Phonetica
Editor: K. Kohler, Kiel

Separatum
Publisher: S. Karger AG, Basel
Printed in Switzerland

Phonetica 1989; 46: 181–196

Categories of Tonal Alignment in English

Janet B. Pierrehumbert, Shirley A. Steele
AT&T Bell Laboratories, Murray Hill, N.J., USA

Abstract. This paper reports the results of an inquiry into the question of category versus continuum in intonation. Variants of the English rise-fall-rise pattern were used to study whether tonal alignment is a categorical or gradient distinction. LPC resynthesis was used to construct a set of stimuli in which the alignment of the F_0 rise-fall varied in small steps. Subjects heard the stimuli in randomized order and imitated what they heard. The position of the F_0 peak relative to the onset of the stressed vowel was measured in each response. Systematic deviations between the peak placement in the stimuli and those in the responses revealed the existence of two categories. We conclude that tonal alignment functions as a binary distinction in English intonation.

Introduction

Topic

In this paper, we report an experimental investigation of the rise-fall-rise intonation patterns of English (a more abbreviated report of the same work appears in Pierrehumbert and Steele [1987]). A fundamental frequency contour for one example of this pattern is shown in figure 1. The sentence is 'Only a millionaire', and the vertical cursor shows the position of the [m] release for 'millionaire'. The transcription indicates the alignment of the segments that follow. Saying this sentence as in the figure would convey the speaker's incredulity, or his uncertainty about how the sentence is related to the discourse [see Ward and Hirschberg, 1985, for a more technical formulation]. The example in figure 1 contrasts with the F_0 contour for the same sentence shown in figure 2. The shape of this second contour is very similar, but the peak now falls on the stressed syllable, and the preceding F_0 minimum is before rather than after the [m] release.

Pierrehumbert [1987] proposes that these two patterns involve a minimal contrast in pitch accent type. In one variant, a low tone (L) is aligned with the stressed syllable, and a high tone (H) trails it; this is the pattern in figure 1. In the other, the H falls on the accent, and the L leads it. Using a diacritic * to represent alignment with the stress, the

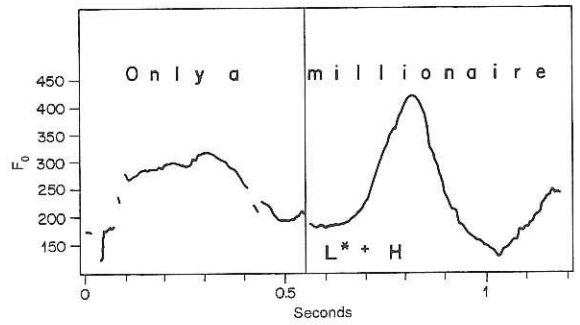

Fig. 1. Fundamental frequency contour of the phrase 'Only a millionaire' spoken with an intonation pattern which indicates incredulity or uncertainty (the L*+H pattern). The vertical cursor is placed at the [m] release in 'millionaire'.

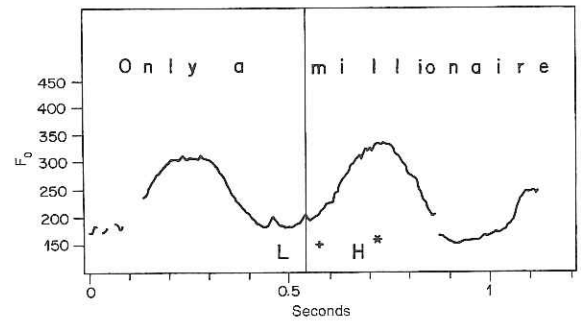

Fig. 2. Fundamental frequency contour of the phrase 'Only a millionaire' spoken with an L+H* intonation pattern. The vertical cursor is placed at the release of the [m] in 'millionaire'.

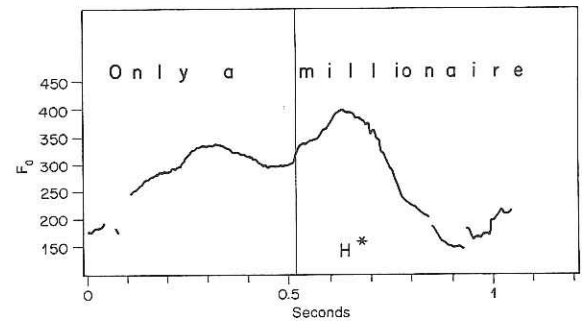

Fig. 3. Fundamental frequency contour of the phrase 'Only a millionaire' spoken with a more neutral intonation (the H* pattern). The vertical cursor is placed at the [m] release in 'millionaire'.

two different pitch accents are then transcribed as L*+H versus L+H*. The position of the unstarred tone with respect to the segmental material is taken to arise from general principles of articulatory coordination and varies considerably depending on the speech rate and the intrinsic length of the segments. Both bitonal accents contrast with the pattern in figure 3, which lacks the distinctive dip before the peak and is transcribed with a plain H*. In all three patterns, the fall-rise portion of the contour arises in the same way, from a sequence of L and H% marking the end of the phrase. (The diacritic % is used to mark a tone which falls right at the phrase boundary, rather than being attracted to a stressed syllable.) Thus, English is proposed to have a three-way distinction among L*+H L H%, L+H* L H% and H* L H%. Readers are referred to Pierrehumbert [1987] for motivation of this decomposition of the melody into characteristics associated with stressed syllables (pitch accents) and characteristics associated with the end of the phrase.

The existence of contours like that in figure 1 has been recognized by many researchers. Bolinger [1958] mentions a subtype of his accent A in which the peak is delayed beyond the stressed syllable. O'Connor and Arnold [1973] describe a rise-fall-rise, contrasting with the plain fall-rise in figure 3. Ladd [1980] discusses a 'scooped' variant of the fall-rise; more recently, the phonological theories of Ladd [1983] and Gussenhoven [1984] use a feature of peak delay to distinguish the contour in figure 1 from those in figures 2 and 3. However, there has not been a consensus about how this pattern is related linguistically to the other two. The particular question which we address here is whether the pattern exemplifies a phonological category of melody, as Pierrehumbert [1987] claims, or whether peak delay is manipulated continuously, as a reflex of some paralinguistic dimension.

The issue of categorical versus continuous distinctions in the intonational system was incisively described in Bolinger [1958]. He points out the need to address the problem experimentally and advances a morphological interpretation for the categories of pitch accent that he identifies. However, subsequent as well as prior descriptions of the intonational system have made widely differing assumptions about which contrasts are categorical and which are continuous, often without detailed justification. On one extreme, treatments such as those of Armstrong and Ward [1926] and Lieberman [1967] claim that English has only two linguistic categories of melody, rising and falling, and that all other variation is due to 'special circumstances' or 'emotional variation'. At the other extreme, theories with four tone levels [Pike, 1945; Trager and Smith, 1951; Liberman, 1975] raise the prospect of a much larger number of categories. The grammar of English melody in Pierrehumbert [1987] generates 28 nuclear configurations and 196 different contours for texts with a prenuclear as well as a nuclear pitch accent. Gradient distinctions in pitch range still play a major role in this theory. In fact, many contrasts which were treated categorically in theories with four tone levels are referred to pitch range in Pierrehumbert's theory, which has only two tone levels. A gradient treatment of range is supported by the 1984 experiment of Liberman and Pierrehumbert. It shows that subjects could successfully follow instructions to produce ten overall pitch ranges, rather than producing utterances clustered into a smaller number of distinct ranges.

Although Pierrehumbert [1987] treats peak delay as a categorical distinction, she gives no substantial grounds for doing so. Indeed Ladd [1980] makes the opposite assumption when he suggests that 'scooping' is a gradient modification of the fall-rise pattern. This assumption receives some further plausibility from work on intonational meaning by Pierrehumbert and Hirschberg [1990]. They suggest that the L*+H and L+H* function similarly in causing the word with the pitch accent to be implicitly compared to a scale of alternatives. (The H* in figure 3 simply adds information without an implicit comparison to alternatives). The L*+H and the L+H* function differently in that the L*+H implies that the value is not necessarily the correct one (the speaker may be uncertain about it, or he may be speaking incredulously), whereas the L+H* implies that the value is correct. Clearly, these two meanings are very closely related and might be taken to set up a paralinguistic continuum.

One main aim in the present work was to resolve the question of whether peak delay acts as a binary distinction or as a continuous dimension of variation. A second aim was to develop an experimental method which might help to put claims about categories in the intonational system on a more solid empirical footing.

Method

In our experiment, subjects heard and imitated stimuli constructed with a continuum of peak delays. If peak delay is a gradient dimension, delays in their responses should also fall along a continuum. (A preferred peak position might cause responses to stimuli at the extremes of the continuum to drift towards the center.) If Pierrehumbert is correct, responses should cluster into two categories.

This experimental method is a variant of the paradigm familiar from studies of categorical perception of speech segments. We have used an imitation task rather than the more commonly used labelling and discrimination tasks, because we were not concernced with the separate analysis of the production and perception systems. In most categorical perception studies, the linguistic analysis is relatively uncontroversial, and the issue is the status of the linguistic description in the psychological system. In our study, we were looking first for evidence about the system of linguistic categories. Determining how these categories are related to characteristics of the perceptual, articulatory or cognitive systems is a matter for further research.

Experimental Procedures

Stimuli

The stimuli were versions of the phrase 'Only a millionaire', in which the location of the F_0 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 the subjects' imitations of it, the main stress fell on the first syllable of the word 'millionaire', not the last. This is an acceptable pronunciation in American English, and subjects reported no difficulty in 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 ms from the end of the [m] in 'millionaire', by 20-ms increments. As the peak was shifted, the

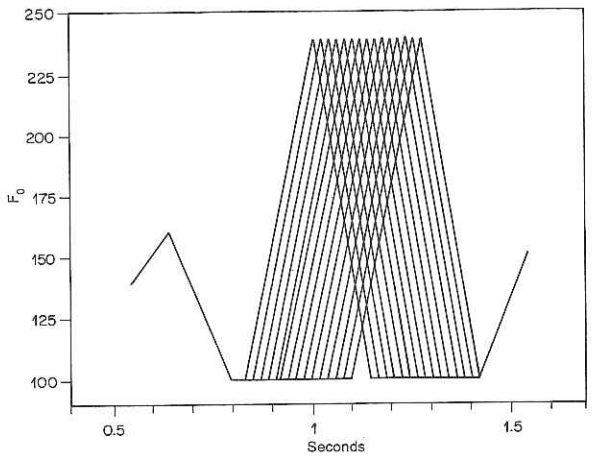

Fig. 4. The superimposed fundamental frequency contours of the 15 aural prompts. The F_0 peak is varied in 20-ms increments.

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. Figure 4 displays the superimposed F_0 contours for all 15 stimuli used.

The particular phrase used was chosen for two reasons. First, it is composed entirely of sonorants, thus avoiding devoicing in the F_0 contour and minimizing consonantal effects in the stressed syllable. Second, its pragmatic interpretation could be sensibly altered by variation in peak delay. The difference in pragmatic interpretation that we had in mind is illustrated in the following situation:

Ms. Jones receives a phone call from a representative of a charitable organization. The representative explains that the organization is launching a special fund-raising campaign, targeting the richest donors, and that Ms. Jones is understood to be a billionaire. Ms. Jones, who is not that rich, corrects him, replying: 'Oh, no. Only a millionaire.' She uses the pattern with an early peak, as in figure 2. The charity representative, astonished to find that his information was incorrect, replies: 'Only a millionaire', using the intonation pattern with a late peak, as in figure 1. In this case, the pattern conveys incredulity.

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 were 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.

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.

Subjects

The subjects were 5 native speakers of American English, 2 females and 3 males. Four of the 5 were naive about the purpose of the experiment. The subjects were: D.T.T., a software engineer; H.L.T., a psychology research assistant; R.L.B., an opto-electronics processing engineer; S.A.S., one of the authors; and T.W.B., a high-school student.

Measurement

The measurement of primary interest is peak delay, defined as the difference between the time of the F_0 peak and the time of the [m] release. This was found by examining visual displays of the F_0 contour and waveform of each response. Time points were established for the release of the [m] into the vowel, for the F_0 peak, for the implosion of the [n] and for the F_0 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. In doubtful cases, the decision was assisted by an examination of the autocorrelation (which typically shows a brief drop at the point of the spectral change) and by playing small sections of the speech. The F_0 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. For the F_0 minimum, the last time point before the rise was selected. The valley preceding the peak was typically very broad, with various irregularities. Thus, the location of the absolute minimum F_0 value showed a great deal of variation and may not be a good index of the location of the L tone.

Although it would have been desirable to further segment the region between the [m] release and the [n] implosion, it did not prove possible to do this reliably. The [l] in this context offers no consistent discontinuity for segmentation, and indeed it was frequently replaced with a high front glide (as is often the case in American English). This outcome was not surprising, since the materials had been designed to make the entire region as vocalic as possible.

A few utterances had to be eliminated from the data set, because the F_0 at crucial points could not be measured. There were no more than 4 such utterances per subject. The data summaries for each stimulus represent in every case at least 28 responses.

Results

Predictions about Peak Delays

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

First, consider a simplified 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 delay in the stimulus. In a graph of median response peak delay against stimulus peak delay, the responses would fall on the line $y = x$. The distributions of response peak delays for each individual stimulus should all have the same shape, regardless of stimulus number. The stimulus number should affect only their location. The overall distribution of responses, being the sum of the individual distributions, should be broad and unimodal.

Next, consider a continuum model in which peak delay is continuously variable but a central value is preferred. In this model, the subject's response to each stimulus would tend to shift towards the center. A graph of median response peak delay against stimulus peak delay should exhibit a dependence on stimulus number, as in the first model, but with values shifted towards the middle. Individual distributions are likely to show a dependence of shape on stimulus number. End-of-continuum stimuli would elicit some responses from the center, but stimuli in the center (the preferred form) would be less likely to elicit responses from the ends. Thus, tight response distributions are predicted for the middle of the continuum, with broader distributions

toward the ends. The overall distribution of responses would again be unimodal, but less broad than under the first model.

In an idealized model with two categories, two peak delays are possible. 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. A graph of the resultant median peak delays is thus predicted to show a stepwise pattern, with a plateau for each category. The overall distribution in the response data should be bimodal. In a somewhat less idealized picture, some of the stimuli in the middle of the continuum are ambiguous, and the subject's response may vacillate. The probability of producing one response or the other would depend on the relative similarity of the stimulus to archetypes for the two categories. The shape of individual distributions for response data is thus predicted to vary by stimulus number, with relatively tight distributions for good instances of each category. For ambiguous stimuli, the distributions should be broader, arising as the sum of sampling from two different distributions. The related median plot would thus exhibit a sloping transitional region between the two plateaus.

Peak Delay Data

In general, the data supported the existence of two intonational categories. Figures 5 and 6 show plots of data for subject T.W.B. (the high-school student). Figure 5 shows a histogram for the peak delays in all of the data. The histogram is obviously bimodal. In figure 6, the median peak delay for responses to each stimulus is plotted against the peak delay in the stimulus itself, with first and third quartiles indicated. The diagonal line shows how the medians

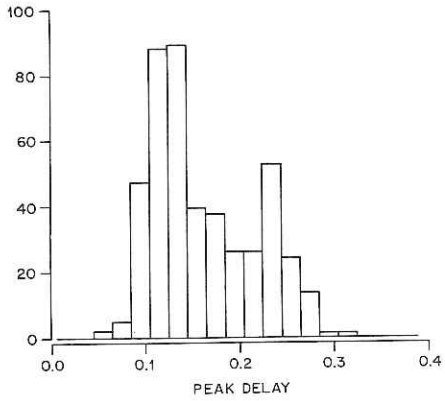

Fig. 5. Speaker T.W.B.; histogram of peak delays for all responses.

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. These deviations are in the direction of the model with two categories and a transition area. That is, for the first 9 stimuli, the peak delay values cluster between 0.1 and 0.15, whereas for the last 4 stimuli, they cluster between 0.2 and 0.25. The responses to stimuli numbers 10 and 11 have intermediate values for the median peak delay.

Histograms of the responses to individual stimuli (not displayed here for lack of space) show that responses to stimuli 1–9 have very little overlap with those for stimuli 12–15. Thus, the responses on the lower and upper arms of the curve in figure 6 are very well separated. None of the 30 responses to stimulus 1 had a peak delay as small as that in the stimulus itself. For the responses to stimulus 9, only 2 out of 30 responses had a delay as great as that in the

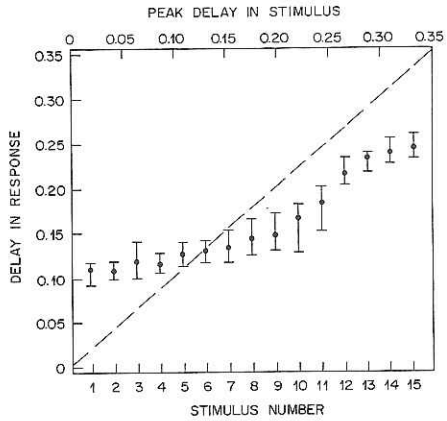

Fig. 6. Speaker T.W.B.; plot of median peak delay for responses to each stimulus against peak delay of the stimulus. First and third quartiles indicated.

stimulus. Thus, the clustering of the data on the lower arm of the curve is a very strong effect. The upper arm also deviates strongly from the predicted y = x line; none of the responses to stimulus 15 has a peak delay as great as in the stimulus. The entire arm is shifted below the prompt curve, however. This might be explained by the subject using a faster speech rate than that used in the stimuli. Not surprisingly, it was not possible to normalize the data for speaking rate in this study, because rate changes produce different durational changes across segments and across speakers.

The broad trend just described may be seen in the quartiles plotted in figure 6. The quartiles also show that the distributions of responses on the upper and lower arms of the curve are relatively narrow; in contrast, those for stimuli 10 and 11 are notably broader. This is what we would expect for stimuli which are ambiguous between two categories. (In particular, the experimenter observed that subjects tended to repeat their imitation of the preceding pattern when faced with an ambiguous prompt.) The overall distribution of responses then probably arises as the sum of sampling from the two distributions for productions belonging to the two categories.

Both arms of the curve display some tendency for the median to track the peak delay in the prompt. That is, both are somewhat tilted. In the corresponding individual stimulus histograms, we see some tendency for a shift in the mode. That is, the shift in the median is not arising merely through progressive contamination of the distribution with samples from the other category. One possibility is that the subject does manipulate peak delay continuously, but within two categories rather than one. A second (and perhaps more likely) interpretation is that the subject had some success in performing a phonetic imitation of the patterns he heard, just as he might have some success in imitating details of another person's idiolect.

Examination of the median plot and the overall histogram indicates that the boundary between the two categories lies in the vicinity of stimuli 10 and 11, rather than halfway through the stimulus continuum. One consequence is that the total number of responses in the late peak category is less than that in the early peak category. This

asymmetry in the response pattern might be attributed to L*+H being a more marked accent type in English than L+H*. Or, some phonetic properties of the original model utterance, which had a relatively early peak, may have been retained in the resynthesized versions as cues for a L+H* pattern. One possible candidate is the exact shape of the rise-fall (which was modelled on that of the original); others include spectral tilt and relative amplitude, either of which may carry secondary information about tonal category.

Figures 7–10 show data for subjects S.A.S. and H.B.T. The results for these subjects were very similar to the results for subject T.W.B. They differ from T.W.B. in the location of the boundary between the two categories.

Figures 11 and 12 show data for subject R.L.B. This data set shows the same tendencies that we saw in the data for the other subjects, but less strongly. In figure 11, the second mode of the histogram is less pronounced than for the first 3 subjects. In figure 12, there is a greater tendency for the median peak delay to track the values for the stimuli. Quartile ranges for individual stimuli show that the behavior of responses to the lower numbered stimuli is quite similar to that of the other subjects, but for responses to the higher numbered stimuli, the distributions are much broader. We believe this happened because of the subject's difficulties with the task. The subject reported that he was hearing more patterns than he could easily reproduce. His responses to the higher numbered stimuli appear to include tokens of both the early peak and the late peak patterns, giving rise to broad distributions. Thus we believe that he was aware of a category difference in

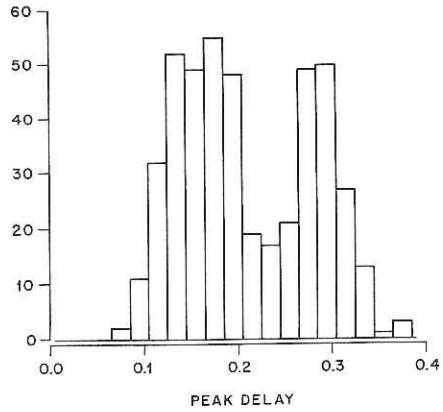

Fig. 7. Speaker S.A.S.; histogram of peak delays for all responses.

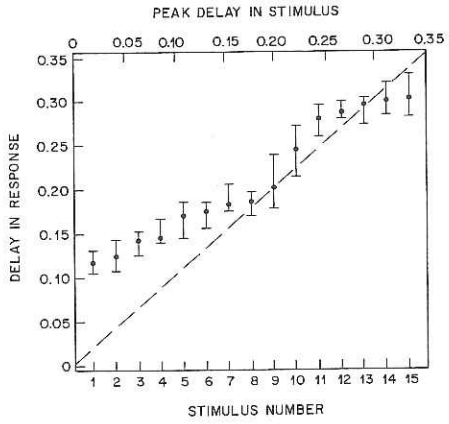

Fig. 8. Speaker S.A.S.; plot of median peak delay for responses to each stimulus against peak delay of the stimulus. First and third quartiles indicated.

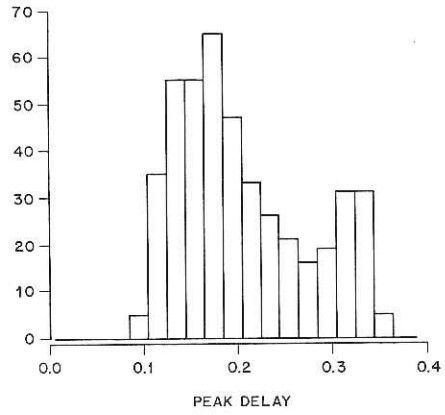

Fig. 9. Speaker H.D.T.; histogram of peak delays for all responses.

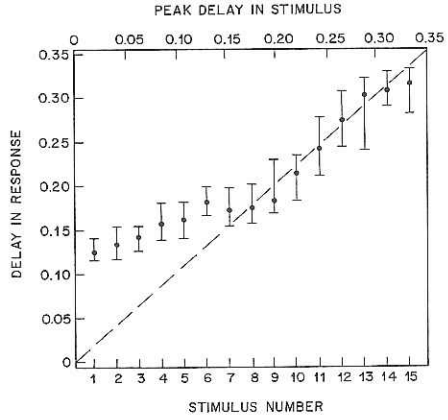

Fig. 10. Speaker H.D.T.; plot of median peak delay for responses to each stimulus against peak delay of the stimulus. First and third quartiles indicated.

the stimuli but did not completely control it in his own speech.

The data for 1 subject, D.T.T., did not conform to those for the other subjects. The histogram for all the data is unimodal, and the median peak delay shows little variation. Individual histograms suggest that stimuli numbers 1 and 2 may have been, in a few cases, interpreted as instances of an early peak category which has not been discussed here. However, the late peak category (with the peak near the [n] of millionaire) does not occur in this subject's speech.

We believe that D.T.T. lacks the L*+H pitch accent, just as a speaker may lack, for example, an /a/-/ɑ/ segmental distinction or the distinctive third person singular in the auxiliary verb 'do'. The continuum theory must also make a special case of D.T.T., since he does not exhibit the substantial range of variation in peak position which is claimed to be used expressively. Presumably D.T.T. would be described as having an unusually strong preference for his central peak position. Thus, D.T.T. does not provide strong evidence for distinguishing between the two proposals.

The L Tone

The early peak and delayed peak variants of the rise-fall-rise showed no significant difference in the F_0 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*.

According to the Pierrehumbert account, the L tone should occur later (relative to the segments) in the delayed peak variant than in the early peak variant. As predicted, a graph of L tone delay (time between the [m] release and F_0 minimum) against peak delay

revealed for each subject a general increase in L tone delay, as peak delay increased.

Figure 13 shows this for speaker T. W. B., and graphs for other subjects were similar. Note that the L tone delay is negative when the F_0 minimum precedes the [m] release. The trend of increasing peak delay is summarized by a lowess regression superimposed on the plot. (The lowess regression is a robust, iterative procedure described in Cleveland [1979]. At each point, its value is determined in a local window of data points. Thus, it can be viewed as a way of smoothing the data so as to bring out the trend.) The vertical dashed line in the plot shows the peak delay for the minimum between the two modes of the overall histogram for the same subject. Thus, it provides an indication of where the boundary between the two categories is. Positive values of the L delay (those which occur after the [m] release) are confined to the L*+H category. However, for this category there are still many utterances in which the F_0 minimum occurred before the [m] release, and for subject R. L. B. this was practically always the case. In general, the L delay values are quite scattered, and there is considerable overlap between the values for the two categories. One consequence is that overall histograms of L delay (not shown here) do not have a conspicuously bimodal form for most of the subjects.

The L delay values are not an entirely accurate reflection of the differential treatment of the low region observed in F_0 contours. In L*+H contours in which the F_0 minimum occurred during the [m], there was an extended low region, only slightly rising, before the sharp rise up to the delayed peak began. In the L+H* contours, the steep portion of the rise began immediately. This meant that the F_0 value at the [m] release was higher for L+H* than for L*+H contours.

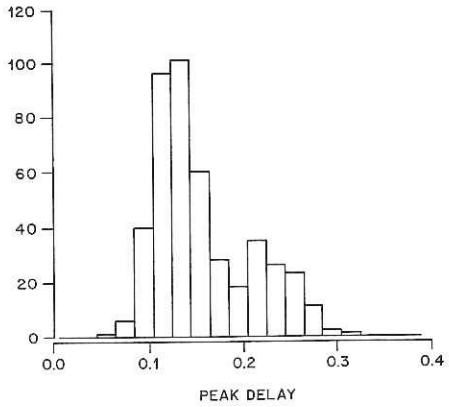

Fig. 11. Speaker R. L. B.; histogram of peak delays for all responses.

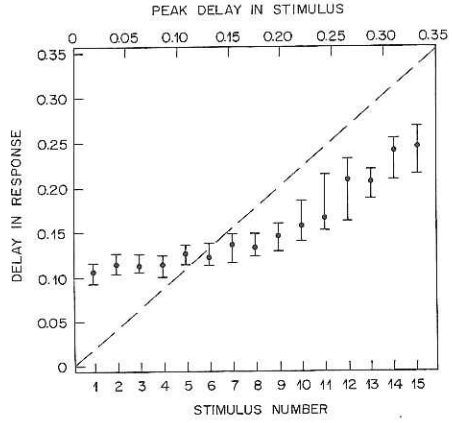

Fig. 12. Speaker R. L. B.; plot of median peak delay for responses to each stimulus against peak delay of the stimulus. First and third quartiles indicated.

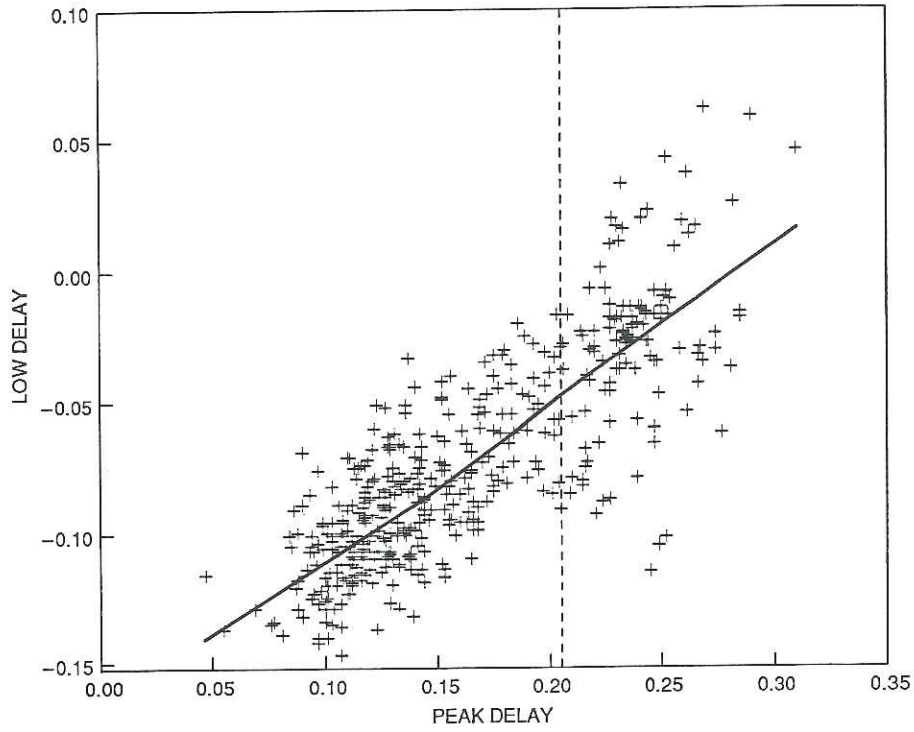

Fig. 13. Speaker T. W. B.; plot of L tone delay gainst peak delay. The vertical dashed line divides the two categories of response; the solid line indicates the trend found by a lowess regression.

Segmental Durations

The relation of the peak delay to the [m]-to-[n] duration is of interest for two reasons. First, it is widely rumored that syllables bearing L tones are longer than equally stressed syllables bearing H tones, although we have been unable to find a published reference. 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. Second, we wished to rule out the possibility that peak delay might be an artifact of durational differences. Recall that the total duration from [m] to [n] was measured instead of the stressed syllable per se, because [l] does not

yield a well-defined measurement point and was, in fact, often missing, with the vowels separated only by a high front glide.

Graphs for the data for each subject showed that the [m]-to-[n] duration did not increase with peak delay. The range of duration variation was about one third the range of variation in peak delay, with little, if any, pattern. Only S.A.S. showed some evidence for longer durations at longer peak delays, but still too small an effect to explain the variation in peak delay.

Discussion

The experimental results are consistent with the theoretical framework of Pierrehumbert [1987]. The contrast between L+H* and L*+H proposed there predicts two categories of peak delay; the delayed peak category is also predicted to show delay of the low-toned region. The duration patterns and the F_0 minima are not predicted to contrast. All four of these predictions are confirmed by the data.

Phonological descriptions proposed in Ladd [1983] and Gussenhoven [1984] also predict a contrast in peak alignment with no concomitant difference in duration pattern. These systems both have a feature of peak delay, which can modify a basic fall-rise to create a pattern in which the F_0 peak actually occurs after the stressed syllable. Gussenhoven [1984] says that the peak delay modification is a gradient feature, but that there may be an 'ideal' target for the delayed value, as well as for the unmodified value. This conjecture is borne out by the bimodal distributions in our subjects' productions. There is perhaps some conceptual difference between proposing two categories and proposing a gradient dimension with two preferred positions. However, it is unclear what experiments for deciding this issue could realistically be performed.

An important difference between Gussenhoven [1984] and Pierrehumbert [1987] is that Gussenhoven treats the extremely low F_0 value before the peak in the delayed variant as an artifact of the peak delay modification. That is, his system provides for contours like those in figures 1 and 3 but not for ones like that in figure 2. It is unclear, in his account, why the F_0 minimum preceding the peak did not vary significantly in our data. Of course this value did not vary in the stimuli, but the responses in general deviated from the stimuli in order to conform to the phonological and phonetic system of English, and so they should have deviated in this regard, also. Ladd [1983] also makes no distinction between nuclear H* and L+H*. He suggests [pers. commun.] that the L+H* cases might be analyzed as involving a phrasing break marked with an L boundary tone or a contrast in the type of the prenuclear accent. O'Connor and Arnold [1973] would rely on the prenuclear phonology to create the three-way distinction shown in figures 1–3. They would view the contour in figure 1 as involving a rise-fall-rise nuclear accent. Figures 2 and 3 would both have a fall-rise accent but would differ in that figure 2 had a falling head (like figure 1), whereas figure 3 had a high head.

Our experiment concerned only the contrast between L+H* and L*+H, and therefore we have no conclusive evidence concerning these alternative formulations. We feel confident from visually examining and listening to the utterances during the measurement process that the early peak and late peak variants did not differ in phrasing. We believe that neither had a phrase break in the middle; the pitch accent on 'only' appeared to be prenuclear, rather than nuclear. A small piece of evidence against the characterization of O'Connor and Arnold arises from the fact that subjects tended to produce 'only' with the same pitch accent type as 'millionaire', even though the F_0 contour on 'only' was not varied in the stimuli. That is, when 'millionaire' was produced with a late peak accent, 'only' tended to have a similar, though less prominent accent. This observation is difficult to accommodate within the framework of O'Connor and Arnold, which provides for a late peak variant only in nuclear position.

It is most interesting to compare our results to those in Kohler's investigations of German intonation [1986, 1987 a]. This is, to our knowledge, the only other work which has applied the experimental techniques of categorical perception to illuminate the linguistic structure of the intonation system. Some previous research, such as Hadding-Koch and Studdert-Kennedy [1964] and Nash and Mulac [1980], reports results of identification tasks for stimuli constructed along intonational continua. Identification tasks, however, automatically force responses into categories. Thus, they alone cannot reveal whether categories exist, but only where the boundaries between categories are. In Kohler's study, a set of tokens of the sentence 'Sie hat ja gelogen' (she's been lying) were synthesized with continuously varying peak placement. Peak placements ranged from before the stressed syllable (that is, in the syllable /gə/ of 'gelogen') up through the end of the stressed syllable /lo:/. These stimuli were used in several experiments. In one, the subjects heard the stimuli played in a continuum. In another, they performed a discrimination task. In a third, they matched the stimuli with their appropriate contexts.

These experiments demonstrated that German has three categories for peak alignment, which the contextualization experiment showed to be related to differences in meaning. Kohler [1987b] gives the three meanings as 'established' (early peak), 'new' (middle peak) and 'emphatic' (late peak). However, there was a difference in the sharpness of the perceptual boundaries between categories. The boundary between early and mid peaks was very sharp and reliable across subjects. The boundary between mid and late peaks appeared to be a softer one.

It is of course difficult to compare phonetic results across languages, since languages can differ at so many levels. We would tentatively suggest that the mid and late peak patterns involve L+H* and L*+H accents, respectively, and that the early peak variant has an H+L* accent. The English H+L* accent, according to Pierrehumbert [1987], has a peak just before the stress and a fall to a phonetically mid level on the stress. When this accent is followed by an L phrasal tone (as in the typical declarative sentence), there is usually a straight fall from the prestress peak to the L phrasal tone. The slope of this fall would ordinarily be less than that after an L+H* or L*+H.

The sharpness of the perceptual boundary between the early and mid peak patterns (relative to the boundary between mid and late peak patterns) might be related to two factors. First, the L+H* and L*+H involve the same tone sequence, whereas the H+L* has a different one. Second, there

may be a difference at the level of meaning. In American English, as we have seen, the L+H* and the L*+H have closely related meanings. The meaning given by Pierrehumbert and Hirschberg [1990] to the H+L* is quite different; the accent implies that the listener should try to relate the accented word to information that is already shared between the speaker and listener. The meanings given by Kohler for the early, mid and late peak patterns make it clear that there are differences between American English and German in the interpretation of the accents. In particular, the late peak variant in German is not used to signal lack of speaker commitment; its meaning is closer to what Gussenhoven [1984] reports for the late peak variant in British English, namely 'significance' or 'non-routineness'. Possibly, all three systems use the L+H accents to signal that the accented word is related to a scale of alternatives; they would differ in what is implied about the specific relation of the word to the scale. The apparent difference in interpretation of the English and German L+H accents may also be related to the boundary tone, which was L% in Kohler's work but H% in the present study as well as in Ward and Hirschberg [1985]. The meaning Kohler gives for the early peak variant is more similar to that just given for the H+L*. Thus, there is at least a possibility that the early peak accent in German, as in English, stands out from the mid and late peak accents in its semantic interpretation.

The experiment reported here was not intended to include the H+L* accent of English. The F_0 peak in the stimuli was never earlier than the stressed syllable. A preliminary continuum which did include such stimuli had to be truncated, because several listeners found the earliest peak placements to be unnatural. We believe that the stimuli in question were unnatural, because the shape of the F_0 contour both before and after the peak was not appropriate for the English H+L* accent. Two subjects (S.A.S. and H.D.T.) have median plots which might be interpreted as being influenced by H+L* interpretations. That is, the upward slope for the medians of the first 3 stimuli might be viewed as a transition between an early peak category and a mid peak category. This interpretation is merely speculative without further data.

Conclusions

In this paper, we have presented data indicating that English speakers have two rise-fall-rise intonation patterns, which differ in how they are aligned with the stressed syllable. We would suggest that the experimental method, in which subjects imitate a stimulus continuum, is a useful tool for investigating the structure of the intonation system.

References

Armstrong, L. E.; Ward, I. C.: A handbook of English intonation (Teubner, Leipzig 1926).

Bolinger, D. L.: A theory of pitch accent in English. Word *14:* 109–149 (1958).

Cleveland, W. S.: Robust locally weighted regression and smoothing scatterplots. J. Am. Statistical Ass. *74:* 829–836 (1979).

Gussenhoven, C.: On the grammar and semantics of sentence accents (Foris Publications, Cinnaminson 1984).

Hadding-Koch, K.; Studdert-Kennedy, M.: An experimental study of some intonation contours. Phonetica *11:* 175–185 (1964).

Kohler, K. J.: Computer synthesis of intonation. Proc. 12th Int. Congr. Acoustics, Toronto 1986, p. A6-6.

Kohler, K. J.: Categorical pitch perception. Proc. 11th Int. Congr. Phon. Sciences. Tallinn, 1987a.

Kohler, K. J.: The linguistic functions of F_0 peaks. Proc. 11th Int. Congr. Phon. Sciences. Tallinn, 1987b.

Ladd, D. R.: The structure of intonational meaning (Indiana University Press, Bloomington 1980).

Ladd, D. R.: Phonological features of intonational peaks. Language *59:* 721–759 (1983).

Liberman, M.: The intonational system of English; MIT doc. diss. (1975; distributed by Indiana University Linguistic Club, Bloomington 1978).

Liberman, M.; Pierrehumbert, J.: Intonational invariance under changes in pitch range and length; in Aronoff, Oehrle, Language sound structure, pp. 157–233 (MIT Press, Cambridge 1984).

Lieberman, P.: Intonation, perception and language (MIT Press, Cambridge 1967).

Nash, R.; Mulac, A.: The intonation of verifiability; in Waugh, van Schooneveld, The melody of language, pp. 219–242 (University Park Press, Baltimore 1980).

O'Connor, J. D.; Arnold, G. F.: Intonation of colloquial English; 2nd ed. (Longmans, London 1973).

Pierrehumbert, J.: The phonology and phonetics of English intonation (Indiana University Linguistics Club, Bloomington 1987).

Pierrehumbert, J.; Hirschberg, J.: The meaning of intonation contours in the interpretation of discourse; in Cohen, Morgan, and Pollack, Intentions in communication. SDF Benchmark Series in Computational Linguistics (MIT Press, Cambridge, 1990).

Pierrehumbert, J.; Steele, S.: How many rise-fall-rise contours? Proc. 11th Int. Congr. Phon. Sciences. Tallinn 1987.

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

Trager, G. L.; Smith, H. L.: Outline of English structure (Battenburg Press, Norman 1951).

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

Received: August 31, 1987
Accepted: August 9, 1989

Janet B. Pierrehumbert
Department of Linguistics
Northwestern University
Evanston, IL 60208 (USA)