Discrimination of speech intonation contour: Evidence for tonetic categories?

A. Faulkner

Speech Research, IBM UK Scientific Centre,
Athelstan House, St Clement Street, WINCHESTER, SO23 9DR

Introduction

Intonation is important in both the production and the recognition of speech. In particular, intonation is related to syntax and semantics, and is an essential component in the synthesis of 'natural' speech. Linguistic theories typically treat intonation as categorical in nature. The American school [1] refer to categories of pitch level, while the British School [2, 3] refer to categories of pitch accent. For example, Halliday [3] proposes categories of primary tone-unit having forms such as *fall, rise, fall-rise*, and *rise-fall*. Although intonation certainly employs degrees of emphasis, or reservation, etc., expressed through a continuum of pitch prominence, it is generally considered that the principal organization of the intonational system is categorical.

A categorical structure is clearly convenient for linguistic analysis, but if such a structure is to be useful in speech synthesis and recognition, this categorical organization should be reflected in categorical perception. The classical notion of categorical perception [4] proposes that speech sounds falling within a single phonetic category can not be discriminated from each other, since pre-categorical acoustic information is supposed to be unavailable to the speech perception system. It is now clear that strictly categorical perception does not occur [5], but rather, stimuli from different categories are markedly more discriminable than stimuli within a single category.

The question of whether the perception of intonation is categorical can thus be cast as the empirical question of whether the discriminability of intonation contour shows peaks in regions where category boundaries are expected. Ainsworth and Lindsay [6] posed essentially this question. They chose to examine discriminability among a set of rather complex 5-section f_0 contours, and their results indeed showed discrimination peaks. Although Ainsworth and Lindsay used a criterion of perceptual discriminability, they were principally concerned to establish boundaries in linguistic rather than perceptual terms. An understanding of the organisation of speech perception demands an psychophysical analysis as well as a linguistic one. A category boundary must be located on some perceptual dimension, and it is important to discover the relation of that perceptual dimension to the physical dimensions of intonation. We might expect the perceptual categorization of intonation contour to depend upon the direction of pitch change, and the rate of that change. Hence, this study analyses discriminability along the physical dimension of f_0 slope.

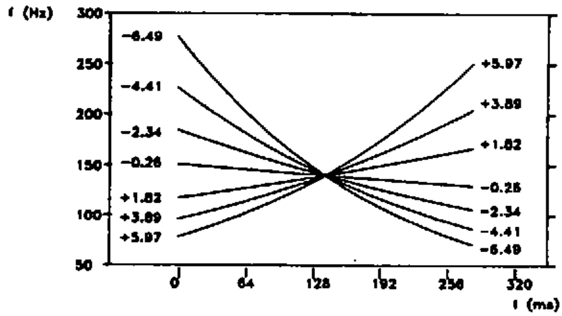
In this context, a comment on the measurement of discriminability may be helpful. An analysis of discriminability requires a bias-free measure of discrimination performance such as that offered by Signal Detection Theory [7,8]. The 'same-different' task frequently used in speech discrimination experiments, including [6], does not meet this criterion, since listeners' willingness to call a pair of stimuli the same may reflect some cognitive or linguistic judgement rather than perceptual evidence. To measure discriminability without bias, we can employ a forced-choice method, in which the listener knows some difference is present, and must locate that difference. Discriminability itself will be measured by the d' statistic [8].

Experiment 1

Method

Stimuli: The stimuli were LPC resyntheses of the word "yes", based upon the natural speech of a male RP speaker. The natural speech LPC analysed using the autocorrelation method [9]. The f_0 contours of the LPC resynthesized stimuli were logarithmic about a fixed mean frequency of 140 Hz, and had slopes of between -7.54 and +7.02 octave/s. Examples of the f_0 contours are shown in figure 1. The total duration of each stimulus was about 390 ms, made up of about 20 ms of prevoicing aspiration, the 282 ms voiced portion, and about 90 ms of unvoiced frication.

Figure 1. The f_0 slope continuum used in experiment 1. The figure shows examples of the f_0 contours. The contours are labelled with their slopes in octave/s.



Apparatus: The original speech was digitized at 10 kHz onto the IBM-PC based speech analysis system SAY [10], and transferred to a host system for LPC analysis and resynthesis. The stimuli were played back through 12 bit DACs at a 10 kHz sample rate, and recorded onto tape after low-pass filtering at 4.5 kHz. During the experiment, the stimulus tapes were reproduced through loudspeakers at a peak SPL of 60 dB at 1m. The listeners were seated about 1.5 m from the loudspeakers, in a quiet carpeted room.

Subjects: Listeners were the author and two students from the IBM UK Scientific Centre, and 8 subjects recruited by the subject panel of the University of Oxford Department of Experimental Psychology. The IBM listeners knew the purpose of the experiment, and were familiar with LPC resynthesized speech. The Oxford listeners were unaware of the purpose of the experiment, and unfamiliar with LPC resynthesized speech.

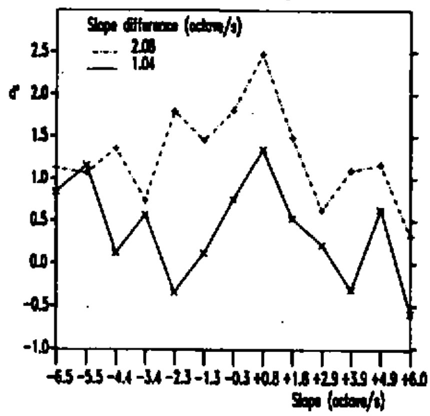
Procedure: The ABX discrimination method was used. Each trial was composed of two successive test stimuli, A and B, whose slopes differed, and a final comparison stimulus

X, which was identical to either A or B. The subjects' task was to choose which of the two stimuli A or B was the same as X. The mean f_0 slope of the A and B stimuli varied from -6.49 to +5.97 octave/s in steps of 1.04. The difference in slope between A and B was 1.04, or 2.08 octave/s. The interstimulus interval was 0.5 s, and the intertrial interval, inclusive of the response interval, was 2.8 s.

Results

The d' statistic was computed for each subject. Discriminability as a function of the mean slope of the A and B stimuli, as measured by the group average of d', is shown in figure 2. A repeated-measures ANOVA showed that larger slope differences led to more accurate discrimination: $F(1,120) = 13.02, p < 0.005$, and that discrimination depended upon the slope; $F(12,120) = 4.42, p < 0.001$. There was also a significant interaction between the slope and the slope difference: $F(1,120) = 2.81, p < 0.002$, largely due to the close to chance performance shown irrespective of slope difference at more extreme slopes. A Newman-Keuls test showed that discrimination was significantly better at a slope of +0.78 octave/s than elsewhere.

Figure 2. d' as a function of f_0 slope
Data from experiment 1. Each point is based on 220 trials.



Discussion

A significant discrimination peak was found for slopes around zero, consistent with categorical perception. The monotonic f_0 continuum seems to show a single category boundary which distinguishes two basic intonational categories, *rise* and *fall*. There is also some degree of continuous perception.

Experiment 2

In the hope of discovering the contours which define other intonational categories, we now consider contours composed of two sections. Here, the first section of the contour will have a fixed rising or falling slope, while the second section will be varied to form the stimulus continuum. Among the set of contours so generated are the *fall*, the *rise*, the *fall-rise*, and the *rise-fall*.

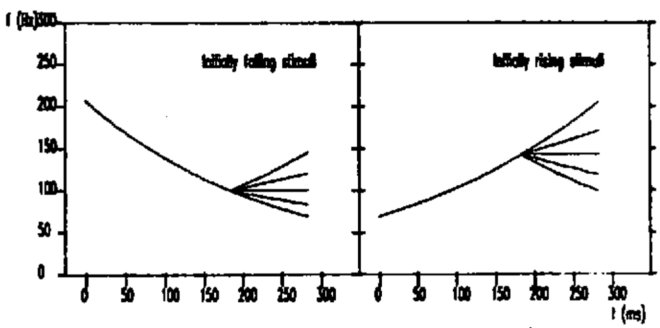
Method

Stimuli: The stimuli were tokens of the word "yes" produced by a software synthesizer [11]. The synthesis parameters were based on an analysis of the natural speech example used in experiment 1. The f_0 contours were composed of a fixed section comprising the initial 185 ms of the voicing, and a variable section comprising the remaining 100 ms of voicing. The fixed initial section had either a falling slope, of -5.32 octave/s, or a rising slope of +5.32 octave/s. The variable section had a slope between -7.64 and +8.31 octave/s. The two stimulus continua are illustrated in figure 3. The fundamental frequency at the temporal centre of the voiced section was always 120 Hz.

Subjects: Subjects were the author and 2 students from the IBM UK Scientific Centre, all experienced listeners, and 11 naive subjects recruited by the Department of Experimental Psychology, University of Oxford.

Procedure: The experiment was run as 2 sub-experiments with minor procedural differences between the two. In both cases, the ABX method was used. For the IBM subjects, the mean slope of the second section of the A and B stimuli was varied in steps of 1.33 octave/s between -6.98 and +4.99 octave/s. The slopes of the A and B stimuli differed by 1.33 or 2.66 octave/s. For the Oxford subjects, the slope was varied between -5.66 and +6.32 octave/s, and since the discrimination of this unpractised group was expected to be less acute, the slope differences were doubled to 2.66 and 5.32 octaves/s.

Figure 3. Two-section f_0 continua used in experiment 2.



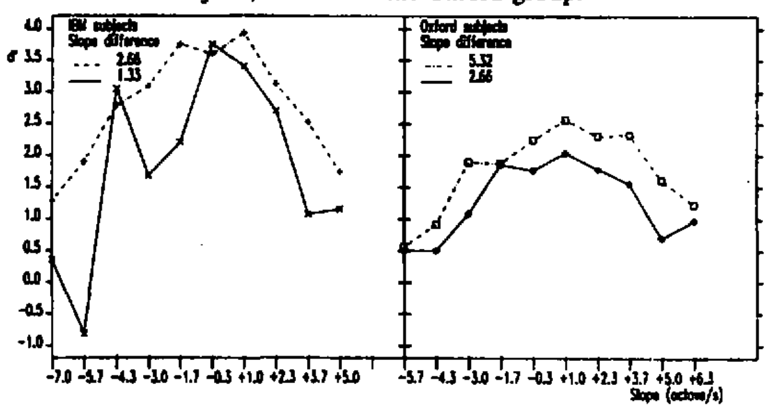
Results

Results for the two sub-experiments will be presented separately. Three of the Oxford subjects failed to show any statistically significant discrimination: their data were excluded. One further Oxford subject failed to respond on a large number of trials, and was also excluded. The discrimination scores from the 3 IBM subjects, and the remaining 7 Oxford subjects, are shown in fig. 4 for the initially falling f_0 contours, and fig. 5 for the initially rising contours. The individual d' scores from the initially rising and initially falling stimuli from the two sub-experiments were analysed in 4 separate repeated-measures ANOVAs.

Initially falling stimuli

IBM subjects: The discrimination function shows a broad peak around a flat f_0 slope. The ANOVA of d' showed significant main effects of slope: $F(9,18) = 12.12, p < 0.001$, and of slope difference: $F(1,2) = 20.13, p < 0.05$, and there was also a significant interaction between slope and slope difference: $F(9,18) = 8.81, p < 0.001$. A Newman-Keuls test showed superior discrimination at slopes of -0.33 and + 1.0 octave/s than at the more extreme slopes of +3.66, +4,99, -5.65, and -6.98 octave/s. Further, discrimination for all slopes between -4.32 and +2.33 octave/s was significantly better than for slopes of -5.65 and -6.98 octave/s.

Figure 4. Initially falling contours
d' as a function of the slope of the second section. Each point is the average of 72 trials for the IBM subjects, and 84 for the Oxford group.



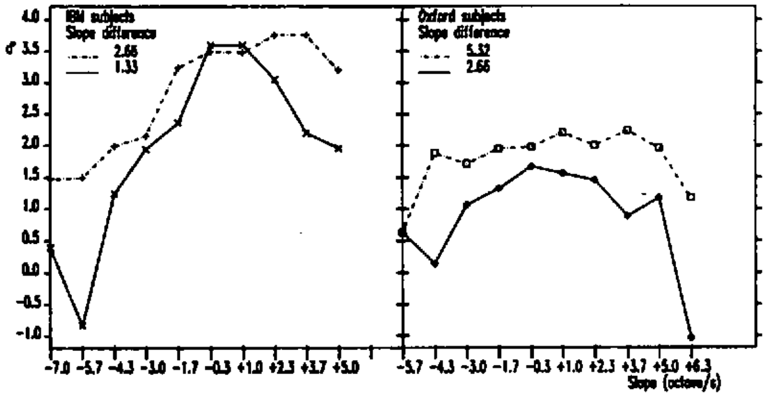
Oxford subjects: The Oxford subjects also showed a discrimination function that peaked around a flat slope. The ANOVA showed significant effects of slope: $F(9,54) = 5.94, p < 0.001$, and slope difference: $F(1,6) = 19.44, p < 0.005$. A Newman-Keuls test showed that d' for slopes between -0.33 and + 3.66 octave/s was greater than at more extreme rising or falling slopes.

Both groups of subjects thus show a significant discrimination peak centred on a slope of between zero and +1 octave/s.

Initially rising stimuli

IBM subjects: The ANOVA showed a significant main effect of slope: $F(9,18) = 20.91, p < 0.001$, and an effect of slope difference: $F(1,2) = 616.04, p < 0.002$. There is no clear evidence of a discrimination peak. A Newman-Keuls test showed d' to be greater for slopes between -1.66 and +5.00 octave/s than for slopes of -2.99 and more extreme negative values. negative values. An inspection of fig. 5 suggests that there may be a discrimination peak around a slope of zero for a slope difference of 1.33 octave/s, but since the interaction between slope and slope difference is not significant, the evidence is not strong.

Figure 5. Initially rising contours
d' as a function of slope of the second section. Each point is the average of 72 trials for the IBM subjects, and 84 for the Oxford group.



Oxford subjects: An ANOVA showed significant main effects of slope: $F(9,54) = 4.87, p < 0.001$, and slope difference: $F(1,6) = 99.55, p < 0.001$, and no significant interactions. A Newman-Keuls test showed that d' was significantly less for the extreme slopes of -5.65 and +6.32 octave/s than for slopes between -3.99 and 4.99 octave/s. The Oxford group, unlike the IBM group, do seem to show a slight discrimination peak for the initially rising stimuli. This peak is, however, very broad, and cannot be located with any confidence.

For the initially rising stimuli, neither group shows a discrimination peak; rather discrimination improves as the f_0 contour becomes closer to a smooth rise.

Discussion

The clearest result of this experiment is the discrimination peak for 2-section f_0 contours which initially fall. The peak occurs where the second part of the f_0 contour is approximately flat, or slightly rising. This discrimination peak could be taken to indicate a category boundary between tonetic categories of *fall* and *shallow fall*, or, in Halliday's terminology, *fall* and a *fall-low rise*.

A similar peak may also be present with an initial rise, but the data do not reliably demonstrate this. Further experiments are required to indicate whether the f_0 slope of the second section is in itself important, or whether the slope of this section relative to the first section determines the location of the discrimination peak.

Summary and Conclusions

1/ The experiments support the claim that pitch variation in intonation contours can be perceived categorically.
2/ Discrimination peaks were found which correspond to a single psychoacoustic category boundary distinguishing *falling* from *rising* f_0 contours.
3/ This single psychoacoustic fall/rise category opposition can account for a variety of contrasts between different forms of intonation contour. These experiments illustrate the function of this opposition in distinguishing firstly a *fall* from a *rise* on a monotonic f_0 slope continuum, and also in distinguishing a *fall* from a *shallow fall* or *fall-low rise* in a two-section contour whose initial section falls. There is suggestive evidence for a categorical distinction between *rise* and a *rise-fall* contours.

References

[1] G. L. Trager and H. L. Smith, 'Outline of English Structure', Battenburg Press, Norman, Oklahoma, (1951).

[2] J. D. O'Connor and G. F. Arnold, 'Intonation of colloquial English', Longman, London, (1961).

[3] M. A. K. Halliday, 'The tones of English', Arch. Linguist., Vol 15, 1-28, (1963).

[4] A. M. Liberman, K. S. Harris, H. S. Hoffman, and B. C. Griffith, 'The discrimination of speech sounds within and across phoneme boundaries', J. Exp. Psychol., Vol 54, 358-368, (1957).

[5] B. H. Repp, 'Categorical perception: Issues, methods, findings': In N. J. Lass (ed.) 'Speech and Language: Advances in basic research and practice, Vol 10', Academic Press, Orlando, 243-335, (1984)

[6] W. A. Ainsworth and D. Lindsay, 'Perception of pitch movement on tonic syllables in British English', J. Acoust. Soc. Am., Vol 79, 472-480, (1986).

[7] D. M. Green and J. A. Swets, 'Signal detection theory and psychophysics', Wiley, New York, (1966).

[8] N. A. Macmillan, H. L. Kaplan, and C. D. Creelman, 'The psychophysics of categorical perception', Psychol. Rev., Vol 84, 452-471, (1977).

[9] J. D. Markel and A. H. Gray Jr., 'Linear prediction of speech', Springer Verlag, New York, (1976).

[10] P. R. Alderson, G. Kaye, S. G. C. Lawrence, and D. A. Sinclair, 'A speech analyser based on the IBM personal computer', Proc. of Institute of Acoustics Autumn Conference, Windermere, (1984).

[11] D. H. Klatt, 'Software for a cascade/parallel speech synthesiser', J. Acoust. Soc. Am., Vol 67, 971-995, (1980).