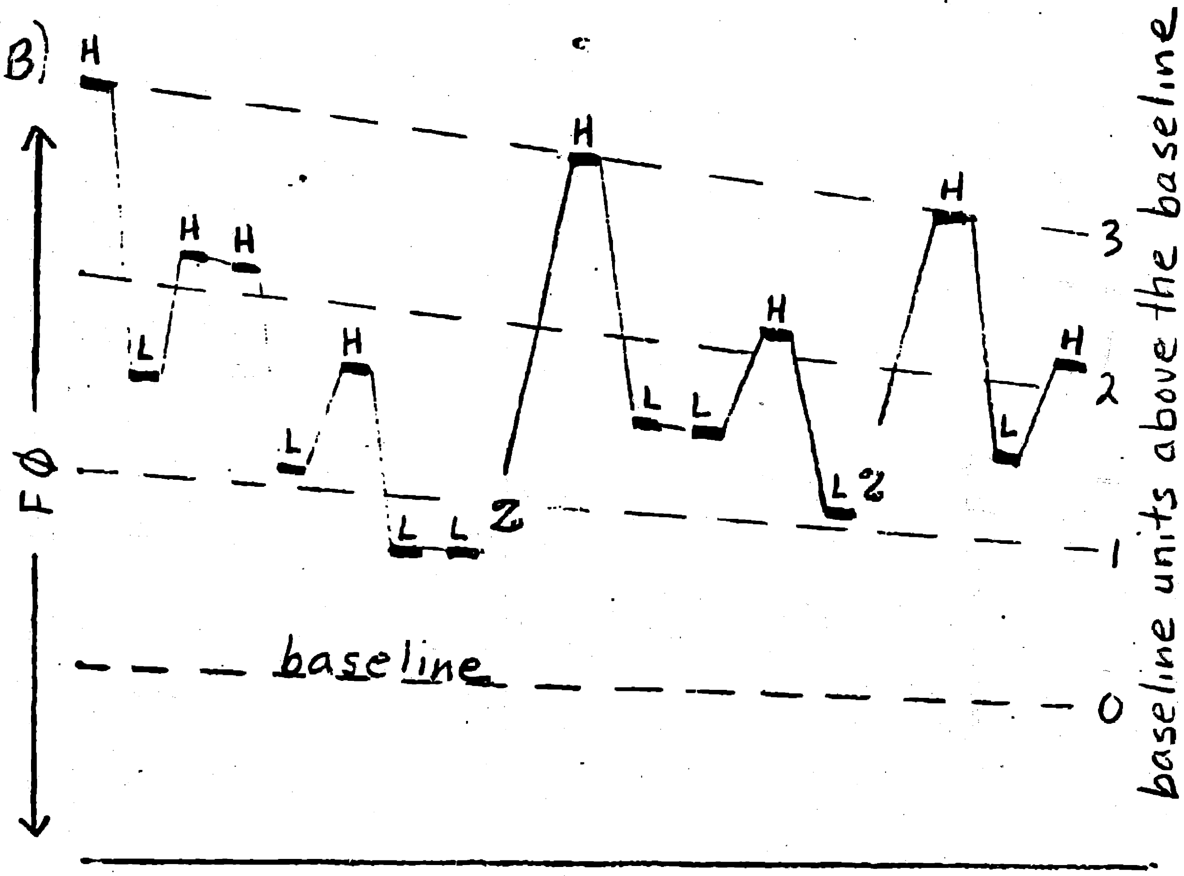
Interpolation

4.16A)



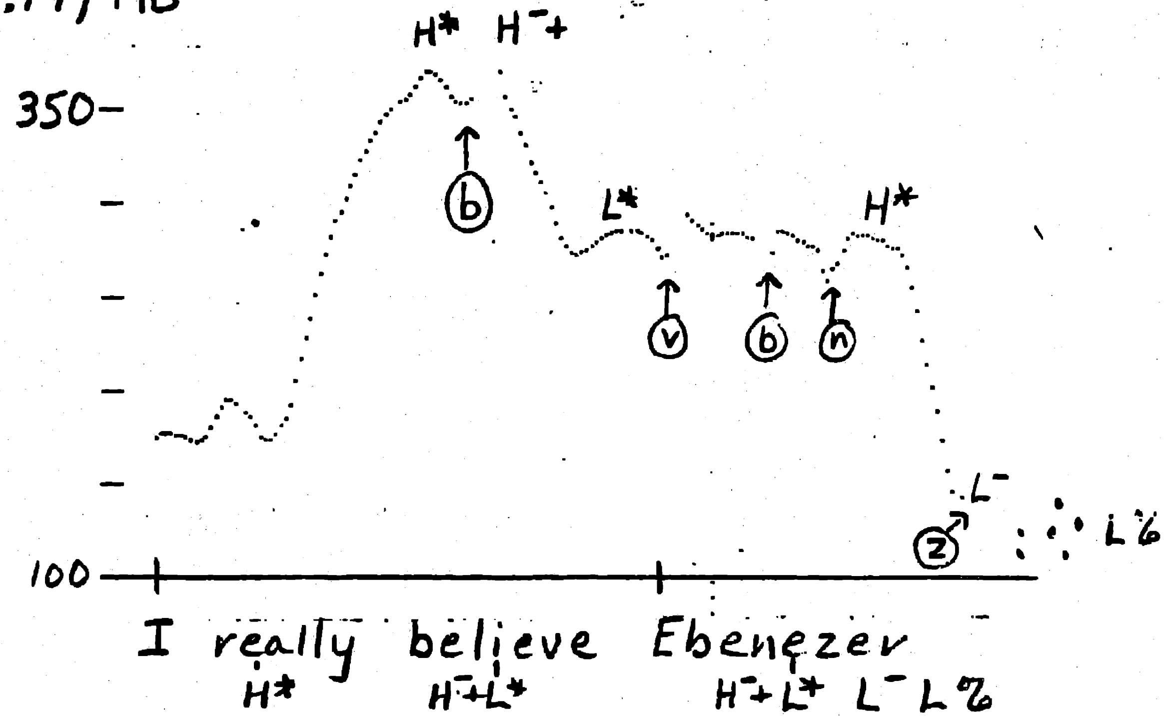
The basic schema for downstep

4.16B)

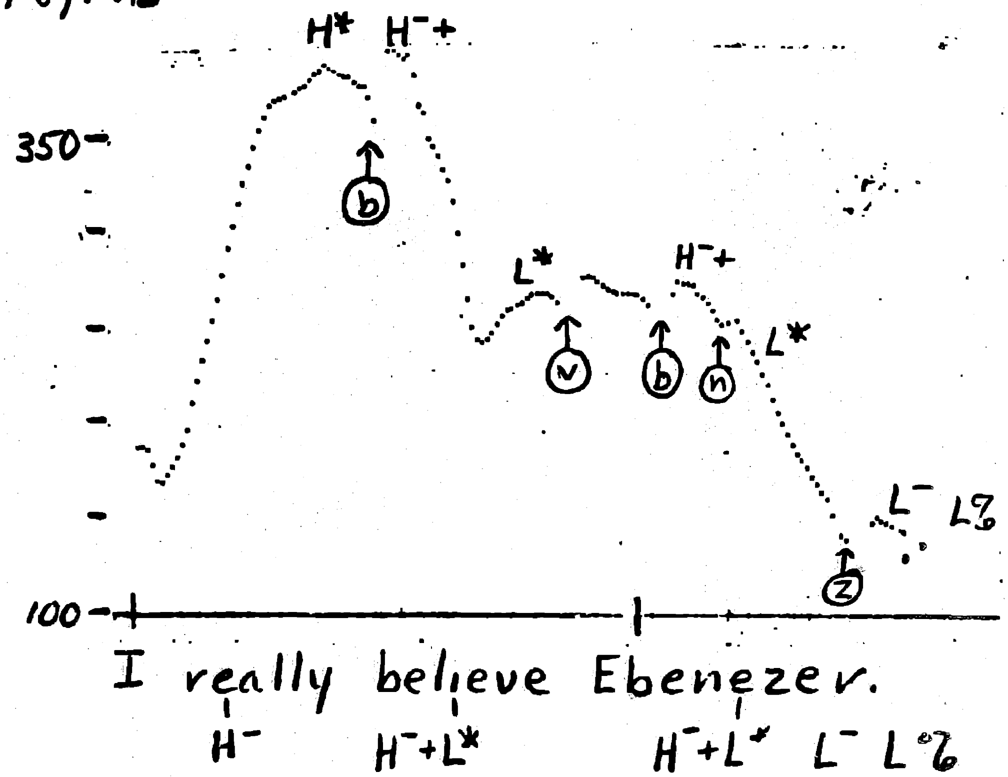


A schematized representation of downstep and declination in Akan. Coefficients relating H and L are arbitrary. Local effects of L% on tonal value are ignored.

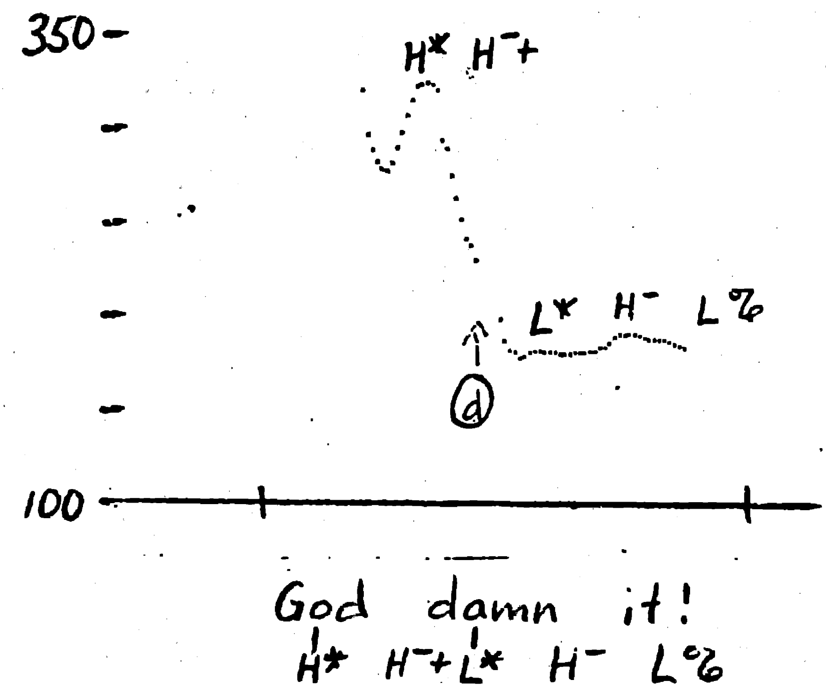
4.17) MB


H* H⁻+
350-
L*
H*
b
v
b
n
L⁻
z
L%
100
I really believe Ebenezer
H*
H⁻+L*
H⁻+L* L⁻ L%

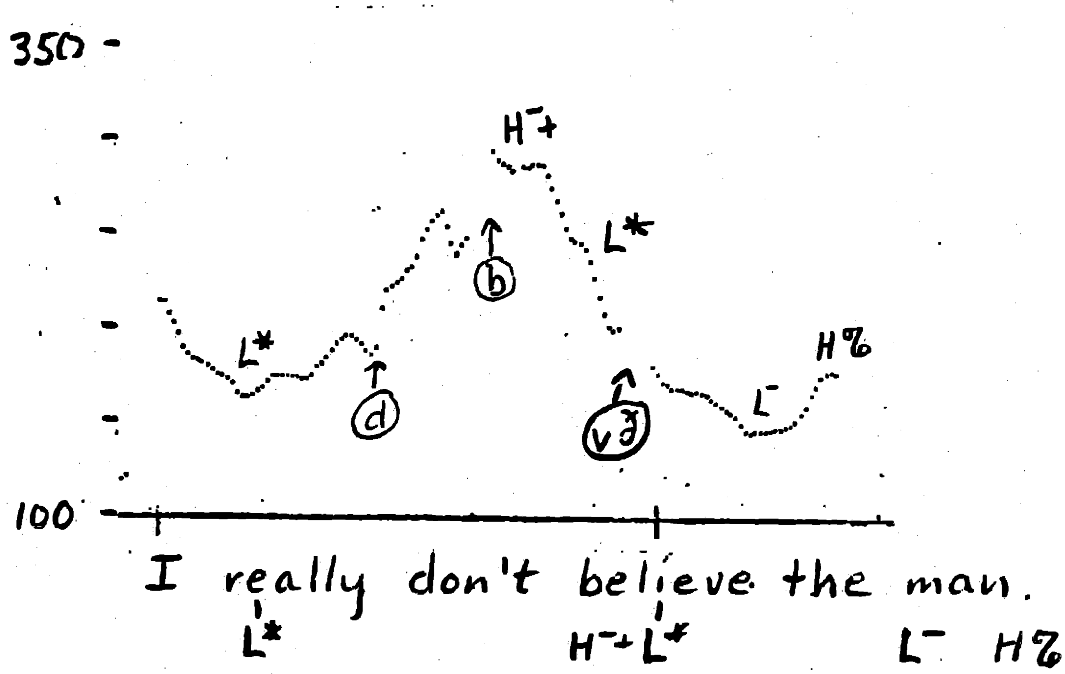
4.18) MB


H* H⁻+
350-
b
L*
H⁻+
L*
v
b
n
L⁻ L%
z
100
I really believe Ebenezer.
H⁻
H⁻+L*
H⁻+L* L⁻ L%

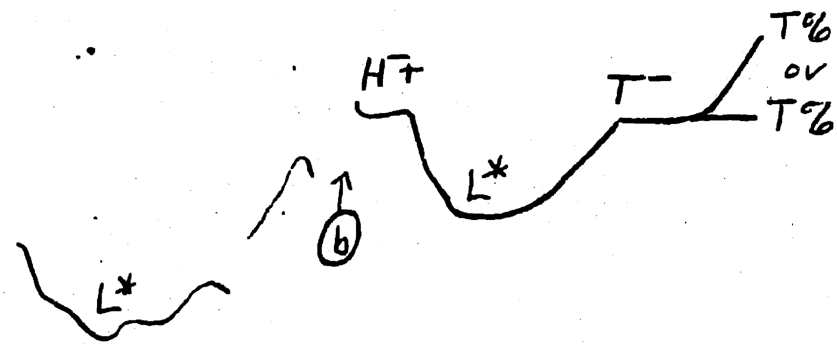
4.19) JBP



4.20) JBP

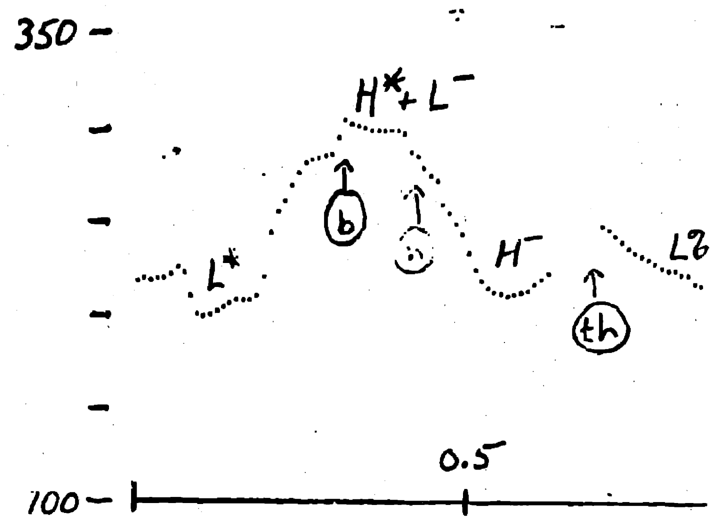


4.21) An impossible intonation:



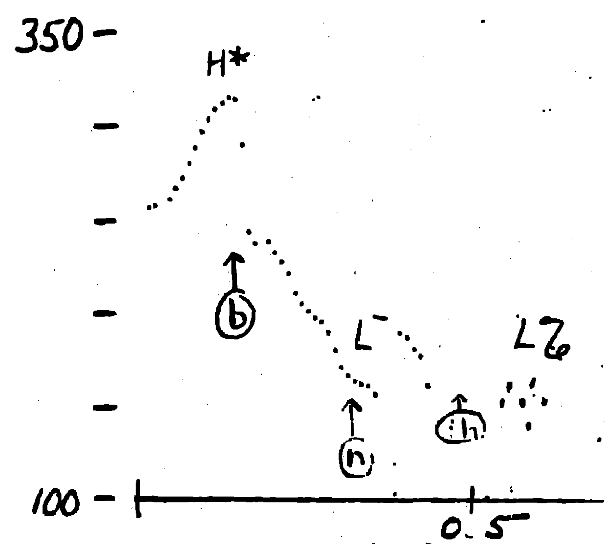
I really don't believe the manager
L* H+L* T^- T%

4.22) JBP



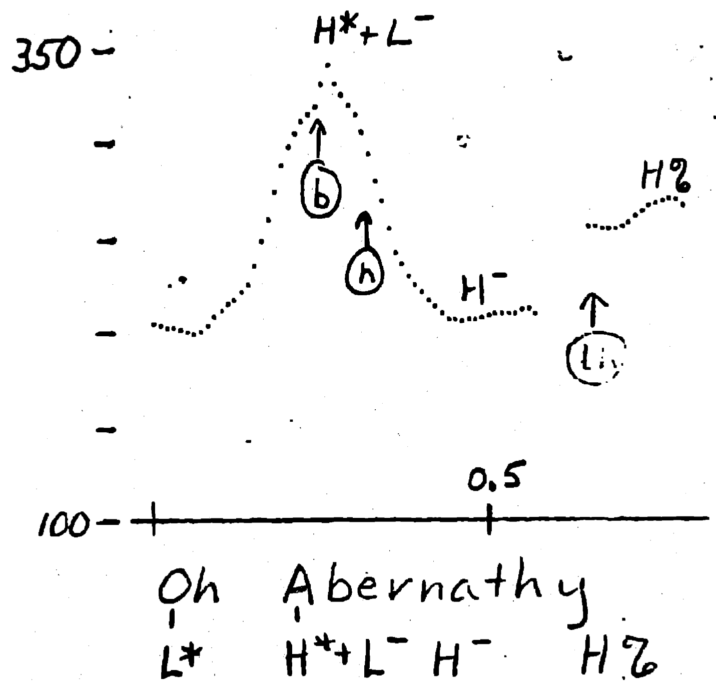
Oh Abernathy
L* H*+L⁻ H⁻ L%

4.23) JBP

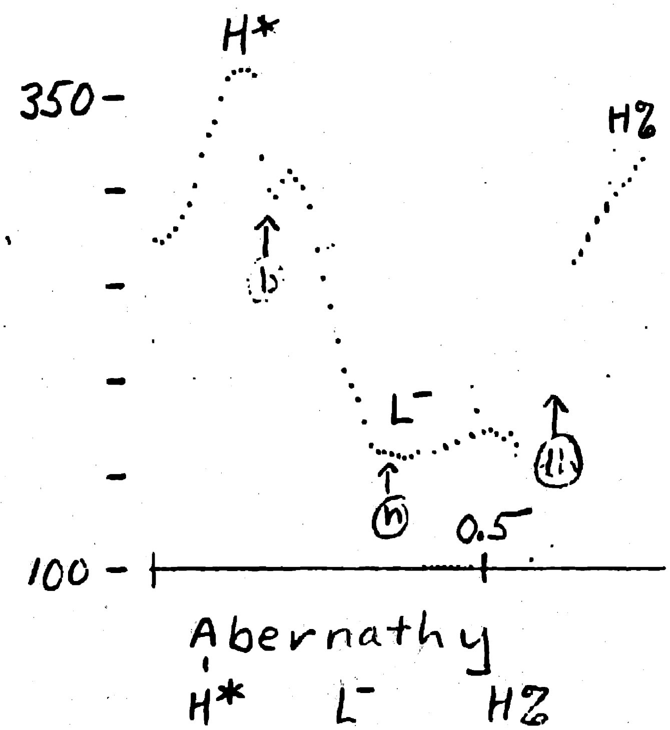


Abernathy
H* L⁻ L%

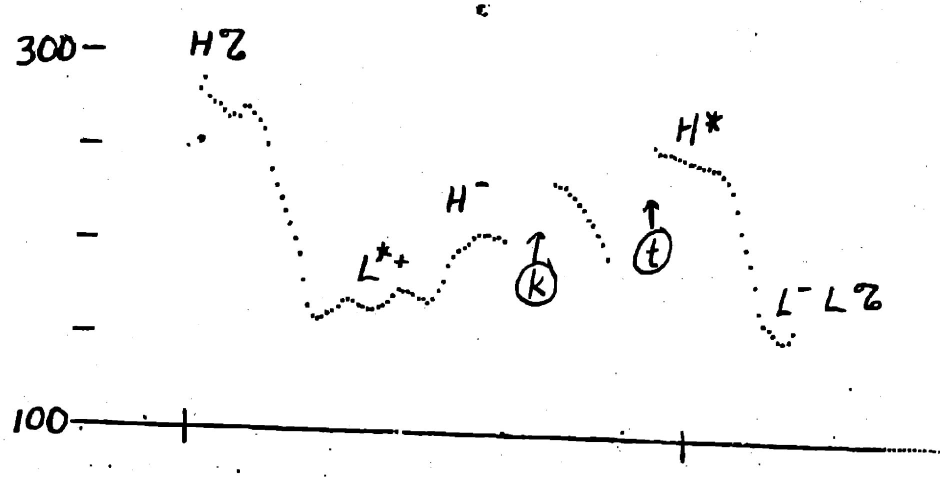
4.24) JBP



4.25) JBP



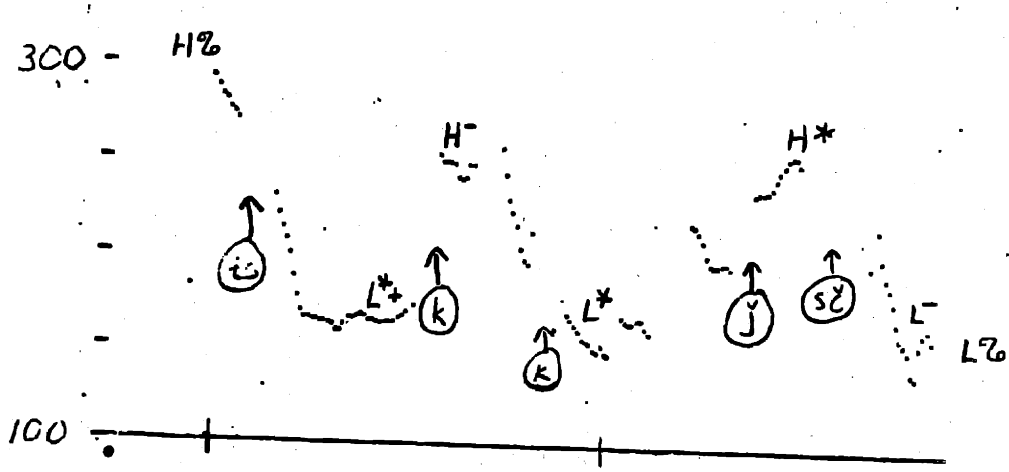
4.26) JBP



Only one in a quintillion!

H% L*+H- H* L- L%

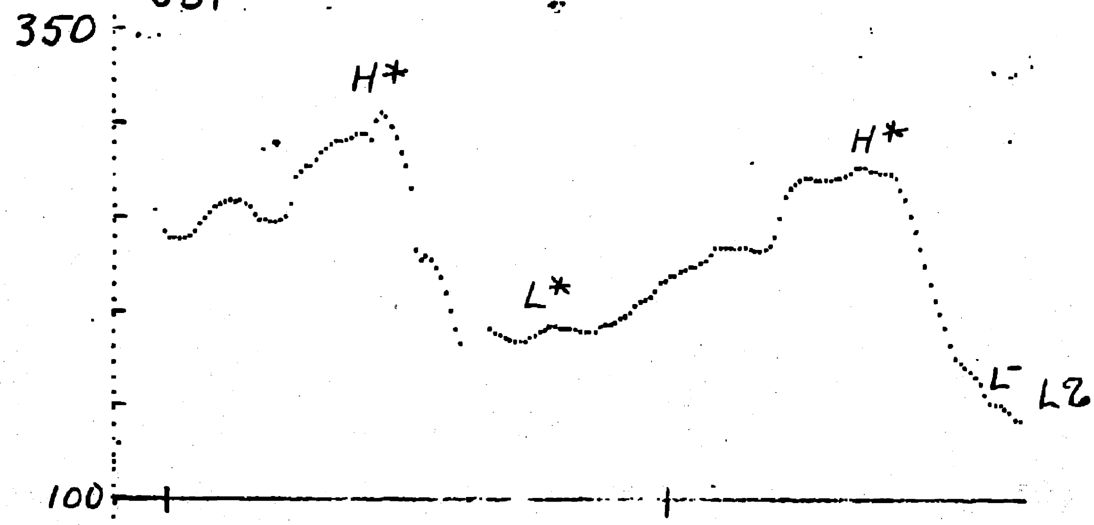
4.27) JBP



That's a remarkably clever suggestion

H% L*+ H- L* H* L- L%

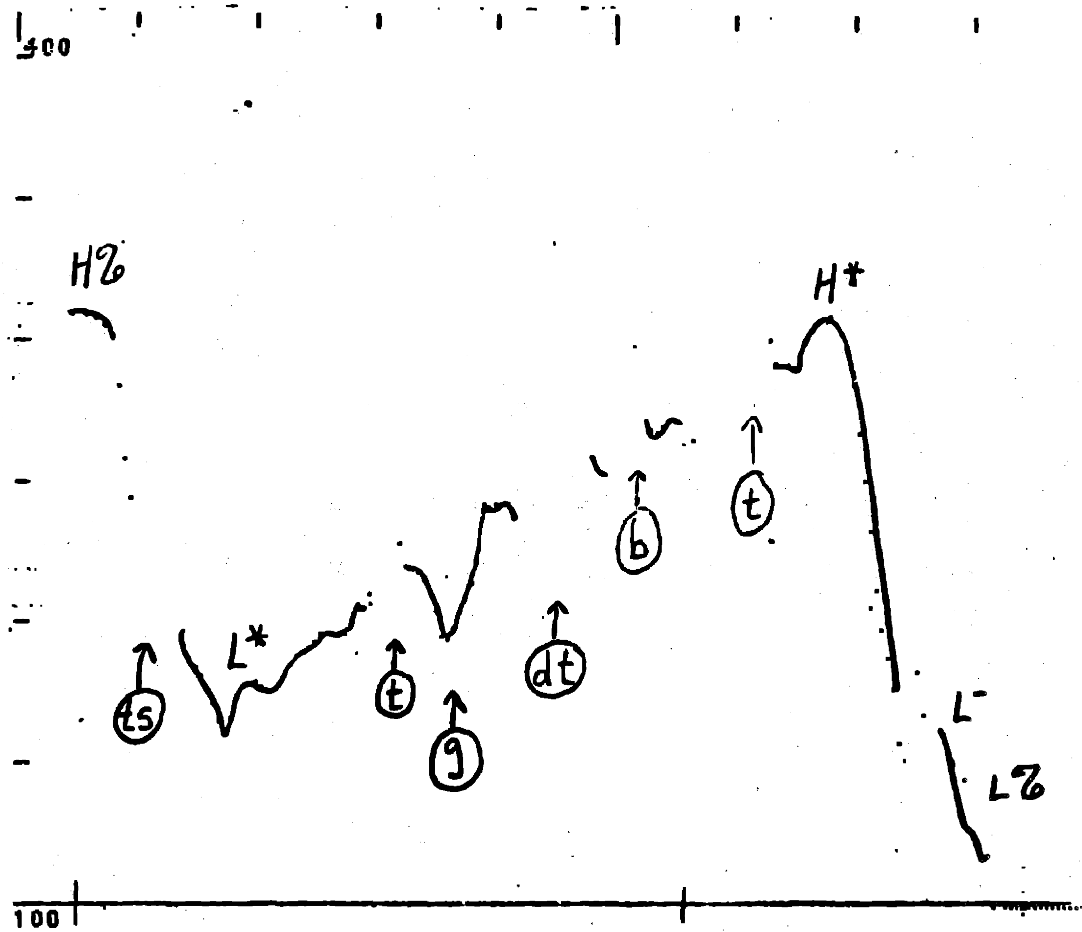
4.28) JBP



And remember to bring along your raingear

!H* L* H* L- L%

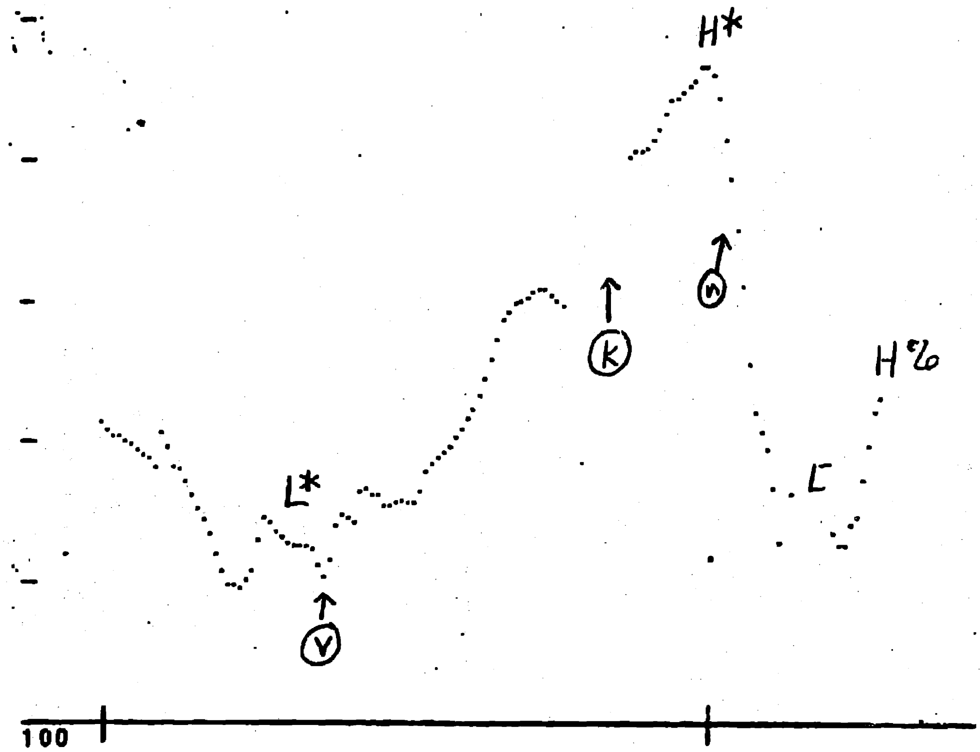
4.29) JBP



It's really too good to be true.

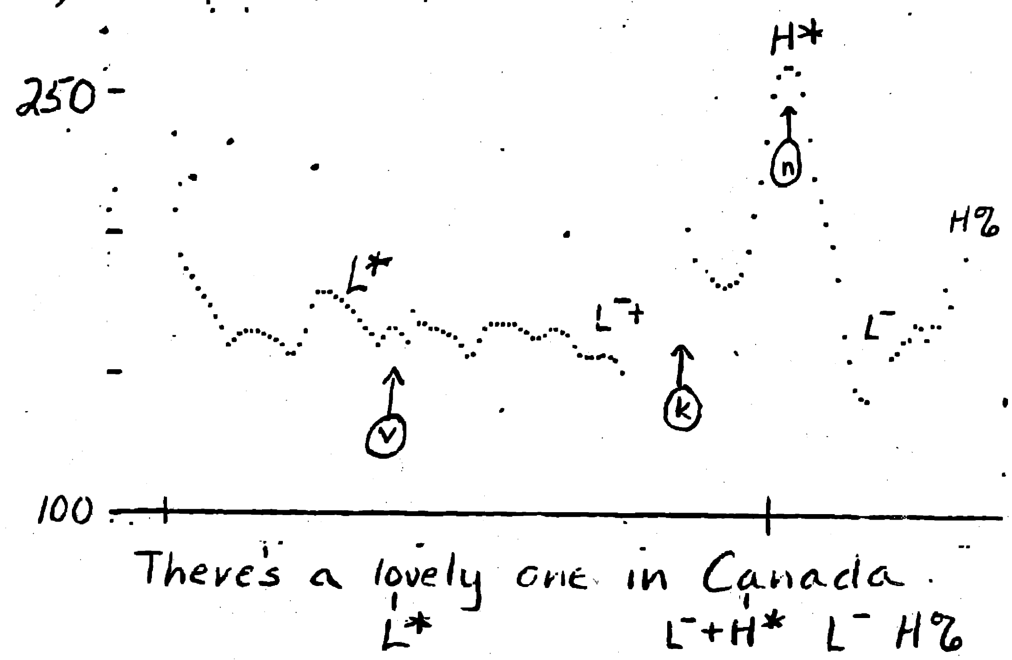
H% L* H* L- L%

4.30) JBP

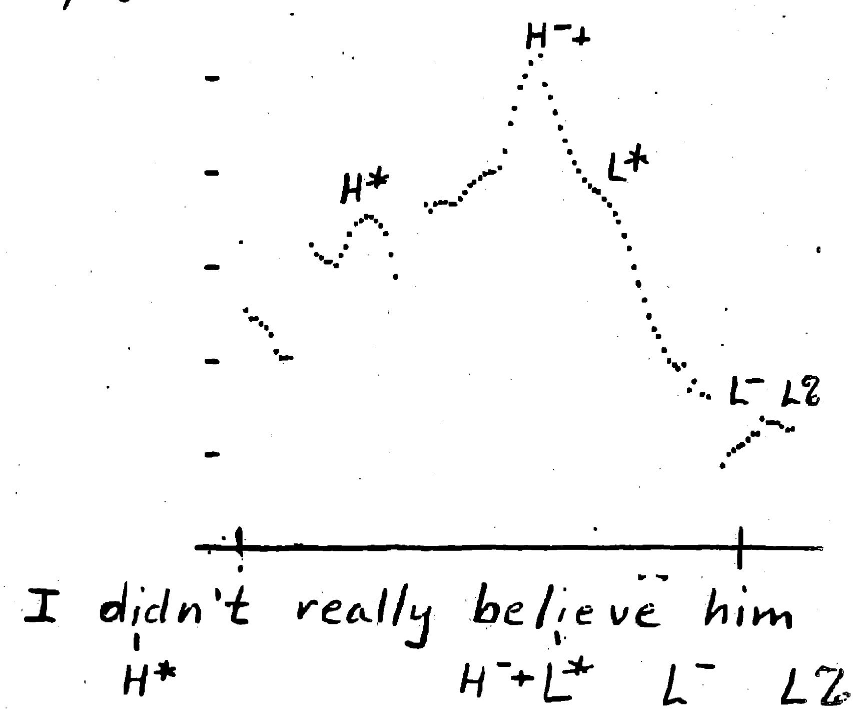


There's a lovely one in Canada
L* H* L⁻ H%

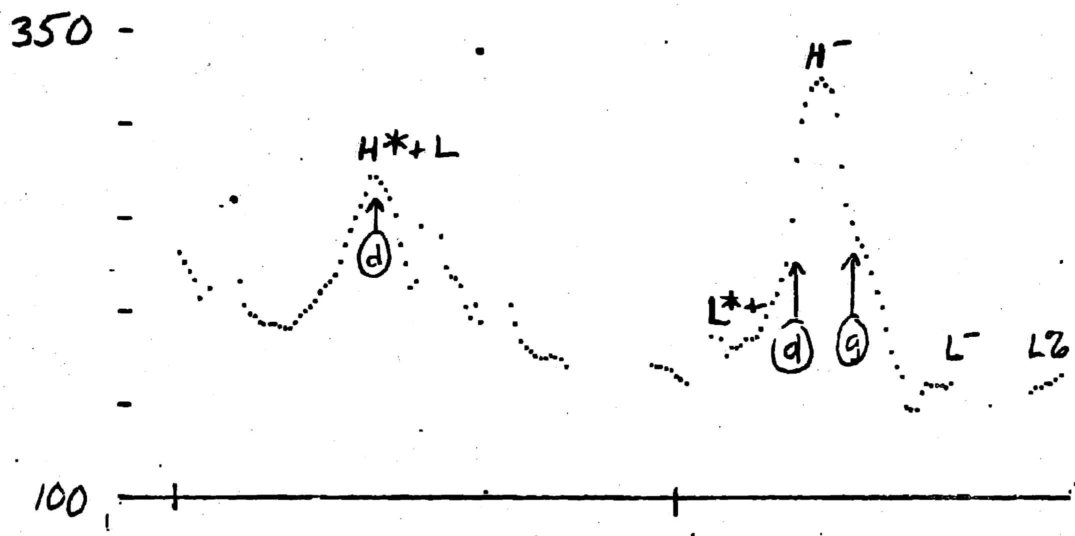
4.31) JBP



4.32) JBP

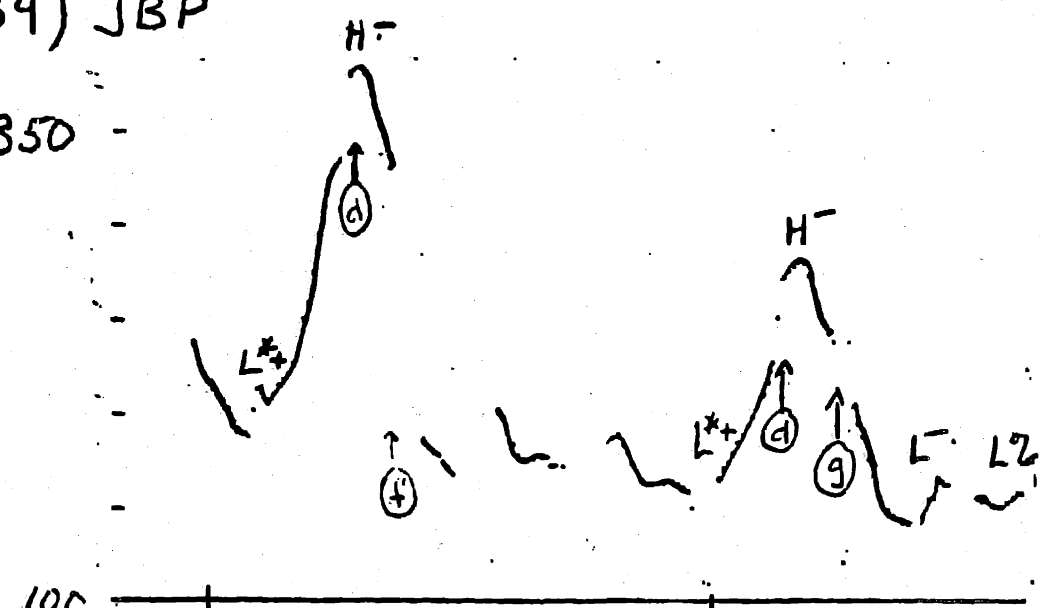


4.33) JBP



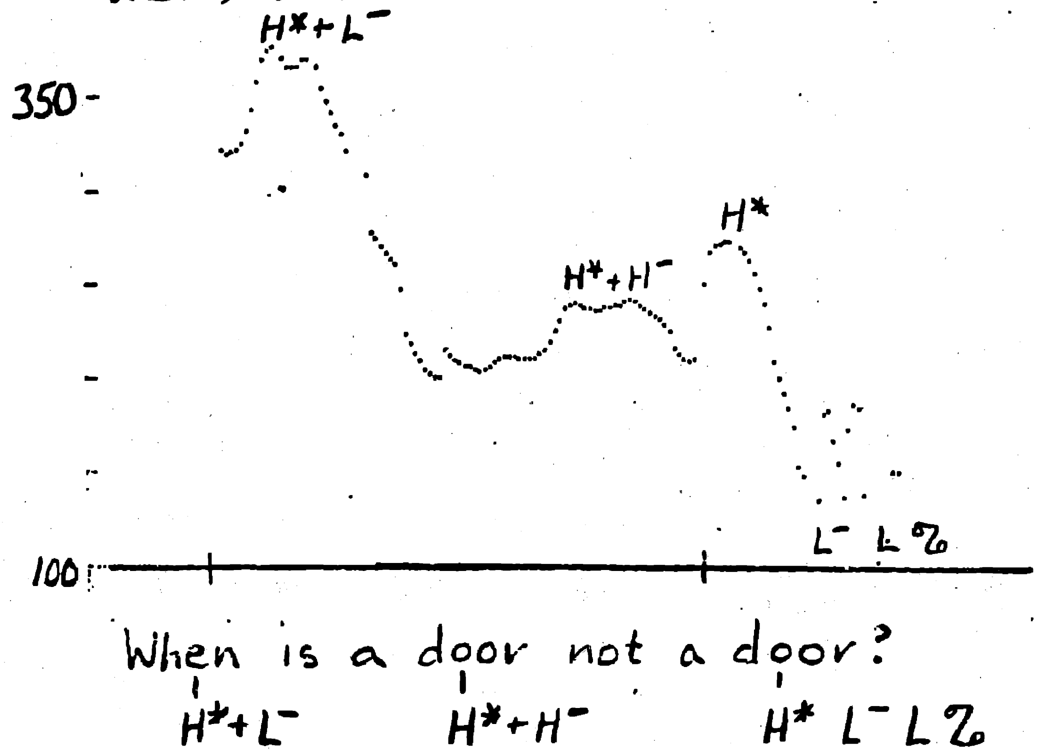
It's a wonderful place to be an undergraduate
H*+L L*+H⁻ L⁻ L%

4.34) JBP

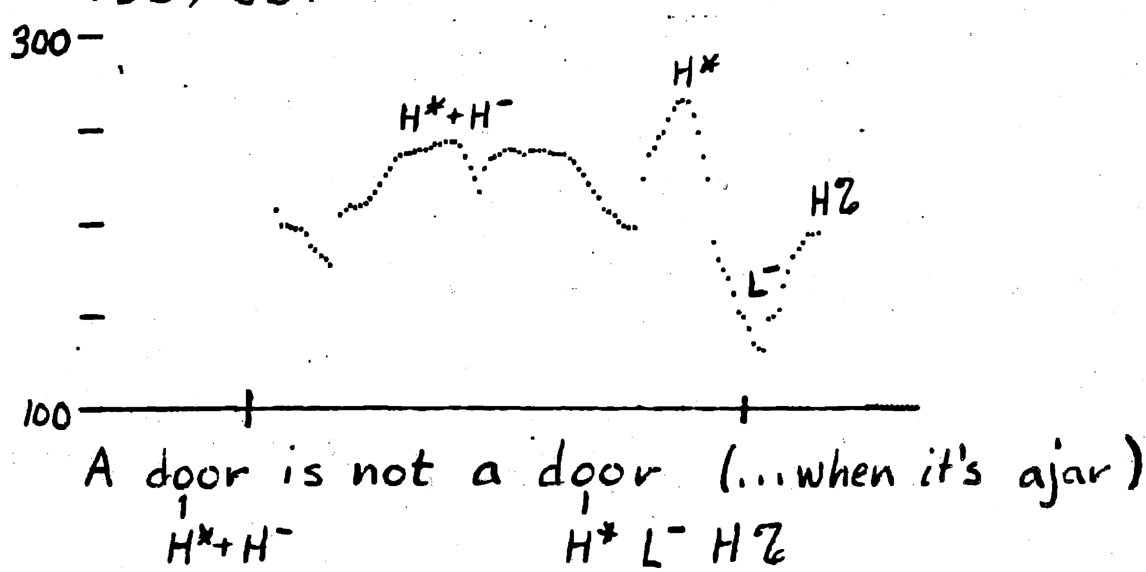


It's a wonderful place to be an undergraduate.
L*+H⁻ L*+H⁻ L⁻ L%

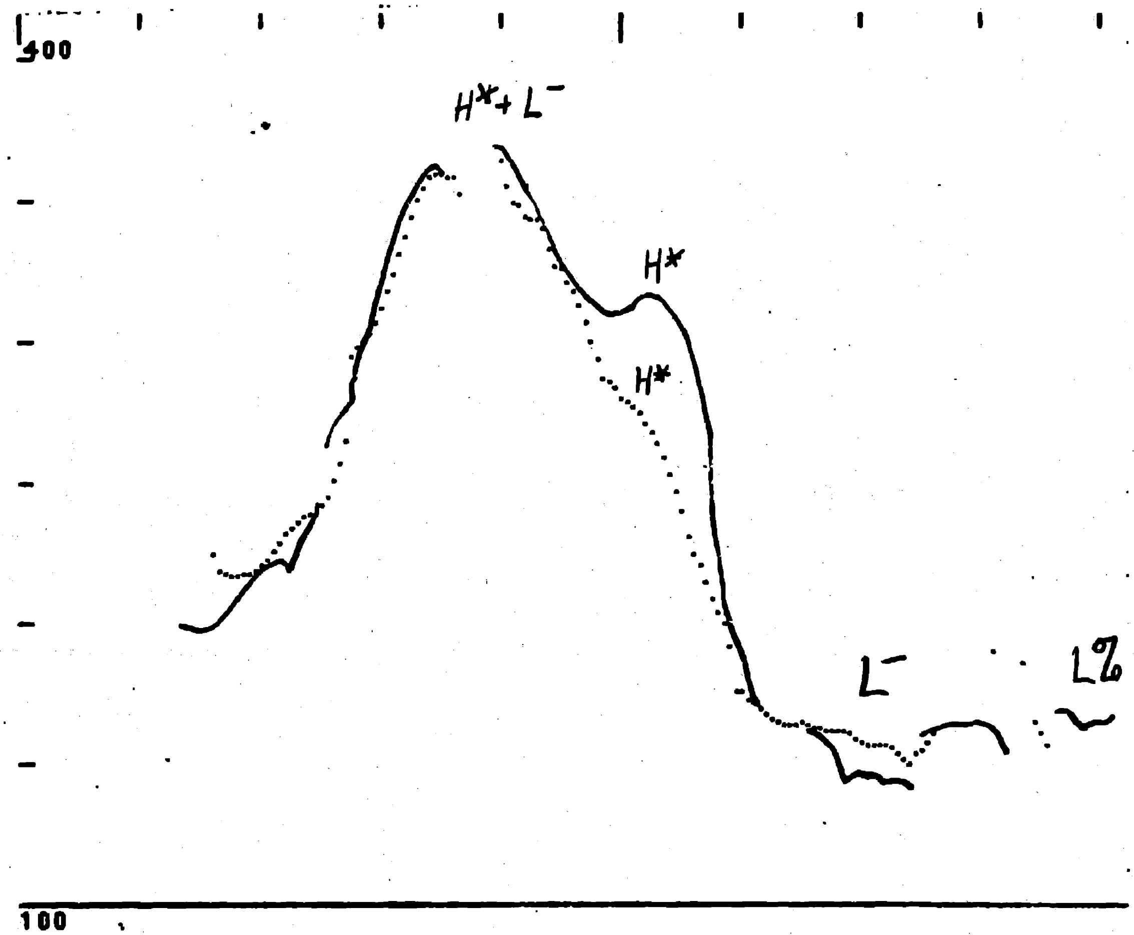
4.35) JBP



4.36) JBP



4.37) MB

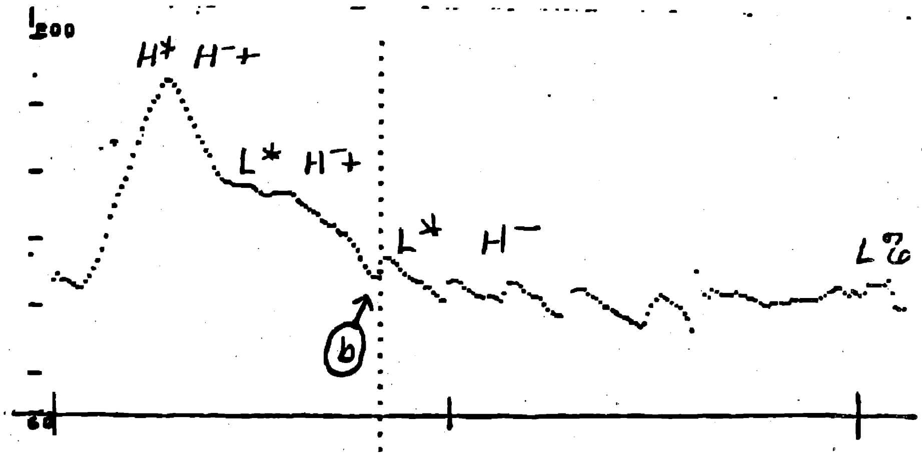


Two F∅ contours for:

I imagine Madeline did it
H*+L H* L⁻ L%

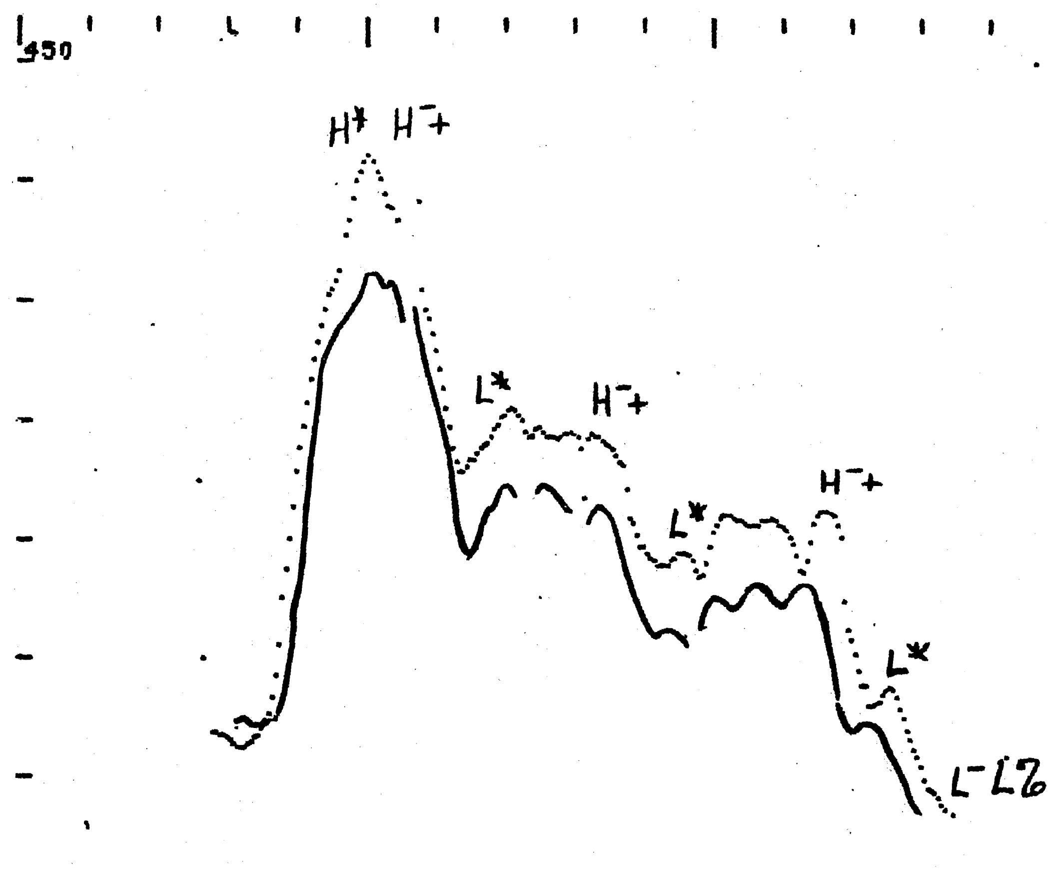
which differ in the prominence on the nuclear accent.

4.38) MYL



The Uruguayan bulldozer driver's union
H* H^-+L* H^-+L* H^- L%

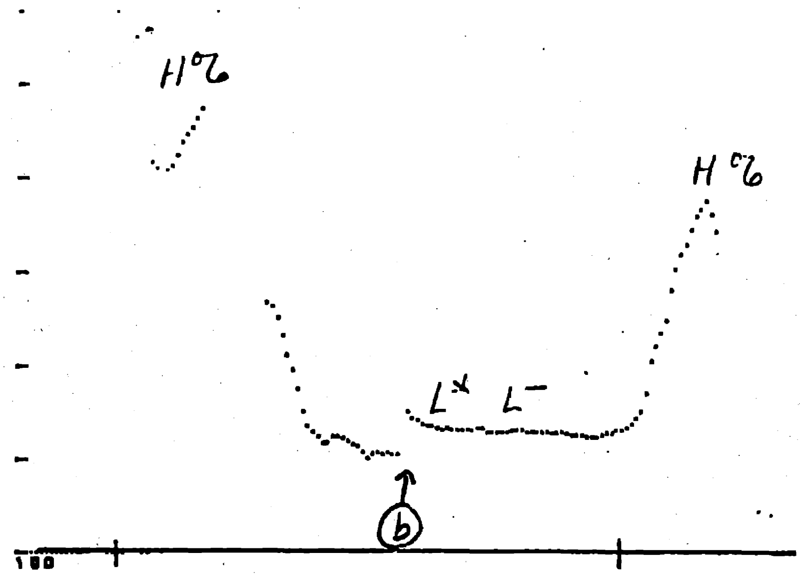
4.39) MB



Two F0 contours for:

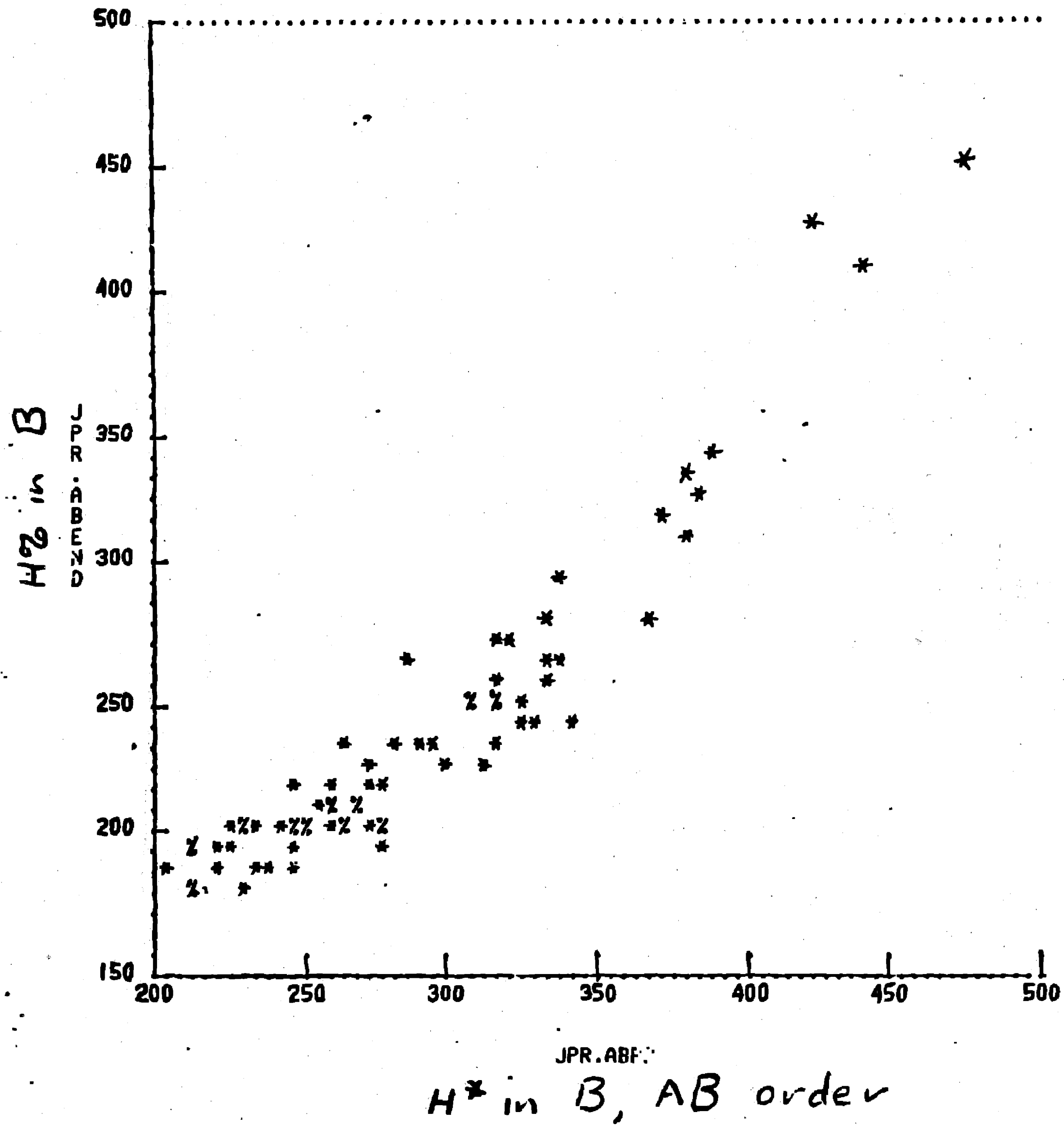
I really believe Ebenezer was a dealer
H* H⁻+L* H⁻+L* H⁻+L* L⁻L%

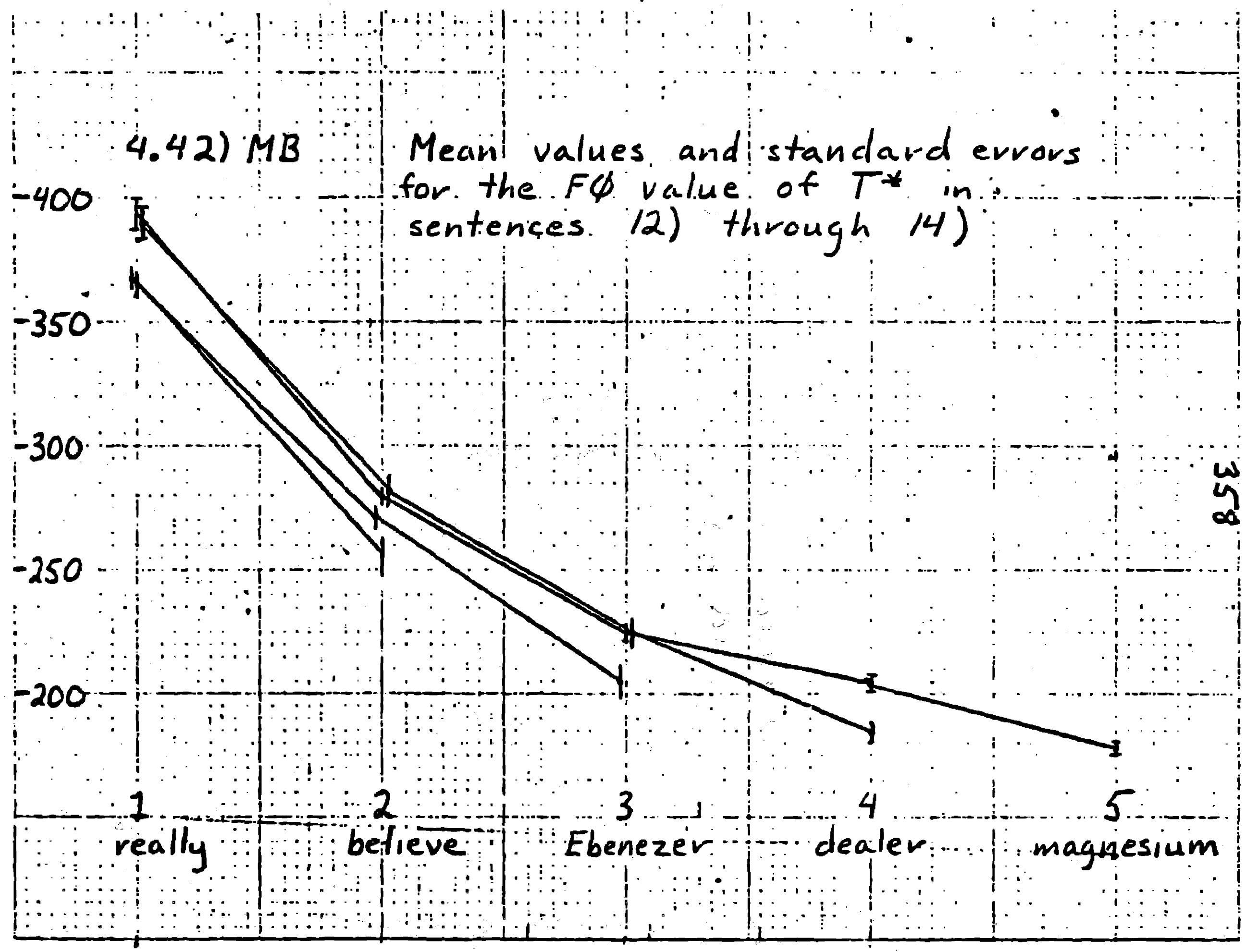
4.40) JBP



That's not a bowling alley
H% L* L- H%

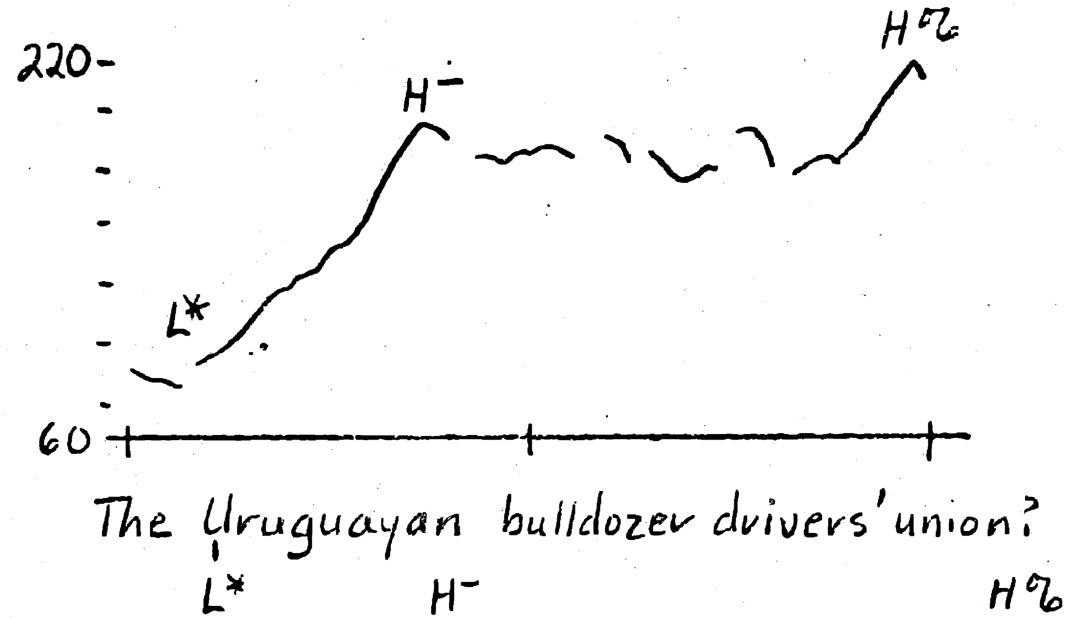
4.41


500
450
400
350
300
250
200
150
JPR.ABEND
H% in B
200
250
300
350
400
450
500
JPR.ABF
H* in B, AB order

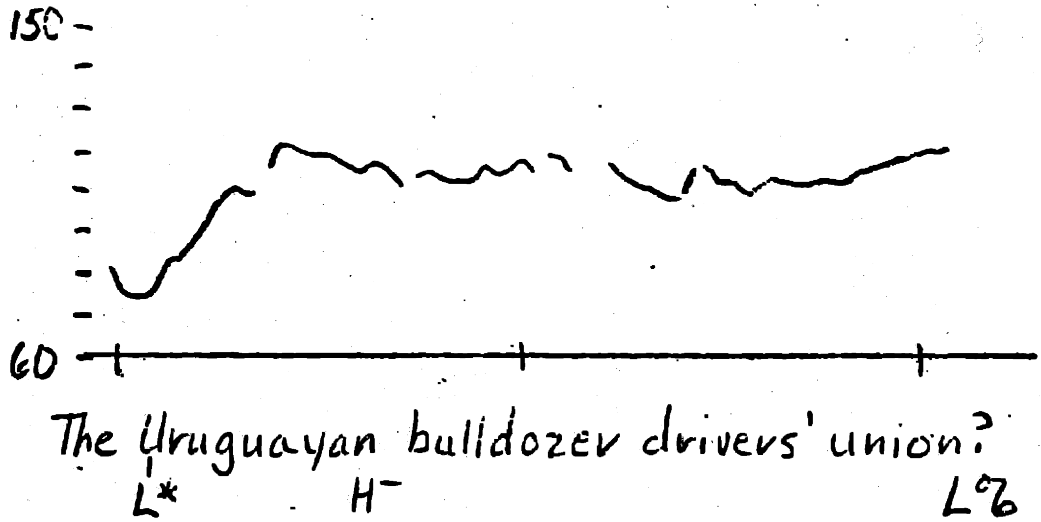


4.42) MB Mean values and standard errors for the F∅ value of T* in sentences 12) through 14)

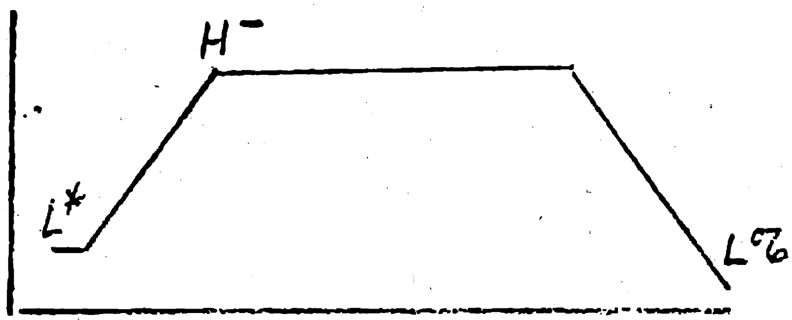
4.43A) MYL



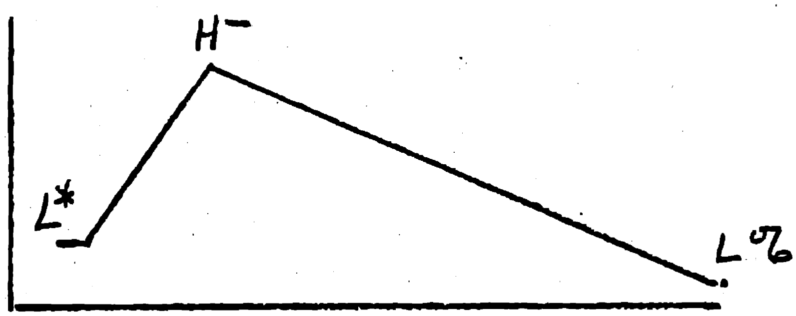
B) MYL



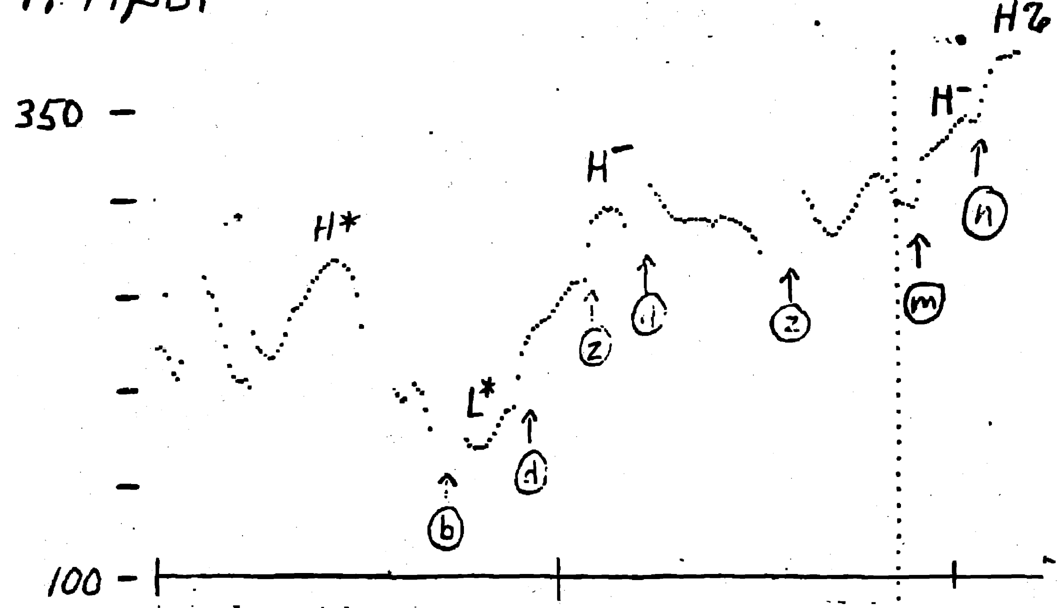
4.43 C) impossible in English



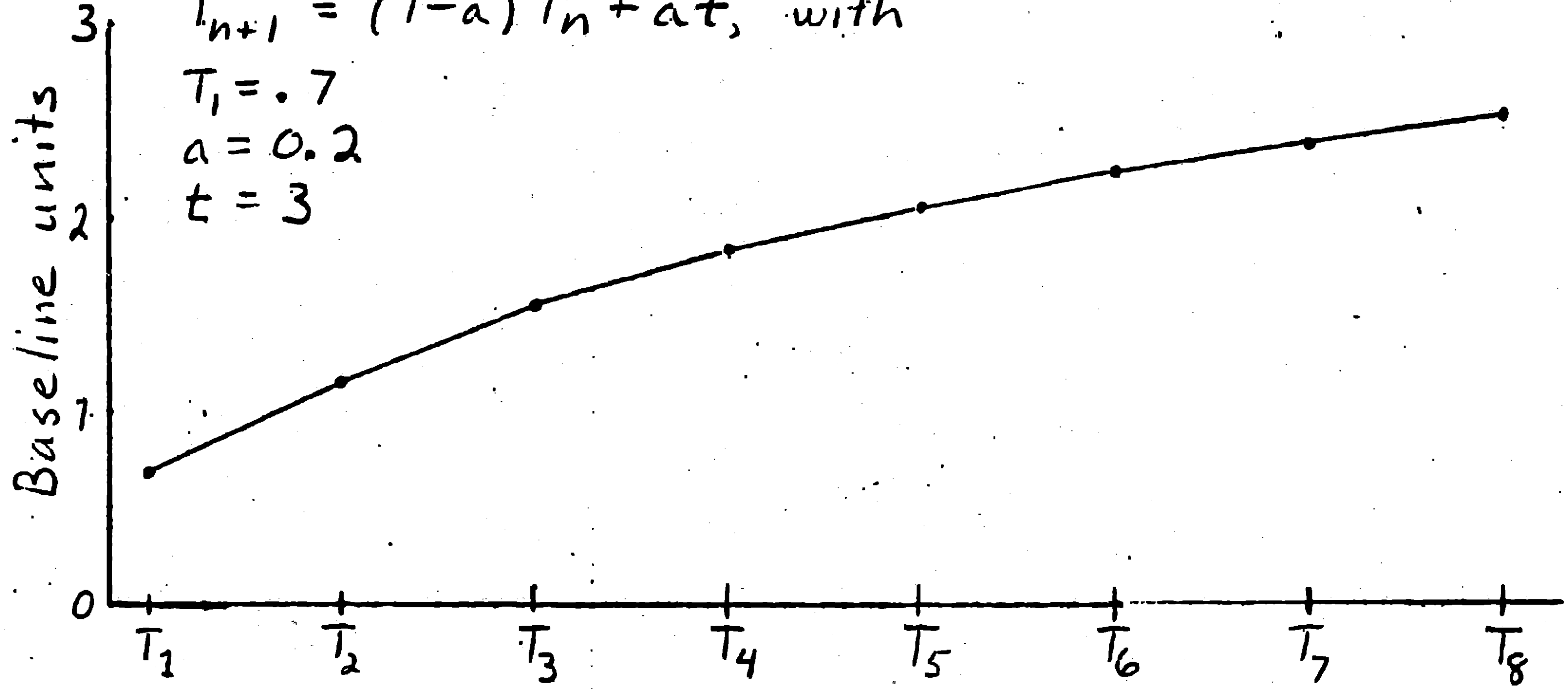
D) also impossible



4.44) JBP

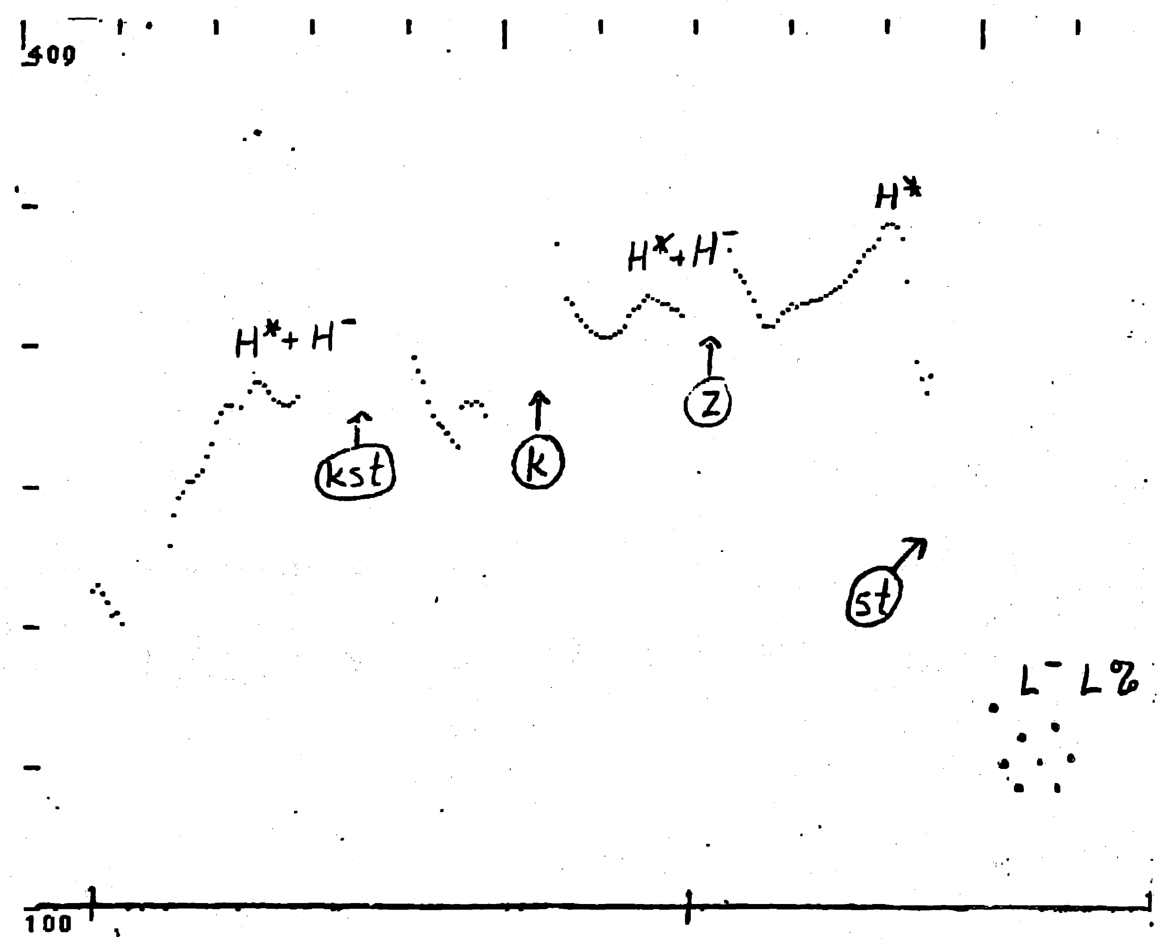


Is he a member of the bulldozer drivers' union, Manny?
H* L* H⁻ H⁻ H%



4.45) Upstepping tones computed by $T_{n+1} = (1-a)T_n + at$, with
$T_1 = .7$
$a = 0.2$
$t = 3$

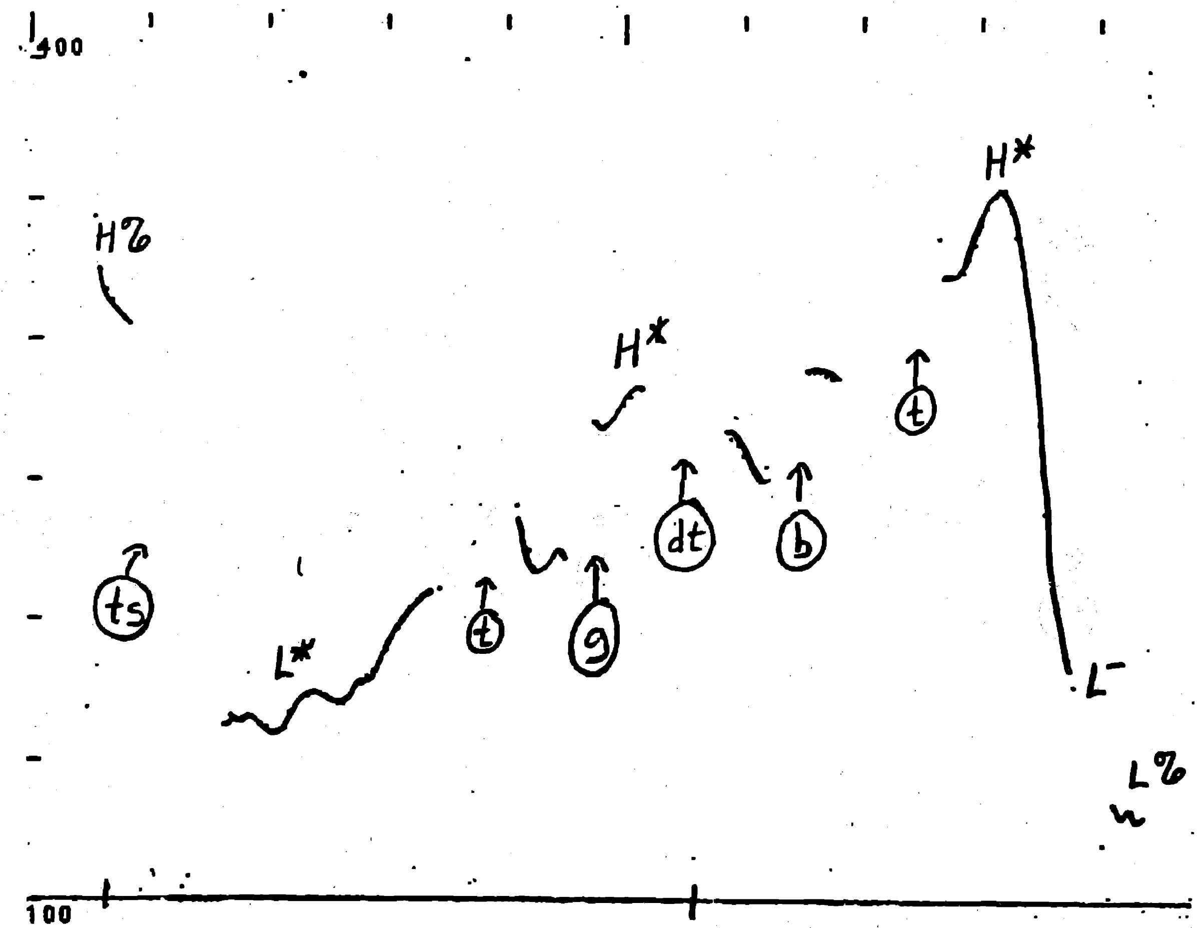
4.46) JBP



It's indexed by the keywords in the abstract.

H*+H-   H*+H-   H* L-L%

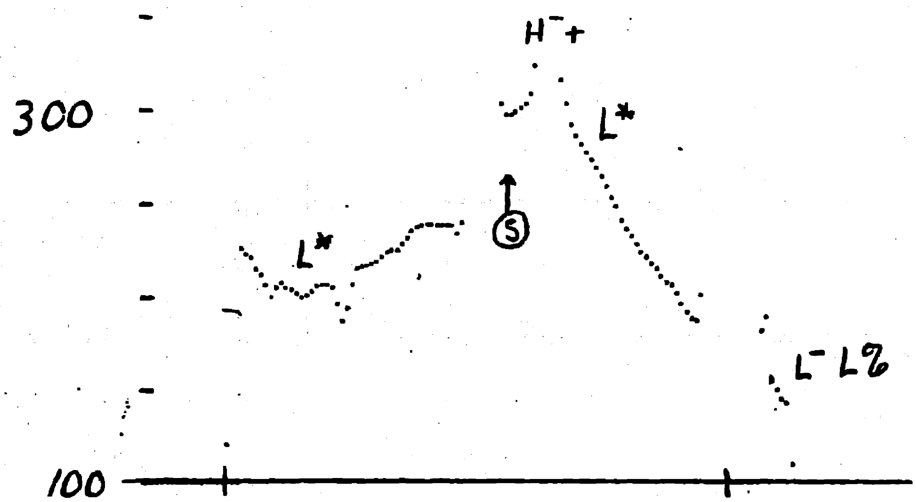
4.47) JBP



It's really too good to be true.

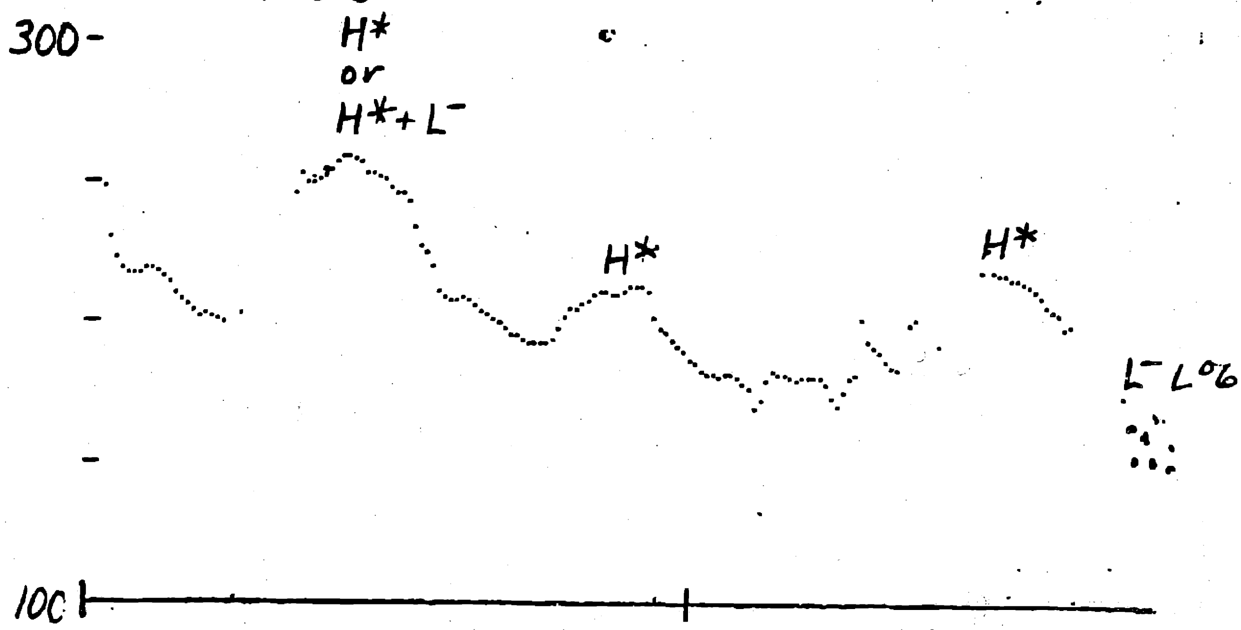
H% L* H* H* L⁻ L%

4.48) JBP



an evanescent rainbow
L* H⁻+L* L⁻ L%

4.49) JBP



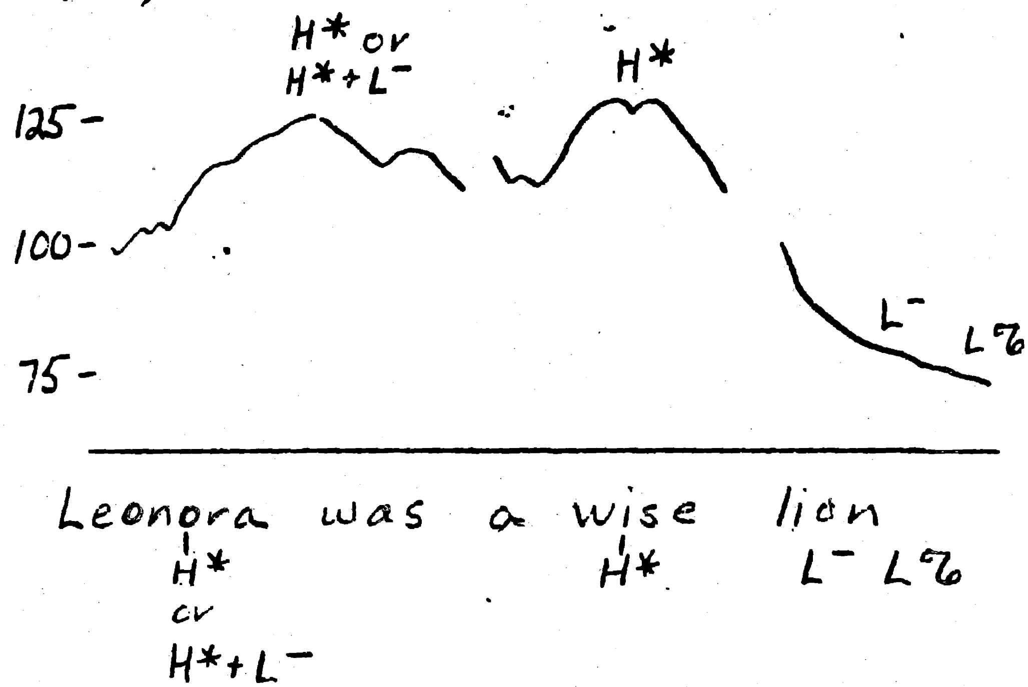
Try occasional moderate agitation.

H^*_1 H^*_2 H^*_3 L^- $L\%$

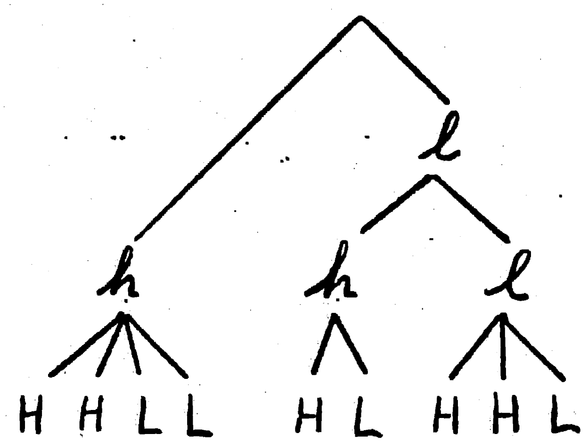
or

H^*+L^-

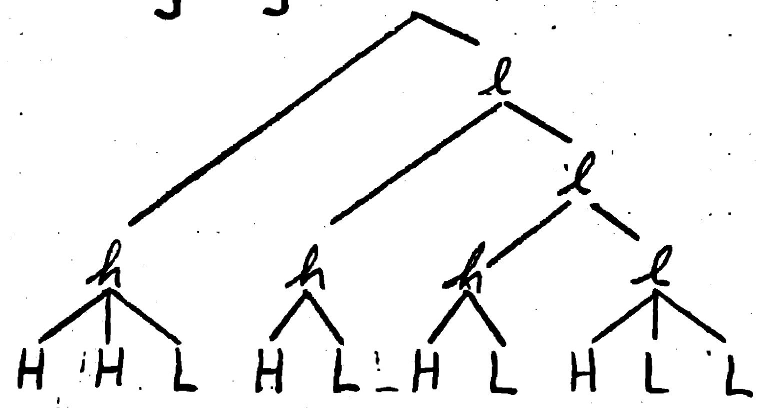
4.50) MYL



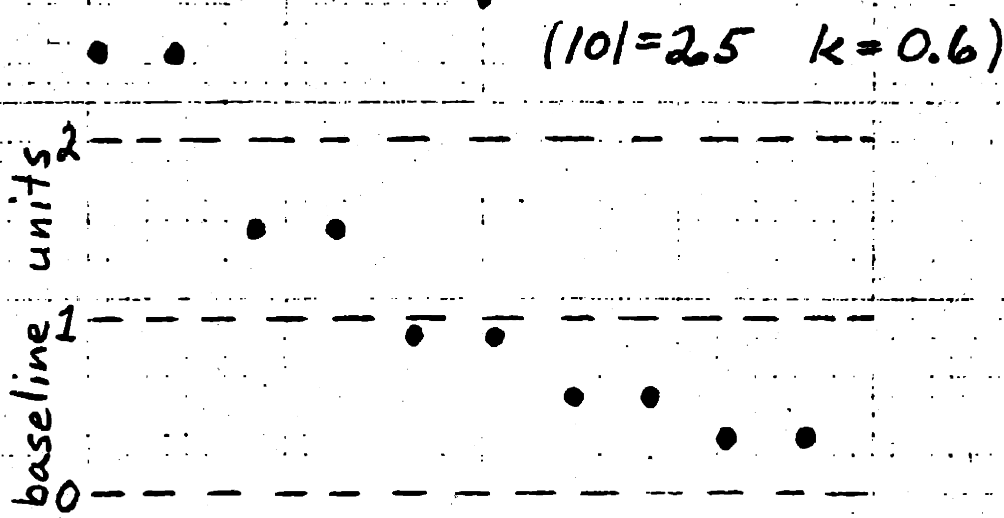
4.51) The hierarchical representation of downstep.



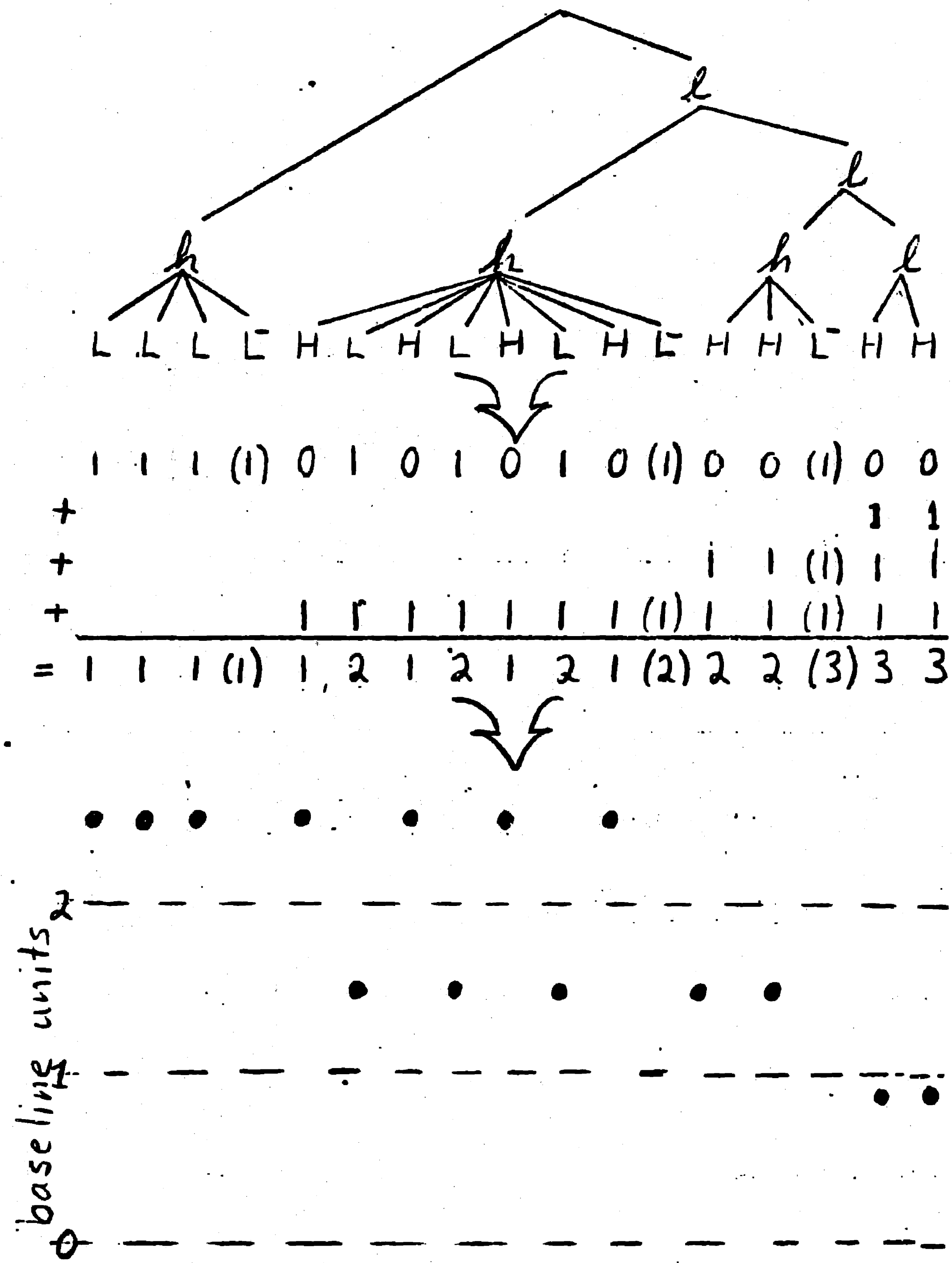
4,52) Deriving tone levels from a hierarchical representation for a language with total downdrift



	0	0	1	0	1	0	1	0	1	1
+								1	1	1
+						1	1	1	1	1
				1	1	1	1	1	1	1
=	0	0	1	1	2	2	3	3	4	4



4.53) Deriving tone levels in Kikuyu
(L⁻ is a floating L tone)



2

4.54) (from Thorsen 1979c)

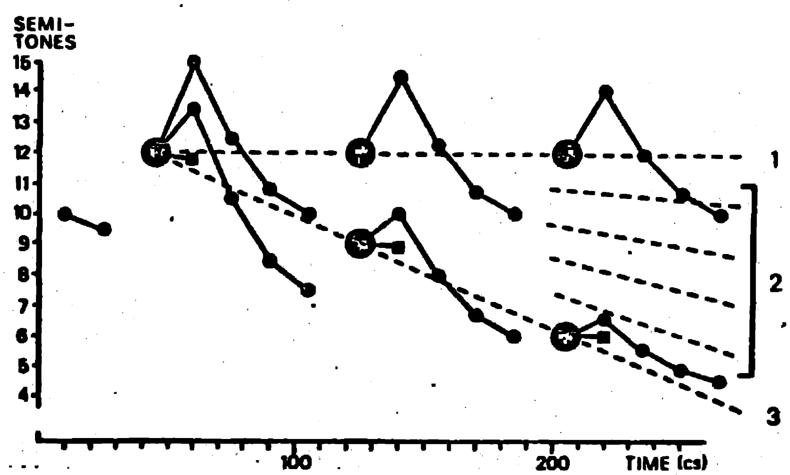
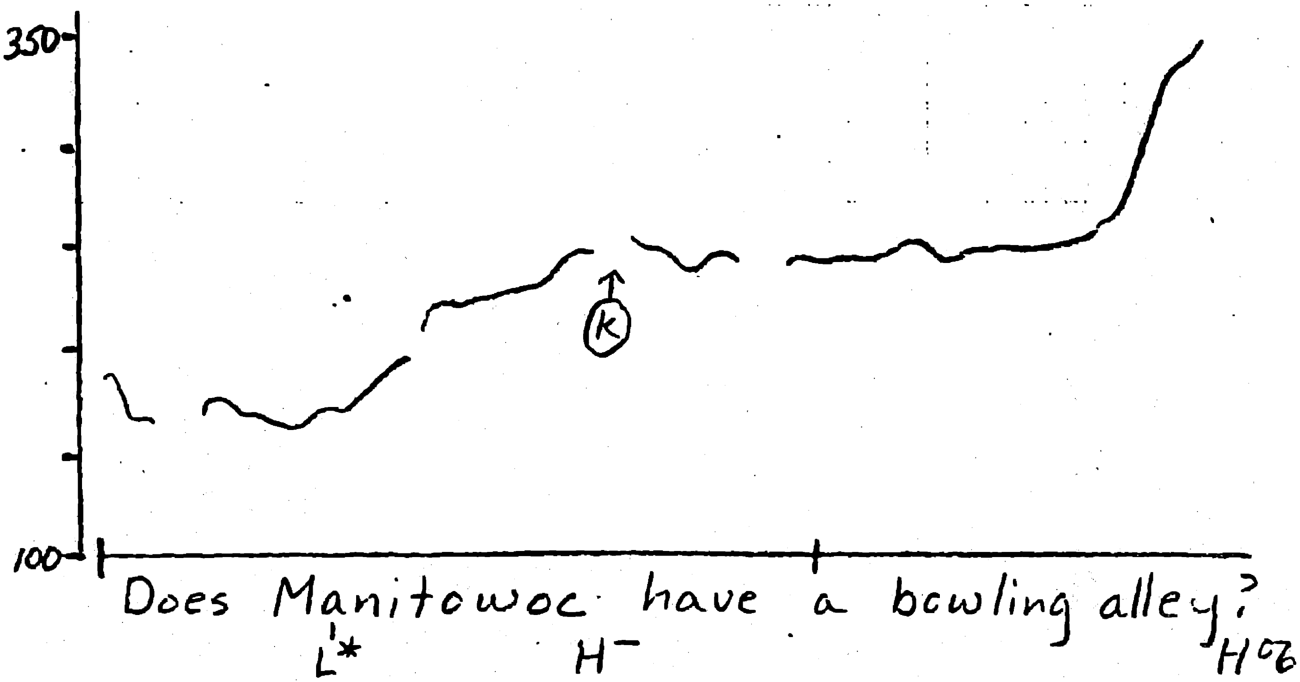


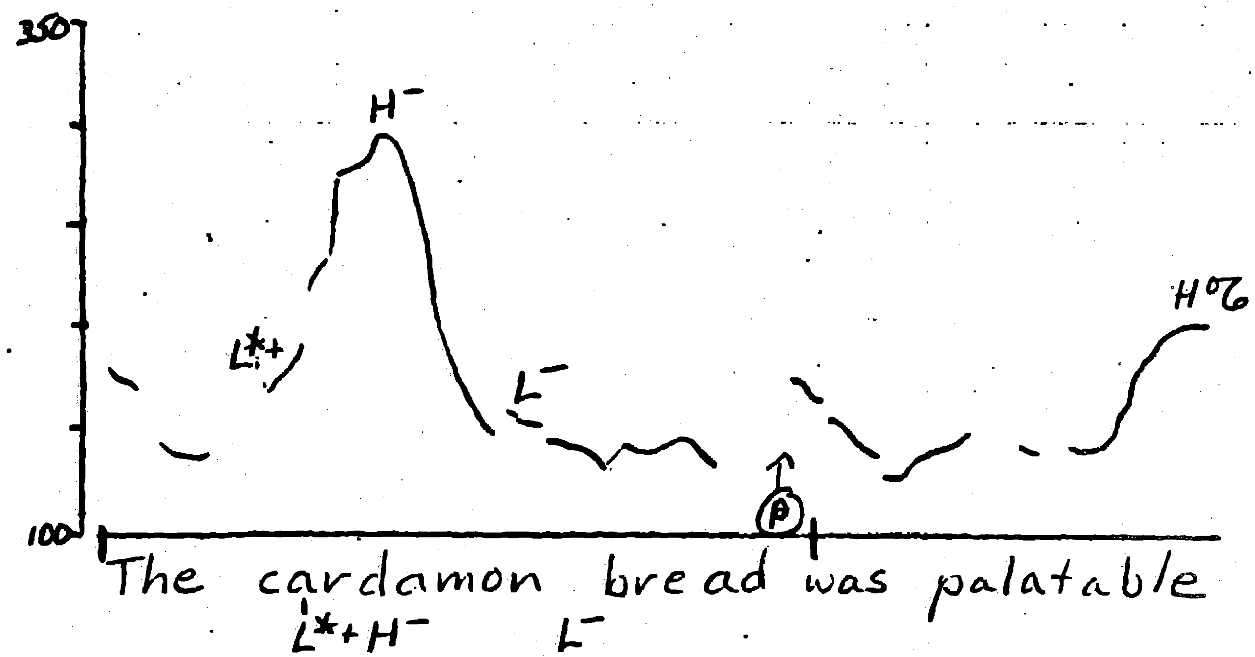
Figure 1

A model for the course of Fo in short sentences in ASC Danish. 1: syntactically unmarked questions, 2: interrogative sentences with word order inversion and/or interrogative particle, and non-final periods (variable), 3: declarative sentences. The large dots represent stressed syllables, the small dots unstressed ones, and the small squares represent an unstressed syllable being the only one between two stressed ones. The full lines represent the Fo pattern associated with stress groups, and the broken lines denote the intonation contours.

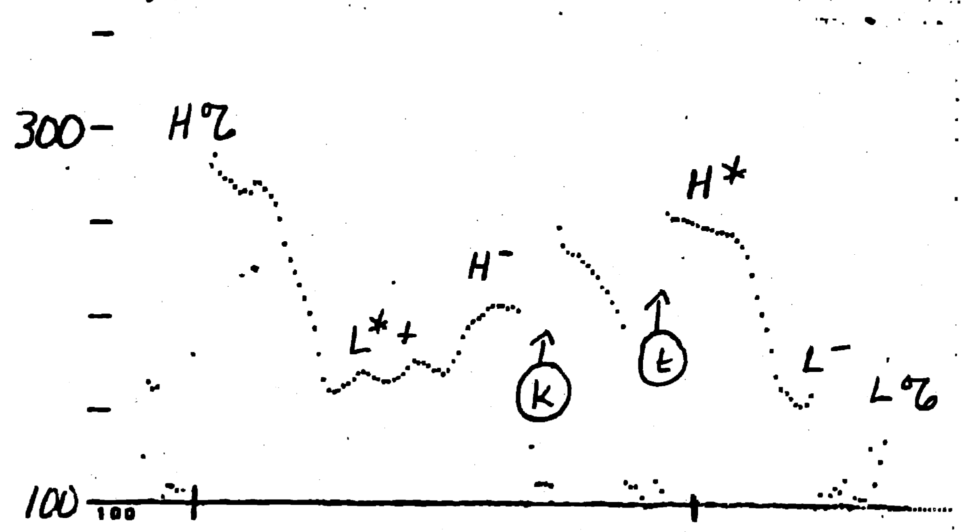
5.1) MB


350
100
Does Manitowoc have a bowling alley?
L*
H⁻
H%

5.2) MB


350
100
H⁻
L*+
L⁻
H%
The cardamon bread was palatable
L*+H⁻
L⁻

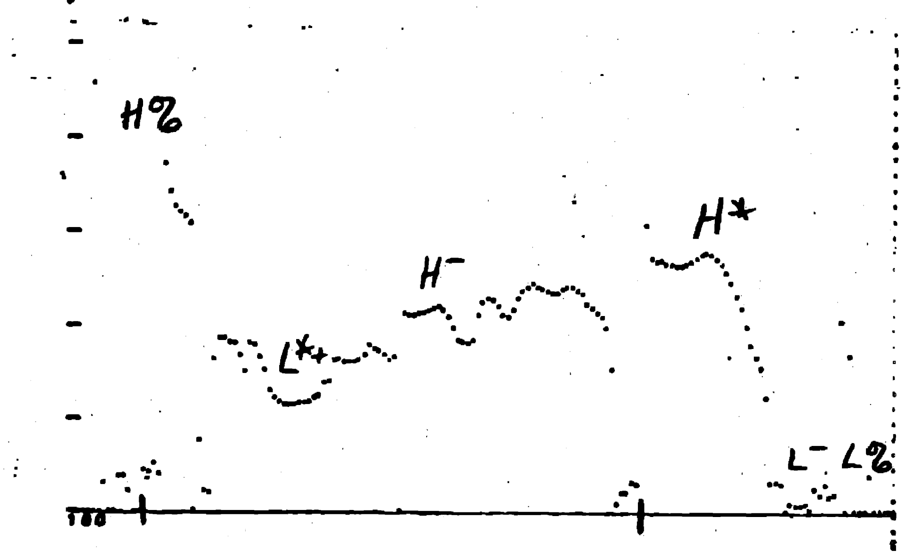
5.3) JBP



Only one in a quintillion

H% L*+H- H* L- L%

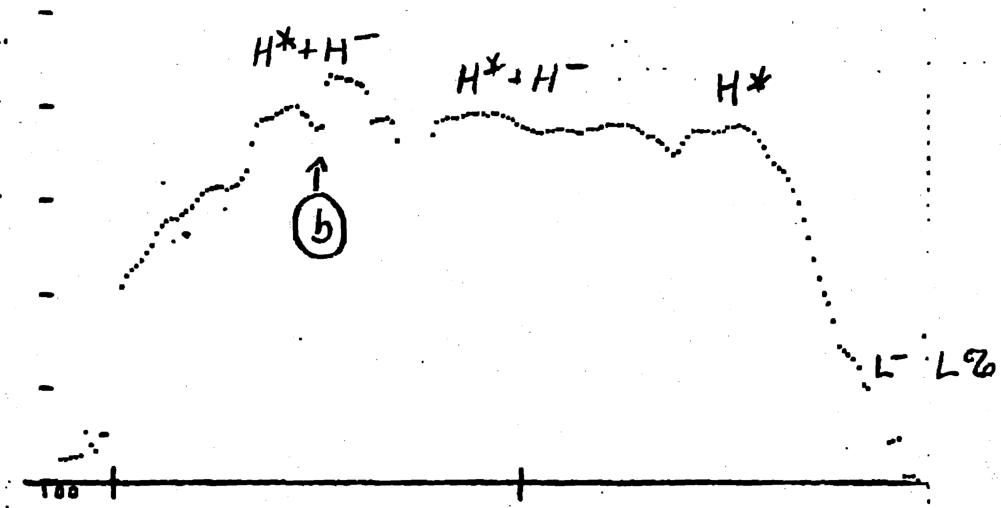
5.4) JBP



But Rob agreed to bring salad.

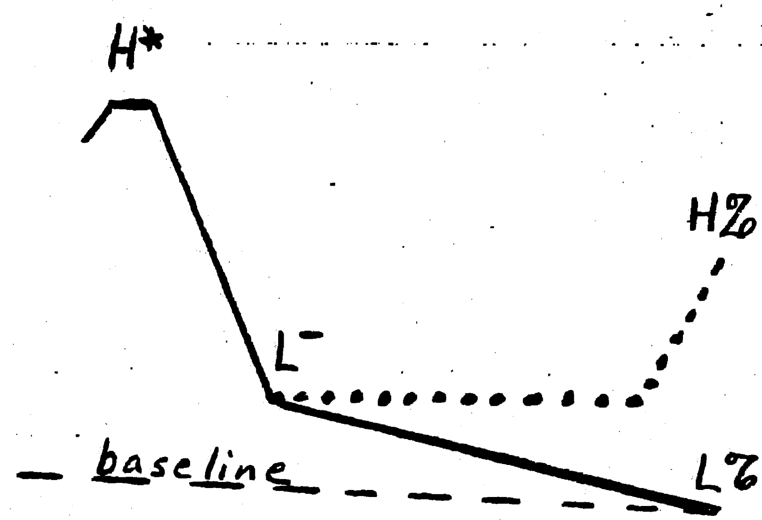
H% L*+H- H* L- L%

5.5) JBP

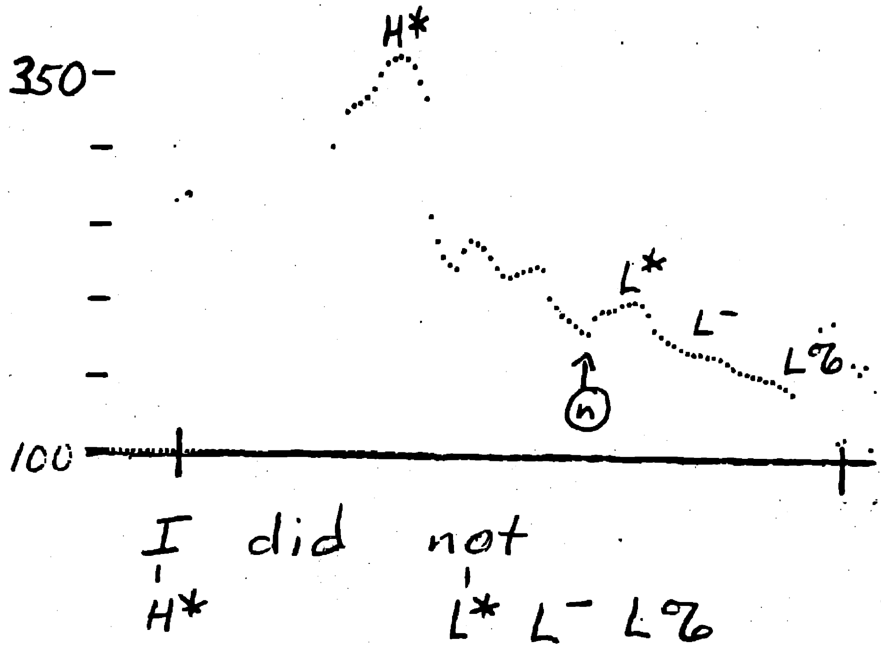


And remember to bring along your raingear.
H*+H⁻ H*+H⁻ H* L⁻ L%

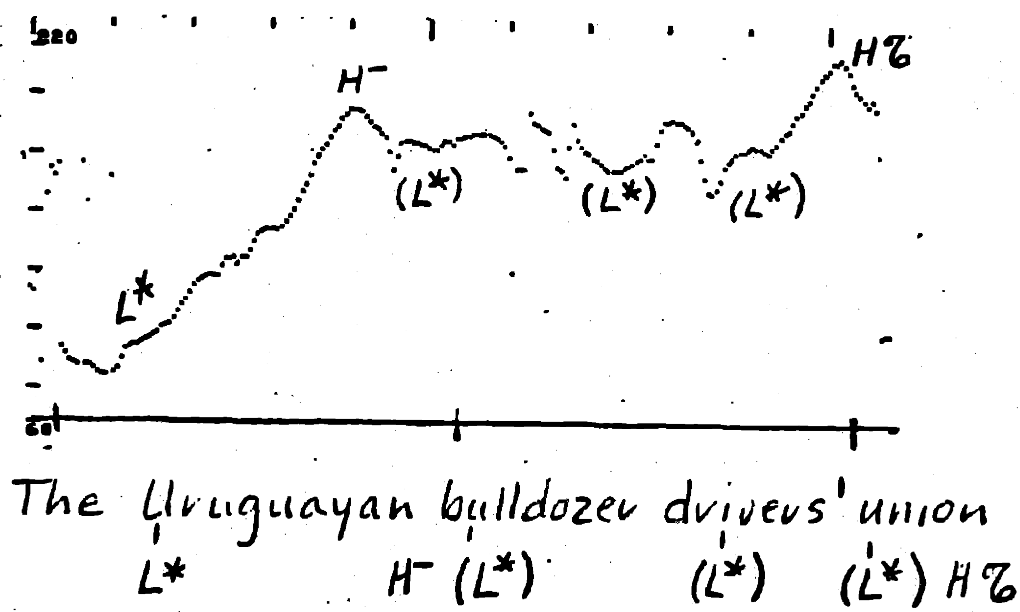
5.6)



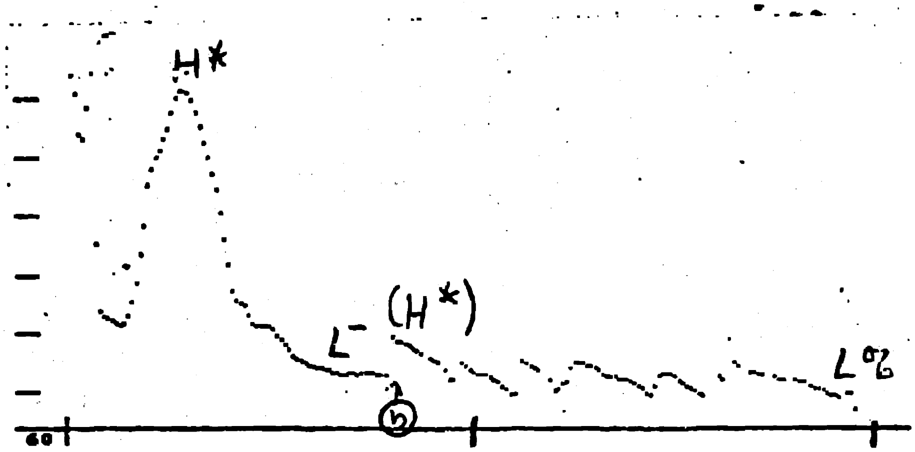
5.7) JBP


H*
350-
L*
L-
L%
n
100
I did not
H*
L* L- L%

5.8) MYL

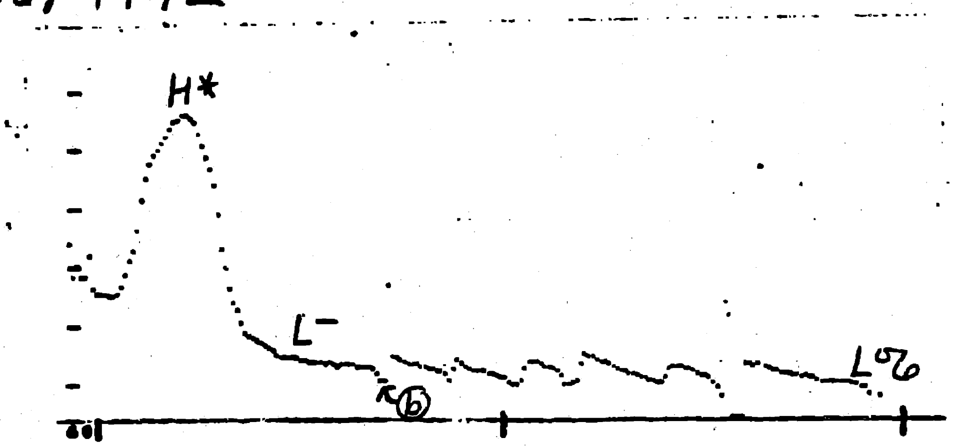

220
H-
H%
(L*)
(L*)
(L*)
L*
50
The Uruguayan bulldozer drivers' union
L*
H- (L*)
(L*)
(L*) H%

5.9) M4L



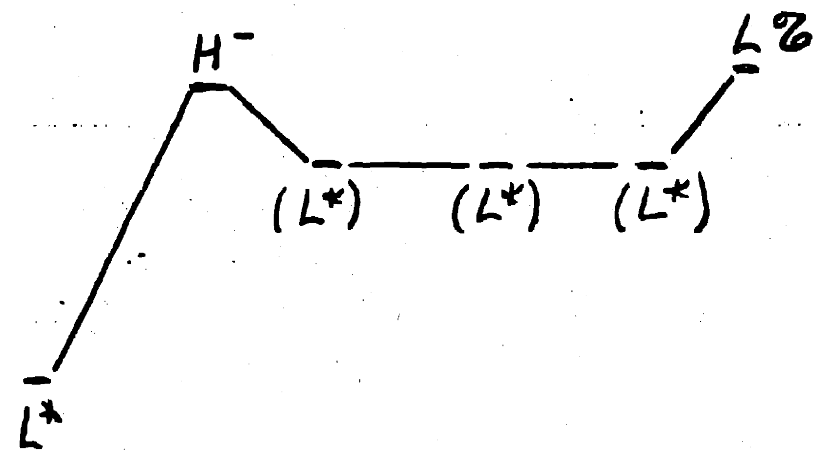
The Uruguayan bulldozer drivers' union
H* L- (H*) L%

5.10) M4L

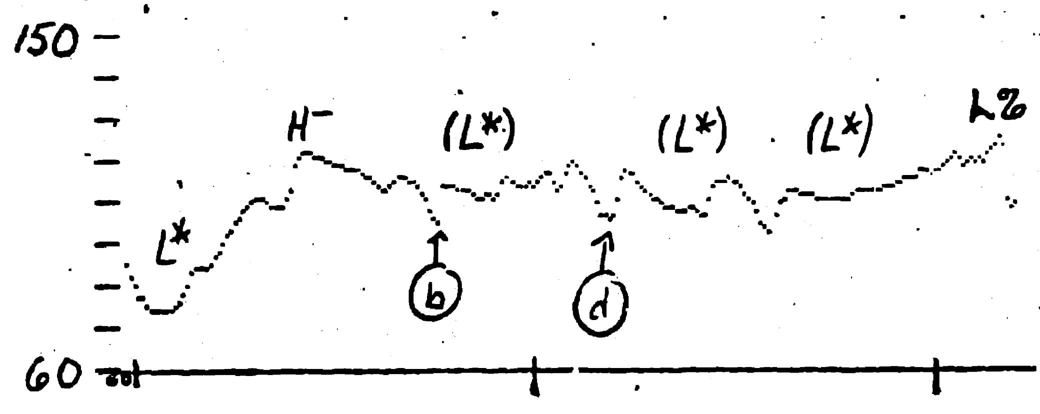


The Uruguayan bulldozer drivers' union.
H* L- L%

5.11)

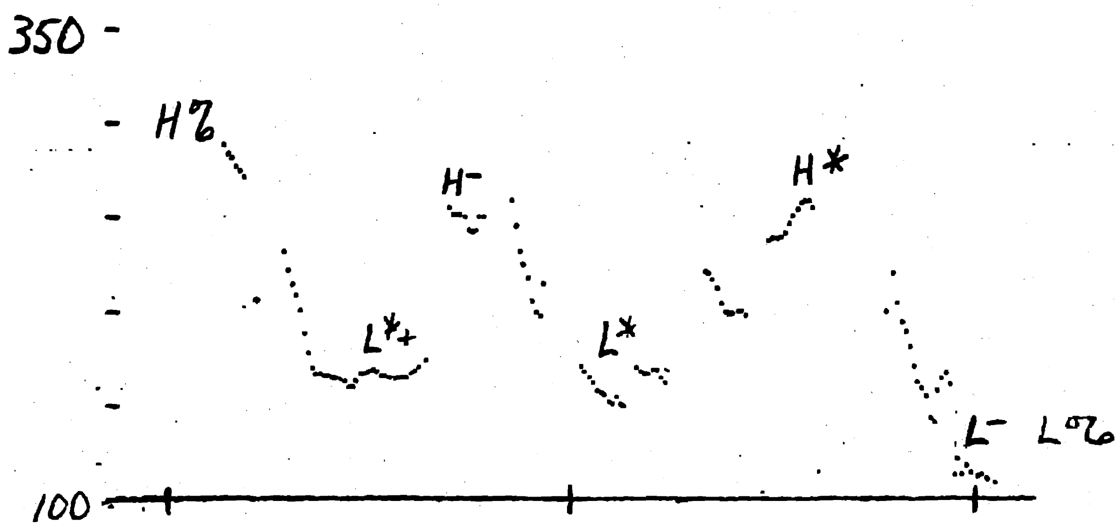


5.12) M4L



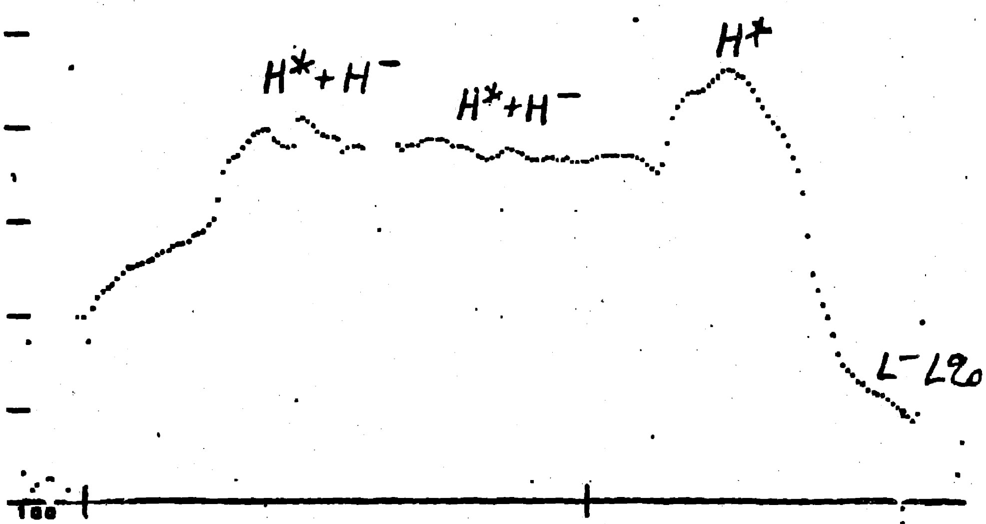
The Uruguayan bulldozer drivers' union.
L* H⁻ (L*) (L*) (L*) L%

5.13) JBP



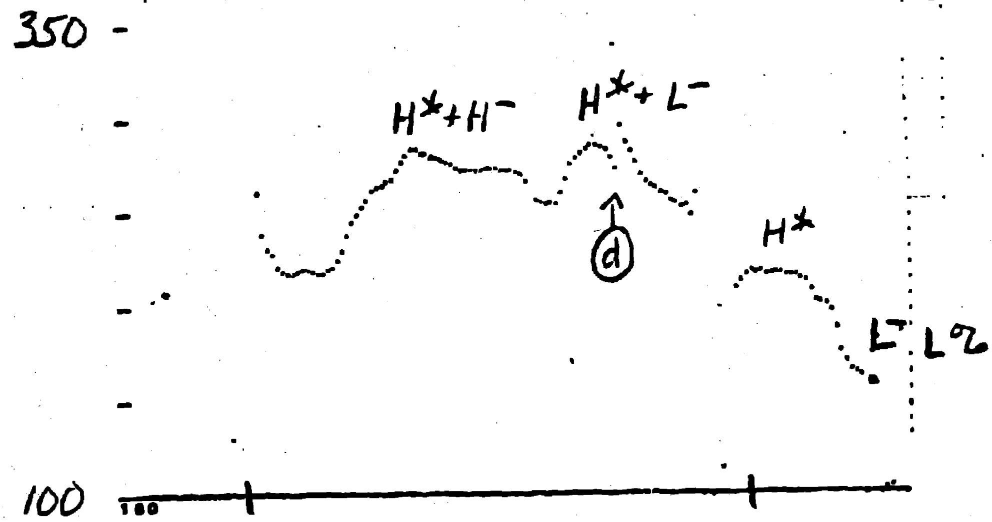
That's a remarkably clever suggestion
H% L*+H L* H* L- L%

5.14) JBP



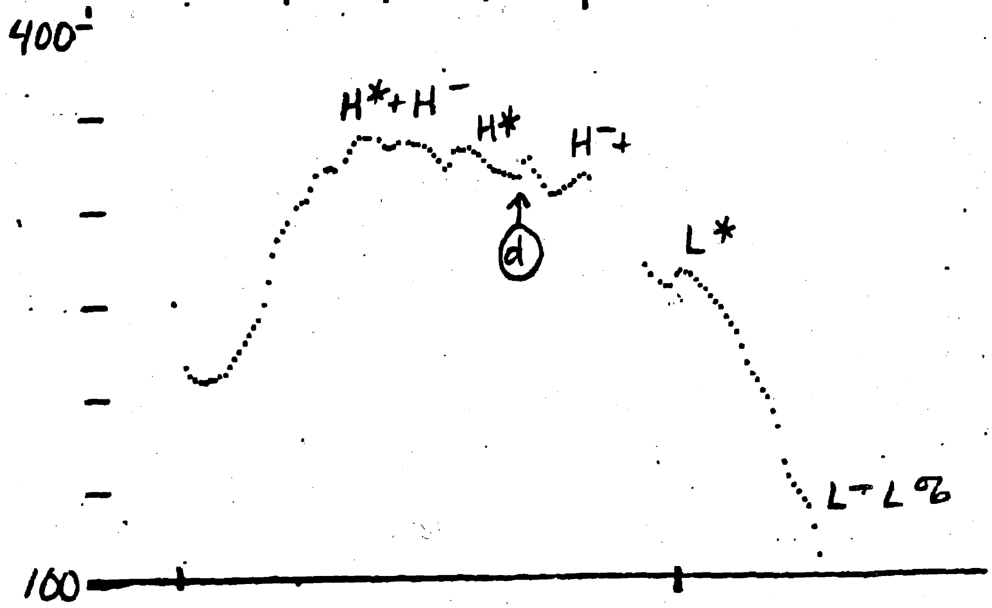
And remember to bring along your raingear
H*+H- H*+H- H* L- L%

5.15) JBP



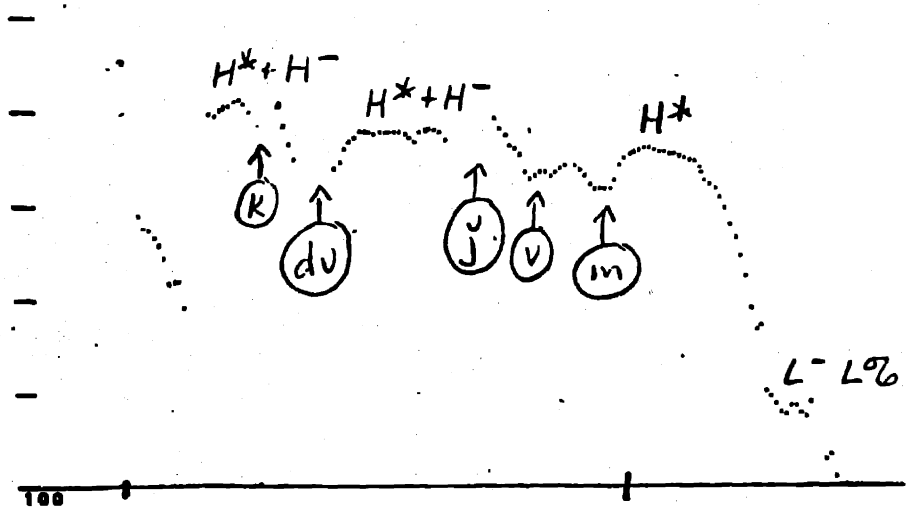
There are many intermediate levels
H*+H⁻ H*+L⁻ H* L⁻L%

5.16) JBP



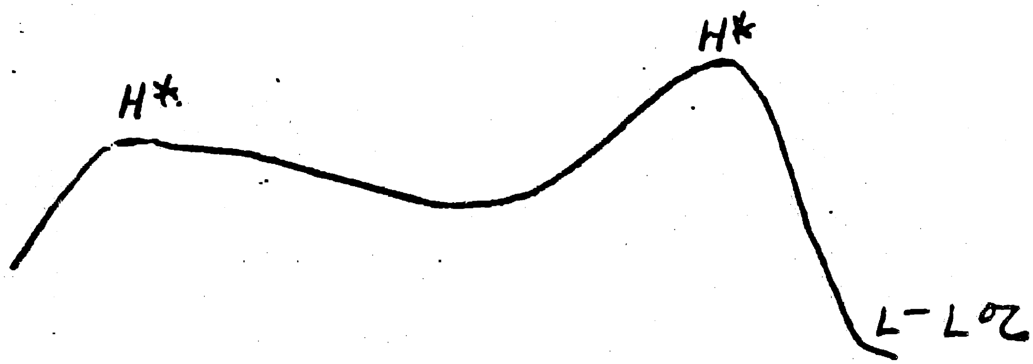
There are many intermediate levels
H*+H⁻ H* H⁻+L* L⁻L%

5.17A) JBP



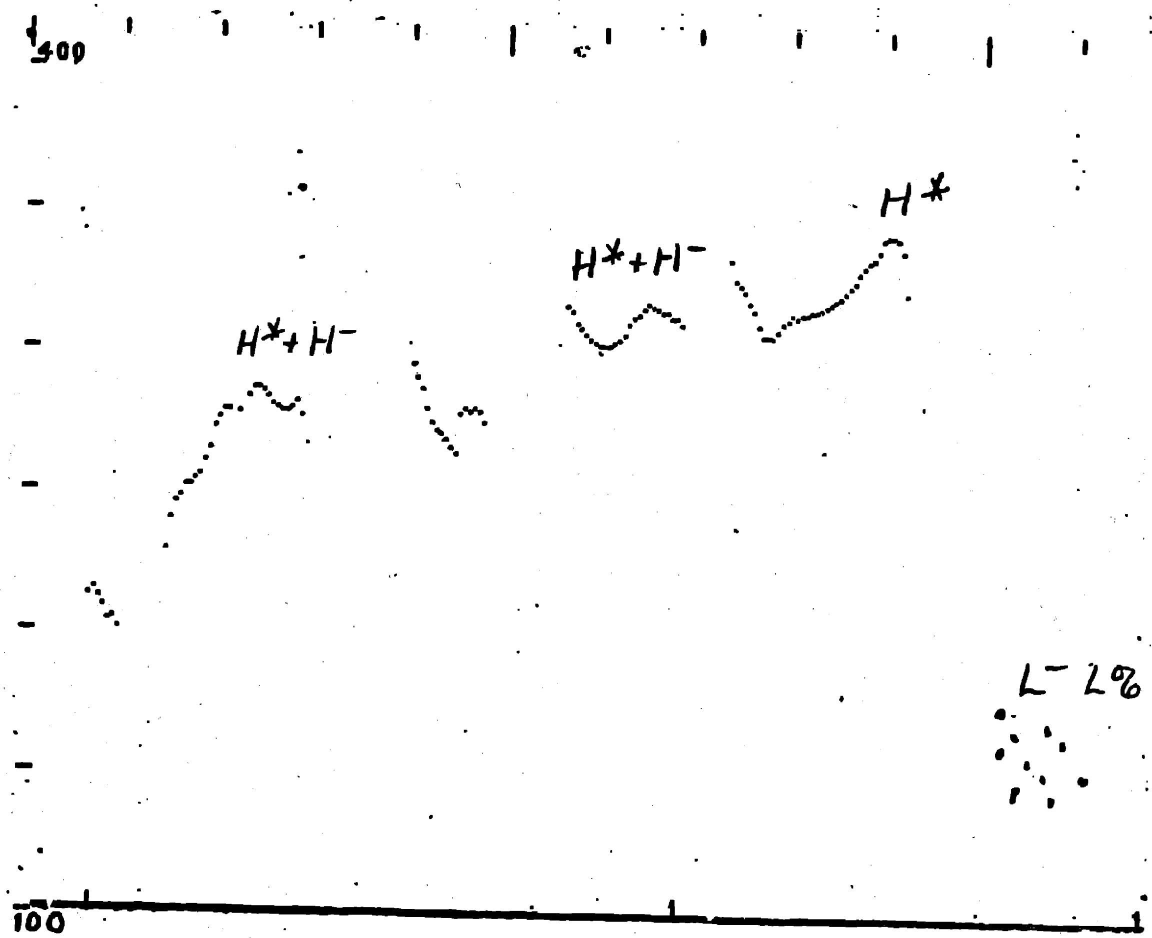
He took advantage of Amanda
H*+H⁻ H*+H⁻ H* L⁻ L%

B)



He took advantage of Amanda
H* H* L⁻ L%

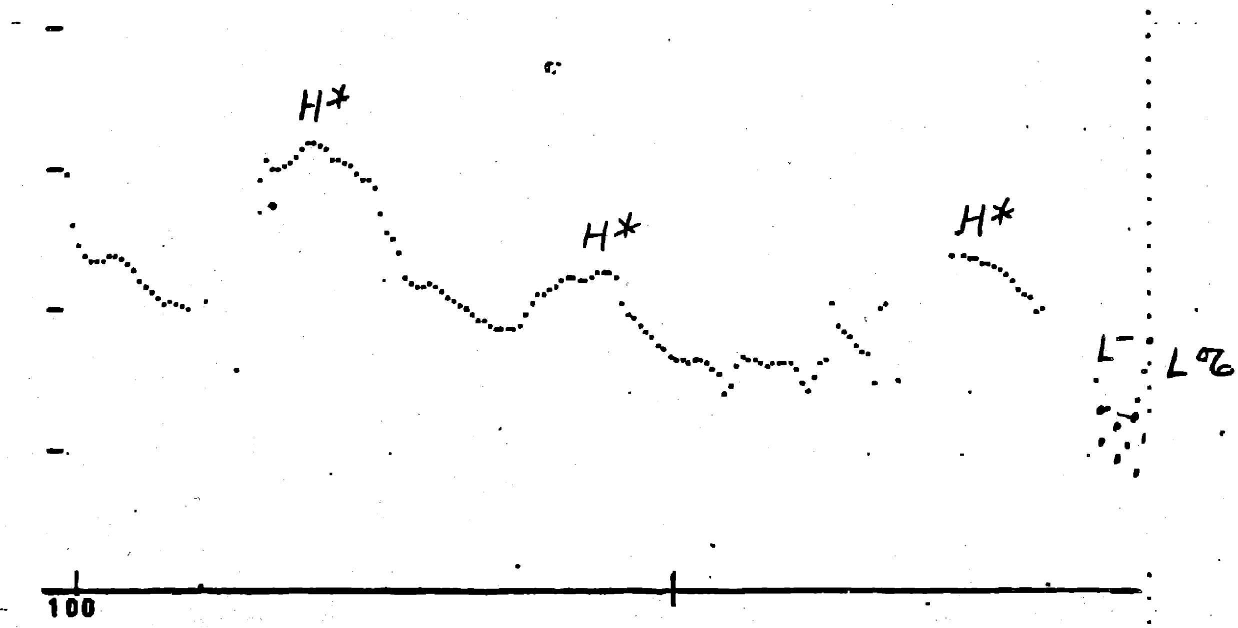
5.18) JBP



It's indexed by the keywords in the abstract.

H*+H⁻ H*+H⁻ H* L⁻ L%

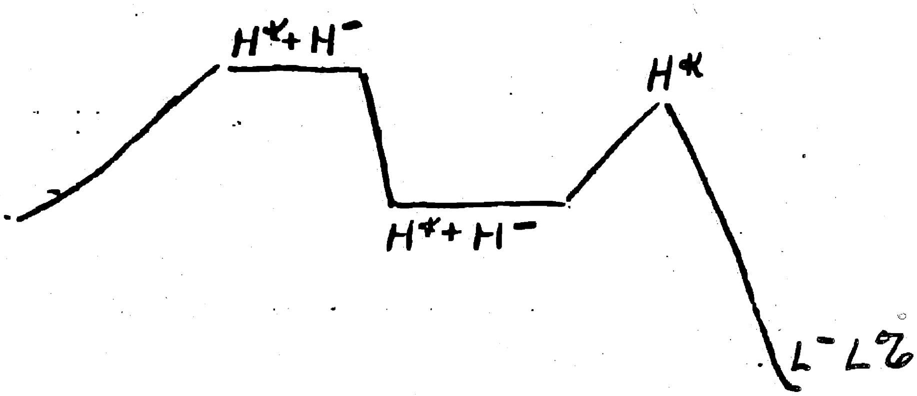
5.19) JBP



Try occasional moderate agitation

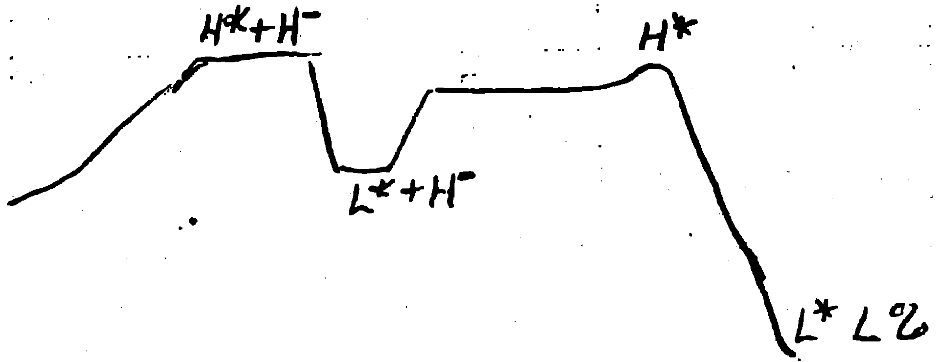
H* H* H* L- L%

5.20)



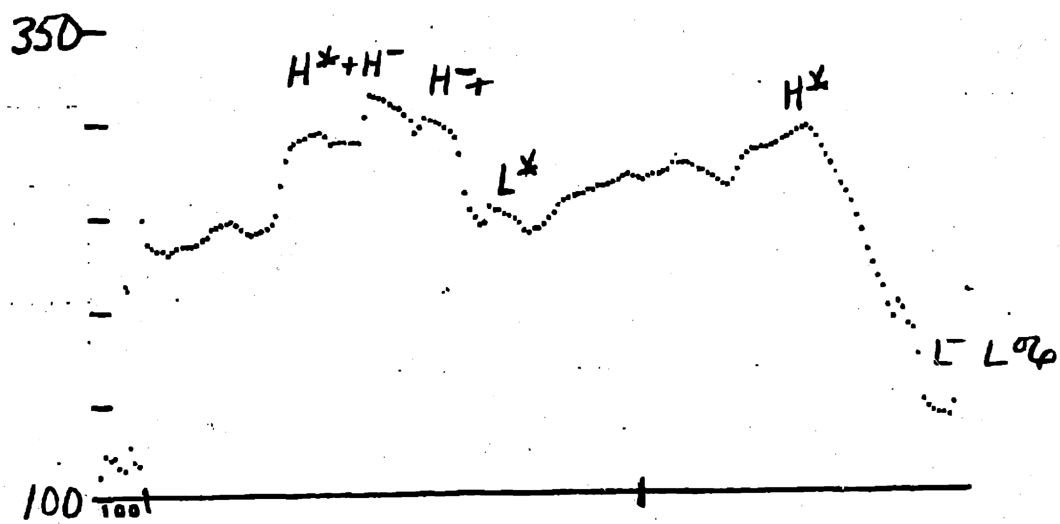
(an impossible pattern)

5.21)



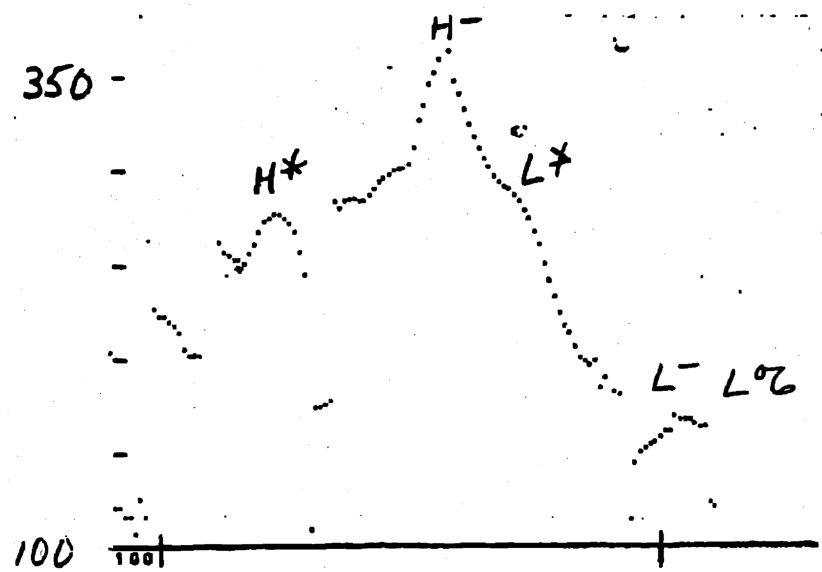
(an impossible pattern)

5.22) JBP



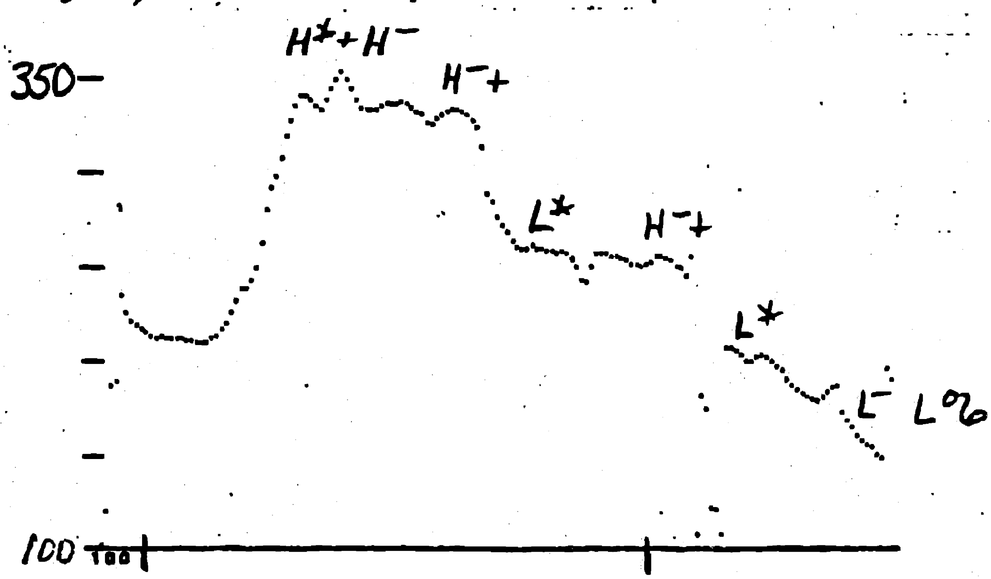
And remember to bring along your raingear
H*+H⁻ H⁻+L* H* L⁻L%

5.23) JBP



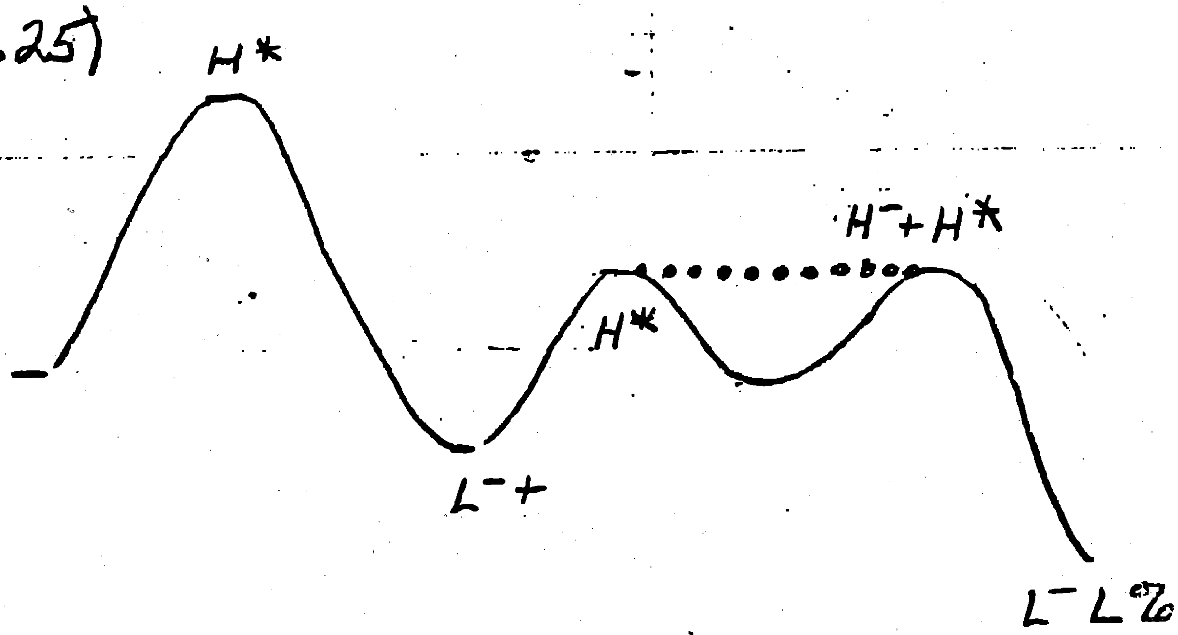
I didn't really believe him.
H* H-+L* L- L%

5.24) JBP



There are many intermediate levels
H*+H- H-+L* H-+L* L- L%

5.25)

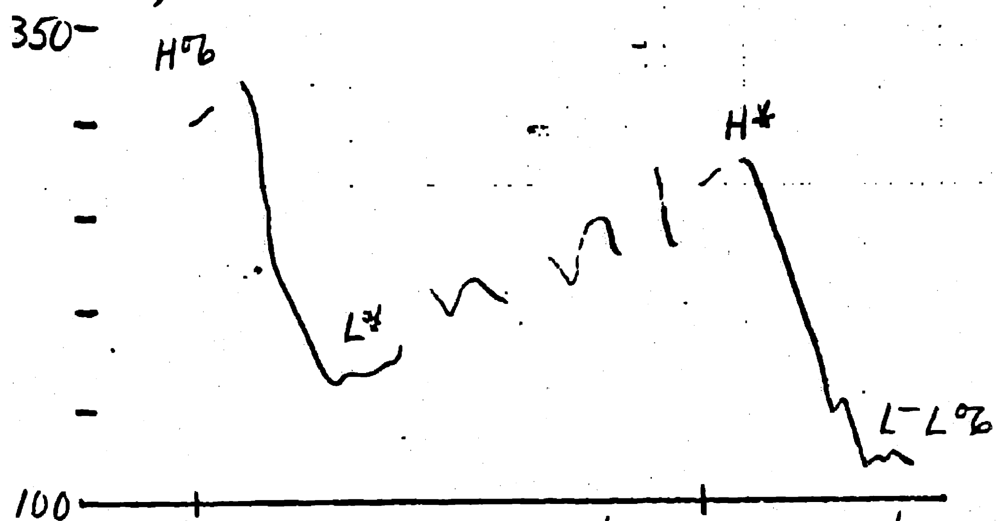


There's a lovely road from Albany to Elmira
H* L⁻+H* H⁻+H* L⁻L%

— a good intonation

···· an odd intonation

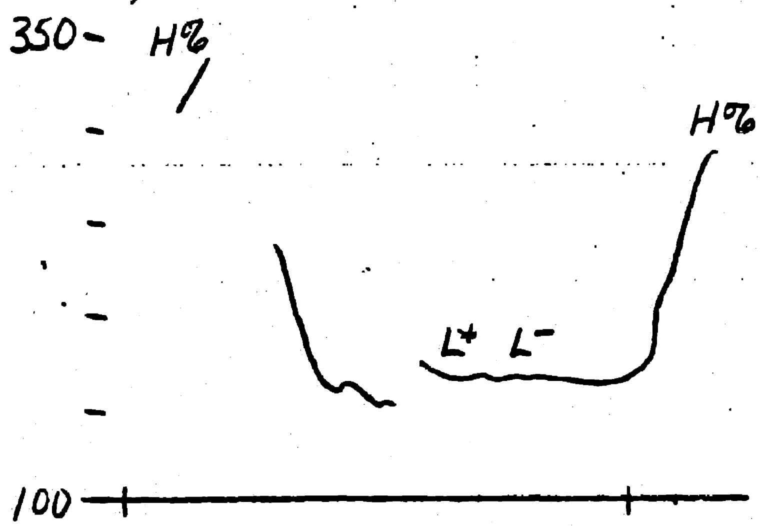
6.1) JBP



That's a remarkably clever suggestion.

H% L* H* L⁻L%

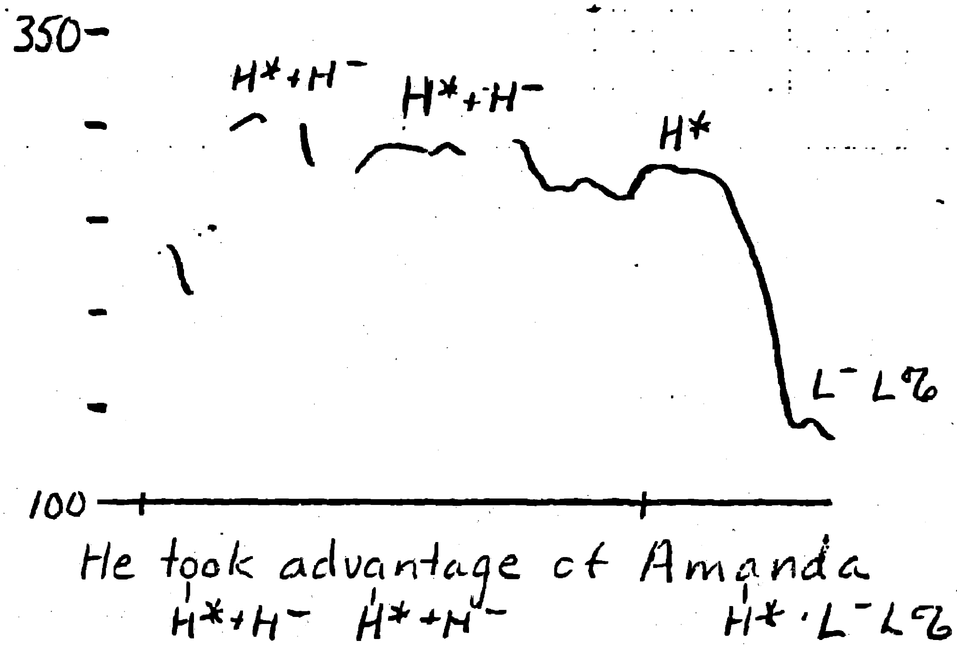
6.2) JBP



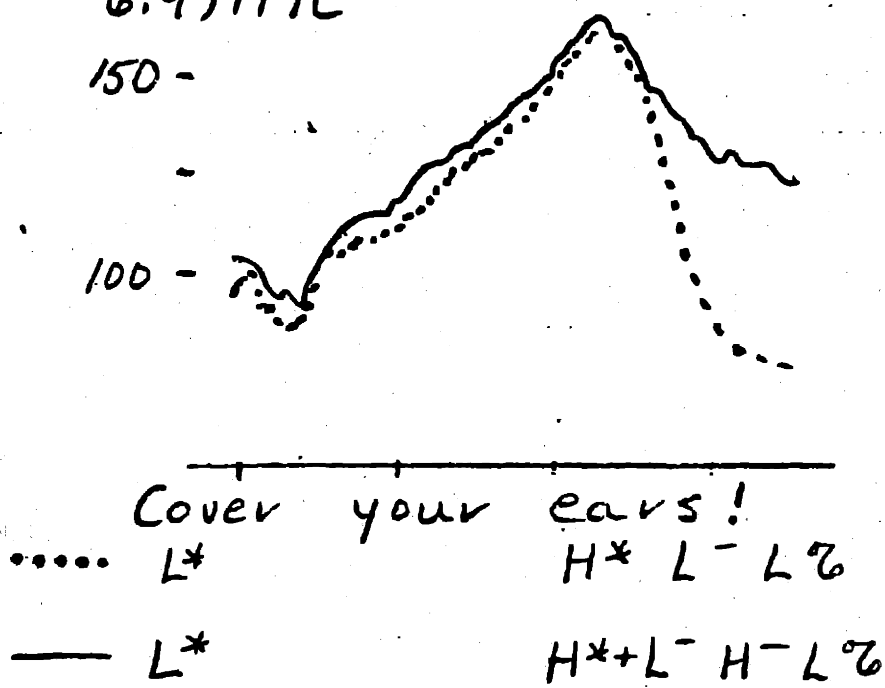
That's not a bowling alley

H% L* L⁻ H%

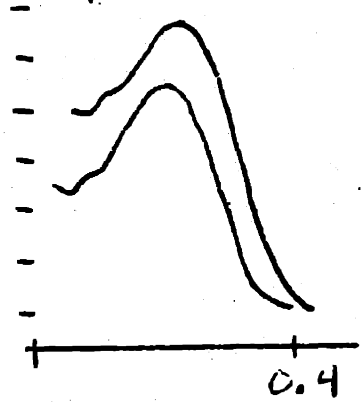
6.3) JBP


350
H*+H⁻
H*+H⁻
H*
L⁻L%
100
He took advantage of Amanda
H*+H⁻ H*+H⁻ H*·L⁻L%

6.4) MYL

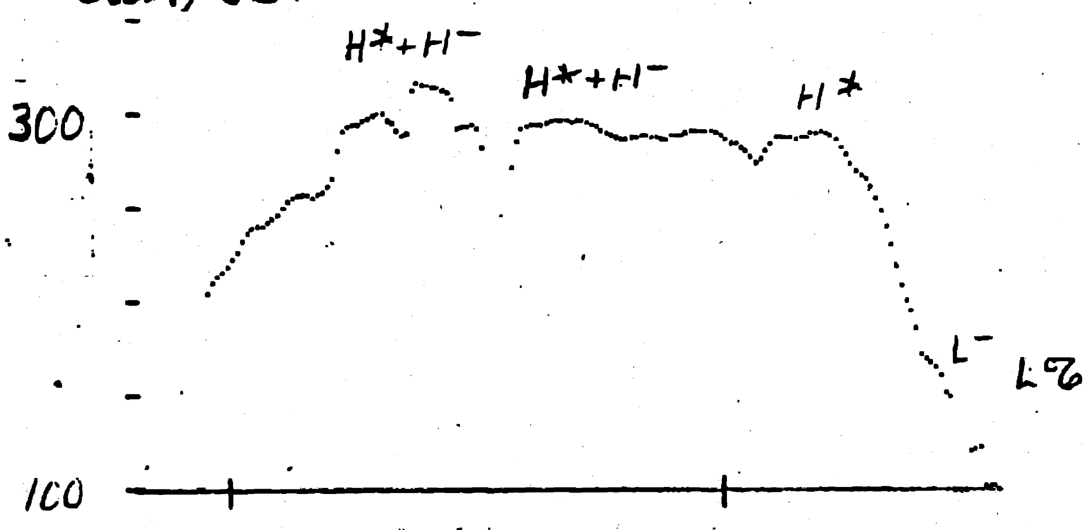

150
100
Cover your ears!
••••• L* H* L⁻ L%
— L* H*+L⁻ H⁻ L%

6.5) MYL



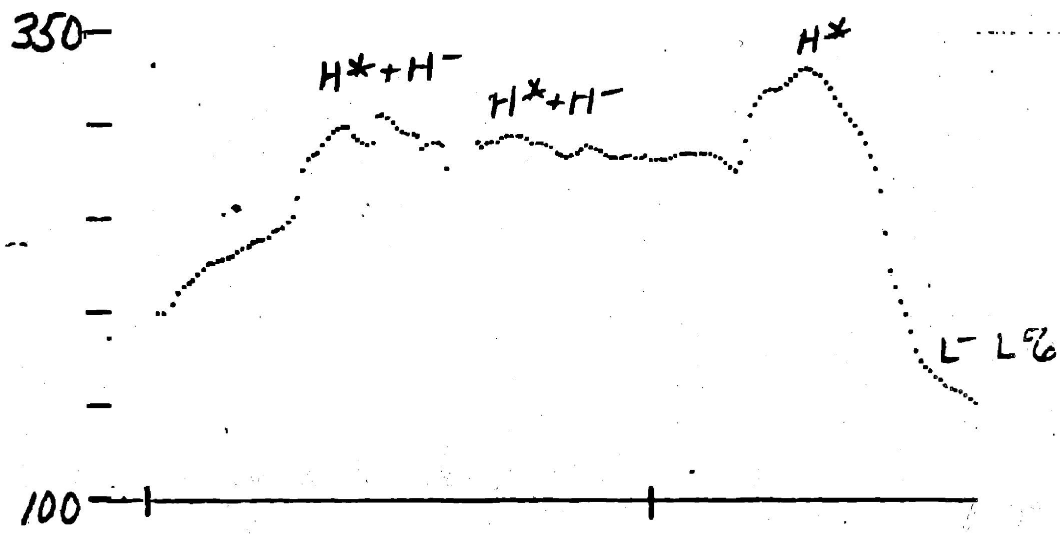
Manny
H* L- L%
in two different pitch ranges.

6.6A) JBP



And remember to bring along your raingear.
H*+H-　　H*+H-　　H* L- L%

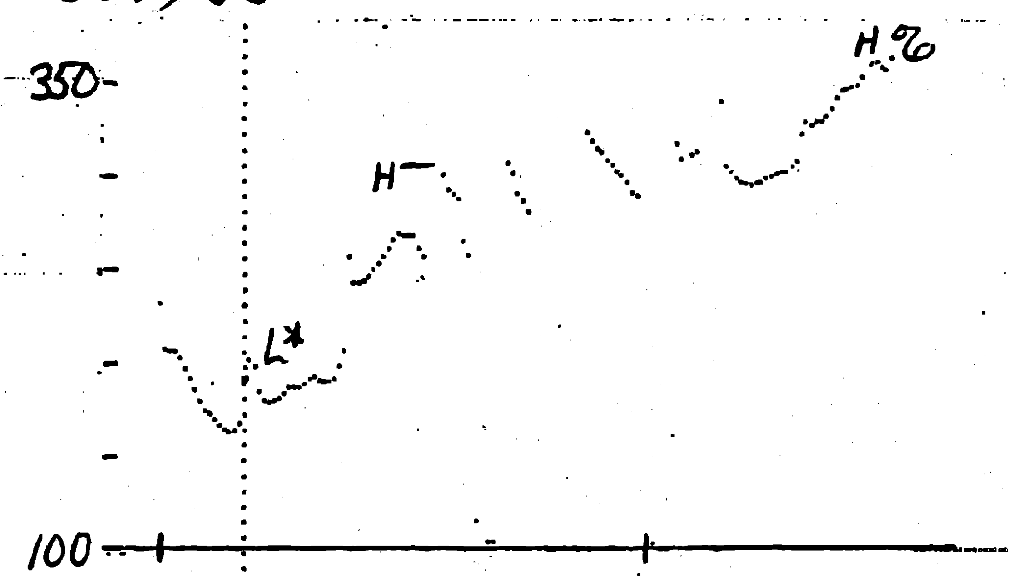
6.6B) JBP



And remember to bring along your raingear

H*+H- H*+H- H* L- L%

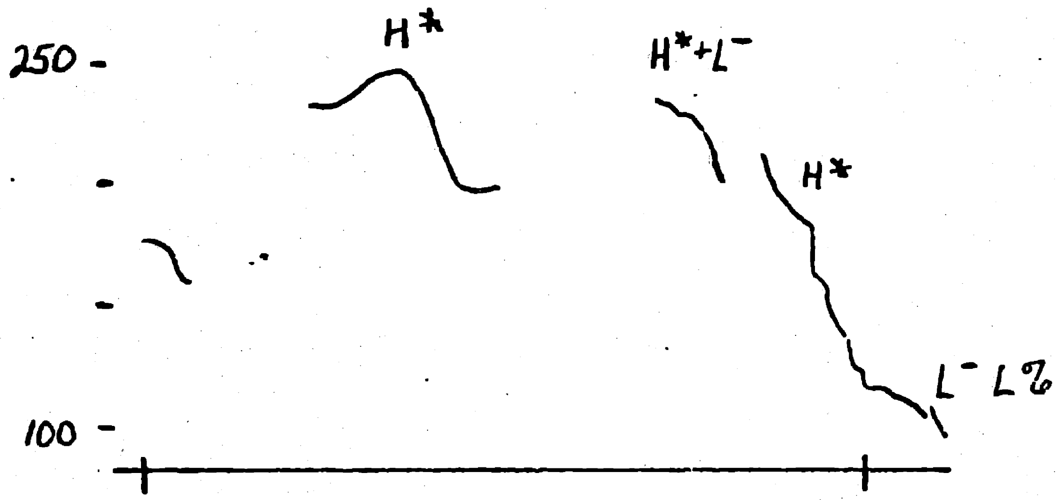
6.7) JBP



Are legumes a good source of vitamins?

L* H- H%

6.8) KXG

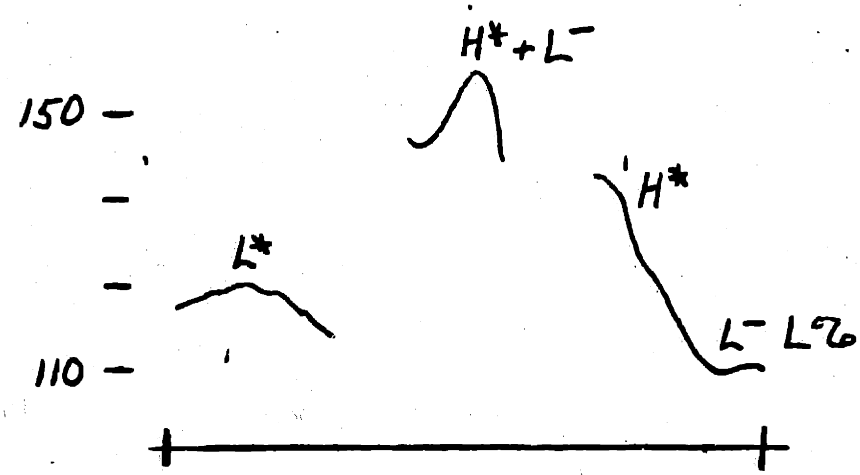


It's spelled with two dots

H* H*+L$^-$ H* L$^-$ L%

(To the ear, "dots" is clearly not deaccented.)

6.9) KS (taken from Maeda 1976)



Almost all farmers

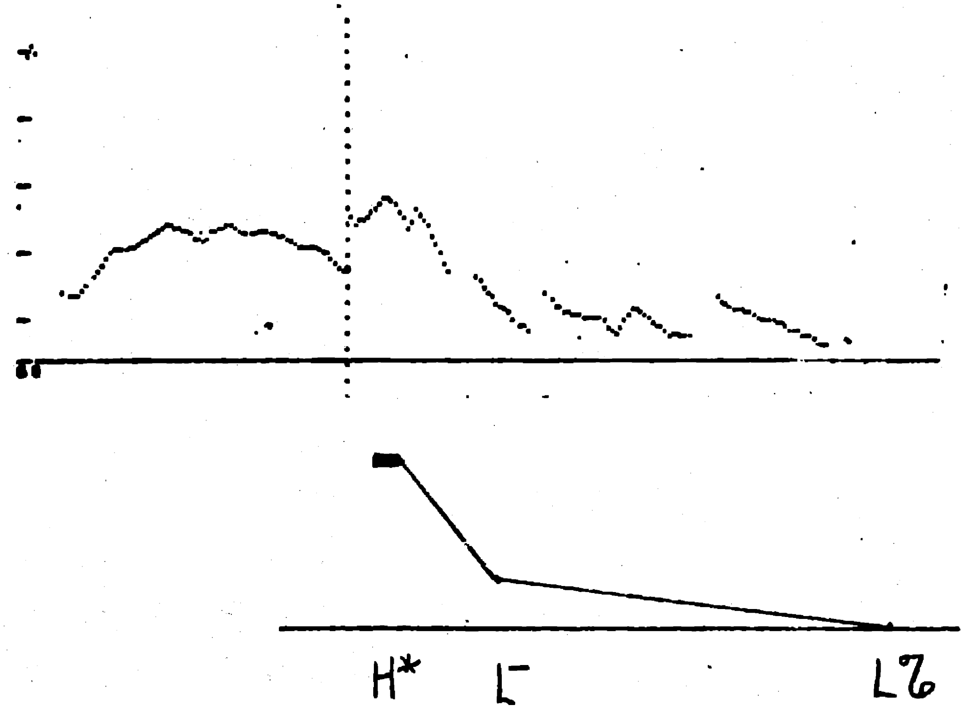
L* H*+L$^-$ H* L$^-$ L%

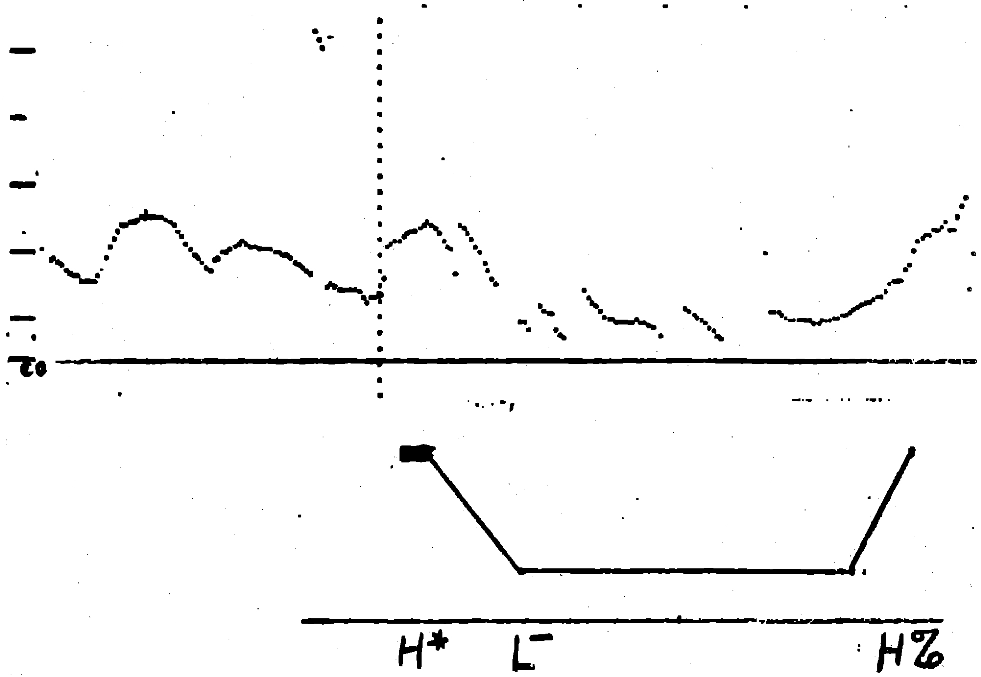
APPENDIX TO THE FIGURES

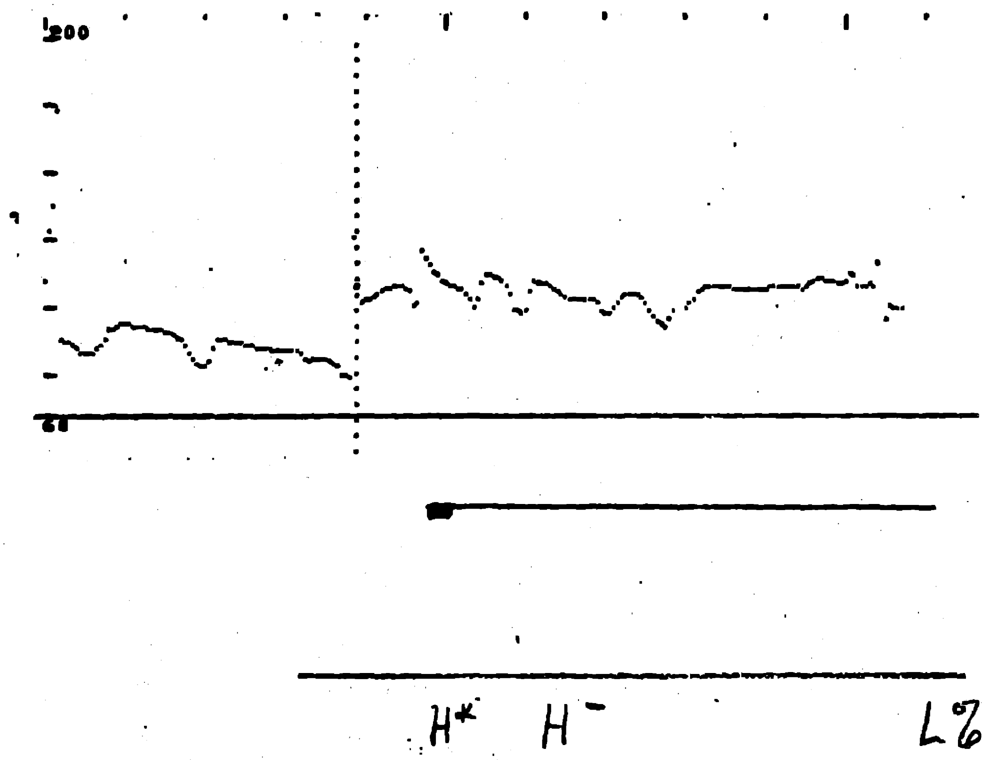
This appendix illustrates all possible combinations of nuclear pitch accent, phrase accent, and boundary tone, and how they are realized by our rules for implementing tones. The phrase used to generate the F0 contours was, "The Uruguayan bulldozer drivers' union." The indicated nuclear accent falls on "bulldozer"; the prenuclear accents vary, since the speaker produced the contours in whatever manner seemed most natural. In some cases, both feet in "Uruguayan" carry an accent.

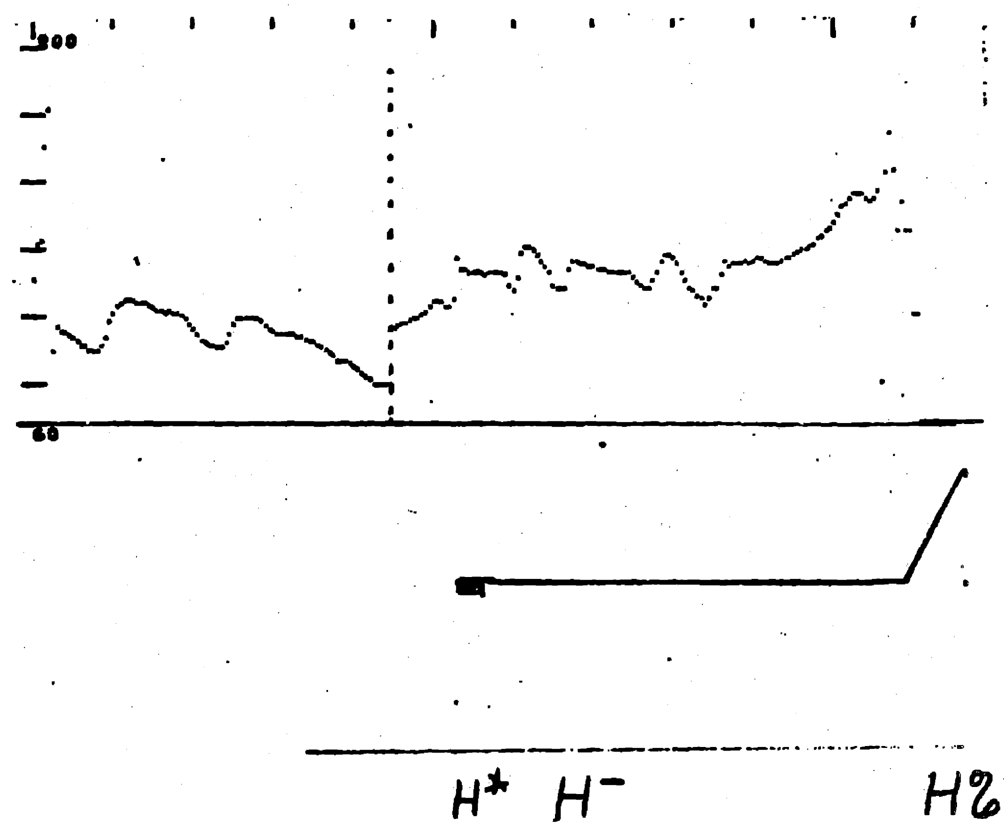
There are 22 different patterns. 6 of the logically possible 28 patterns are omitted because the implementation rules neutralize them with other forms. These are H^*+L^- L^- L% and H^*+L^- L^- H%, which are neutralized with H^* L^- L% and H^* L^- H%, respectively, and all four cases of H^*+H^-, which is neutralized with H^* for either phrase accent.

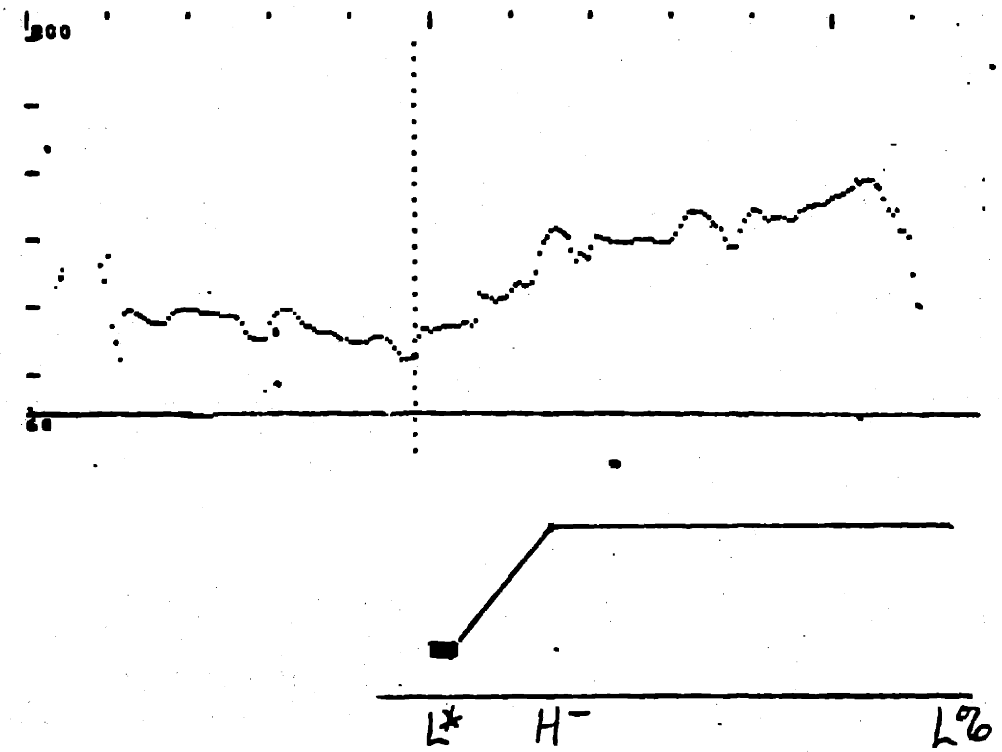
In each F0 contour, the vertical dotted line is located on /b/ in "bulldozer". In the schematized patterns, a bar marks the location of the syllable with nuclear stress.

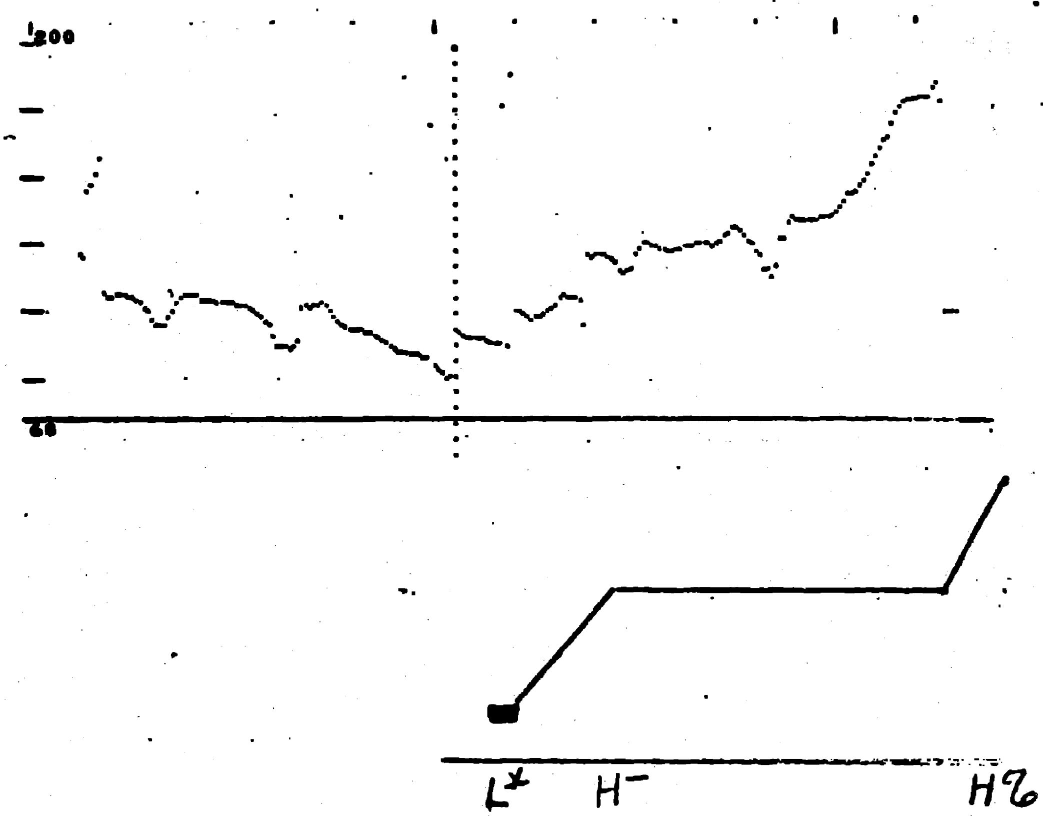

H*
L⁻
L%

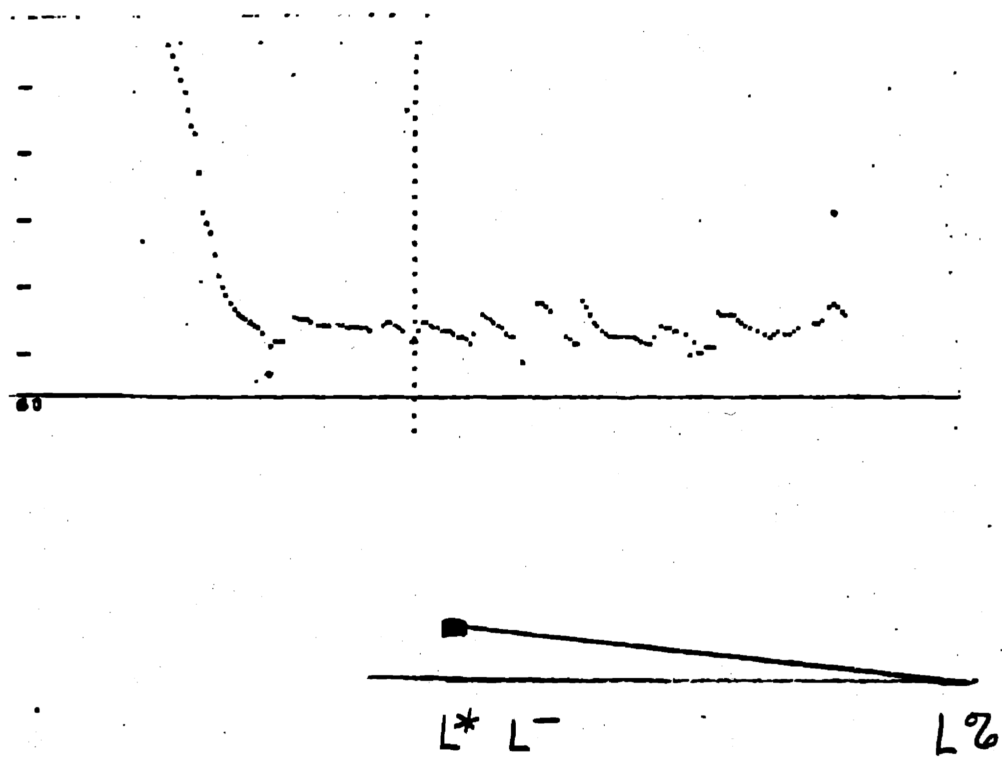

H*
L⁻
H%

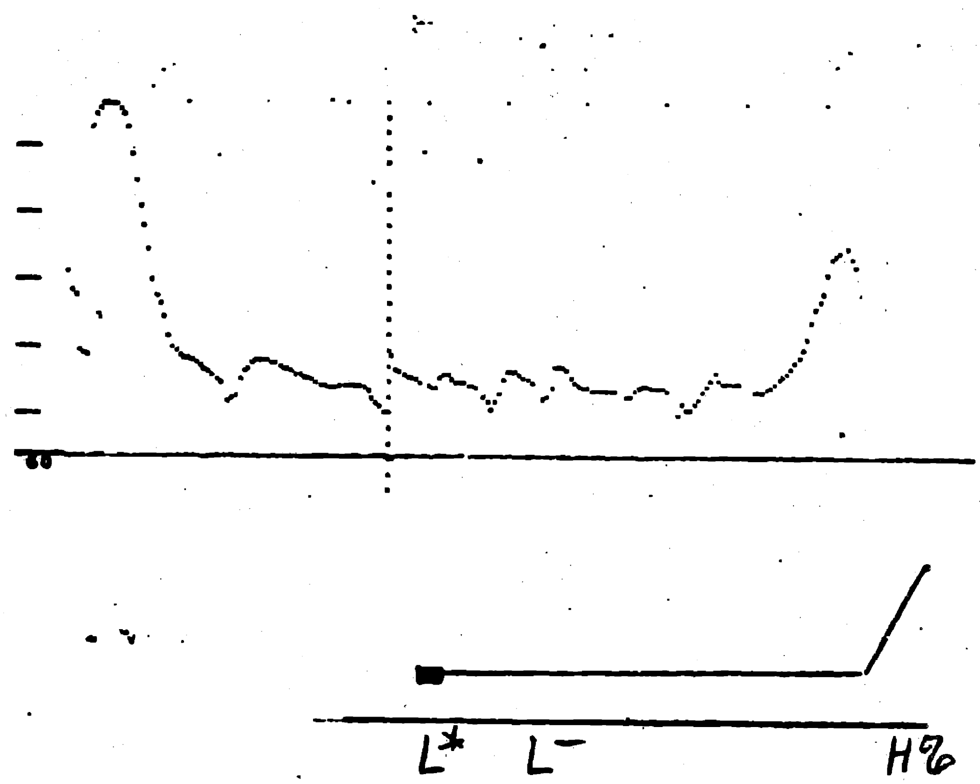

200
60
H* H- L%

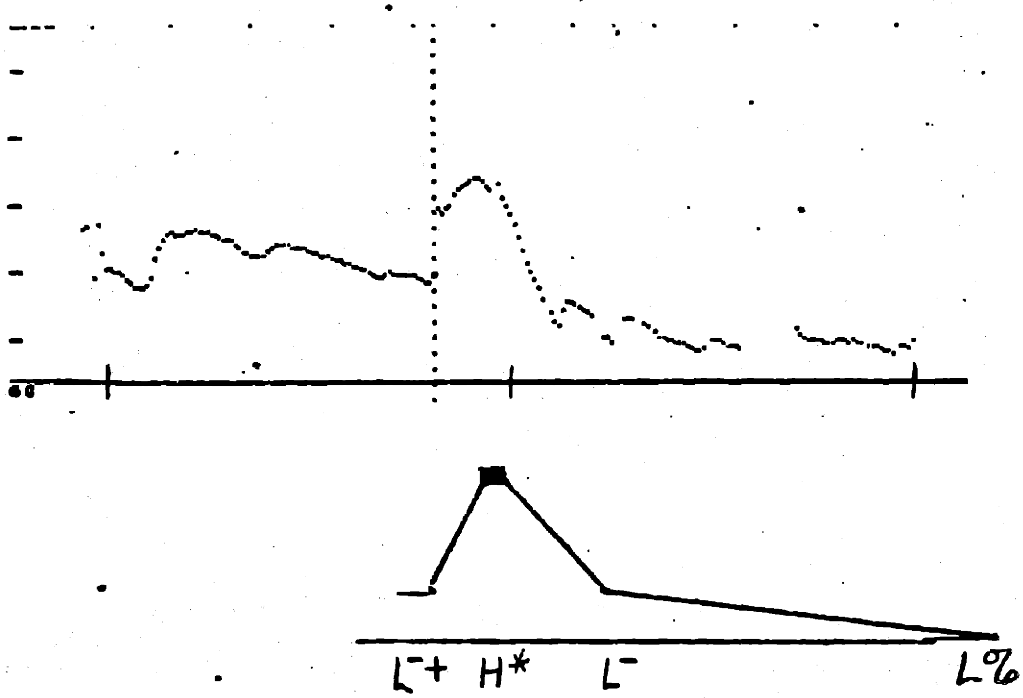

200
60
H* H- H%

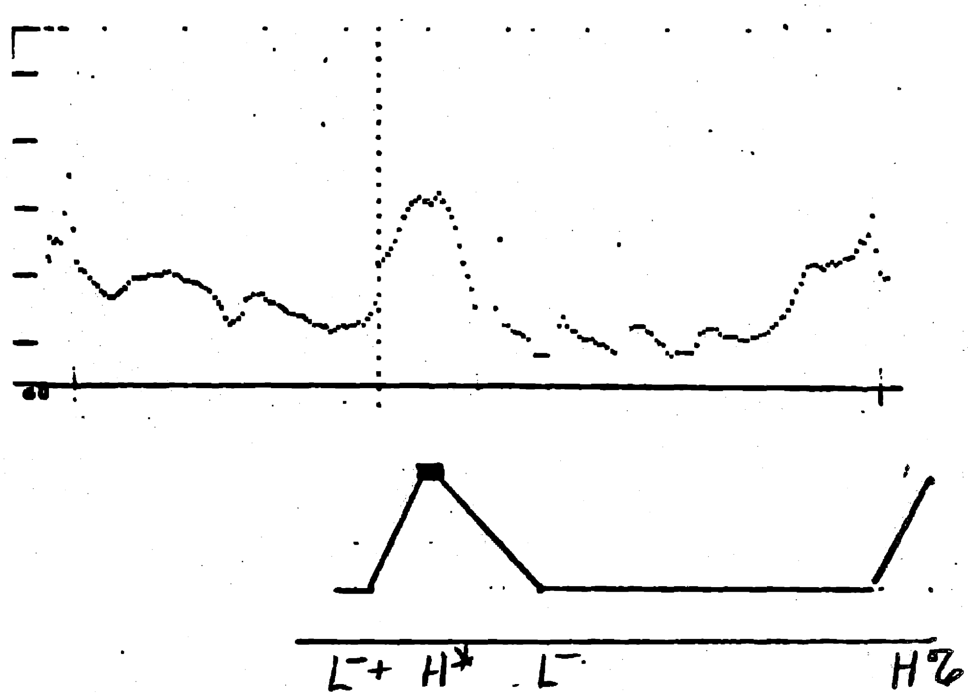

200
60
L* H- L%

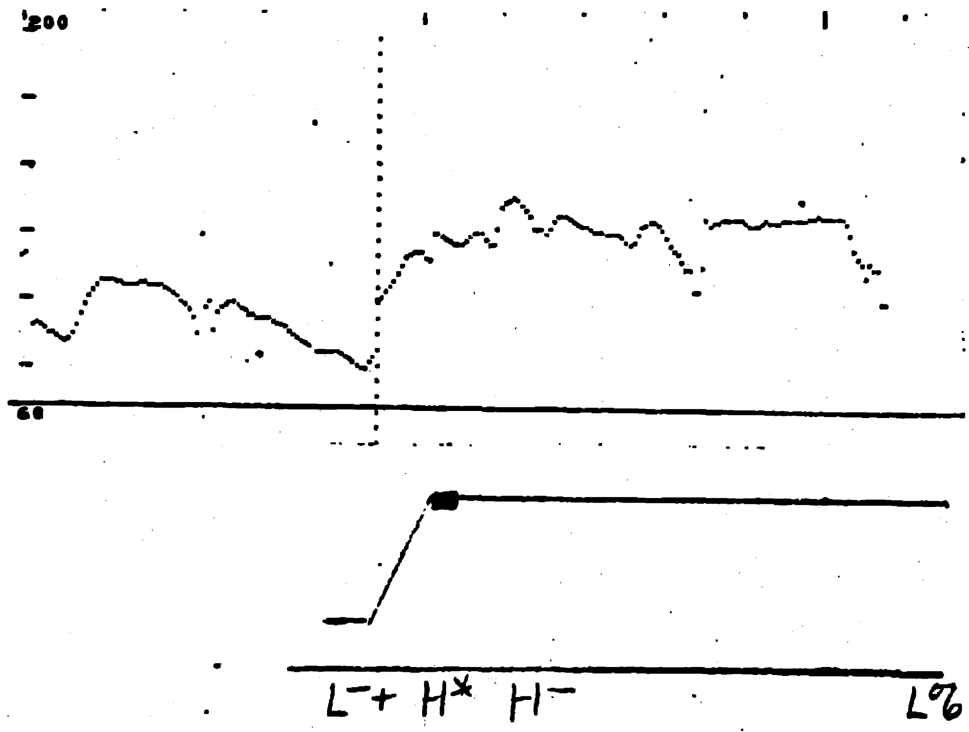

200
60
L* H- H%

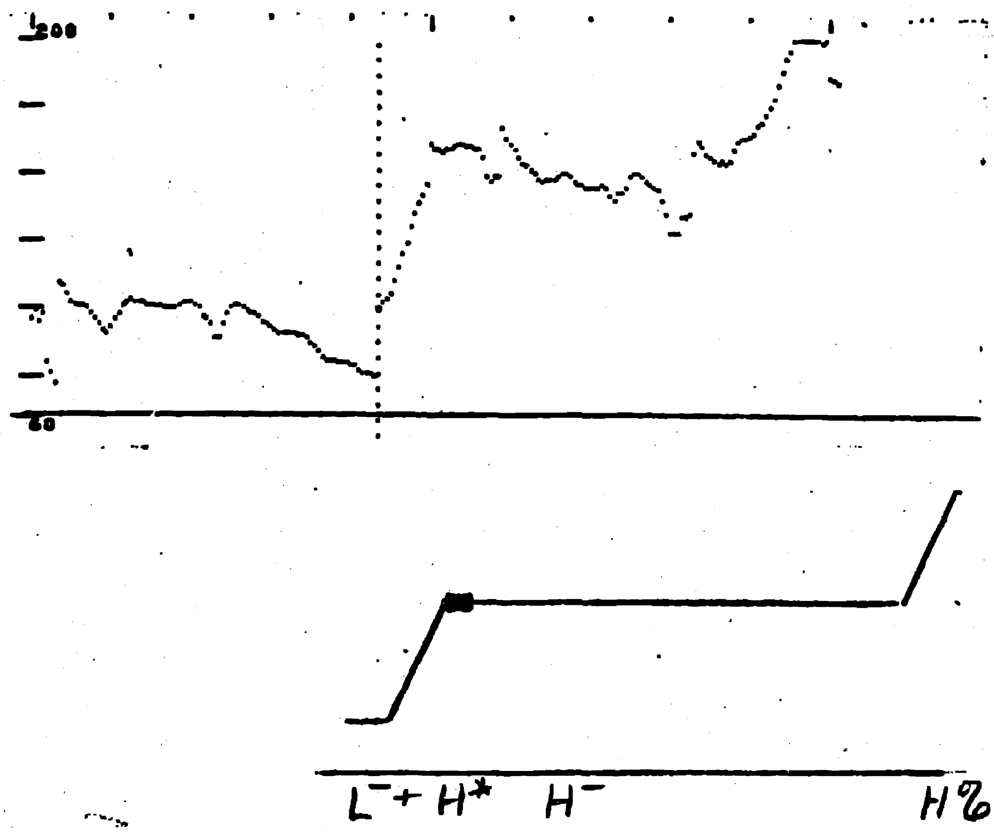

L* L⁻
L%

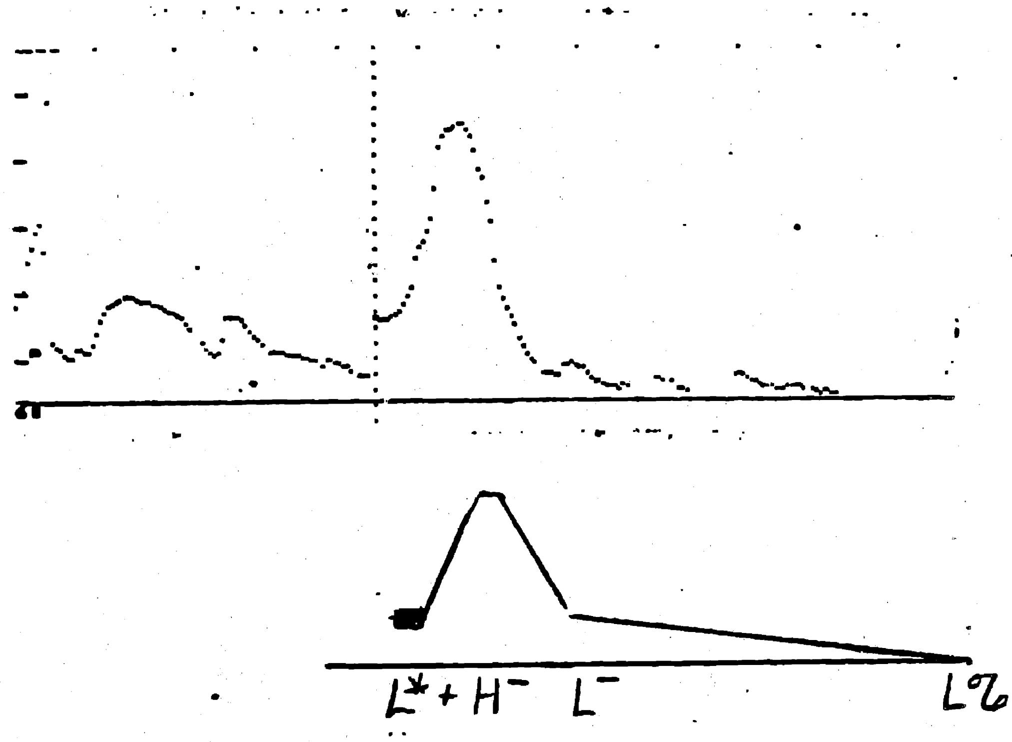

L* L⁻
H%

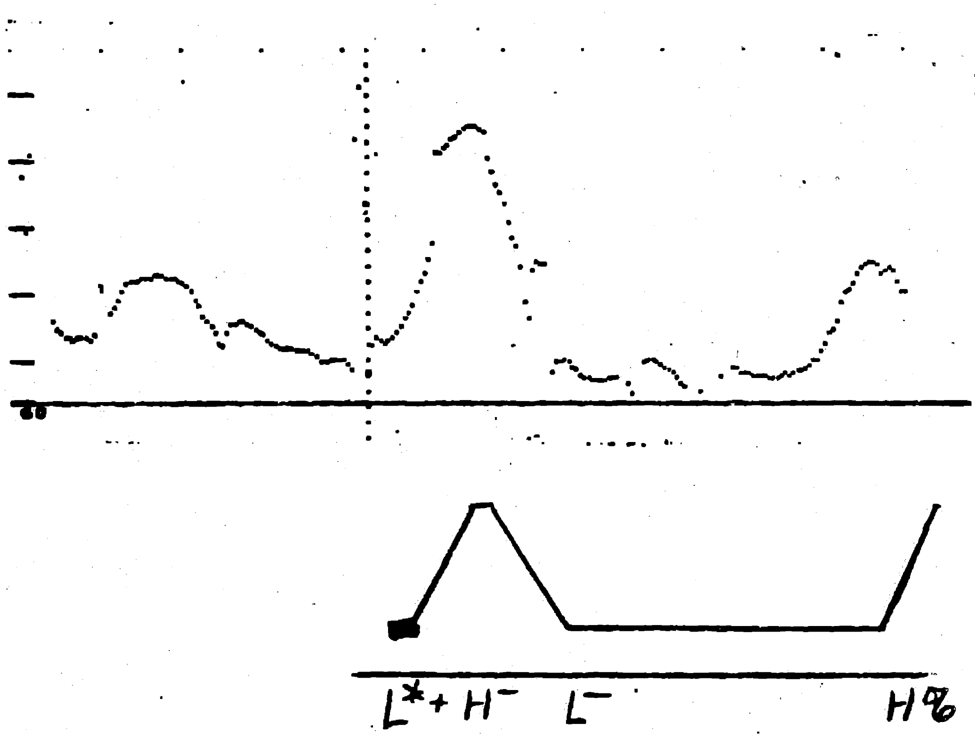

L⁻+ H* L⁻ L%

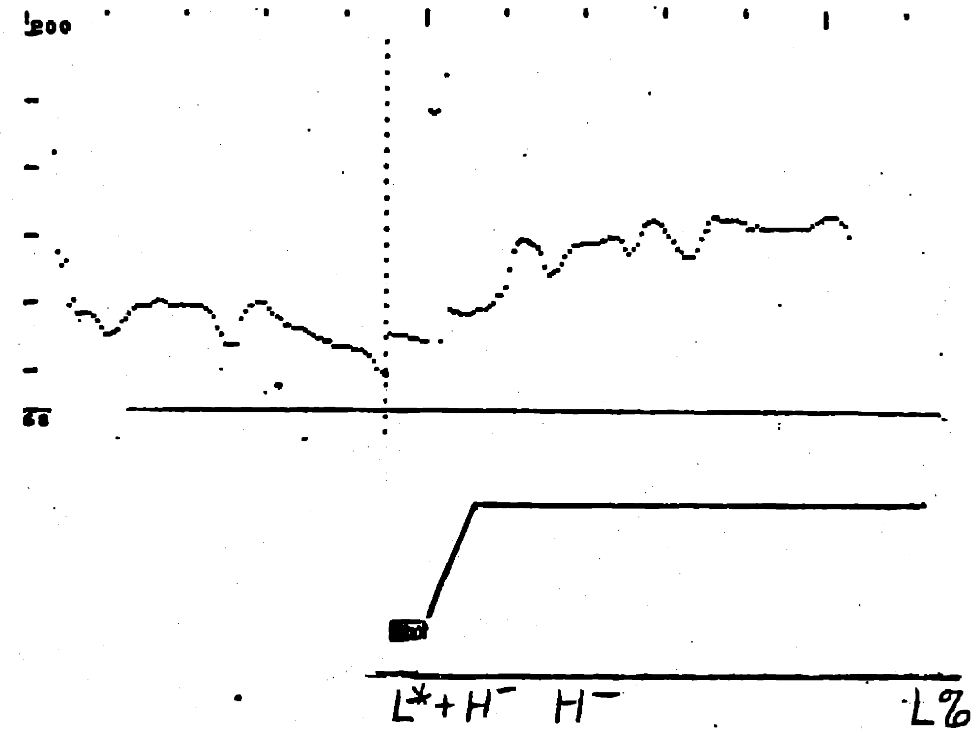

L⁻+ H* L⁻ H%

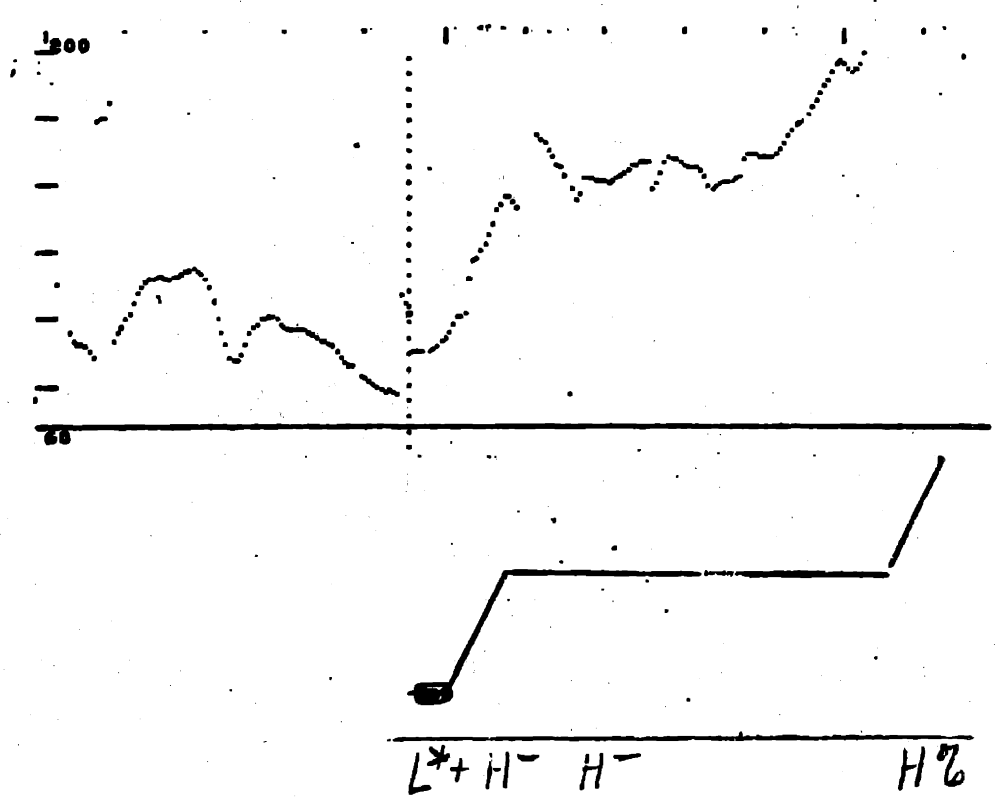

200
60
L⁻+ H* H⁻ L%

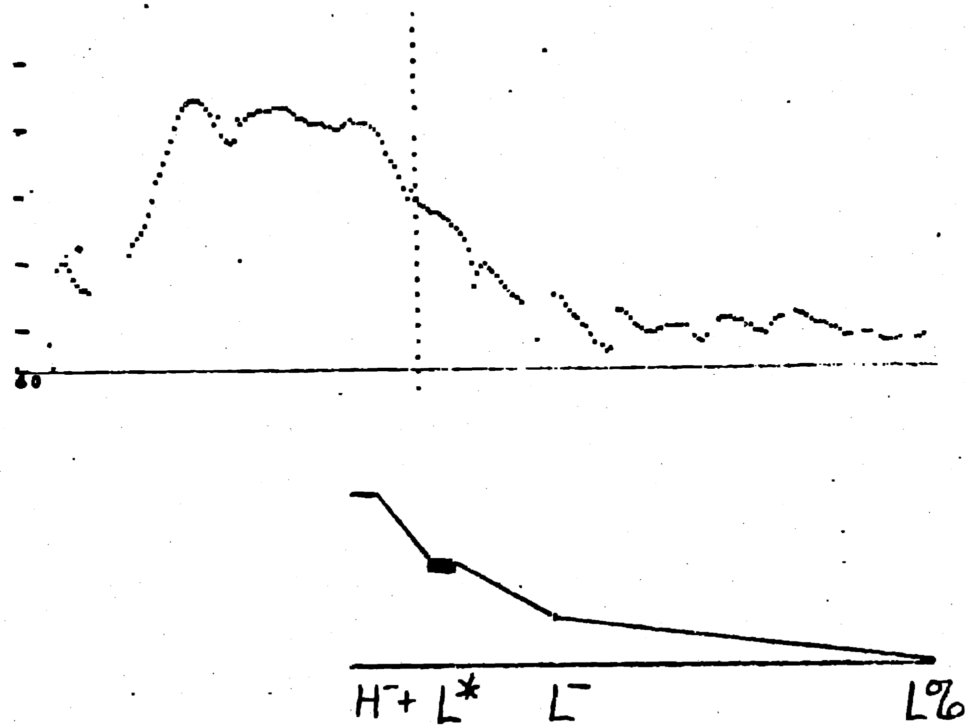

200
60
L⁻+ H* H⁻ H%

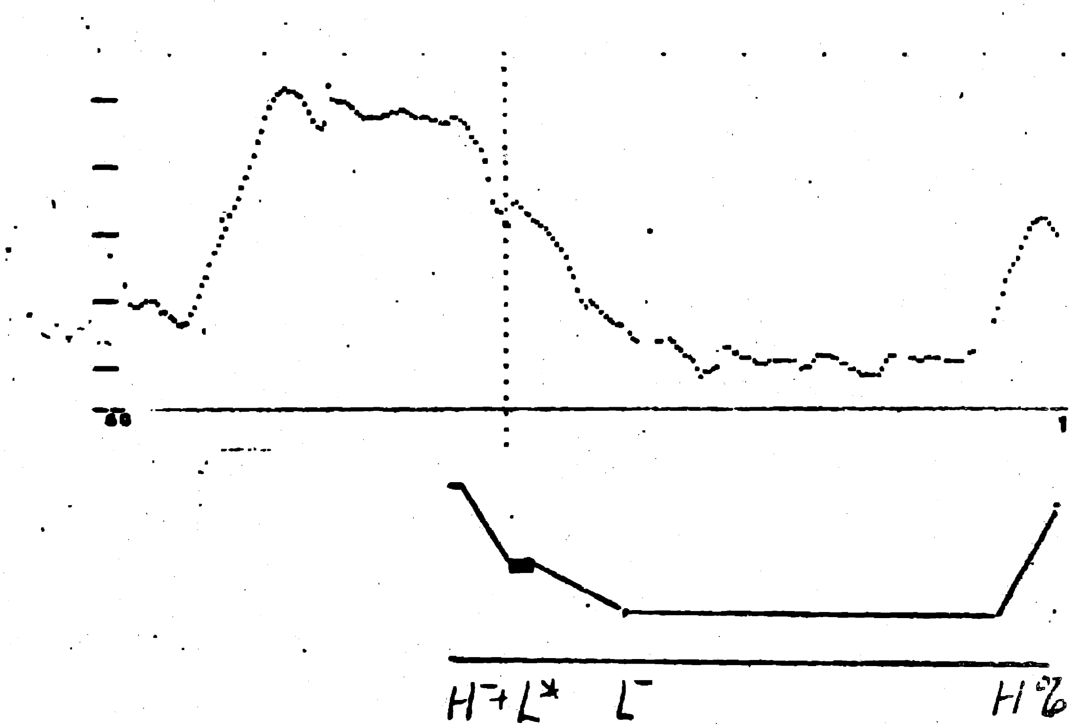

L*+H⁻ L⁻ L%

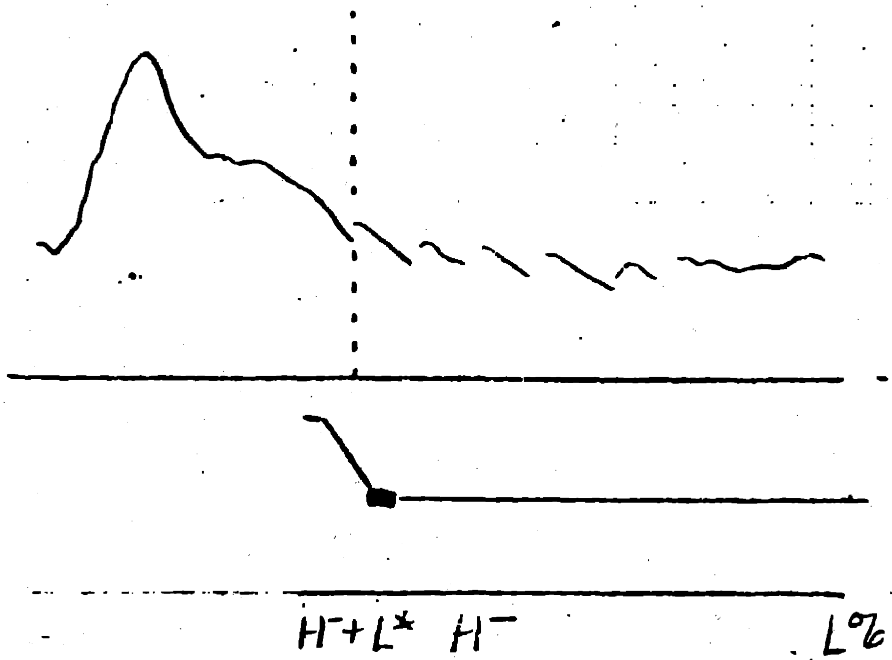

L*+H⁻ L⁻ H%

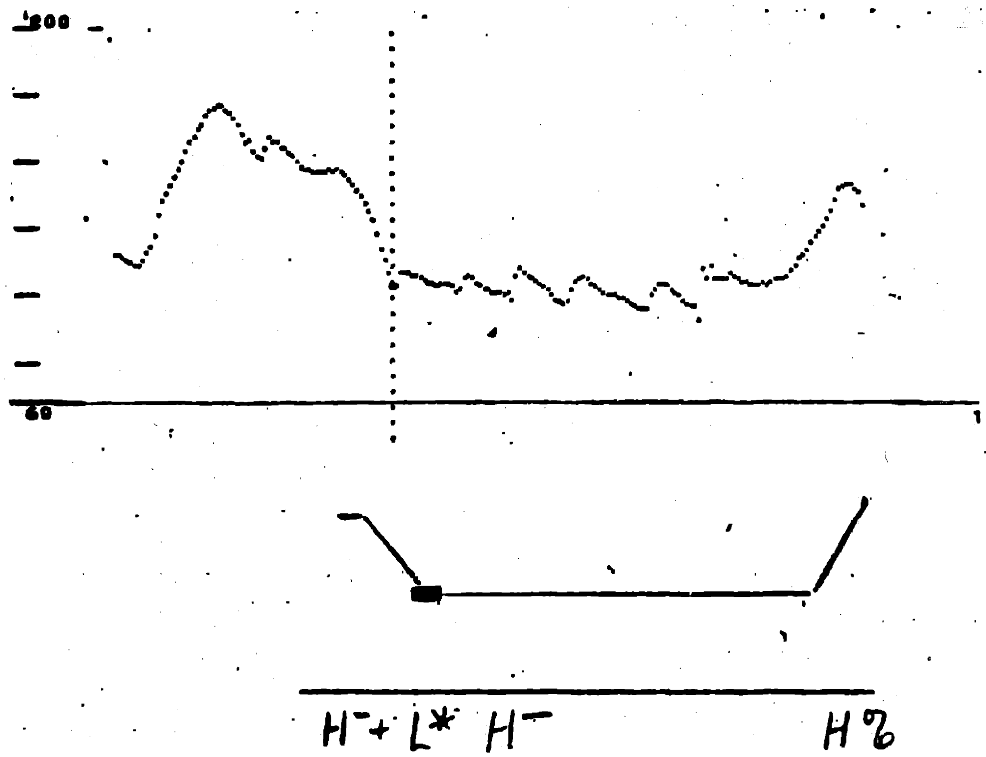

200
L*+H⁻ H⁻ L%

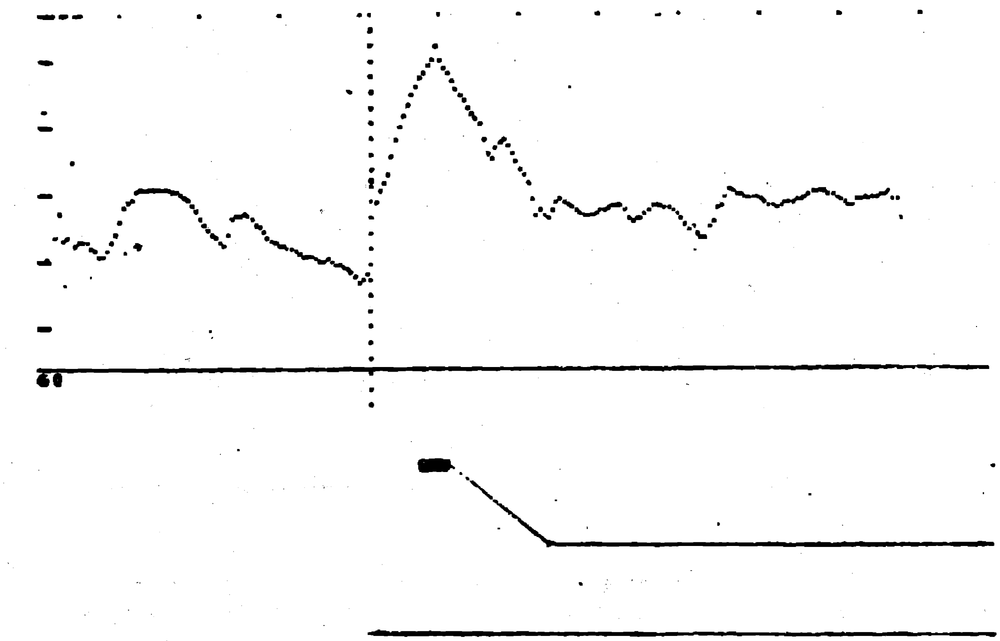

200
L*+H⁻ H⁻ H%


H- + L* L- L%

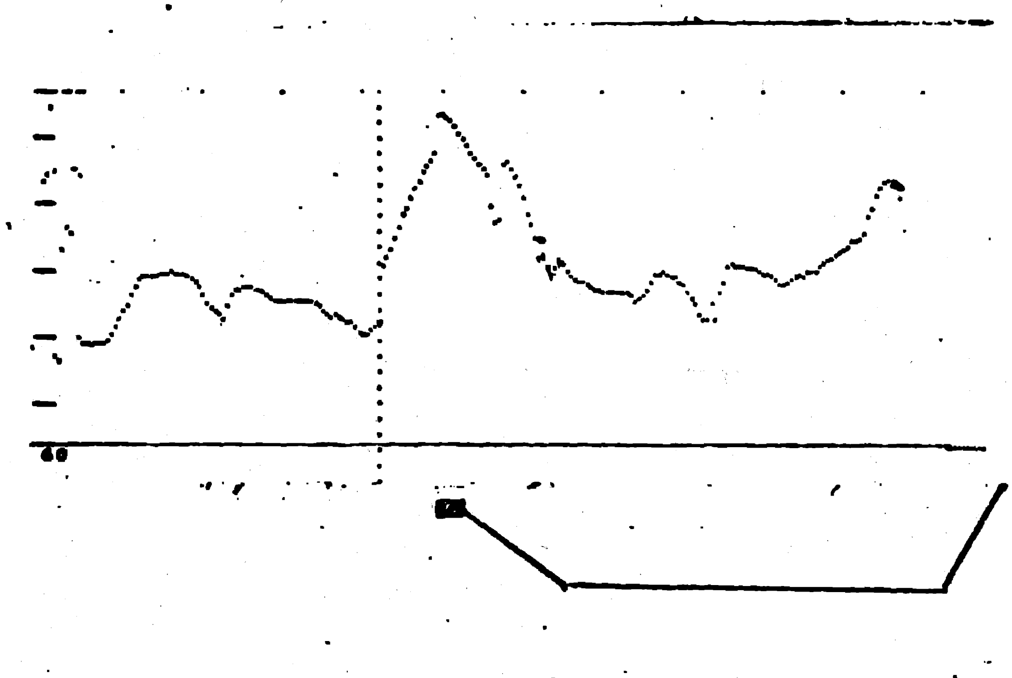

H- + L* L- H%


H-+L* H-
L%


H-+L* H-
H%



H^*+L^- H^- $L\%$



H^*+L^- H^- $H\%$