IDENTIFICATION AND DISCRIMINATION OF HALLIDAY'S PRIMARY TONES.

W.A. Ainsworth and D. Lindsay

Dept. of Communication and Neuroscience, University of Keele,
Keele, Staffordshire, England.

INTRODUCTION

Halliday [1] suggested that one way of describing the intonation of English was to divide each utterance into a series of tone groups. Each tone group could then be divided into the 'pretonic', the syllables which preceded the most prominent syllable, and the 'tonic', the remainder of the tone group. Furthermore, he suggested that it was the pitch change on the tonic which determined the type of the utterance. If the tonic consisted of a single syllable it could be one of five types, and the pitch change associated with it would follow one of five time courses. These pitch changes were called the primary tones. They are shown in Table 1 together with their intended meanings.

Table 1. Pitch changes associated with the primary tones.

Tone	Pitch change	Semantic label
1	Fall	Statement
2	Rise or fall-rise	Question
3	Level - slight rise	Weak statement
4	(Rise) - fall - rise	Reservation
5	(Fall) - rise - fall	Emphatic statement

The purpose of the present experiments was to synthesise a set of utterances having intonation contours which included those described in Table 1, and to try to identify the tones by means of the semantic labels. In addition, it was proposed to measure discrimination functions for the tones and to investigate the correlation between the discrimination and identification functions.

SYNTHESIS OF STIMULI

The stimuli consisted of sixty versions of the carrier word 'yes' with different intonation contours. They were generated by a synthesis-by-rule system [2]. Each stimulus consisted of six sections as described in Table 2. The first section contains formant transitions appropriate for a /j/ sound before an /ɛ/ vowel. During this section the fundamental rose from 120 Hz to 150 Hz. The next four sections consisted of the vowel /ɛ/. During the first 200 ms the fundamental remained constant at 150 Hz. In the next section the fundamental changed to a frequency P1 which could take one of six values in the range 100 to 215 Hz. The fundamental then changed to 100 Hz during the next section which was 100 ms long. In the last vowel section, also 100 ms in duration, the fundamental changed to a frequency P2 which could take one of ten values in the range 50 to 215 Hz. The final section consisted of a voiceless /s/ sound.

Table 2. Duration and fundamental frequency of each section of the stimuli.

Phoneme	Duration (ms)	Final FO (Hz)
	-	120
/j/	100	150
/ɛ/	200	150
/ɛ/	10	P1(100 - 215)
/ɛ/	100	100
/ɛ/	100	P2(50 - 215)
/s/	100	-

IDENTIFICATION EXPERIMENTS

The object of this experiment was to investigate whether listeners could identify the stimuli by means of the semantic labels shown in Table 1. Two experienced listeners took part. They were presented with the sixty stimuli ten times in different random orders. They were asked to classify each stimulus by pressing one of five switches labelled as shown in the last column of Table 1.

The response patterns obtained were similar for both listeners. Tone 1 was selected when the stimulus contained low values of both P1 and P2. Tone 2 was heard with low values of P1 and high values of P2. Intermediate values of both P1 and P2 caused tone 3 to be chosen. Tone 4 was perceived when both P1 and P2 were high, and tone 5 when the stimulus had high values of P1 and low values of P2.

The centroids of each of the response areas in the P1-P2 response space were calculated. From the values of P1 and P2 for these centroids the change in fundamental associated with each tone was calculated. The results are shown in Table 3.

Table 3. Fundamental frequency changes associated with each tone derived from the identification experiments.

Tone	Fundamental frequency change
1	Fall of 66 Hz
2	Fall of 50 Hz, rise of 88 Hz
3	Fall of 50 Hz, rise of 9 Hz
4	Rise of 19 Hz, fall of 69 Hz, rise of 38 Hz
5	Rise of 27 Hz, fall of 103 Hz

It will be seen that the fundamental frequency changes correspond more or less exactly with the pitch changes suggested by Halliday [1]. Tone 1 implies a statement, and is represented by a fall. Tone 2 implies a question, and is represented by a fall-rise. The first option, a simple rise, was precluded by the stimuli which all contained fundamentals starting at 150 Hz and passing through 100 Hz. Tone 3, a weak statement, is represented by a fall followed by a slight rise. This is slightly different from that suggested by Halliday (level followed by a slight rise) again because of the 100 Hz fixed point. Tone 4, reservation, is represented by rise-fall-rise, and tone 5, emphatic statement, by rise-fall.

IDENTIFICATION AND DISCRIMINATION OF HALLIDAY'S PRIMARY TONES.

DISCRIMINATION EXPERIMENTS

The fact that the tones could be classified by identifying them by means of semantic labels suggests the possibility of categorical perception of nuclear intonation. One way of testing this hypothesis is by measuring discrimination functions [3]. The idea is that pairs of tones which span a perceptual boundary should be easier to discriminate than pairs which belong to the same perceptual category. This would lead to peaks in the discrimination function which correspond to boundaries in the identification function.

A preliminary experiment was carried out to determine an appropriate step size for the difference between stimulus pairs. If this is too large all the pairs will be discriminable, but if it is too small, none of them will.

Nine listeners took part in the main discrimination experiments. They heard six sets of stimuli ten times each. Two sets had P1 fixed at a frequency of 120 or 157 Hz and P2 varying between 50 and 215 Hz, and four sets had P2 fixed at 63, 84, 120 or 157 Hz and P1 varying between 100 and 215 Hz. The listeners were asked to decide whether the members of each pair of sounds were the same or different, and to press an appropriately labelled switch. They were not told the purpose of the experiment.

Table 4. Obtained and predicted discrimination functions (%) for fixed values of P1

P2 frequency (Hz)	P1 = 120 Hz		P1 = 157 Hz	
	Obtained	Predicted	Obtained	Predicted
52	7	13	11	20
63	28	7	35	7
74	74	13	73	20
84	80	80	72	53
103	67	13	60	40
120	44	40	36	27
138	44	47	38	33
157	53	80	50	47
177	45	20	45	47

The results are shown in the 'obtained' columns in Tables 4 and 5. For a fixed P1 peaks occurred for values of P2 of 84 Hz and 157 Hz. For a fixed value of P1 they occurred for a value of P2 of 138 Hz. The positions of these peaks occurred at approximately the positions of the boundaries obtained in the identification experiments.

CORRELATION

In order to investigate the categorical perception hypothesis in more detail, discrimination functions were predicted from the identification functions by means of a formula given by Liberman et al [3]. The results are shown in the 'predicted' columns of Tables 4 and 5. Although the predicted values of discriminability differ from the ones obtained directly, the peaks occur at about the same frequencies and a high correlation is suggested.

Table 5. Obtained and predicted discrimination functions (%) for fixed values of P2.

P1 frequency (Hz)	P2 = 63 Hz		P2 = 84 Hz		P2 = 120 Hz		P2 = 157 Hz	
	O	P	O	P	O	P	O	P
103	46	13	36	17	26	27	15	7
120	42	0	41	7	29	27	26	53
138	85	27	86	33	81	53	70	53
157	83	87	80	93	86	40	80	60
177	52	0	59	0	56	47	54	27

The correlations between the obtained and predicted data were tested in two ways. Pearson's correlation coefficient was calculated and is shown in Table 6. The correlations were significant for both the values of P1 at the $p<0.001$ level, and for all the values of P2 at the $p<0.001$ level except for P2 = 84 Hz where it was significant at the $p<0.001$ level.

One of the limitations of Pearson's method is that it assumes that the distributions of data being correlated are similar. This is not necessarily the case, so a non-parametric method, Spearman's, was also employed. The correlations were again shown to be significant.

Table 6. Correlation between obtained and predicted discrimination functions according to (a) Pearson's method and (b) Spearman's method.

Fixed frequency (Hz)	(a) Pearson's method			(b) Spearman's method		
	n	r_s	p<	n	r_s	p<
P1 = 120	81	0.284	0.01	81	0.287	0.05
P1 = 157	81	0.320	0.01	81	0.330	0.01
P2 = 63	45	0.522	0.001	45	0.641	0.001
P2 = 84	45	0.471	0.01	45	0.515	0.001
P2 = 120	45	0.654	0.001	45	0.629	0.001
P2 = 157	45	0.530	0.001	45	0.591	0.01

CONCLUSIONS

It has been shown that the shapes of the five primary tones of Halliday can be determined from the results of an identification experiment using semantic labels. Discrimination functions for the five tones have been measured and have been shown to have a significant positive correlation with the discrimination functions predicted from identification data. This result suggests that the perception of the tones is categorical.

Acknowledgements

The work was supported in part by the Joint Speech Research Unit, Cheltenham, England.

REFERENCES

[1] M.A.K. Halliday, 'The tones of English', Archivum Linguisticum, vol. 15, 1-28, (1963).

[2] W.A. Ainsworth, 'Performance of a speech synthesis system', Int. J. Man-Machine Studies, vol. 6, 493-511, (1974).

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