Reaction Time as an Indicator of Discrete Intonational Contrasts in English

Aoju Chen

Centre for Language Studies
University of Nijmegen, the Netherlands
aoju.chen@let.kun.nl

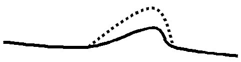
Abstract

This paper reports a perceptual study using a semantically motivated identification task in which we investigated the nature of two pairs of intonational contrasts in English: (1) normal High accent vs. emphatic High accent; (2) early peak alignment vs. late peak alignment. Unlike previous inquiries, the present study employs an on-line method using the Reaction Time measurement, in addition to the measurement of response frequencies. Regarding the peak height continuum, the mean RTs are shortest for within-category identification but longest for across-category identification. As for the peak alignment contrast, no identification boundary emerges and the mean RTs only reflect a difference between peaks aligned with the vowel onset and peaks aligned elsewhere. We conclude that the peak height contrast is discrete but the previously claimed discreteness of the peak alignment contrast is not born out.

1. Introduction

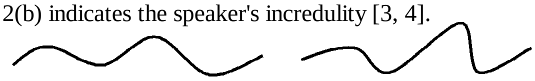
It has been well recognised that the nature of intonation is two-fold. That is, intonation is discrete or categorical as well as gradient or continuous (see e.g. [1] for more discussion). The discreteness of intonation refers to the fact that a given pitch contour can unambiguously have a given interpretation. Gradience in intonation has been suggested for pitch range and the alignment of pitch peak (H*) or peak valley (L*). Take the contour H*L L% for example. It can be varied either in the F0 of H* or in the alignment of H* without losing its identity as H*L L% or the corresponding interpretation, although variations in H* can signal different degrees of a given meaning independent of what is conveyed by the contour. However, ambiguity between discreteness and gradience in intonation has frequently been observed. The difficulty in telling discrete intonational contrasts from gradient intonational contrasts reflects the differences in intonation models proposed for English. Two ambiguous cases in English that have been under investigation are (1) Peak height contrast: the distinction between normal High accent and emphatic High accent [2]; (2) Peak alignment contrast: the distinction between early peak alignment and late peak alignment [3].

The difference between normal and emphatic High accents can be associated with different meanings, as shown in (1), in which the low peak suggests 'an everyday occurrence' while the high peak suggests 'an unusual experience' [2].



(1) The alarm went off.

The distinction between early peak alignment and late peak alignment is illustrated in (2). In 2(a), the peak occurs inside the vowel of the accented syllable while the peak in 2(b) occurs in the following unstressed syllable. The difference between early and late peak alignments can be associated with different meanings as well. 2(a) can be interpreted as a genuine question while 2(b) indicates the speaker's incredulity [3, 4].



(2) A. Only a millionaire B. Only a mi llio naire
L+H* L H% L*+ H L H%

Different experimental approaches have been undertaken to examine the nature of these intonational contrasts. Section 2 reviews the methodology and the findings in previous investigations. In section 3, we propose a Reaction Time approach as a solution to the drawbacks in previous investigations. In section 4, we try to determine the nature of the peak height contrast and the peak alignment contrast in the light of present findings.

2. Previous investigations

2.1. Investigating the nature of peak height contrast

Ladd and Morton [2] employed the classical categorical perception (CP) paradigm deriving from studies of the CP of phonemes to investigate the peak height contrast as illustrated in (1). The classical CP paradigm consists of a forced-choice identification task and a discrimination task. Unlike the forced-choice identification task in the classical CP paradigm, Ladd and Morton's identification task is a semantically motivated one. In their identification task, subjects listened to stimuli generated from the utterance *The alarm went off* along a peak height continuum, and judged for each stimulus which of the two interpretations provided was more likely the intended meaning of the utterance. Their discrimination task is similar to the one used in the classical CP paradigm. An identification boundary occurred at 144.9 Hz in their male-voice stimuli. However, they failed to observe a discrimination peak at the identification boundary. As the essence of the classical CP paradigm is that discrimination is the easiest at the identification boundary and most difficult within identification categories, Ladd and Morton concluded that the distinction between normal high accent and extra high accent was not categorically perceived. They suggested that the presence of an identification boundary might be an artefact of the forced-choice identification task.

However, we argue that the absence of a discrimination peak at the identification boundary does not necessarily mean that the peak height contrast is not categorical. As made clear by Ladd and Morton, the CP paradigm may be unsuitable for investigating the categoricality of peak height contrast. The CP paradigm relies on the incapability of listeners to detect differences between two stimuli taken from the same category. However, subjects appeared to be equally capable in perceiving fine differences in f0 across the continuum. It is exactly this capability in perceiving fine F0 variation that makes the CP paradigm completely inadequate in examining CP of peak height contrast. Therefore, the nature of the distinction between normal and emphatic High accents still remains indeterminate.

2.2. Investigating the nature of peak alignment contrast

With respect to the distinction between early peak and late peak as illustrated in (2), empirical evidence for the discreteness of this distinction has emerged from Pierrehumbert and Steele’s production study [3]. Thirty repetitions of each of the 15 peak conditions along the alignment continuum implemented on the utterance *Only a millionaire* realised with a fall-rise-fall contour were presented as audio prompts to subjects in a random fashion. Subjects were asked to imitate each prompt. The authors argued that if the subjects were able to reproduce the continuum in their imitation, peak alignment difference must be gradient. However, if subjects' imitations were to fall into two categories, peak alignment difference must be categorical. They found that by and large the distribution of peak alignments was bimodal in the imitation data and therefore concluded that the distinction between early peak alignment and late peak alignment was discrete.

However, as the semantic aspect of intonation was discarded in [3], it is not clear whether the discrete alignment contrast found in production can indeed be associated with the binary meaning distinction as discussed in section 1. It has been speculated that a higher peak takes a longer time to reach than a lower peak and is therefore likely to be aligned later [1]. This is also readily observable in Pierrehumbert and Steele's Figures 1 and 2, which show two contours obtained from natural productions of the utterance *Only a millionaire* and are used to illustrate the difference in the alignment between L+H* and L*+H. In detail, the peak height of L+H* is approximately 300 Hz while the peak height of L*+H is approximately 400 Hz. This implies that the peak alignment contrast found in [3] alone may not be held responsible for the binary meaning distinction. In order to testify this implication and obtain a thorough understanding on the discreteness of the peak alignment contrast found in Pierrehumbert and Steele's production data, it is necessary to examine the nature of the peak alignment contrast in a context where the meaning of intonation is taken into account.

3. Present approach

As a solution to the predicaments in previous investigations, we propose a Reaction Time (RT) approach to examine the nature of peak height contrast and the discreteness of peak alignment contrast. The RT approach employs a semantically motivated identification task, similar to the one used in [2]. Two variables are measured: (1) response frequencies; (2) mean RTs for identification. If the identification categories emerging from the response frequencies are not task-induced but linguistically real, we will expect that the within-category stimuli are comparable in terms of cognitive load and therefore will trigger similar mean RTs for identification. However, on the whole, the within-category stimuli are expected to be cognitively less demanding than the across-category stimuli. As a result, the within-category stimuli will trigger shorter mean RTs than the across-category stimuli. An outcome related to this is that a mean RT peak will occur at the identification boundary.

The validity of this method is further fed by findings we have recently retrieved regarding RTs to identification within and across phonetic categories, /ba/ vs. /pa/ [5]. It was found that subjects are slowest for the stimulus at the phonetic boundary but fastest for stimuli within phonetic categories. As pointed out by Pisoni and Tash, 'Reaction Time is a positive function of uncertainty, increasing at the phonetic boundary where identification is least consistent and decreasing where identification is most consistent…'

The RT measurement has proved to be a sensitive on-line method to examine the appropriateness of intonation [6]. However, it has not been used before to investigate the categorical perception of intonation. The second aim of the present study is therefore to develop an on-line experimental method that can help to tell a gradient intonational contrast from a discrete contrast.

3.1. Stimuli

Two continua were produced: (1) peak height continuum; (2) peak alignment continuum. In view of the low pitch range our continuum fell into, it seemed reasonable that the peak height continuum was derived from a linear scale. As to the peak alignment continuum, because it is not clear how the human auditory system processes the time-related variation and studies concerned with the perception of alignment have adopted a linear scale, we have chosen to use a logarithmic scale here for a complementary purpose (Alice Turk, pc).

One carrier-utterance was composed for each continuum: (1) for peak height continuum: *The alarm's gone off*, similar to one used in [2]; (2) for peak alignment continuum: *To Birmingham*. Natural productions of these utterances served as the source utterances for our stimuli. They were read by a male native speaker of British English. Speech manipulation was performed by means of Praat [http://fonsg3.let.uva.nl/praat/].

The 'alarm' sentence was assigned the falling contour H* L H%. H* was realised as a 30 ms-high plateau starting 30 ms after the CV boundary of /la:/, preceded by a 120 ms rise and followed by a 120 ms fall. The peak height continuum starts at 106Hz and continues till 196 Hz in 6-Hz steps. These gave us 16 peak heights. Other pitch points were assigned fixed values through out the continuum, as illustrated in Figure1.

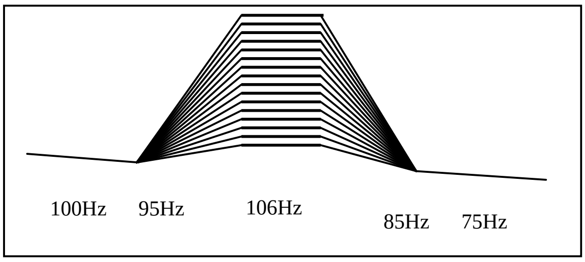


Figure 1: Peak Height continuum

The 'Birmingham' sentence was assigned the fall-rise-fall contour. The peak was realised as a 170Hz-pitch point. The steps (Y) were derived from the equation in (1), in which X_0 is 180ms, the time value of the CV boundary of the stressed syllable, N equals 1, 2, 4, 8, 16, 32, and 64, and X is 3ms.

$$Y= X_0 + N \times X_1 \qquad (1)$$

Including X_0, this gave us an eight-step peak alignment continuum as illustrated in Figure 2. The alignment continuum crosses two syllables, as in [3]. Other pitch points were assigned fixed pitch values and time values.

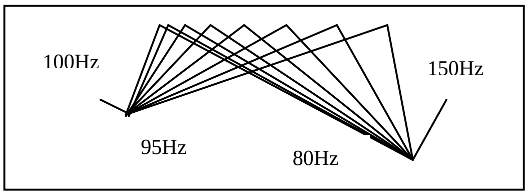


Figure 2. Peak alignment continuum

3.2. Experimental procedure

The experiment was set up by means of the psychology software E-prime version 1.0 [http://www.pstnet.com] and conducted in the Phonetics Laboratory of University of Edinburgh. Four native speakers of American English and four native speakers of British English participated in the experiment. Stimuli were presented to subjects over headphones in a sound-attenuated booth. Subjects were instructed to pay attention to the intonation of the stimuli and decide which interpretation was the more likely interpretation for each stimulus by pressing the corresponding button. A practice session was given prior to the experiment proper, to get subjects used to the experimental task. In the first experimental session, subjects judged the stimuli generated from the peak alignment continuum. In the second experimental session, subjects judged stimuli generated from the peak height continuum. Each peak variation was presented to subjects five times in a randomised order.

A timer with 1 ms accuracy was activated at the beginning of each stimulus and the RTs were recorded from the beginning of each stimulus until a response was given. The experiment was set up in such a way that the next stimulus was presented only when a response was given. However, subjects were instructed to press the buttons as quickly as they could, but not before the end of the utterance.

The two interpretations provided for the alignment continuum are (A) ' Is Birmingham the place you are moving to? ', the ' genuine (question) ' interpretation; (B) ' Do you really mean you are going to move to Birmingham? ', the ' incredulous ' interpretation. The two interpretations provided for the peak height continuum are (A) ' everyday occurrence '; (B) ' unusual experience ', as in [2].

Data of the response frequencies and the RT were automatically recorded in E-prime.

4. Statistical analysis and results

4.1. Peak height continuum

The response frequencies for each interpretation from all the eight subjects are shown in Figure 3. The response frequency-curves suggest that the first six peak heights form the category 'normal High accent' and the last six peak heights form the category 'emphatic High accent'. Peak heights 7, 8 and 9 appear to form the 'dynamic zone' of the continuum, where a steep decrease in the response frequencies for the 'everyday occurrence' interpretation but a steep increase in the response frequencies for the 'unusual experience' interpretation take place. We refer to them as the across-category peak heights.

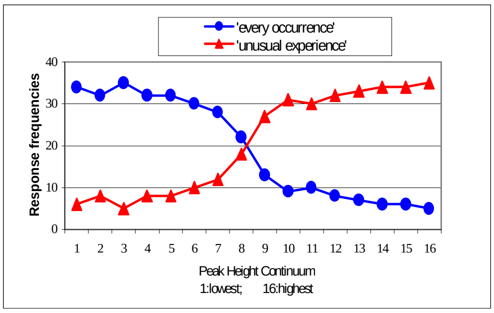


Figure 3. Response frequencies of the peak height continuum

To determine the boundary between categories, a linear regression analysis was performed with the response frequency as the predictor on four data points (peak heights 7, 8, 9, 10), which form the only near-straight line of the response curve for the interpretation 'unusual experience'. The analysis of linear regression shows that the response frequency can serve as an accurate predictor to the interpretation boundary (R2 = 0.981; F(1, 3) = 103.714, p = .01). The equation in (2) was used to calculate the location of the interpretation boundary. The value of b_0 is 5.23 (t =15.426, p =.004). The value of b_1 is 0.149. X was 18.5, the middle point between the lowest and highest response frequencies of the four data points.

$$Y = b_0 + b_1 \times X \Rightarrow 5.23 + 0.149 \times 18.5 = 7.99 \approx 8 \qquad (2)$$

The boundary predicted by the model is therefore peak height 8, i.e. 148 Hz in our male-voice stimuli, similar to the identification boundary in [2].

The mean RTs for identification, obtained by subtracting the duration of the utterance from the mean RTs recorded by E-prime, are shown in Figure 4.

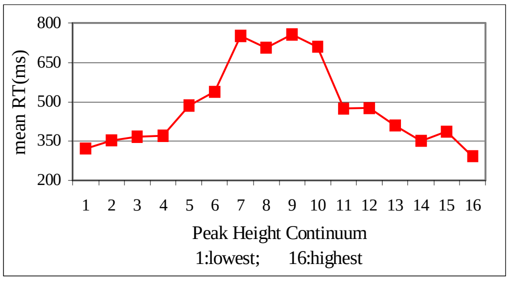


Figure 4. Mean RTs of the peak height continuum

Inspection of the mean RTs shows first that by and large subjects are slowest for stimuli at peak heights 7, 8, 9 and 10, which are across the identification categories, and fastest for the other stimuli, which are within identification categories. However, instead of a mean RT peak at peak height 8, the identification boundary, there appears to be a mean RT high plateau with two weak mean RT peaks at peak heights 7 and 9. Because it has been observed that American English has a smaller mean pitch range than British English, it is likely that the RT peak occurs earlier for the American English speaking subjects than for the British English speaking subjects. As a consequence, when collapsing the results from the two groups of subjects, the two RT peaks would manifest themselves as a high RT plateau with two weak peaks. Examination of the RT data obtained from the American English speaking subjects and the RT data from the British English speaking subjects confirmed this speculation. The RT peak occurs at peak height 7 in the 'American English data' but at peak height 9 in the 'British English data'. The difference in the mean pitch range of the two varieties of English is therefore a plausible explanation for the absence of a RT peak at peak height 8 and the presence of the two-peaked RT high plateau. Hence, on the whole, the response frequencies and the mean RTs show that the identification categories emerging from the interpretation task are linguistically real and the distinction between normal high accent and emphatic high accent is of a discrete nature.

4.2. Peak alignment continuum

Data of the response frequencies indicate that subjects opted more often for the 'genuine' interpretation than for the 'incredulous' interpretation across the continuum, as shown in Figure 5. This suggests that differences in the alignment do not appear to induce the binary meaning difference. In other words, peak alignment cannot be held solely responsible for the meaning difference at issue, provided that the two interpretations proposed for (2) in [3] and [4] are valid.

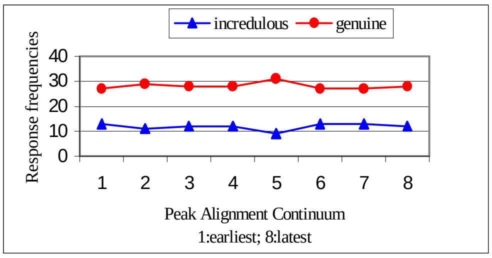


Figure 5. Response frequencies of the peak alignment continuum

The mean RT curve suggests that subjects were marginally quicker at stimuli with the peak aligned with the vowel onset of the stressed syllable and that of the following syllable than at stimuli with the peak aligned elsewhere, as shown in Figure 6. Moreover, the mean RTs of the alignment continuum are considerably longer than the mean RTs of the height continuum. The overall long RTs reflect the difficulty in interpreting intonational meaning when peak height does not contribute to the meaning and alignment is the only variable. On the whole, the discreteness of the peak alignment contrast supported by Pierrehumbert and Steele's production data is not born out in our on-line perception data.

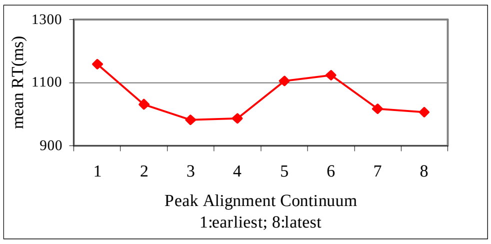


Figure 6. Mean RTs of the peak alignment continuum

5. Discussion and conclusions

Data obtained from the stimuli representing the peak height continuum clearly show that the distinction between normal High accent and emphatic High accent is discrete. By contrast, data obtained from the stimuli representing the peak alignment continuum do not support the discreteness of the distinction between early and late peak alignments claimed in [3]. In the light of present findings, we argue first that the binary meaning difference proposed for L*+H L H% and L+H* L H% by [3] and [4] is probably due to the co-variation of peak height and peak alignment rather than peak alignment alone in English. Second, the discreteness of the peak alignment contrast observed in production appears to be different in nature from the discreteness of the peak height contrast observed in perception. The former does not trigger a binary meaning difference associated with the contour pair in question while the latter does. A question arises as to how to represent these two types of discreteness in an intonation model such as the autosegmental-metrical model [7]. It is beyond the scope of the present paper to discuss the implications of these findings for representing intonational contrasts. But this issue is of great relevance to intonational studies and needs to be addressed properly in future research.

To conclude, by combining the response frequencies with the mean RTs in the semantically motivated identification task, we can distinguish the task-induced identification categories from linguistically real identification categories. Short mean RTs for within-category identification and long mean RTs for across-category identification are essential properties of linguistically real identification categories, in addition to the presence of a mean RT peak at the identification boundary. Making use of the attribute of RT as 'a positive function of uncertainty' [5], the present approach can determine the nature of intonational contrasts in a more accurate way than the conventional off-line methods.

Acknowledgements: I am grateful to Tim Mills for feedback on the stimuli, Carlos Gussenhoven, Bob Ladd, Bert Remijsen and Alice Turk for suggestions on the experimental design, Bob Ladd for discussions on the results, Toni Rietveld and Frans van der Slik for help with the statistical analysis, Jean-Pierre Sprengers for providing me with a home during the writing process and Carlos Gussenhoven for useful feedback on this paper.

6. References

[1] Gussenhoven, C., "Discreteness and gradience in intonational contrasts", *Language and Speech, Vol. 42, 1999: 283-305*.

[2] Ladd, D. R., "The perception of intonational emphasis: continuous or categorical?", *Journal of Phonetics Vol. 25, 1997: 313-342*.

[3] Pierrehumbert, J. B. and Steele, S. A., "Categories of tonal alignment in English", *Phonetica, Vol. 46, 1989: 181-196*.

[4] Ward, G. and Hirschberg, J., "Implicating uncertainty: the pragmatics of fall-rise", in *Language, Vol.* 61, *1985*: *747-776 .*

[5] Pisoni, D. B. and Tash, J., "Reaction times to comparisons within and across phonetic categories", *Perception & Psychophysics, Vol. 15(2), 1974: 285-290.*

[6] Birch, S. and Clifton, C. Jr., "Focus, Accent, and Argument Structure Effects on Language Comprehension", *Language and Speech, Vol. 38(4), 1995: 365-391*.

[7] Pierrehumbert, J. B., "Categories of tonal alignment in English", *Phonetica, vol. 46, 1989: 181-196.*