Proceedings of The Institute of Acoustics 4A24

PERCEPTUAL GROUPING OF SPEECH SOUNDS

C.J. DARWIN

LABORATORY OF EXPERIMENTAL PSYCHOLOGY UNIVERSITY OF SUSSEX.

INTRODUCTION

A persistent problem in the psychology of perception is to discover what conditions must be satisfied for the listener to accept that different frequency components come from the same sound source. Voiced speech, like any periodic sound, has frequency components that are integral multiples of its period, and it has been known for some time that a single sound will be heard when components that are multiples of a common fundamental but which lie in different frequency bands are lead simultaneously to opposite ears (1,4). When the two ears do not receive the same periodicity, subjects report hearing two sounds located one at each ear. The question remains, however, whether the timbre of each of the latter sounds is determined solely by the frequency components reaching the ear in question. Cutting (2) claims that it is not. He played the first formant of a voiced-stop-vowel syllable to one ear of his listeners and the second and third formants to the other ear, and found no decrease in subjects' ability to hear the original consonant when the ears received different periodicities, although they invariably heard two sounds. This result suggests an interesting distinction between mechanisms responsible for deciding "how many" sounds and those that determine "what" sounds- it also questions the use of periodicity in the latter but not the former. Cutting's experiment failed to use an appropriate control, since he did not ask subjects to identify the stop-consonant on the basis of the second and third formants alone, so the following experiment was run to check the efficacy of periodicity in perceptual grouping.

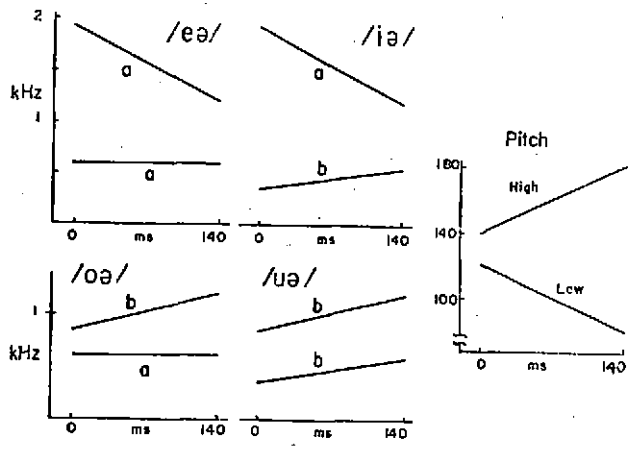

FIGURE 1

The four formants used in the experiment, combined into the four dipthong categories used as responses. On the right are the two pitch contours used.

PERCEPTUAL GROUPING OF SPEECH SOUNDS

METHOD

Two first-formants (F1) and two second-formants (F2) were chosen so that their combinations made four plausible English diphthongs /eə,iə,oə,uə/ (Figure 1). Eight subjects listened binaurally and then dichotically to a number of different combinations of these formants with different or identical pitch contours on the formants. The full list of conditions is shown as part of Tables 1 & 2, with the exception of an isolated formants condition in which subjects heard only a single formant at a time, binaurally. Every token in each of the 10 conditions was heard 5 times (randomized within binaural and dichotic). The subjects (after practice in identifying the four diphthong categories) were told to tick the appropriate category boxes on their response sheets for all the sounds they heard that could be categorised as one of the four diphthongs; they were also told to write down the total number of sounds they heard (i.e. the number of ticks plus any other uncategorisable sounds).

RESULTS

Figure 2A shows the number of diphthong responses given to the isolated formants and to their various binaural pairings. There is essentially the same pattern of responses whether the binaural pairs are on the same pitch or on different pitches. The important point is that the responses to the pairs cannot, in general, be predicted from the responses to the individual formants presented in isolation. This result validates the experiment as a test of timbre fusion.

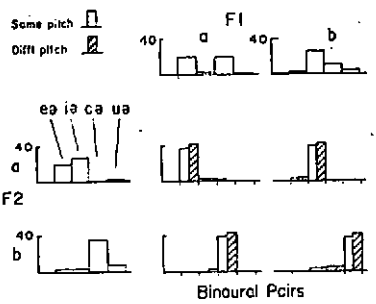

Fig 2A

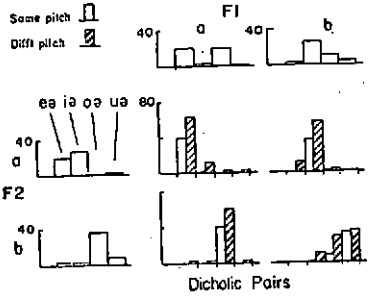

Fig 2B

FIGURE 2

At the top and sides of each figure are shown the total number of diphthong responses given to each of the four formants when played in isolation. The remainder of each figure shows for binaural (Fig 2A) or dichotic (Fig 2B) presentation the total number of diphthong responses for formants presented in the pairs shown in Figure 1. Pairs presented on the same pitch and those presented on different pitches are distinguished by light and dark bars on the histogram.

PERCEPTUAL GROUPING OF SPEECH SOUNDS

Table 1 shows that although the total number of sounds heard increases slightly when binaural formants are on different pitches, the proportion of trials with at least one response corresponding to the fused percept is no lower.

A similar pattern of results is found, confirming Cutting's findings, when formant pairs are presented dichotically. There is a marked increase in the number of sounds heard when the ears receive different periodicities but again no reduction in the probability that at least one of the responses will be the fused category.

TABLE 1

Results for two formants presented binaurally or dichotically.

	Pitch	Av. No. of sounds	% "fused"	% with "fused"	Example Stimulus	
Binaural	Same	1.07	87.7	92.8	F1aL + F2b1	
	Diff	1.33	78.1	94.1	F1aL + F2bH	
					Left	Right
Dichotic	Same	1.06	87.0	91.3	F1aL	F2bH
	Diff	1.92	70.1	90.6	F1aL	F2bH

The same conclusion can be drawn from the results of the more complex stimulus conditions (Table 2) where four formants are simultaneously presented. When two pairs of formants are presented binaurally, each pair on a different pitch, there is no suggestion that subjects are more likely to report the categories formed by the pairs which have the same pitch. Additionally, in the split formant conditions, there is no evidence that a common pitch can override a common ear in determining which pairs of formants will be categorised. There is only very weak evidence - an increase of 10% - that a pair of formants presented to one ear will be categorised together more often if it differs in pitch from the pair on the other ear than if it has the same pitch as the other ear. Rather than reflecting pitch relations, the detailed pattern of results is dominated by particular formants on particular pitches, the low F1 on the high pitch being particularly weak.

TABLE 2

Results for four formants presented binaurally or dichotically.

		Av. No. of sounds	% grouped by pitch	ear	Example stimulus	
Binaural	Same	1.56	-	-	F1aL+F1bL+F2aL+F2bL	
	Diff	1.81	43.1	-	F1aL+F1bH+F2aH+F2bL	
					Left	Right
Dichotic	Same	1.54	-	57.6	F1aL+F2aL	F1bL+F2bL
	Diff	1.98	67.0		F1aL+F2aL	F1bH+F2bH
Split-Formant		2.00	47.4	52.6	F1aL+F2bH	F1bH+F2aL

PERCEPTUAL GROUPING OF SPEECH SOUNDS

DISCUSSION

Our results so far have confirmed Cutting's and support his conclusion that a common pitch provides little justification for perceptually grouping together two formants. However, both my experiment and Cutting's used short sounds - mine lasted only 140 ms. I am sure that a common pitch does contribute to perceptual grouping for longer sounds. My own observations, which have proved difficult to demonstrate experimentally because of the lability of the grouping effects, are that when sounds lasting over about 500 ms are used, then a pair of formants presented dichotically with a different pitch on each ear will pull apart to give two veridical timbres. For example, with F1b on one ear and F2b on the other ear, an initial impression of /uə/ corresponding to timbre fusion gives way to /iə/ (the dominant response to F1b) plus some other sound like /oə/ (the dominant response to F2b) when the pattern is continuously repeated about 4 times. Such decomposition does not occur when the dichotic pair has the same pitch. Binaurally, repetition gives rise to a more distinct impression of two sounds, but perhaps surprisingly, the timbre does not appear to separate veridically. In terms of our previous example /uə/ is still the dominant percept. Repetition of the split-formant condition causes the formants to group by ear rather than by pitch.

The effects have only been investigated subjectively by myself and much more, careful experimentation is needed before they can be confirmed reliably. But there is an interesting analogy with another psycho-acoustic fusion effect which has received quite extensive study. Fellows(3) has recently found evidence that the central pitch phenomenon (6) is heard much more easily for short than for long sounds, and claims that this might explain the failure of some workers to confirm the existence of central pitch (5).

The general picture emerging from Cutting's and my results is that it takes the perceptual system some time to use harmonic structure to separate the timbres of sounds arriving at opposite ears, and it is questionable whether it can do this at all for sounds which do not differ in location, or in any of the other dimensions which can be used to permit grouping, such as onset time. This bias towards hearing sounds from a common source despite the lack of a common harmonic structure should perhaps be welcomed since for all unvoiced speech, and indeed for a substantial part of what is normally classified as voiced speech a regular harmonic structure is not found throughout the first few formants.

Acknowledgements This work was supported by grant GR/A/08024 from the SRC.

REFERENCES

(1) BROADBENT,D.E. and LADEFOGED, P. (1957) JASA, 29, 708-710.

(2) CUTTING, J.E. (1976) Psychological Review, 83, 114-140.

(3) FELLOWS, J. (1979) Unpublished B.Sc. project. Dept. of Psychology, Univ. of Reading.

(4) FLETCHER, H. (1929) Speech and Hearing. New York: van Nostrand.

(5) HALL, J.W. and SODERQUIST, D.R. (1975) JASA, 58, 1257-1261.

(6) HOUTSMA, A.J.M. and GOLDSTEIN, J.L. (1972) JASA, 51, 520-9.