HOW MANY RISE-FALL-RISE CONTOURS?

Janet B. Pierrehumbert and Shirley A. Steele

AT&T Bell Laboratories,
600 Mountain Ave, Murray Hill, NJ 07974
U.S.A.

ABSTRACT

This paper reports an experimental study of the rise-fall-rise intonation patterns of English. Variants of the rise-fall-rise contour differ from each other in the alignment of the F0 valley and peak with the text. The phonological analysis of this difference is controversial and is part of a wider debate on the function of category and continuum in intonation. Variants of the rise-fall-rise pattern have been viewed as either (1) falling into two categories, each with a distinct pragmatic meaning, or (2) occurring along a continuous dimension of peak delay and pragmatic meaning. This issue was addressed by examining how speakers imitate stimuli varying continuously in the alignment of the F0 rise-fall.

In the stimuli for the experiment, the alignment of the F0 rise-fall was varied in small steps by using LPC coding and resynthesis. Subjects heard the stimuli in randomized order and imitated what they heard. Peak delays in the responses were found to cluster in two groups, thus differing systematically from the peak delays in the stimuli. This result is readily explained by a model with two categories.

1. INTRODUCTION

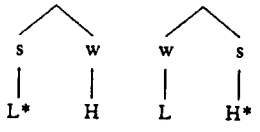
1.1 Topic

In this paper, we report an experimental investigation of the rise-fall-rise intonation patterns of English. Variants of the rise-fall-rise contour differ from each other in the alignment of the F0 valley and peak with the text. Figure 1 exemplifies the F0 contour of this pattern with an early peak and Figure 2 shows a longer peak delay. (In both Figures, a vertical line shows the location of the [m]-release.) The pattern in Figure 2 conveys speaker incredulity or uncertainty (see Ward and Hirschberg [7], [8]), while that shown in Figure 1 typically marks contrast or correction.

The phonological analysis of the rise-fall-rise intonation pattern is controversial and is part of a wider debate on the function of category and continuum in intonation. Peak delay variants of the rise-fall-rise pattern may be viewed as either (1) falling into some number of categories, each with a distinct pragmatic meaning, or (2) occurring along a continuous dimension of peak delay and pragmatic meaning. A formulation of the first view is found in Pierrehumbert [5] and a statement of the second in Gussenhoven [1].

According to Pierrehumbert, there are two different rise-fall-rise patterns, differing in how the low-high pitch accent is aligned with respect to the stressed syllable. In one, the high (H) is aligned with the stressed syllable (Figure 1). In the other, the low (L) falls on the stressed syllable (Figure 2). Using a diacritic "*" to represent alignment with the stress, the two patterns are transcribed L+H* and L*+H and are called "bitonal accents" because two tones are required to describe the accent. In both patterns, the fall-rise after the accent is explained by a L H sequence which marks the end of the phrase.

In Pierrehumbert's theory, the "*" diacritic for tones is analogous to stress for syllables. In metrical stress theory, as laid out in Liberman and Prince [4], the stress pattern of a word like "Peter" has a relatively strong syllable followed by a relatively weak syllable. In "repeat" the strength relation is reversed. Pierrehumbert's bitonal accents are treated in the same way, with a strength relation among the tones, as illustrated below:



The single tone accents have a status corresponding to that of stressed monosyllables, such as "Pete".

This approach contributes to a broader picture in which tones participate in a hierarchical organization, which controls their alignment and phonetic realization. This broad picture is first proposed in Liberman [2] and is amplified and revised in Pierrehumbert and Beckman [6].

Gussenhoven asserts that English does not have a binary distinction, but a continuous dimension of peak delay. He suggests ([1], p. 218) that while there may be a preferred ("ideal") position along this continuum, a wide range is possible. His treatment of peak delay is thus analogous to the treatment of continuously variable overall pitch range in Liberman and Pierrehumbert [3]. Peak delay in pitch accents is treated as a paradigmatic feature, while word stress is viewed as a matter of metrical organization.

Our aim in the present work was to empirically investigate the category/continuum question in the rise-fall-rise intonation contour.

1.2 Method

In our experiment, subjects heard and imitated randomized rise-fall-rise contours constructed along a continuum of peak delay. If subjects perceive a continuum, the response peak delays should be approximately continuous. (A preferred peak position might cause responses to stimuli at the extremes to drift towards the center). If subjects hear categories, responses should cluster into discrete groups.

This experimental method is a variant of the paradigm familiar from studies of categorical perception of speech segments. Our study is, to our knowledge, the first application of such methods to the study of syntagmatic features. We have used an imitation task rather than the more commonly used labelling and discrimination tasks, because we were not concerned with separate analysis of production and perception systems. In most categorical perception studies, the linguistic analysis was relatively uncontroversial (e.g., English /b/ versus /p/); at issue was the status of linguistic description in the psychological system. In our study, we looked for evidence about the system of linguistic analysis; the relationship of any categories to perceptual or articulatory systems is a matter for later research.

2. EXPERIMENTAL PROCEDURES

2.1 Stimuli

The stimuli were versions of the phrase, "only a millionaire," in which the location of the F0 peak was incrementally moved from a relatively early to a relatively late position in the accented word. This was done by recording a natural production of the sentence and using LPC coding and resynthesis to produce a systematic set of variants. In both the original recording and in subjects' imitations of it, main stress fell on the first syllable of the word "millionaire", not the last. This is an acceptable pronounciation in American English and subjects reported no difficulty using it.

The shape of the rise-fall-rise pattern in the stimuli was established by making a piece-wise linear approximation to the rise-fall-rise pattern of the original recording. Peak positions varied between 35 and 315 msec from the end of the [m] in "millionaire", by 20 msec increments. As the peak was shifted, the durations of the rise and fall were kept constant. In the stimulus with the greatest peak delay, the peak occurred just before the end of the [n]. These bounds were established by asking several naive listeners to evaluate the naturalness of the stimuli. Stimuli at the ends of the continuum which the listeners felt to be unacceptable in English were eliminated.

The particular phrase used was chosen for two reasons. First, it is composed entirely of sonorants, thus avoiding devoicing in the F0 contour and minimizing consonantal effects in the stressed syllable. Second, its pragmatic interpretation could be sensibly altered by variation in peak delay. The early peak variant could be used to assert that the speaker does not feel very rich, while the late peak variant would be appropriate for an incredulous rejoinder.

Note that this meaning distinction could be viewed as either categorical (assertion versus incredulity) or continuous (along a dimension of degree of speaker commitment). The meaning difference was not discussed with the subjects; thus, using a sentence with two potential pragmatic interpretations did not prejudice the experimental outcome.

For each subject, data was collected in at least two sessions, using a real time data collection program. Subjects were told that they would hear a series of aural prompts, and that the phrase was always the same but the intonation varied. They were asked to listen carefully to each token and then imitate what they had heard. If subjects were not satisfied with a particular response or wanted to hear the prompt again, they could tell the experimenter, and the token would be repeated. No time limit was imposed for responses.

2.2 Randomization

The 15 versions of the prompt were randomized in blocks; then in each of two sessions, 15 different randomized blocks were presented to the subject, for a total of 225 tokens per session or 450 tokens in all. Thus, each of the 15 tokens was repeated 30 times.

2.3 Subjects

The subjects were five native speakers of American English, two females and three males. Four of the five were naive about the purpose of the experiment. The subjects were: DTT, a software engineer; HLT, a psychology research assistant; RLB, an opto-electronics processing engineer; SAS, one of the authors; and TWB, a high school student.

2.4 Measurement

The measurement of primary interest is peak delay, defined as the difference between the time of the F0 peak and the time of the [m]-release. This was found by examining displays of the F0 contour and waveform of each response. Time points were established for the release of the [m] into the vowel, for the F0 peak, for the implosion of the [n], and for the F0 minimum preceding the peak. The transitions into and out of the nasal consonants could in general be found with great accuracy, due to the abrupt change in the waveform occurring at these points. The F0 peaks were fairly narrow, and thus their location provides a good index of the location of the H tone. In cases where several time points shared the same maximum value, the earliest was selected. A few utterances had to be eliminated from the data set because the F0 at crucial points could not be measured. There were no more than four such utterances per subject. The data summaries for each stimulus represent in every case at least 28 responses.

3. RESULTS

3.1 Predictions about peak delays

In order to make the discussion of the data more transparent, let us first explain some idealized experimental results for different models of the intonation system.

First, consider a continuum model in which peak delay is continuously variable and all possible peak delays are equally preferred. In this model, the subject should, on the average, faithfully reproduce the peak delays in the stimulus. A plot of response peak delay against stimulus peak delay whould thus follow the line $y = x$. The overall distribution of responses, being the sum of the individual distributions, should be broad and unimodal, approaching the rectangular distribution of peak delay in the stimulus set.

Next, consider a continuum model in which peak delay is continuously variable but a central value is preferred. In this model, the responses tend to shift towards the center. Individual distributions are likely to show a dependence of shape on stimulus position in the continuum. The responses to stimuli at the ends of the continuum are likely to include tokens from the center of the continuum. But stimuli in the center (exhibiting the preferred form) would be less likely to elicit responses from the ends. Overall distribution would still be unimodal.

In a model with two categories, there are two preferred values of peak delay. The subject perceives each stimulus as an instance of the pattern with the closest peak delay value. He then produces an instance of that pattern. Thus, the overall distribution in the response data should be bimodal. A graph of response peak delay against stimulus peak delay should show a sigmoid shape. In the idealized case, the transition from the lower to the upper arm is abrupt. Real data however, usually exhibit a transitional region which arises when the stimuli seem ambiguous and the subject vacillates in his response.

3.2 Peak delay data

In general, the data supported the existence of two intonational categories. Data plots for three subjects are displayed in Figure 3. Figure 3A shows plots of data for subject TWB (the high school student). The histogram for the peak delays is obviously bimodal. On the plot of median response peak delay against stimulus peak delay, the diagonal line shows how the medians would behave if the subject had faithfully reproduced what he heard. It is clear that there are substantial deviations between the stimuli and the responses. For the first 9 stimuli, the peak delay values cluster between .1 and .15 seconds, whereas for the last 4 stimuli, they cluster between .2 and .25 seconds. The responses to stimuli number 10 and 11 have intermediate values for the median peak delay.

Figure 3B shows data for subject HBT. The results for this subject, and for subject SAS, were very similar to the results for subject TWB. HBT and SAS differ from TWB in the location of the boundary between the two categories.

Figure 3C shows data for subject RLB. This data set shows the same tendencies that we saw in the data for the other subjects, but less strongly. The second mode of the histogram is less pronounced than for the first three subjects. Also, on the response versus stimulus peak delay plot, there is a greater tendency for the median peak delay to track the values for the stimuli. We believe this tendency arose because of the subject's difficulties with the task. The subject reported that he was hearing more patterns than he could easily reproduce. His pattern of responses to the lower numbered stimuli is quite similar to that of the other subjects, but his responses to the higher numbered stimuli appear to include tokens of both the early peak and the late peak pattern, giving rise to broad distributions. Thus we believe that he was aware of a category difference in the stimuli but did not completely control it in his own speech.

The data for one subject, DTT, did not conform to that for the other subjects. The histogram for all his data is unimodal, and the median peak delay shows little variation. We believe that DTT lacks the L*+H pitch accent. The continuum theory must also make a special case of DTT, since he does not exhibit the substantial range of variation in peak position which is claimed to be possible. Presumably DTT would be described as having an unusually strong preference for his central peak position. Thus, DTT does not provide strong evidence for distinguishing between the two proposals.

3.3 Other characteristics of the data

The early peak and delayed peak variants of the rise-fall-rise showed no significant difference in the F0 minimum value preceding the peak. Thus, we are confident that subjects produced instances of the L+H* and the L*+H, not instances of plain H*, which would exhibit a much higher F0 minimum preceeding the peak. According to several measures, the L tone occurs later (relative to the segments) in the delayed peak variant than in the early peak variant, as predicted by the Pierrehumbert model. We omit supporting graphs, for lack of space.

Individual histograms of responses to each stimulus were examined. In general, these were broader in the transitional region of each sigmoid than on the arms. This result is predicted by a two-category model, since the distribution of responses to an ambiguous stimulus arises as a mixture of sampling from the distributions for the two categories. We are in the process of determining whether individual distributions can be quantitatively modeled in this way.

The relation of the peak delay to duration pattern is also of interest. It is sometimes claimed that syllables bearing L tones are longer than equally stressed syllables bearing H tones. Since the L*+H accent and the L+H* accent place opposite tones on the stressed syllable, there might be an effect of this sort. It is also important to rule out any possibility that the peak delay data might be an artifact of durational differences. In the phrase we used, it was impossible to measure the duration of the stressed syllable per se. The [l] in "millionaire" does not yield a well-defined measurement point; in fact it was often missing, with the syllables separated only by a [y] glide. However, an increase in the duration of the stressed syllable should be reflected in an increase in the total duration of from [m] to [n], which was easily measured.

The [m]-to-[n] duration did not increase with peak delay. The range of variation in duration is about one third the range of variation in peak delay. There is little pattern to it, and what pattern exists is not consistent across subjects. Only SAS shows some evidence for longer durations at longer peak delays. But for her, the effect is still far to small to explain the variation in the peak delays.

4. CONCLUSION

Four out of five subjects support the existence of two categories of rise-fall-rise. The L tone shifted rightward with the peak, as predicted. No significant durational effects were found. Thus, the results support a taxonomy in which alignment functions as a binary linguistic distinction.

REFERENCES

[1] Gussenhoven, C.: (1984) *On the grammar and semantics of sentence accents*. (Foris Publications, Cinnaminson, N.J.).

[2] Liberman, M.Y.: (1975) *The intonational system of English*. MIT Ph.D. diss.; published by Garland, New York, 1979.

[3] Liberman, M.Y.; Pierrehumbert, J.: (1984) Intonational invariants under changes in pitch range and length," Aronoff and Oerle, eds., *Language Sound Structure*. (MIT Press, Cambridge, MA).

[4] Liberman, M.Y.; Prince, A.: (1977) On stress and linguistic rhythm. *Linguistic Inquiry 8*: 249-336.

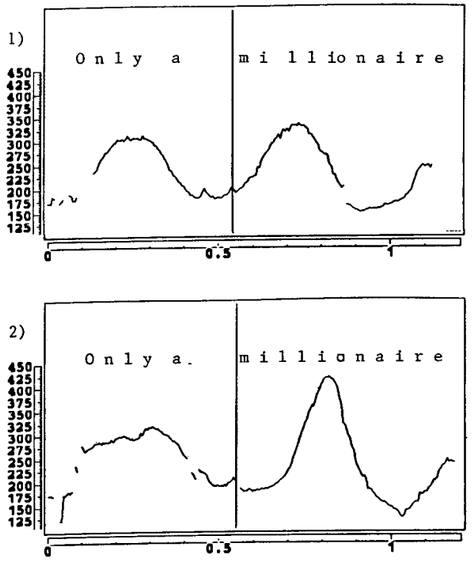
[5] Pierrehumbert, J.: (1980) The phonology and phonetics of English intonation. MIT Ph.D. diss.

[6] Pierrehumbert, J.B.; Beckman, M.E.: (forthcoming) *Japanese tone structure*. Linguistic Inquiry Monograph Series.

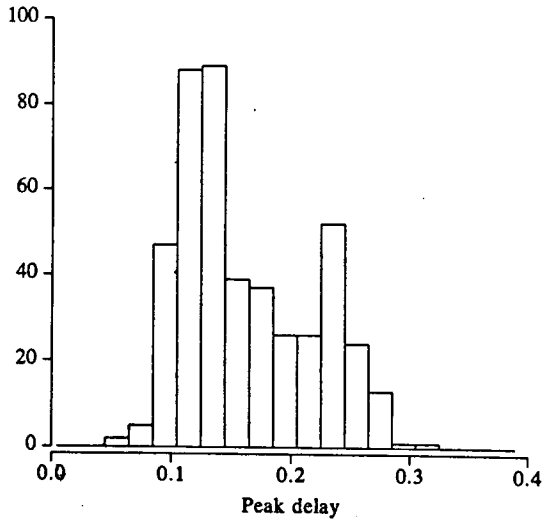
[7] Ward, G.; Hirschberg, J.: (1985) Implicating uncertainty: the pragmatics of fall-rise intonation. *Language* 61:4 747-776.

[8] Ward, G.; Hirschberg, J.: (1986) Reconciling incredulity with uncertainty: a unified account of the L*+H L H% intonational contour. Linguistic Society of America Annual Meeting.

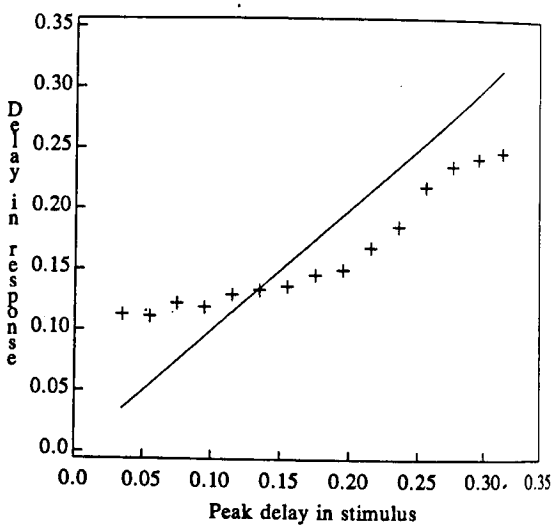
FIGURES



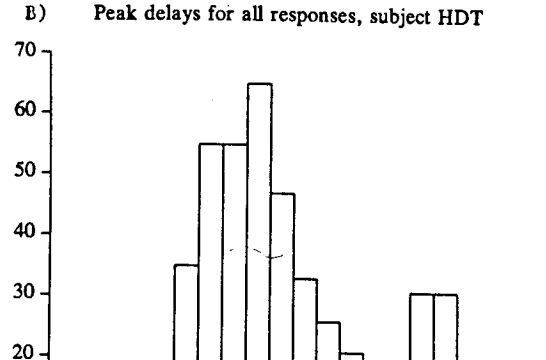
3A) Peak delays for all responses, subject TWB



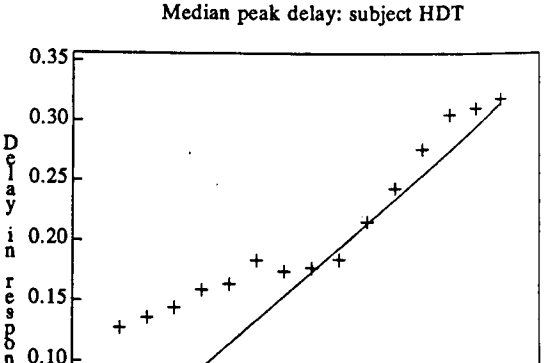
Median peak delay: subject TWB



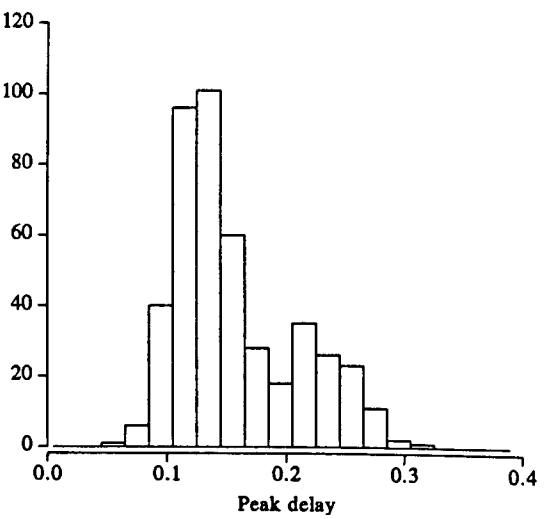
B) Peak delays for all responses, subject HDT



Median peak delay: subject HDT



C) Peak delays for all responses, subject RLB



Median peak delay: subject RLB