JUDGEMENT OF THE VOWEL COLOUR OF NATURAL AND ARTIFICIAL SOUNDS

JOHN MORTON and ALAN CARPENTER
Medical Research Council, Applied Psychology Research Unit, Cambridge

Five vowel sounds, each produced in five different ways, were presented to subjects with widely differing experience of tasks of this type for identification.

The vowels were the last five of the standard set of eight cardinal vowels, and were presented to subjects with no experience of phonetics as belonging to the words *bat, bart, bought, boat* and *boot*. Each vowel was (a) spoken by a phonetician, (b) synthesized by adding individual harmonics in direct imitation of the spoken sounds, (c) represented by two pure tone components only, (d) synthesized by a talking machine (P.A.T.) and (e) imitated by selecting harmonics with a low-pass filter.

All these sounds were categorised more or less "correctly" by skilled subjects; and, although there were wide and systematic differences in accuracy with untrained subjects, these performed consistently better than chance.

The experiment showed (a) that the selection of harmonics according to formant theory is not the only, and perhaps not always the best means of synthesizing isolated vowels, (b) that even sounds which are poor representations of vowels can still be categorised with some consistency even by untrained subjects and (c) that a simple "percent correct" score is less sensitive than other measures.

INTRODUCTION

The purpose of this experiment is to investigate the relation between various physical properties of sound stimuli and the perception of these sounds as vowels. This subject has been studied since the time of Helmholtz, Paget and Miller, and it is now generally accepted that the vowel colour of a sound is completely specified by the position on the frequency scale of two or three formants. This formant theory is quite appropriate to theoretical analyses of the operation of the vocal tract, and fits very well most of the current phonetic classifications of speech sounds.

On the other hand, very little is known about the perception of complex sounds, and about what physical qualities enable sounds to be identified as speech, and categorised. Further, analyses of natural speech not infrequently show the formant structure to be either poorly defined, or absent. For example Ladefoged (1959, p. 50) made a detailed analysis of 31 sets of cardinal vowels spoken by experienced phoneticians and failed to locate the positions of one or more of the first three formants in 26 versions of /u/, 12 each of /o/ and /ɔ/, and 16 of /ɑ/. These vowels were all accepted as good vowels by the phoneticians who produced them, and by other impeccable authorities.

Talking machines such as the Parametric Artificial Talker (P.A.T.) (Lawrence, 1953) produce very realistic speech in which the vowels have clearly specified and formed formants. It appears certain that formant structure is sufficient to enable sounds to be accepted and identified as vowels. Yet the present hypothesis is that formant structure may not be necessary for vowel colour to be perceived. Direct confirmation of this was sought in an experiment described in another paper (Carpenter and Morton, 1962). The present experiment, while making a fairly wide review of sounds made in different ways, is more directly aimed at solving the experimental problem of the best way in which those sounds may be presented for judgement.

Except to phoneticians, the identification of isolated vowel sounds is a very artificial task. It is necessary to present this task in a way which produces results which are both consistent and unambiguous and also relevant to the perception of normal speech. In an earlier experiment (Carpenter, 1962) the first part of this was achieved but probably not the second. In that experiment, spoken and synthetic sounds were compared using an ABX framework, i.e. sounds were presented in groups of three and the subject was required to say which of the first two the third more resembled. Synthetic vowel sounds were associated with the appropriate spoken vowels in spite of experimental conditions which could have led to them being associated otherwise. Nevertheless the sounds could be directly compared with each other as arbitrary noises, and without them being taken as representative of the sounds of speech, either remembered from the subjects' background experience or provided in the experiment.

Experiment I

In the present experiment each sound is presented by itself, and a form of absolute judgement is called for ; that is, the subjects' background knowledge of the language is used. The subjects were asked to select the most appropriate category from a specified list, in each case. For those who were phoneticians the list consisted of phonetic symbols ; for untrained subjects it was a list of words from which the sounds could have come.

Materials

The experiment compares five ways of shaping the harmonic structure of five vowel sounds. The vowels were the primary cardinal vowels numbered 4 to 8, i.e.

/a/ as in *bat*
/ɑ/ as in *bart*
/ɔ/ as in *bought*
/o/ as in *boat*
/u/ as in *boot*

Each of these was produced in the following five ways.

I *Natural vowels.* Spoken by a phonetician.[1]

II *Complete imitation.* Individual harmonics were added in direct imitation of the spectra of the natural vowels.

III *Peaks only.* This was a simplified version of II including only the most prominent components. Faced with the hypothesis that the analysis of vowel sounds by ear involves some method of locating the amplitude peaks which locate the formants, we were curious as to whether presenting the harmonics nearest to the peaks of the first two formants would enable the vowel to be identified.

IV *P.A.T.—two formants.* A set of vowels produced by P.A.T., an artificial talking machine which operates by passing a stream of pulses into two resonators which represent the two primary vocal tract resonances—the formants. These sounds were included to provide a comparison between two kinds of artificial source.

V *Low-pass filter—no formants.* A set of noises produced by passing the output from a pulse source through a low-pass filter with cut-off points determined empirically by the experimenters to give the closest imitation of the five vowels. Such sounds have no formant structure, in that these spectra show no amplitude peaks, and thus whatever vowel quality they possess cannot be attributed to the discrimination of such peaks.

The spectral composition of these twenty-five sounds, as confirmed by Sonagraph analysis, are shown in Table 1, and a diagram showing the five versions of the vowel /ɑ/ is given in Fig. 1.

An electronic gate, open for 0.8 sec., and with the on and off transitions smoothed by a time constant of 0.01 sec., was used to switch the stimuli, which were recorded on magnetic tape, using a Vortexion recorder and a tape speed of $7\frac{1}{2}$ i.p.s. A separate good quality loudspeaker was used to reproduce the stimuli in the experimental runs.

The various sounds differed in reproduced sound level over a range of almost 10 db. When a measurement was made in a typical case the louder sounds gave readings of 70 db on a sound level meter with flat response.

Subjects

A total of 46 subjects took part in the experiment. They covered a very wide range of skill, and fell conveniently into three classes.

A. Phonetically trained subjects. This group comprised 7 phoneticians, 6 phonetics students, and 7 speech therapy students, all provided by the Phonetics Department of the University of Edinburgh.

B. Subjects unfamiliar with phonetics but familiar with experimental situations. This group consisted of 11 experimental psychologists, all members of the staff of the Applied Psychology Research Unit.

[1] *Recordings kindly provided by Dr. P. Ladefoged.*

Table 1

Specification of the Stimuli

Source	Stimulus	Relative harmonic amplitude 1	2	3	4	5	6	7	8	9	10	11	12	13	Formant position for source IV (cps.)	Cut-off for source IV in experiment II
I II	1	8	12	14	15	15	24	12	20	20	16	1				
III					24			27	27							
IV		3	13	21	28	11	17	27	17	15	1				720, 1,250	2,400
V		32	32	32	31	31	30	30	30	29	28	28	28	27		
I II	2	10	14	22	17	24	23	2								
III				19		30										
IV		6	18	28	19	20	23	19	15	11	7	7	1		615, 950	1,700
V		32	32	32	31	31	30									
I II	3	20	26	27	14	13	4	1								
III				30		22										
IV		9	24	29	23	23	17	14	13	9					510, 810	1,200
V		32	32	32	31											
I II	4	15	27	4	1											
III			30		22											
IV		9	29	19	18	13	6.5	10	3						400, 650	850
V		32	32													
I II	5	28	8													
III		30														
IV		21	29	18	13	9	2								250, 610	600
V		32														

Note : The harmonics are multiples of 180 cps. As the arbitrary reference level for the db scale was not strictly controlled, comparisons from one horizontal line to another are illegitimate.

C. Completely untrained subjects, typical of the average population of the country. In this group there were 9 ratings from the Royal Navy, all in their late 'teens or early twenties, and 6 post-office workers in middle age.

Method

Subjects performed the experiment in groups depending upon their availability. Group size ranged from three to six.

Subjects of group A were told : " You are about to hear a series of real and

Morton, John, Applied Psychology Research Unit, Cambridge, ***Judgement of the vowel colour of natural and artificial sounds*** , Language and Speech, 5:4 (1962:Oct./Dec.) p.190

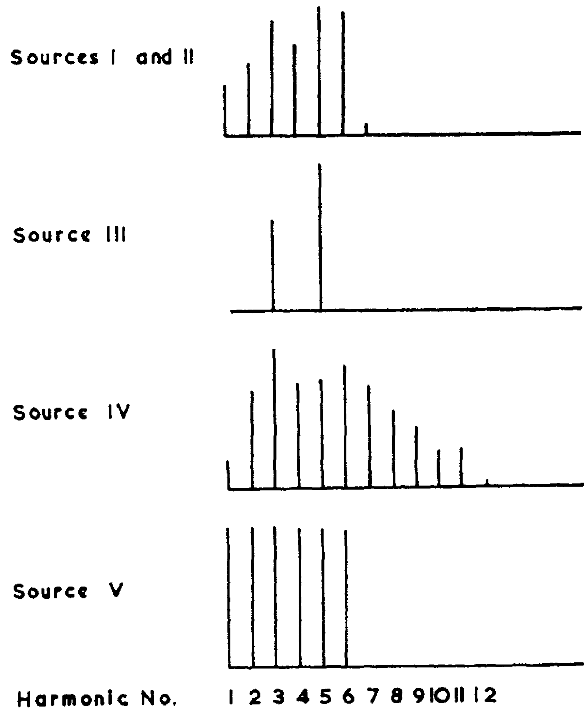

Fig. 1. Representation of the different versions of /ɑ/. The height of the lines represents the relative intensity of the harmonics.

imitation versions of the last five cardinal vowels. Please write on this prepared answer sheet the phonetic symbol appropriate to the vowel which you hear in each case."

With the subjects of groups B and C more explanation was necessary. They were told : " You will hear a series of sounds which represent the vowels contained in the words written on your answer sheet. Some of the sounds were actual spoken vowels, others are imitations. In each case please write in the successive spaces on the paper, the number of the word from which the sound could have come. If you are in doubt, guess ! ". The words were : (1) *bat ;* (2) *bart ;* (3) *bought ;* (4) *boat ;* (5) *boot.* The words were read out to the subjects, and the principle of a steady vowel, with no initial or final transition, was demonstrated for each of the vowels.

Results

Table 2 summarises all the results. For each group of subjects and for each sound source the results were summed on a matrix of which the rows represent the five stimulus vowels, and the columns the responses given. For each subject the first block of 25 responses, one to each stimulus, was discarded, leaving four responses to each of the 25 sounds.

Because of the unequal numbers in the three subject groups, Table 2 gives the *percentages* of the responses to each stimulus to facilitate comparison.

Morton, John, Applied Psychology Research Unit, Cambridge, ***Judgement of the vowel colour of natural and artificial sounds*** , Language and Speech, 5:4 (1962:Oct./Dec.) p.190

TABLE 2

Percentage of Responses to each Stimulus

SOURCE I

STIMULUS	/a/	/ɑ/	/ɔ/	/o/	/u/
GROUP A	Trained Subjects (N = 20)				
1	**88**	12			
2	1	**51**	48		
3		1	**78**	21	
4			1	**95**	4
5					**100**
GROUP B	Psychologists (N = 11)				
1	**36**	52	11		
2		**55**	36	9	
3	5		**82**	14	
4		2	36	**59**	2
5	2	5	5	23	**66**
GROUP C	Naive Subjects (N = 15)				
1	**10**	63	13	10	3
2	15	**58**	15	8	3
3	10	17	**45**	25	3
4	5	13	45	**20**	17
5	5	5	10	30	**50**

SOURCE II

STIMULUS	/a/	/ɑ/	/ɔ/	/o/	/u/
GROUP A					
1	**85**	15			
2	3	**81**	16		
3		6	**74**	20	
4			3	**94**	3
5					**100**
GROUP B					
1	**41**	43	5	7	
2	5	**77**	14	5	
3	2	11	**59**	25	2
4	2	5	20	**61**	11
5	2		5	9	**84**
GROUP C					
1	**20**	58	7	10	5
2	27	**40**	18	12	3
3	12	20	**27**	28	13
4	8	13	13	**37**	28
5	2	3	10	32	**53**

SOURCE III

STIMULUS	/a/	/ɑ/	/ɔ/	/o/	/u/
GROUP A					
1	**37**	44	11	5	3
2	1	**11**	60	24	4
3			**55**	37	8
4			9	**60**	31
5					**100**
GROUP B					
1	**27**	52	7	9	5
2	2	**32**	45	14	7
3	7	11	**50**	16	16
4	5	2	41	**23**	30
5	2	7	2	23	**66**
GROUP C					
1	**42**	15	23	5	15
2	25	**13**	25	13	23
3	18	10	**37**	12	23
4	20	3	17	**28**	32
5	3	5	12	20	**60**

SOURCE IV

STIMULUS	/a/	/ɑ/	/ɔ/	/o/	/u/
GROUP A					
1	**74**	22	1	3	
2	59	**30**	11		
3	5	19	**75**	1	
4	3		8	**89**	1
5	4	4	15	74	**4**
GROUP B					
1	**50**	36	14		
2	50	**34**	11	5	
3	7	32	**34**	27	
4	2	20	34	**41**	2
5	7	16	25	50	**2**
GROUP C					
1	**25**	55	8	10	2
2	28	**43**	10	13	5
3	13	43	**15**	25	3
4	10	35	27	**22**	7
5	2	27	23	32	**17**

SOURCE V

STIMULUS	/a/	/ɑ/	/ɔ/	/o/	/u/
GROUP A					
1	**66**	28	6		
2	3	**76**	18	4	
3		1	**55**	44	
4			1	**58**	41
5				1	**99**
GROUP B					
1	**39**	55		7	
2		**43**	41	14	2
3		2	**55**	41	2
4			11	**27**	61
5			7	9	**84**
GROUP C					
1	**25**	48	5	15	7
2	7	**47**	25	17	5
3	7	8	**37**	35	13
4	7	7	15	**37**	35
5	2	8	12	20	**58**

TABLE 3

Mean Percentage on Diagonal

SOURCE	GROUP A	GROUP B	GROUP C	SIGNIFICANT DIFFERENCES using Mann-Whitney test between Groups
I	82.3	59.5	36.7	A > B > C
II	86.8	64.5	35.3	A > B > C
III	52.8	39.5	36.0	
IV	54.3	32.3	24.3	A > B
V	70.8	49.5	40.7	A > B
Wilcoxon test between sources :	I, II > III, IV, V V > III, IV	I, II, V > III, IV II > V	I, III, V > IV	

Note : The maximum is 100% ; a completely random performance would give 20%.

The theoretically ideal matrix would have 100% of the responses to each stimulus on the main diagonal indicated in the figure. On inspection it is apparent that the trained subjects have more consistent results than the other groups and that the psychologists are better than the naive subjects in this respect.

The mean percentage of the responses falling on the diagonal for each group and source are given in Table 3. The differences between groups and between sources given in the table were calculated using the Mann-Whitney and Wilcoxon (two-tail) tests respectively. Group A perform significantly better than the other groups on all except source III (tones) but group B are better than group C for sources I and II. The human sound (source I) and its direct imitation (source II) give the best results for the first two groups, with the low-pass filter (source V) well above the other two. Naive subjects, on the other hand, on this measure have approximately the same score for all sources except P.A.T. (IV) which is lower.

Since the sounds are on a phonetic continuum it is possible to take into account the 'amount' of deviation of the 'incorrect' responses. Thus, to the stimulus /ɑ/, /ɔ/ is more nearly correct than /o/. Accordingly for an alternative score, responses in the diagonals adjacent to the main diagonal are given a weighting of two and three respectively, and responses in the corners furthest from the main diagonal have a weight of four. The two scores obtained in this way, from above the diagonal and below the diagonal, give an estimate of the extent of the deviation when added, and indicate the direction of the deviation when subtracted. An example of the arithmetic is given in the Appendix.

These more sensitive deviation scores, given in Table 4, separate groups B and C significantly for all sources, and also affect the ordering of the scores for group C, where sources I, II and V are now significantly better than sources III and IV.

Inspection of Table 2 indicates that the responses for source IV are biased below the diagonal, and the deviation direction scores, seen in Table 5, show that the

TABLE 4

Mean Deviation per Subject

SOURCE	GROUP A	GROUP B	GROUP C	SIGNIFICANT DIFFERENCES using Mann-Whitney test between Groups
I	88	250	442	A < B < C
II	68	221	458	A < B < C
III	307	428	571	A , B < C
IV	283	466	589	A < B < C
V	157	293	456	A < B < C
Wilcoxon test between sources :	I, II < III, IV, V V < III, IV	I, II, V < III, IV	I, II, V < III, IV	

Note : The higher the score, the more the responses deviate from the correct ; a random performance would give an average score of 1,000.

TABLE 5

Deviation Direction

SOURCE	GROUP A	GROUP B	GROUP C
I	82*	38	4
II	42	55	52
III	287*	142	121
IV	−191*	−238*	−193*
V	145*	221*	146*

* Significant by the Wilcoxon test at $P < 0.01$.

Note : This is a measure of any systematic shift above or below the main diagonal. A negative score indicates a shift towards the bottom left hand corner of the response matrix in Table 2. A random performance would give an average score of 0.

tendency is consistent and significant for all groups. Experiment 2, to be described later, tests a possible explanation of this result.

An alternative way of testing the results utilises information theory statistics. (The theory of this application is given in Attneave, 1959, Chapter 3.) In the first place we can calculate the value of $\hat{T}$, which gives us a measure of the information concerning the stimuli, communicated by the responses. With a five-choice situation the maximum information would be $\log_2 5 = 2.322$ bits. All the scores, given in Table 6, are significantly better than zero at the 1% level (using the test given by Attneave, p. 63). In addition, for group C the ordering of the scores corresponds more nearly to the other groups than with the 'percent correct' score (Table 3), in that the scores for sources III and IV are now below those of the other three sources. In this respect, the $\hat{T}$ score resembles the deviation score ; both these measures take into account the distribution of the 'incorrect' responses.

TABLE 6

Information Transmitted (in bits)

SOURCE	GROUP A	GROUP B	GROUP C
I	1.73	0.95	0.38
II	1.75	0.95	0.34
III	0.99	0.46	0.18
IV	0.94	0.37	0.15
V	1.40	0.66	0.35

Note : All the scores are significant at $P < 0.01$. A random performance would give a score of 0.

TABLE 7

$H(x,y)_{max} - \hat{H}(x,y)$

SOURCE	GROUP A	GROUP B	GROUP C
I	1.76	1.08	0.49
II	1.76	1.00	0.38
III	1.15	0.55	0.27
IV	1.25	0.63	0.36
V	1.43	1.08	0.41

Note : A random performance would give a score of 0. The maximum score is 2.32. All the scores are significant at $P < 0.01$.

The $\hat{T}$ score would be maximum with the theoretically ideal stimulus - response configuration shown in Fig. 2(a), but not with configurations such as those given in Fig. 2(b) and (c). We must however take into consideration the possibility that for any source, two of the stimuli might sound to the subjects like the same vowel, which might result in a matrix such as 2(b). Alternatively, the shift below the diagonal for P.A.T. (source IV) may for some reason be systematic, which would result in a matrix such as 2(c). When we refer to ' correct ' and ' incorrect ' responses in these experimental situations we claim that our subjects ought to perceive sounds in the way we intend them to. However they may be consistent with respect to a particular stimulus in a way other than that intended. The consistency of the subjects' responses may be tested by using another information measure, $[H(x,y)_{max} - \hat{H}(x,y)]$ (see Attneave, *op, cit.*). The scores by this measure are given in Table 7. The important difference between this measure and the others is that for all groups the scores for source IV are now greater than for source III. (Unfortunately no method of testing the significance of such differences is available.) This is added evidence that the shift below the diagonal for source IV is systematic.

Interpretation of the results

There are two points which may qualify the results to a slight extent.

(a) The vowels from sources I and II were cardinal, and differ slightly from the sounds normally used in the words given as categories to the subjects of groups

		Responses				
(a)		X				
			X			
	Stimuli			X		
					X	
						X
(b)		X				
			X			
				X		
					X	
					X	
(c)		X				
		X				
			X			
				X		
					X	

An X indicates that all the responses to a particular stimulus were the same.

Fig. 2. Examples of stimulus-response matrix.

B and C. This is unlikely to have had any great effect on the classifications given by these subjects, but there may have been some effect.

(b) The P.A.T. vowels (source IV) were made using two formant simulators only, and these connected in parallel. With these filters in series and a few other technical differences the machine can now produce much more convincing vowels than those used in this experiment. The results for source IV are therefore not a fair reflection of P.A.T.'s present ability to synthesize vowels.

With these points in mind the following conclusions may be drawn :

(1) Irrespective of the scoring method employed, the variation between the five sources was less than between the groups of subjects. It is clear that the choice of who shall listen to and attempt to categorise speech sounds will influence the results more than the method of manufacturing the sounds. The majority of experiments on synthetic speech have employed highly trained subjects who have been given time to consider their decisions analytically. The validity of applying the results of such experiments to the normal listening situation where relatively untrained listeners have to make instantaneous decisions is, to say the least, questionable.

(2) The supposed control condition, the natural vowels of source I, led to less than perfect performance even by the most skilled subjects. (The seven phoneticians averaged 90% for this source.) This is unlikely to be due to imperfections in the recording and reproducing apparatus, although it is a point which could be confirmed by further experimentation. Neither is it likely to be due to individual speech idiosyncracies in the speaker ; the sounds were taken from a cardinal vowel record which had been checked and accepted by some of the most irreproachable opinions in phonetics. It is more likely that this remaining inaccuracy represents an irreduceable minimum of ambiguity in classifying this particular group of sounds, *when taken together with artificial sounds of such diversity of character.*

(3) The result of the previous experiment is confirmed in this different context. Sounds made in careful imitation of a spoken sound (source II) are as recognizable as the originals (source I).

(4) The results from the low-pass filter source (source V) show that it is not necessary to have peaks in the frequency spectrum of a complex sound for it to have a consistently classifiable vowel colour. This point has been investigated in a further experiment which will be reported in a separate paper.

(5) Although the responses to source III are more random than to the other sources, there is still a significant amount of information in them with regard to vowel colour, in spite of the view expressed by some of the subjects that these sounds were " just like a fog-horn."

Experiment II

It was noted above that the responses for source IV (P.A.T.) were consistently placed below the diagonal in Table 2. A possible explanation of this phenomenon might be found in the fact that the P.A.T. vowels contained energy at higher frequencies than any of the other sources. The low-pass filter results (source V) indicated that having energy present at higher frequencies resulted in the displacement of judgements in the direction from /u/ to /a/ with the five vowels used. If then the P.A.T. vowels are being judged in the context of the other sounds, the additional energy would result in the kind of displacement which was found (cf. Broadbent and Ladefoged, 1960). It has been noted that the P.A.T. vowels, while not completely consistently categorised, do not lead to such a bias in the results when they are played by themselves.

To investigate this point further, we re-recorded the experimental tape passing the P.A.T. vowels through a low pass filter set at a cut-off frequency chosen to make the highest frequency present correspond more nearly to the other sources. The cut off frequencies are given in the right hand column of Table 1. This modified tape was played subsequently to the members of group B (11 psychologists) under the same conditions as before.

Results

The percentage matrices are given in Table 8, with the previous results for the group given for comparison. The matrices were analysed as before with results given in Tables 9-13 corresponding to Tables 3-7 above. The group B results for the first experiment are also given for comparison.

It will be seen that performance was better for all sources on nearly all the measures, implying some general increase in ability to categorise the sounds consistently as a result of practice.

Of particular interest is the comparison between the scores in Table 11 which gives a measure of the direction of deviation from the diagonal.

The scores for the P.A.T. (source IV) show a significant decrease in the size of

TABLE 8

Percentage of Responses to each Stimulus (Group B)

STIMULUS	SOURCE I					II					III					IV					V				
RESPONSES	/a/	/ɑ/	/ɔ/	/o/	/u/	/a/	/ɑ/	/ɔ/	/o/	/u/	/a/	/ɑ/	/ɔ/	/o/	/u/	/a/	/ɑ/	/ɔ/	/o/	/u/	/a/	/ɑ/	/ɔ/	/o/	/u/
EXPERIMENT I																									
1	**36**	52	11			**41**	43	5	7		**27**	52	7	9	5	**50**	36	14			**39**	55		7	
2		**55**	36	9		5	**77**	14	5		2	**32**	45	14	7	50	**34**	11	5			**43**	41	14	2
3	5		**82**	14		2	11	**59**	25	2	7	**11**	**50**	16	16	7	32	**34**	27			2	**55**	41	2
4		2	36	**59**	2	2	5	20	**61**	11	5	2	41	**23**	20	2	20	34	**41**	2		2	11	**27**	**61**
5	2	5	5	23	**66**	2		5	9	**84**	2	7	2	23	**66**	7	16	25	50	**2**			7	9	**84**
EXPERIMENT II																									
1	**55**	41	2	2		**48**	50	2			**41**	52	2	2	2	**36**	45	7	11		**43**	50		7	
2		**77**	20	2		7	**75**	11	7		7	**25**	41	20	7	23	**57**	14	7		5	**64**	18	14	
3		14	**75**	11			11	**71**	18		5	7	**52**	27	9	2	20	**71**	7		2	11	**52**	34	
4	2	5	16	**75**	2			16	**82**	2		7	20	**34**	39		5	43	**45**	7			18	**55**	27
5			5	11	**84**			2	16	82		2		7	**91**			34	55	**11**	2			2	**95**

TABLE 9

Mean Percentage on Diagonal (Group B)

SOURCE	EXPERIMENT I	EXPERIMENT II
I	59.5	73.2
II	64.5	71.4
III	39.5	48.6
IV	32.3	44.1
V	49.5	61.8
Wilcoxon test between sources :		I, II $>$ III, IV, V
		V $>$ III, IV

For notes, see Table 3.

TABLE 10

Mean Deviation per Subject (Group B)

SOURCE	EXPERIMENT I	EXPERIMENT II*
I	250	155
II	221	153
III	428	327
IV	466	357
V	293	226
Wilcoxon test between sources :		I, II $<$ III, IV, V
		V $<$ III, IV

* All scores significantly better than Experiment I at $P < 0.01$ on Wilcoxon test.

TABLE 11

Deviation Direction (Group B)

SOURCE	EXPERIMENT I	EXPERIMENT II
I	38	23
II	55	45
III	142	185
IV	−238	−89*
V	221	130*

* Significantly lower than Experiment I at $P < 0.01$ on Wilcoxon test.

TABLE 12

Information Transmitted (in bits) Group B

SOURCE	EXPERIMENT I	EXPERIMENT II
I	0.95	1.24
II	0.95	1.25
III	0.46	0.75
IV	0.37	0.61
V	0.66	1.01

For notes, see Table 6. All scores are significant at $P < 0.01$ level.

TABLE 13

Group B

$H(x,y)_{max} - \hat{H}(x,y)$

SOURCE	EXPERIMENT I	EXPERIMENT II
I	1.08	1.29
II	1.00	1.31
III	0.55	0.82
IV	0.63	0.85
V	1.08	1.12

For notes, see Table 7. All scores are significant at $P < 0.01$.

the negative score (using the Wilcoxon test). This change can also be seen in the response matrices of Table 8.

This predicted result agrees with our conclusions from the results of the low-pass filter source (V) ; the presence or absence of energy at certain frequencies alters the vowel quality of a sound although the formant peak positions are unchanged. Broadbent and Ladefoged (1960) showed that vowel judgement of a sound depends on the formant positions of the vowels in the context. The present result suggests that this context effect may also be described in terms of the pattern of energy distribution over the whole frequency scale. With the five vowels used in this experiment the most important parameter of this pattern appears to be the upper limit.

Ladefoged (1959, p. 1) says, ". . . it is probable that vowels such as /e, ɛ, a, ɑ/ are best specified in terms of the relation between the pitches of their first two formants and the pitches of the first two formants of other vowels of this general type spoken by the same speaker. This is not true of vowels such as /i, u, o, ɔ/. These vowels often cannot be conveniently analysed as having two or three formants ; nor . . . are they adequately specified in terms of formant frequencies (or pitches) ".

The results reported here seem to indicate the possibility of an extention of the relational theory of vowel recognition to include all vowels.

Such a theory may well use acoustic parameters other than the position of formant peaks.

Appendix

Example of Calculation of the Deviation Scores
Group A, Source IV

		Responses					
	Stimuli	/a/	/ɑ/	/ɔ/	/o/	/u/	Score × weight = sum
Score × weight = sum	1	**74**	22	1	3		0 × 4 = 0
160 × 1 = 160	2	59	**30**	11			3 × 3 = 9
20 × 2 = 40	3	5	19	**75**	1		1 × 2 = 2
7 × 3 = 21	4	3		18	**89**	1	35 × 1 = 35
4 × 4 = 16	5	4	4	15	74	**4**	
Sub-total = 237							Sub-total = 46

Deviation score = Sum of sub-totals = 46 + 237 = 283.
Deviation direction score = Difference of sub-totals = 46 − 237 = −191.

References

Attneave, F. (1959). Application of Information Theory to Psychology (New York).

Broadbent, D. E. and Ladefoged, P. (1960). Vowel judgements and adaptation level. *Proc. Royal Soc., B*, 151, 384.

Carpenter, A. (1962). Discrimination of vowel sounds synthesized by harmonic addition. *Language and Speech*, 5, 151.

Carpenter, A. and Morton, J. (1962). The perception of vowel colour in formantless complex sounds. *Language and Speech*, 5, 205.

Ladefoged, P. (1959). The perception of vowel sounds. Ph.D. Thesis, University of Edinburgh.

Ladefoged, P. (1960). The nature of vowel quality. Revista do Laboratório de Fonética Experimental, Coimbra, 5, 73.

Lawrence, W. (1953). The synthesis of speech from signals which have a low information rate. In 'Communication Theory' (ed. Willis Jackson) London.