Carpenter, Alan, Applied Psychology Research Unit, Cambridge, ***The perception of vowel colour in formantless complex sounds*** , Language and Speech, 5:4 (1962:Oct./Dec.) p.205

THE PERCEPTION OF VOWEL COLOUR IN FORMANTLESS COMPLEX SOUNDS*

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

Complex sounds were manufactured by passing the output of a buzz source through band-pass filter networks. The resulting stimuli were played both to phoneticians and to phonetically naive subjects who were asked to judge the vowel colour of the sounds. The responses showed that complex sounds are categorised consistently with regard to their vowel colour in spite of the absence of formant peaks.

The following additional conclusions can be drawn :

(1) The judgement of vowel colour is relatively unaffected by such features of the stimulus as variability in length between 300 and 700 msec. and the temporal qualities of the on-off transitions.

(2) Vowel sounds of the back group can be simulated with a low-pass filter ; and increasing the cut-off frequency shifts judgement in the direction from " high " to " low ", i.e. from /u/ to /ɑ/, and then /a/.

INTRODUCTION

It was reported in a previous experiment (Morton and Carpenter, 1962), that sounds produced by passing the output of a ' pulse source ' through a low-pass filter, which we will term ' square spectra ', were consistently identified as vowels. This result is particularly interesting since nearly all present work on human recognition and machine recognition and synthesis of vowels is based on an analysis of the vowels by the positions of their formants, with the implicit assumption, rarely stated, that the ear and the auditory nervous system together perform some kind of peak-picking operation when identifying vowels. In the previous experiment some small amount of evidence was obtained in favour of such a theory in that sounds consisting merely of the individual harmonics closest to the first two formant peaks did provide some information as to the original vowels even when these were cardinal (i.e. phonetic reference points rather than natural English vowels) and with phonetically naive subjects. However, the consistency of the perception of these sounds was very much lower than the consistency of identification of the " square spectra " which have no peaks.

The present experiment aimed to investigate more thoroughly the vowel colour of " square spectra ", using a wider range of sounds than was employed in the previous experiment.

* *This research was sponsored by the Army Personnel Research Committee. We are grateful to Mr. Donald Broadbent for his encouragement and to Mr. David Abercrombie, Head of the Phonetics Department in the University of Edinburgh for his advice and for the hospitality of his department.*

Copyright (c) 2007 ProQuest LLC
Copyright (c) Kingston Press

Carpenter, Alan, Applied Psychology Research Unit, Cambridge, ***The perception of vowel colour in formantless complex sounds*** , Language and Speech, 5:4 (1962:Oct./Dec.) p.205

TABLE 1

The Stimuli Used in the Experiment

STIMULUS NUMBER	CUT-OFF FREQUENCIES		PREVIOUS IDENTIFICATION
	Low-pass	*Band-pass*	
1.	300	—	/u/
2.	435	—	/o/
3.	600	—	
4.	850	—	/ɔ/
5.	1,200	—	/ɑ/
6.	1,700	—	
7.	2,400	—	/a/
8.	4,800	—	
9.	225	1,700 – 2,400	/y/
10.	225	2,400 →	/i/
11.	—	2,400 →	
12.	—	1,700 – 4,800	

MATERIALS

The twelve stimuli used are listed in Table 1. All were produced by filtering a pulse signal with a fundamental frequency of 180 cps. The stimuli fell into two groups.

(a) Eight sounds produced by passing the pulse signal through a low-pass filter with cut-off frequencies ranging from 300 to 4,800 cps. This group included the sounds previously identified as the vowel sounds /a/, as in *bat ;* /ɑ/, as in *cart ;* /ɔ/, as in *bought ;* /o/, as in *boat ;* and /u/, as in *boot.*

(b) Four further sounds were produced. Two of these contained the output from a low-pass filter and a further set of harmonics either band limited or low limited. These two sounds had been discovered empirically as being perceived respectively as /y/, as in the French *plume,* and /i/ as in *beet.* The last two sounds were similar to these, but without the low group of harmonics. The low-pass filter used for these sounds had a cut-off rate of only 20 db per octave. That for the upper group of harmonics and the sounds of group (a) had an attenuation of $>$ 80 db at a frequency ten per cent greater than the cut-off frequency.

In Table 1 will be found a list of the sounds, the numbers by which they will be referred to, their method of production and their previous identification. Fig. 1 gives the relative amplitude of the various harmonics in the sounds.

We expected those stimuli in the first of the above groups which had not been previously used to be judged as being intermediate in the series. Stimuli 8, 11 and 12 we expected to be called non-vowel.

Each of the twelve stimuli was repeated 5 times, making a set of 60 which was arranged in 5 randomised blocks. The whole set was recorded twice on magnetic

Copyright (c) 2007 ProQuest LLC
Copyright (c) Kingston Press

Carpenter, Alan, Applied Psychology Research Unit, Cambridge, ***The perception of vowel colour in formantless complex sounds*** , Language and Speech, 5:4 (1962:Oct./Dec.) p.205

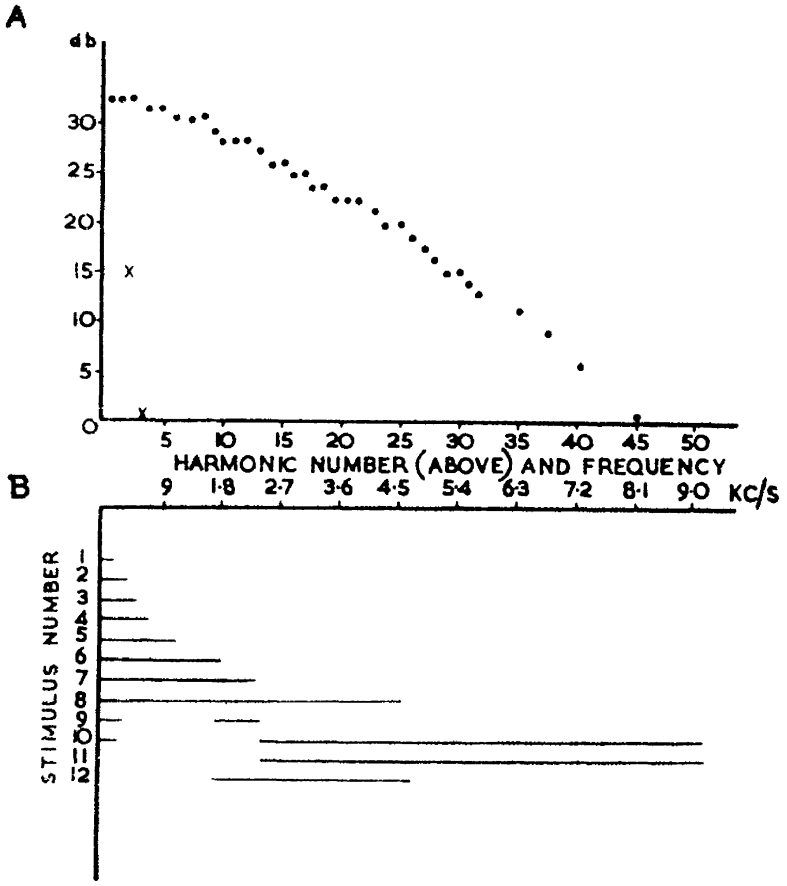

Fig. 1. A. The relative intensity of the harmonics in the stimuli as measured by harmonic analyser. 'x' denotes the level of the 2nd and 3rd harmonics in stimuli nos. 9 and 10. The noise level was −30 db on the same scale. B. Schematic indication of which harmonics were present in each stimulus. The abscissa is common to both parts.

tape, using a Vortexion tape recorder and a speed of 7.5 inches per second. On the first tape, Tape 1, each sound was repeated three times on each presentation, the switching being by hand. This had two effects ; firstly the sounds had a duration which varied unsystematically between 300 and 700 milliseconds, and secondly they had abrupt and rather inconstant onset and cessation, depending on the instant at which the key was operated. The interval between repetitions was about 0.5 sec.

On Tape 2, the switching was done electrically via a gate which had a "distortionless" transition from off to on and vice versa lasting about 0.01 sec. This removed the 'click' quality from the transitions but left them subjectively well defined in time. The gates were controlled by timers, so that the duration of the stimuli was 0.4 sec. with an interval of 0.6 sec. between the two repetitions of each sound. On both tapes there was an interval of 10 sec. between the presentation of one sound and the next.

The records were reproduced by good quality speaker systems. For one group of untrained subjects a 15 in. speaker in a large non-resonant cabinet was used ; for

Copyright (c) 2007 ProQuest LLC
Copyright (c) Kingston Press

Carpenter, Alan, Applied Psychology Research Unit, Cambridge, ***The perception of vowel colour in formantless complex sounds*** , Language and Speech, 5:4 (1962:Oct./Dec.) p.205

the other untrained group and all the phoneticians a full range electrostatic loud-speaker was used. There was no control of the listening room or of the number of people present. It appears that quite wide variations in acoustic environment make little difference to auditory judgements of this type, although it is hoped to clarify this point in a further experiment. Similarly there was no control over the loudness level. The various sounds differed in this respect over a range of approximately 10 db ; the loud sounds gave readings of 70 db on a sound level meter (flat response) in a typical case when measurement was taken.

Both tapes were played to two groups of subjects, trained and untrained. All subjects heard Tape 2 and then Tape 1.

Results

(1) *Trained subjects*

Ten phoneticians from Edinburgh and Cambridge listened to both tapes in groups of up to three. They were instructed to transcribe the sounds as regards vowel colour but were permitted to judge them as having no vowel quality.

The results are presented in Tables 2 and 3. The first block of sounds on each tape was treated as practice, and so each stimulus was judged four times by each phonetician, giving a total of 40 responses to it per tape. In these tables, certain of the notational categories are grouped together. Where this occurs, the second named of the pair of symbols was used not more than three times in all, and it seemed simpler to pool the results for these phonetically adjacent categories rather than further complicate the table. For similar reasons any diacritics (i.e. minor phonetic modifications to the primary judgement) which were used have been ignored with the exception of judgements of nasality. The numbers of nasal diacritics which were used are given in brackets in the tables.

The results largely confirm the predictions made.

(a) As the cut-off of the low-pass filter is increased (i.e. from stimulus 1 to stimulus 8), the judgement moves progressively from /u/ towards /a/ as if the phonetic continuum corresponded to a continuously increasing upper frequency limit for the presence of energy. This tendency is shown more strikingly in Fig. 2. There is some scatter of judgements, but this scatter is largely limited to adjacent response categories, and reflects differences between judges and not individual inconsistencies.

(b) The phoneticians agreed with us in the assignment of stimuli 9 and 10 as /y/ and /i/ respectively. The agreement with respect to stimulus number 9 was almost embarrassing, the only dissident being one person who was 'set' for English vowels and did not consider the French vowel /y/, judging the sound to be non-vowel in quality. In discussion she admitted that the sound was in fact /y/.

Copyright (c) 2007 ProQuest LLC
Copyright (c) Kingston Press

Carpenter, Alan, Applied Psychology Research Unit, Cambridge, ***The perception of vowel colour in formantless complex sounds*** , Language and Speech, 5:4 (1962:Oct./Dec.) p.205

TABLE 2

Results for Phoneticians—Tape 1

RESPONSES	STIMULI											
	1	2	3	4	5	6	7	8	9	10	11	12
u ʊ	20(4)	10(1)			1							
o		23(3)	15(5)	7(1)	1							
ɔ		1	16(5)	16(8)	4(1)	1						
ɑ ɒ			7	16(2)	27(3)	7(1)	1	1				
a æ					2(1)	17(6)	19(8)	11(9)				3
ɛ œ							1	2(1)				1
e ø		1(1)					1(1)			3	5	5
ɪ										2	2	
i	1								1	25(1)	19	1
y									35(3)	7(1)	4(2)	1(1)
ɐ					1	1	1	1				
ɜ						5	3	3				
N	19	5	2	1	4	9	14	22	4	3	10	29

10 subjects × 4 judgements = 40 responses per stimulus.
Figures in brackets refer to judgements of nasality.

TABLE 3

Results for Phoneticians—Tape 2

RESPONSES	STIMULI											
	1	2	3	4	5	6	7	8	9	10	11	12
u ʊ	19(1)	13(2)		1								
o	1	25	29(3)	15(3)	1	1						
ɔ			8(4)	16(5)	8(1)	1						
ɑ ɒ			3	5	22(2)	8(2)	3	4				
a æ		1(1)			2(1)	12(6)	12(3)	6(1)				
ɛ œ						3(1)	4(2)					3(1)
e										3	5	7
ɪ										1	1	1
i	4									26	21	1
y	3								36	7	5	4
ɜ					1	4	3	4				
ə												
N	13	1	0	3	6	11	18	26	4	3	8	23

10 subjects × 4 judgements = 40 responses per stimulus.
Figures in brackets refer to judgements of nasality.

Copyright (c) 2007 ProQuest LLC
Copyright (c) Kingston Press

Carpenter, Alan, Applied Psychology Research Unit, Cambridge, ***The perception of vowel colour in formantless complex sounds*** , Language and Speech, 5:4 (1962:Oct./Dec.) p.205

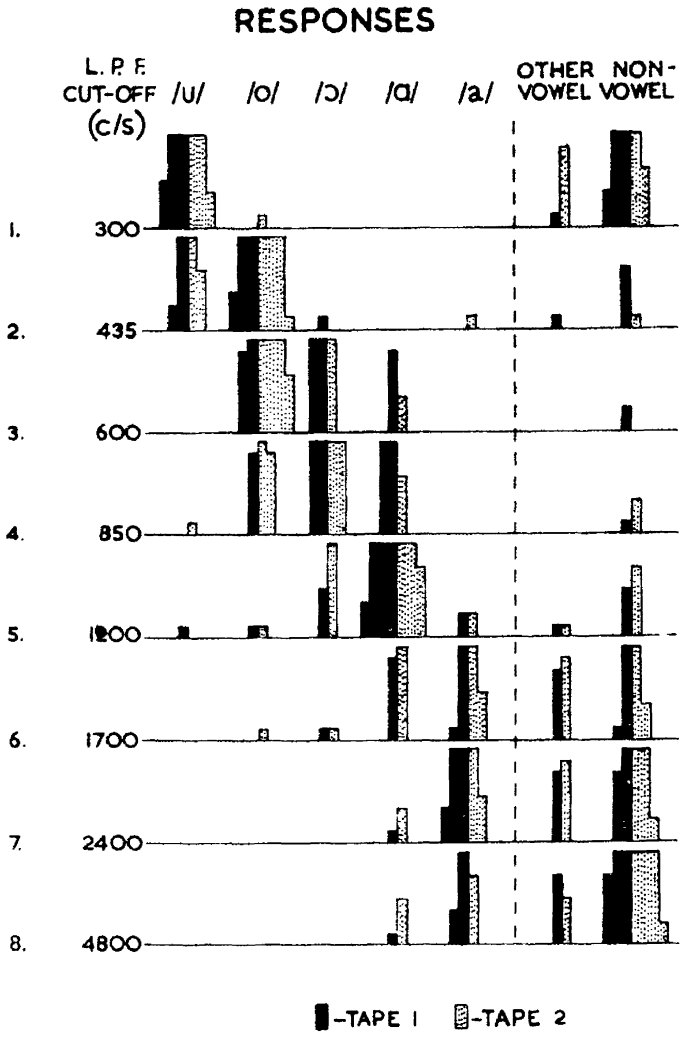

Fig. 2. Results for phoneticians. The shift in response from /u/ to /a/ as the cut-off frequency of the low-pass filter is increased. In the key the areas correspond to 5% of the responses to any stimulus.

(c) Stimuli 8 and 12 were predominantly classed as non-vowel, but stimulus 11, which to us was as non-vowel as these, was frequently judged to be /i/. This sound had in common with stimulus 10 all the harmonics above 1,800 cps, and the addition of the fundamental frequency from whatever source—perception of the envelope frequency, non-linearity of the ear, or distortion in the reproducing equipment—was apparently sufficient to produce the /i/ sensation.

(d) The most surprising aspect of the results was the great similarity between the results on the two tapes. Tape 1 produced more judgements of nasality than Tape 2 (66 as opposed to 39) but except that a few of the subjects reported an initial /b/ on the sounds on Tape 1, there appears to be little difference between the two results in spite of vigorous complaints about the intrusiveness of the switching on this tape.

Copyright (c) 2007 ProQuest LLC
Copyright (c) Kingston Press

Carpenter, Alan, Applied Psychology Research Unit, Cambridge, ***The perception of vowel colour in formantless complex sounds*** , Language and Speech, 5:4 (1962:Oct./Dec.) p.205

The ' naturalness ' of the sounds

After they had heard both tapes and given their judgement on vowel colour, the phoneticians were asked their opinion on the " goodness " of the sounds. Although not specifically told so, all were aware that the sounds were artificial, yet there was good agreement that about half the sounds were " good " vowels, but that the naturalness of these varied from bad machine imitations to a human speaker in the next room. It was agreed that such sounds could be judged for naturalness and for vowel colour independently.

(2) *Untrained subjects*

The tapes were also played to R.N. ratings who may be regarded as being phonetically completely naive. They were given the following instructions : " You are about to hear some sounds. Some of the sounds are vowel sounds. The other sounds are nonsense sounds. Will you please indicate in the appropriate space your judgement of the sounds in the following way : —

(1) If you can recognize one of the following vowels, enter the appropriate symbol in the relevant space.

bat	— a	boot	— oo
bart	— ah	beet	— ee
bought	— or	bit	— i
boat	— o	bet	— e

(2) If you think the sound is a vowel, but either you are not certain which one, or if it is not one of the given vowels, enter ' V '.

(3) If you think the sound is not a vowel, enter ' N '." (Note : /y/ was not included as a possible response since none of these subjects could be expected to realize the existence of nonEnglish vowels let alone categorise them.)

When we were reasonably sure that they had understood the instructions, a practice set of 24 stimuli was played to them before proceeding to the test proper.

The results are given in Tables 4 and 5. In Table 4 is summarised the results for 15 subjects judging all 5 blocks of sounds on Tape 1. In Table 5 are the results on Tape 2 for a different 11 subjects.

In spite of a greater scatter of judgements, which was caused partly by differences between subjects and partly by the inconsistency of individuals, the general trend of responses to stimuli 1 to 8 is the same as was found for the phoneticians. (See also Fig. 3.)

Comparing the results for Tapes 1 and 2 it will be seen that a greater proportion of ' N ' responses were given to Tape 1, but apart from shifts in the responses to stimuli 5 and 8 the vowel quality of the sounds was judged to be much the same. This is in spite of the fact that different groups of subjects were used for the two tapes.

With these subjects, stimulus 9, nominally /y/, was classed as either /u/ or /i/ (i.e. as other high vowels) in 34% of the responses, and as nonsense in a further 25%.

Copyright (c) 2007 ProQuest LLC
Copyright (c) Kingston Press

Carpenter, Alan, Applied Psychology Research Unit, Cambridge, ***The perception of vowel colour in formantless complex sounds*** , Language and Speech, 5:4 (1962:Oct./Dec.) p.205

TABLE 4

Results for Naive Subjects—Tape 1

RESPONSES	STIMULI											
	1	2	3	4	5	6	7	8	9	10	11	12
u	19	21	6	17	6	2	1	1	11			
o	8	14	28	9	13	4	1		3		2	1
ɔ	6	5	15	19	21	9	7		6	1	1	2
ɑ	4	11	8	10	16	25	22	9			2	2
a	1	2	4	2	5	12	8	13	6		1	1
ɛ	1				1	3	8	21	6	16	17	16
ɪ	1	2	3		1		2	7	6	12	16	29
i	2	1				1	2	7	14	40	28	17
V	5	4	4	5	4	7	7	6	2	2	3	
N	28	15	7	13	8	12	17	11	21	4	5	7

TABLE 5

Results for Naive Subjects—Tape 2

RESPONSES	STIMULI											
	1	2	3	4	5	6	7	8	9	10	11	12
u	24	20	12	8	3	4	1	3	6			
o	4	11	11	6	8	1	1	1	4			2
ɔ	4	3	22	23	10	3	2	1	2	1	2	
ɑ	2	3	3	10	26	24	29	21	3	1		2
a		1	2	3	2	9	8	10				6
ɛ	3	2	1	1	1	1	5	4	7	9	17	12
ɪ	2	2	1			3			7	15	15	13
i	3	2						3	14	25	18	12
V	3	3	1		1	4	1	3	1	1	1	1
N	10	8	2	4	4	6	8	9	11	3	2	7

DISCUSSION

Thus far it seems possible to claim that complex sounds can have a fairly consistent vowel quality, even when there are no peaks in the harmonic structure. On first glance this result might lead us to conclude that the auditory system does not analyse vowel sounds by picking out peaks in some way analogous to the methods employed in machine analysis of speech.

It might be argued against this that given sufficient motivation, or suitable instructions, a subject would be prepared to categorise *any* complex sound in terms of a vowel system, (the ' consistency ' and ' accuracy ' of the judgements depending on the discriminability of the sounds and the communality of the vowel systems of subject and

Copyright (c) 2007 ProQuest LLC
Copyright (c) Kingston Press

Carpenter, Alan, Applied Psychology Research Unit, Cambridge, ***The perception of vowel colour in formantless complex sounds*** , Language and Speech, 5:4 (1962:Oct./Dec.) p.205

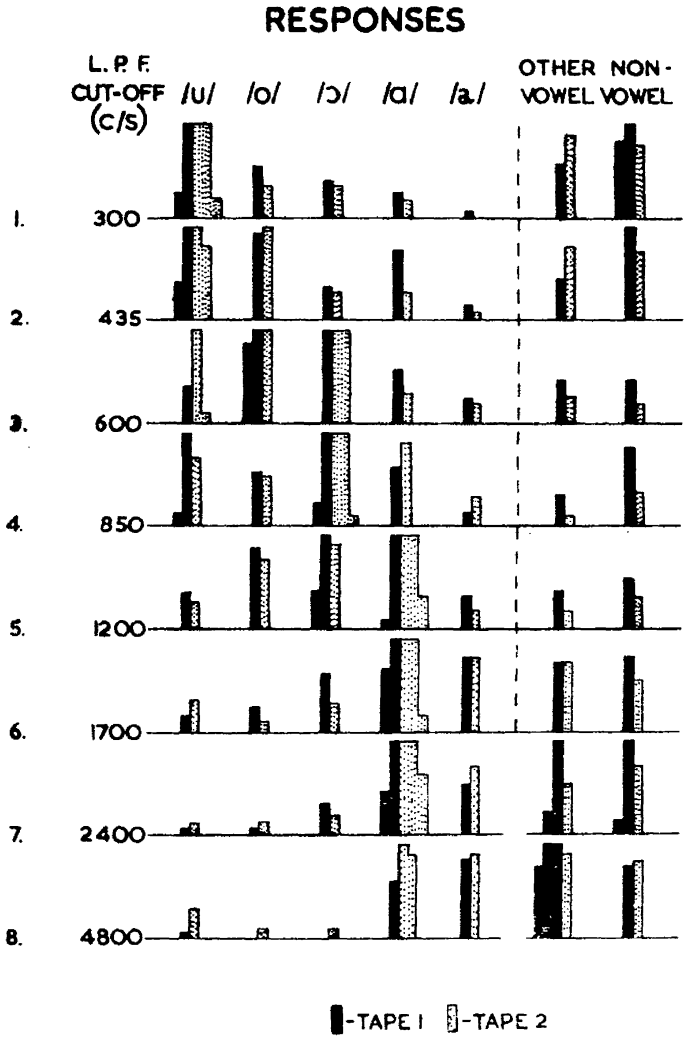

Fig. 3. Results for untrained subjects. The shift in response from /u/ to /a/ as the cut-off frequency of the low pass filter is increased. In the key the areas correspond to 5% of the responses to any stimulus.

experimenter), but this would not necessarily mean that the mechanisms through which this is achieved are the same as the mechanisms employed in listening to real (i.e. human) vowels.

In the absence of sufficient neurological knowledge it is difficult to counter such an argument directly, but it can equally well be applied to the judgement of machine-made formant-type vowels. The fact that human vowels are, for any individual, reasonably discriminable on analysis in terms of the formant positions alone, is a phenomenon relating to the method of production of the vowels, and it does not follow that the mechanism of speech recognition proceeds in a similar way. In addition, the first half of the hypothetical argument in the previous paragraph implies that some sort of

Copyright (c) 2007 ProQuest LLC
Copyright (c) Kingston Press

Carpenter, Alan, Applied Psychology Research Unit, Cambridge, ***The perception of vowel colour in formantless complex sounds*** , Language and Speech, 5:4 (1962:Oct./Dec.) p.205

comparison can be made between classes of stimuli coded in terms of formant positions, and stimuli which cannot be coded in this way. This would require some system of mapping from one classification system to another, which not only seems a little wasteful but implies that a classification of vowels exists neurologically in terms other than those corresponding to formant theory.

A final point to be considered is that we normally expect vowel sounds to be human in quality, and if a sound lacks naturalness, that is if it lacks those aspects of human vowel sounds which enable us to distinguish between the vowels of one person and another, it is possible that we initially reject it as being a vowel and only then consider the vowel colour in the abstract, although the ' vowelness ' of the sound may have been complete. In other words we could postulate an interaction between the separable attributes of naturalness and vowel colour which ought to be considered before drawing conclusions about the human auditory system from experiments on the judgement of artificial vowels.

Reference

Morton, J. and Carpenter, A. (1962). Judgement of the vowel colour of natural and artificial sounds. *Language and Speech*, 5, 190.

Copyright (c) 2007 ProQuest LLC
Copyright (c) Kingston Press