THE
QUARTERLY
JOURNAL OF
EXPERIMENTAL
PSYCHOLOGY
Published for the Experimental Psychology Society
Editors A. Mike Burton and Simon Killcross
Psychology Press

The Quarterly Journal of Experimental Psychology Section A
Human Experimental Psychology
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713683590

Perceptual grouping of speech components differing in fundamental frequency and onset-time

C. J. Darwin [a]
[a] Laboratory of Experimental Psychology, University of Sussex, Brighton

Online Publication Date: 01 May 1981

To cite this Article: Darwin, C. J. (1981) 'Perceptual grouping of speech components differing in fundamental frequency and onset-time', The Quarterly Journal of Experimental Psychology Section A, 33:2, 185 — 207

To link to this article: DOI: 10.1080/14640748108400785
URL: http://dx.doi.org/10.1080/14640748108400785

Quarterly Journal of Experimental Psychology (1981) **33A**, 185–207

PERCEPTUAL GROUPING OF SPEECH COMPONENTS DIFFERING IN FUNDAMENTAL FREQUENCY AND ONSET-TIME

C. J. DARWIN

*Laboratory of Experimental Psychology, University of Sussex, Brighton BN*1 9*QG*

Are there general auditory grouping principles that allow the sounds of a single speaker to be grouped together before phonetic categorisation? Four experiments are reported on the use made of a common fundamental frequency or a common starting time in grouping formants together to form phonetic categories. The first experiment shows that the perception of a vowel category is unaffected by formants being excited at different fundamentals or starting at 100-ms intervals. The second and third experiments show no effect of a different fundamental on the combination of the timbres of pairs of formants presented either binaurally or dichotically to form diphthongs. Onset-time also has no effect with binaural presentation. The fourth experiment finds both an effect of grouping formants by a common fundamental using formant trajectories that do not overlap in frequency, and also an effect of onset-time. Neither a common fundamental nor common onset-time is either a necessary or a sufficient condition for formants to be grouped into a common speech category, although they can be shown to exert an influence. Both these variables exert a considerable influence on the number of sounds that subjects report hearing, even under conditions where they do not influence the reported speech category, indicating a dissociation between mechanisms concerned with "how many" sound sources there are, and those concerned with "what" a source consists of.

Introduction

A long-standing problem in perception is how we isolate the speech of a single speaker from competing sounds. The conventional wisdom subscribed to either implicitly or explicitly by speech theorists (see Darwin, 1976) is that there are low-level processes operating on simple attributes of sound that group together into a single channel (Broadbent, 1958) or stream (Bregman, 1978) frequency components that share certain properties. These processes can then present to those processes subsequently involved with phonetic categorisation a bundle of sounds containing the output from a single source or speaker. The idea is attractive since it relieves the categorisation process of a substantial burden, allowing, for example, spectral templates (e.g. Klatt, 1980) to be used as recognition primitives. Yet the experimental evidence for it is not overwhelming.

0272–4987/81/020185+23 $02.00/0

Four areas of research have contributed evidence. First, selective attention is clearly more efficient when the voices to be distinguished are spatially separated (Cherry, 1953), differentially filtered (Egan, Carterette and Thwing, 1954) or of a different sex (Broadbent, 1952; Treisman, 1960). Second, the automatic separation of simultaneous voices has exploited differences in location (Mitchell, Ross and Yates, 1971) and in voice fundamental (Parsons, 1976). Third, using repeating patterns of simple or complex tones, Bregman and his colleagues have identified a number of parameters that influence whether a particular frequency will group with others to give a complex percept. These include rate, frequency and fundamental frequency change and onset-time. A sequence of tones will split into two perceptually separate groups if by so doing, it reduces substantially the rate of frequency change between tones that are adjacent within a perceptual group (Bregman and Campbell, 1971; Heise and Miller, 1951; Miller and Heise, 1950). A common timbre, or harmonic composition, also exercises some influence on which successive sounds are grouped together (Dannenbring and Bregman, 1978). Frequency components that start and stop at the same time are more likely to be grouped together to give a richer timbre (Bregman and Pinker, 1978) and less likely to group with other tones of similar frequency presented before and after them (Dannenbring and Bregman, 1978) than when their onsets differ. Conversely, it is easier to identify the presence of a frequency component of a complex when it starts at a different time from the rest of the complex.

Fourth, experiments on synthetic speech sounds have suggested a dual role that the fundamental frequency (F_0) of speech might play in grouping sounds together. During voiced speech, the vocal cords generally vibrate at a slowly and smoothly changing frequency corresponding to the pitch of the voice. This quasi-periodic vibration is then filtered by the vocal tract to give the sound the timbre characteristic of the vowel or whatever is being articulated. The spectrum of such voiced speech thus consists of a series of harmonics each of which has a frequency that is an integral multiple of F_0 and an amplitude that varies with the vowel. The harmonics have greatest amplitude when they are close in frequency to the resonant frequencies of the vocal tract or *formant* frequencies. Vowel-like sounds can be synthesised by filtering a pulse-train of variable fundamental frequency and amplitude through simple resonant filters centered on the formant frequencies for a particular vowel. The two putative roles for fundamental frequency are to group consecutive sounds together into the continuing speech of a single talker, and to group together the harmonics from different formants of one talker to the exclusion of harmonics from other sound sources. If the fundamental frequency of synthetic speech is made to alternate rapidly between two values, two different voices are heard where only one is heard when the fundamental is monotonous or changes smoothly and slowly (Darwin and Bethell-Fox, 1977; Nooteboom, Brokx and de Rooij, 1978). Continuity of fundamental frequency thus serves to integrate events occurring at different times into the speech of a single talker. The second role, that of grouping together simultaneous harmonics, is suggested by experiments by Broadbent and Ladefoged (1957). They played separate formants of a synthetic sentence to the two ears of their subjects. When the formants had the same fundamental, a large majority of subjects reported hearing a single voice; but when the formants had different fundamentals almost all the listeners heard more than once voice. This result is compatible with the idea that those harmonics that are multiples of a common fundamental are perceptually grouped together as the frequency components likely to have come from a single speaker or sound source.

Although these experiments and signal processing techniques suggest signal properties that could be used in perceptual grouping, none of them provide direct evidence on the contribution of such grouping processes to phonetic categorisation in the perception of normal speech. The major aim of this paper is to examine what role onset-time, spatial location and a common fundamental play in grouping together frequencies into a phonetic category. In particular, it investigates whether formants which differ in some or all of these properties are precluded from being integrated into a common phonetic percept. If, for example, the two formants of a vowel are on different fundamentals, or start at different times, do you still hear the vowel, and /or do you hear the timbre of each formant as a separate percept?

The experiments reported here were provoked by Cutting (1976) revealing an apparent ambiguity in Broadbent and Ladefoged's results. Using syllables in which the first formant was led to one ear and the second and third formants to the other, Cutting found that subjects' identifications of the initial stop-consonant's place of articulation were unimpaired by exciting the two sets of formants with different fundamentals, despite subjects reporting two sounds. Cutting's result suggests a dissociation of "what (timbre)" and "where" (or "how many?") reminiscent of Deutsch's musical scale illusion (Deutsch and Rolls, 1976). Thus, although Broadbent and Ladefoged's data clearly demonstrate the importance of a common fundamental in determining *how many* voices are heard, they do not speak to the question of what quality, timbre or phonetic category those voices carry.

All the experiments reported here are concerned with whether a common fundamental determines which frequency components will be taken together to form a perceived category. If only those harmonics which share a common fundamental can be grouped together to form a complex, then vowels whose formants are excited at different fundamentals should be less intelligible than vowels with formants synthesised on the same fundamental. The first experiment tests this hypothesis and finds no effect of fundamental, neither does it find any effect of varying the onset and offset times of the formants, although both these variables have a clear effect on the number of sounds that subjects report hearing. The lack of any effect of fundamental frequency leads to a more sophisticated paradigm in Experiments II and III. They use four different formant trajectories, which in different combinations give different diphthong percepts. But again the experiments show no tendency for formants on the same fundamental to be grouped together either with binaural (Experiment II) or with dichotic (Experiment III) presentation. The final experiment does find an effect of grouping by fundamental. It uses four widely spaced formants from which one particular combination of three gives the syllable /ru/, while another combination gives /li/. Subjects are more likely to hear the syllable corresponding to the formants that have a common fundamental than the one that does not, although the other syllable is still heard on a significant proportion of occasions.

Experiment I

This experiment investigates whether vowels synthesised with their formants on different fundamentals and/or with their formants starting or stopping at different times are any harder to classify phonetically than those with each formant on the same fundamental and simultaneous onsets and offsets.

Speech synthesis procedure

All the sounds used in the following experiments were produced by filtering a train of clicks through parallel digital second-order filters. The resonant frequencies of the filters and the period and intensity of the click train entering each filter were under dynamic program control, but the bandwidths of the filters were fixed at 50, 60 and 100 Hz for formants 1, 2 and 3 respectively. The digital signals (generated on a PDP-12) were output at 10 kHz through 12-bit DACs low-pass filtered at 3·5 kHz and recorded on tape.

Stimuli

Ten three-formant vowels were synthesised in each of 16 (8×2) configurations. In half the configurations the formants had the same fundamental and in half the three formants had different fundamentals (120, 133 and 146 Hz on F_1, F_2 and F_3 respectively). Within each set of fundamental frequency conditions the formants differed in whether their onsets and offsets were simultaneous or staggered. The eight different conditions of stagger are listed in Table I. In two of these configurations the formants started and stopped simultaneously and lasted either 300 ms (condition SS-300) or 500 ms (SS-500) excluding 10-ms rise/fall time. In the remaining six conditions all three formants were always present simultaneously for 300 ms, but they differed in whether the lowest (L) or highest (H) formant started first (first letter) or stopped last (second letter). When the formants were staggered they came on (or went off) at 100-ms intervals either in the order F_1, F_2, F_3 or the reverse. The formants could also start or stop simultaneously (S).

Method

Subjects were first trained to identify the two basic configurations (those with simultaneons onsets and offsets) with all formants on the same F_0. They were given response sheets on which they had to cross through the word whose vowel corresponded to the vowel sound they heard most clearly. The words were "heed, hit, head, had, hard, hot, hood, who, her, hub". Ten subjects individually heard the 20 stimuli seven times in a random order over headphones in a quiet cubicle. Before the trials began and after each fifth trial in the first 50 trials they heard a voice pronouncing each word from the response set followed by the two synthetic vowel sounds. After 50 trials the subject scored his answers and then continued with another 50 trials in a similar manner but with the demonstration after every ten trials rather than after every five. If the subject scored more than 75% correct on the last 20 trials he was used in the main experiment. Two subjects out of the ten tested failed.

In the first part of the main experiment the eight remaining subjects listened to three different randomisations of the 160 stimuli. Their task was to identify the clearest vowel they heard, again by scoring through the appropriate word on their answer sheet. They were told that they would hear the previous sounds along with altered versions of them. They were told to guess if unsure.

In the second part of the main experiment the subjects listened to the same stimuli again, but this time they were asked simply to indicate for each stimulus whether they heard one or more sounds by scoring through either "1" or "2" on their answer sheet.

Results

The number of correct vowel identifications and the average number of sounds heard according to stagger condition and F_0 are given in Table I. Analysis of variance found no influence of any of the experimental variables on the accuracy with which the vowel categories were identified.

TABLE I

Percentage of correct identifications of vowels with formants in various conditions of F_0 and stagger in Experiment I

	Fundamental frequency	
	Same	Different
HS	74 (1·9)	72 (2·0)
LS	75 (1·5)	75 (1·8)
LH	70 (1·9)	73 (2·0)
HL	64 (1·9)	70 (2·0)
SH	73 (1·8)	77 (2·0)
SL	71 (1·3)	71 (1·7)
SS-300	77 (1·0)	75 (1·6)
SS-500	73 (1·0)	76 (1·7)

The average number of sounds reported is given in parentheses (see text for description of conditions).

The number of sounds heard, however, was influenced by all the variables in the experiment. More sounds were heard when the fundamentals of the formants were different than when they were the same ($P<0·001$). More sounds were heard when the onsets or offsets of the vowels were staggered than when they were not ($P<0·001$ in both cases). In the conditions where either the onset or the offset of the formants were staggered, there was a significant effect of the order in which the formants came on or went off, so that more sounds were heard when the higher formants started before the lower ($P<0·005$) and when they continued beyond the lower ($P<0·001$).

Discussion

Although putting formants on different fundamentals or staggering their onsets or offsets by 100 ms clearly increases the number of sounds that subjects report hearing, none of these manipulations has any reliable effect on their ability to categorise three-formant vowels. This result is clearly compatible with Cutting's suggestion that there is a dissociation between the decisions of how many sounds are present, and what are their phonetic categories. So although our subjects were hearing more than one sound in many of the conditions, they were still apparently able to integrate the formant information over the separate formants to hear the vowel category that was defined by all three together.

This conclusion though is not quite water-tight, since it is possible in both this experiment and in Cutting's that subjects were able to come up with the correct phonetic percept simply by hearing the formants separately. Neither experiment explicitly controlled for this possibility.

Experiment II

This experiment uses an improved design to investigate this question. It exploits the fact that reasonable syntheses of three different English diphthongs can be made from combinations of two first formant trajectories and two second formant trajectories whose timbres in isolation do not sound like the diphthongs formed by their various combinations. If subjects still hear the appropriate diphthong category when pairs of formants are on different fundamentals, then we can be much more sure that they are in fact combining timbres across formants.

Method

Stimuli

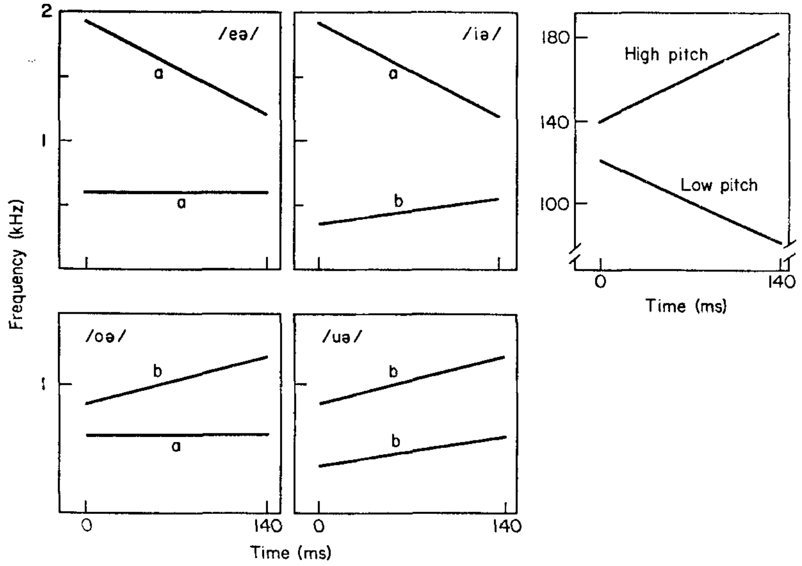

FIGURE 1. Formant and fundamental frequency trajectories used in Experiments II and III. The combination F_{1b}, F_{2b} was not used in Experiment II.

The three diphthong categories used in this experiment were produced by combining two different first formants with two second formants. These sounds (plus a fourth category used in the next experiment) are illustrated in Fig. 1, together with the two F_0 contours which each formant could be excited by. The second formants were synthesised at -5 dB from the first formants on a software parallel-formant synthesiser. In the simultaneous condition just these sounds were used, but in the staggered condition the first formant was extended forwards by 100 ms with a constant resonant frequency and F_0, so that it started 100 ms before the second formant.

The three diphthongs /eə/, /iə/, /oə/ were synthesised with the four possible F_0 contour combinations (both formants high or low, one formant high and the other low). The four isolated formants on either F_0 contour were also used.

There was thus a total of 40 sounds (the isolated second formants being played twice as often as the other sounds since they were identical in the simultaneous and staggered conditions).

Procedure

Six subjects, students in the laboratory, listened to a tape that had five tokens of each sound in a random order. For each token they had to decide whether "ear", "air" or "oar" was the best label for the main sound they heard and whether the main sound and each additional sound they heard was best characterised as "speech", "distorted speech" or "non-speech". They were allowed to stop the tape between trials if they required more than the recorded 4-s interval.

Results

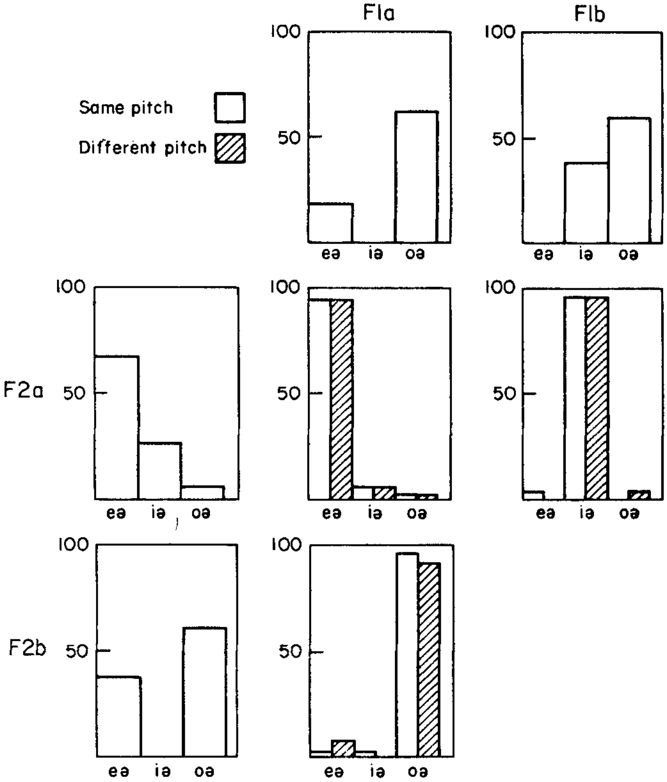

FIGURE 2. Percentage of trials on which each dipthong response was given to the separate formants and their binaural combinations in Experiment II.

Subjects had no difficulty assigning to the intended categories two-formant diphthongs with both formants starting simultaneously, despite minimal training with the basic sounds, whether the two formants were on the same F_0 (95·6% correct) or on different fundamentals (94·9% correct). These rates are much higher than would be predicted from identification of the isolated formants, as can be seen from Figure 2 and Table II. This point is particularly clear for the diph-

Table II

Percentage of trials on which each sound used in Experiment II was categorised as "ear", "air", or "oar", depending on formant fundamental and relative formant onset-times

	Single formants			
	1a	1b	2a	2b
Response				
Ear	0	38	27	0
Air	38	1	68	24
Oar	62	61	5	80

	Simultaneous pairs					
	1b+2a		1a+2a		1a+2b	
Pitch response	Same	Different	Same	Different	Same	Different
Ear	97	97	5	5	2	0
Air	3	0	93	93	2	8
Oar	0	3	2	2	97	92

	Staggered pairs					
	1b+2a		1a+2a		1a+2b	
Pitch response	Same	Different	Same	Different	Same	Different
Ear	97	95	2	3	2	0
Air	2	5	93	75	0	0
Oar	2	0	5	22	98	100

thong "ear". Neither of its component formants (1b, 2a) is identified by itself predominantly as "ear", yet when the two formants are presented simultaneously, regardless of whether they are on the same F_0 or not, "ear" is reported overwhelmingly. This experiment thus shows that the timbres, rather than the categories of the individual formant patterns are combining. Although the constituent formants of the other two diphthongs resembled these sounds more than was the case for "ear", the number of appropriate fusion responses for the combined sounds was again much higher than the responses to the isolated formants would indicate. All six subjects showed overall more fusion responses than predicted from their responses to the isolated formants. Staggering the onsets of the formants did not significantly reduce the number of fusion responses even when the formants were on different fundamentals.

Despite the absence of any effect of F_0 or formant synchrony on subjects' categorisations of the sounds, these variables did have a marked effect on the number of sounds heard.

Figure 3 shows the percentage of trials on which a particular quality of additional sound was reported (isolated formants are excluded since they were almost never heard as being more than one sound). Analysis of variance on the scores shown in Figure 3 reveals that more additional sounds were heard when the formants were

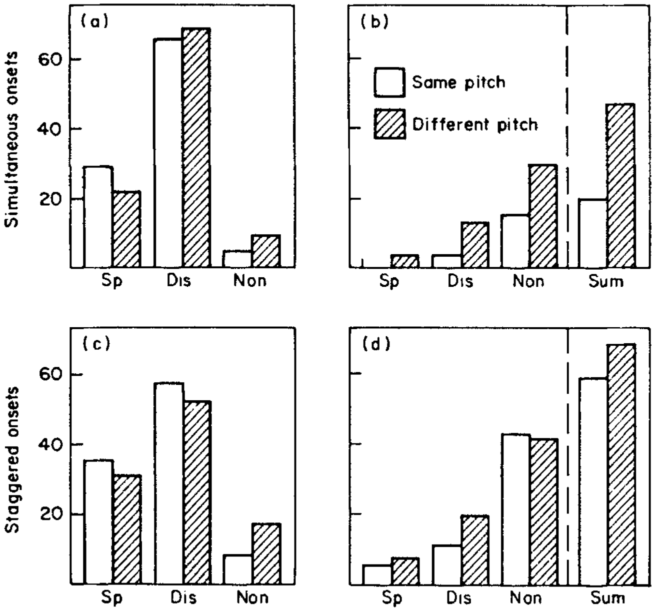

FIGURE 3. Number of speech quality judgements made in Experiment II to either the main sound heard (a) (c), or to any additional sound heard (b) (d).

staggered than when they were simultaneous [$F(1,5)=16{\cdot}6$, $P<0{\cdot}01$], and when the formants were excited on different fundamentals than by the same F_0 [$F(1,5)=14{\cdot}3$, $P<0{\cdot}025$], but there was no interaction between these two variables. There was a weak interaction between staggering and speechlikeness [$F(2,10)=5{\cdot}39$, $P<0{\cdot}05$] due to the additional sound being heard as more speech-like when the first formant started first than when the formants were simultaneous. This may be related to isolated first formants being heard as relatively speech-like. The main sound was generally rated as being distorted speech for the formant pairs and for the isolated first formants but as non-speech for the isolated second formants. There is a small shift away from speech-likeness when the formants are played on different fundamentals [$F(2,10)=5{\cdot}0$, $P<0{\cdot}05$] and a change in speech-likeness with stagger which is in different directions for different vowels [$F(4,20)=5{\cdot}80$, $P<0{\cdot}005$]; "ear" and "air" get less speech-like whereas "oar" becomes more speech-like.

Discussion

Experiment II provides a better test of whether subjects are combining formants with different fundamentals than did the first experiment. As in the earlier experiment subjects are hearing a category that corresponds to both formants even though they are on different fundamentals and/or start at different times. Moreover, this experiment also replicates the earlier finding that these manipulations are sufficient to cause subjects to hear multiple sound sources.

The next experiment looks to see whether the same holds true when formants are presented dichotically rather than binaurally.

Experiment III

Method

Stimuli

The same formants as in the simultaneous condition of the previous experiment were used, but the experimental design was improved since it had been realised that the combination of formants 1b and 2b gave a plausible realisation of the diphthong /uə/, as in "doer" (one who does). The basic formants remained identical.

The two main conditions were whether the formants were presented binaurally (with identical sounds coming to the two ears) or dichotically. Again, pairs of formants (example configurations are given in Table III) and isolated formants were played to subjects, and in addition combinations of four formants with various assignments of formants to ears and fundamentals. These appear, along with example stimulus combinations, in Table IV. There were two binaural conditions, one where all four formants were played on the same F_0, and one where two formants were on one F_0 and two on another. In addition there were three dichotic conditions, in two of them each ear received a pair of formants appropriate to a diphthong, with the two ears either being excited by the same or on different fundamentals; in the third dichotic condition (called split-formant), two of the formants from the dichotic, different-F_0 condition were switched across the ears, so that although each ear received a first and a second formant, they were on different fundamentals. This last condition tests whether formants will tend to group by F_0 rather than by ear. For instance, in the example given in Table IV grouping the formants by ear would give a precept of "oar" and "ear", whereas grouping them by F_0 gives "air" and "(d)oer".

TABLE III

Average number of sounds reported, percentage of total response corresponding to the fused formant category, percentage of trials on which at least one response was the fused category and example formant configurations for two-formant sounds in Experiment III

F_0	Average number of sounds	Fused (%)	With fused (%)	Example stimulus
Binaural				
Same	1·07	87·7	92·8	$F_{1a}L+F_{2b}L$
Different	1·33	78·1	94·1	$F_{1a}L+F_{2b}H$
				Left Right
Dichotic				
Same	1·06	87·0	91·3	$F_{1a}L$ $F_{2b}L$
Different	1·92	70·1	90·6	$F_{1a}L$ $F_{2b}H$

Procedure

The procedure was similar to the previous experiment except that subjects were asked to indicate one of the four diphthong categories for every sound they heard that they could classify in this way, and also to write down the *total* number of sounds that they heard on each trial. This procedure sacrifices information on speech-likeness for simplicity for the subjects, allowing them to concentrate more on what categories they heard.

Eight subjects (staff and students of the laboratory) initially heard a demonstration of the four diphthongs with the F_0 the same on both formants, once with the low and once with the high F_0. They were then given three tokens of each of these sounds in a random order to identify, which all did perfectly. They then did the main experiment, first with binaural presentation of five tokens of each of the different sounds (150 trials in all); after a short pause they took the dichotic condition (which used the same tape) after being told that now they might hear different sounds in the two ears.

Results

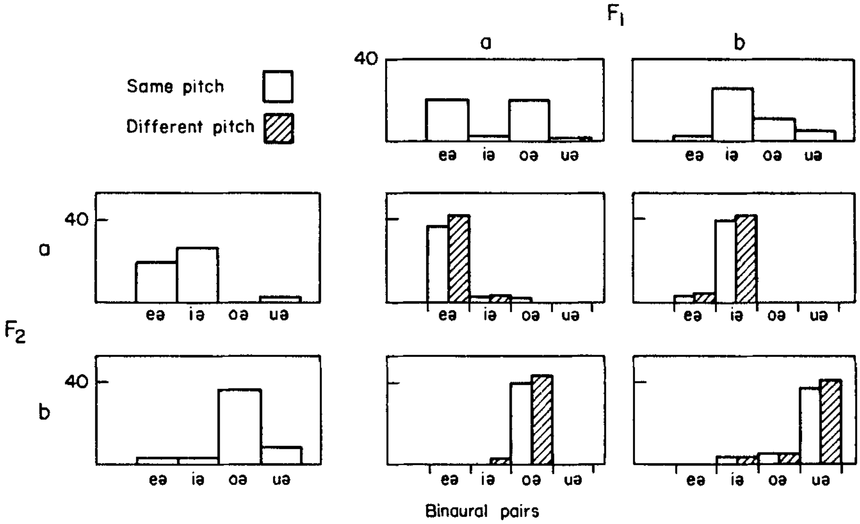

FIGURE 4. Average number of trials (out of 40) in Experiment III on which a particular diphthong response was made to binaural presentation of either a single formant, or to a pair of formants. Responses to the individual formants appear round the edge of the matrix, with responses to their combinations appearing at their intersection within the matrix.

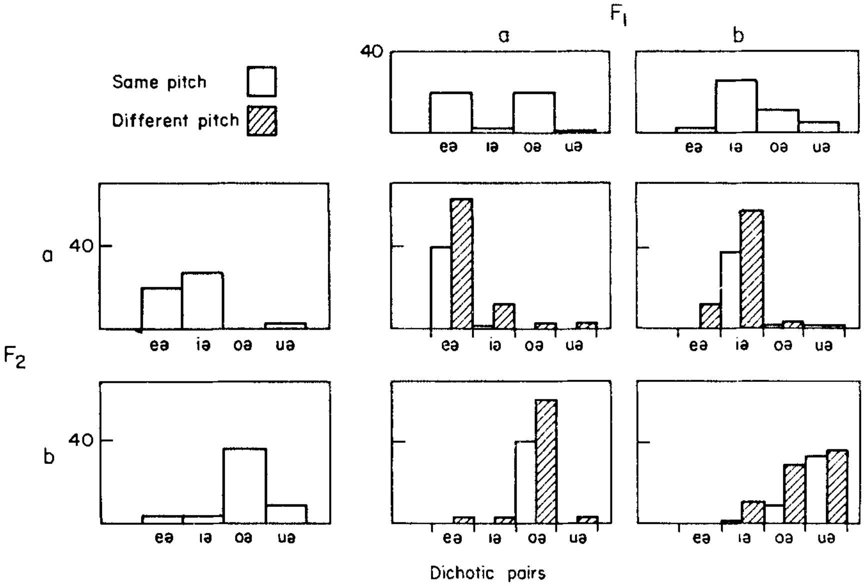

FIGURE 5. As Figure 4, but for dichotic presentation, with different formants being played to different ears.

Figure 4 shows the number of diphthong responses given to the isolated formants andto their various binaural pairings. Figure 5 does the same for dichotic pairings. As in the previous experiment the pattern of responses to the formant pairs cannot be predicted from that to the isolated formants. The best example of this is for the diphthong /uə/. Very few "(d)oer" responses were given to either F_{1b} or F_{2b} (its constituent formants) alone, but they formed the overwhelming majority when both formants were present. Again we have a valid test of the fusion of timbre.

Table IV

Average number of sounds reported, percentage of responses reflecting grouping by F_0 and/or by ear, and example formant configurations for four-formant sounds in Experiment III

		Mean	Grouped by				
		number	F_0	ear			
	F_0	of sounds	(%)		Example stimulus		
Binaural	Same	1·56	—	—	$F_{1a}L+F_{1b}L$	+	$F_{2a}L+F_{2b}L$
	Different	1·81	43·1	—	$F_{1a}L+F_{1b}H$	+	$F_{2a}H+F_{2b}L$
					Left		Right
Dichotic	Same	1·54	—	57·6	$F_{1a}L+F_{2a}L$		$F_{1b}L+F_{2b}L$
	Different	1·98	67·0		$F_{1a}L+F_{2a}L$		$F_{1b}H+F_{2b}H$
Split-formant	Different	2·00	47·4	52·6	$F_{1a}L+F_{2b}H$		$F_{1b}H+F_{2a}L$

Table IV shows that although the total number of sounds heard increases slightly when binaural formants are on different fundamentals, the proportion of trials with *at least one* response corresponding to the fused percept is no lower (94·1 %) than when F_0 is the same (92·8%) on both formants.

A similar pattern of results is found, confirming Cutting's claims, when the formants are presented dichotically. Now, although there is a marked increase (from 1·06 to 1·92) in the total number of sounds heard when the ears receive different rather than the same fundamentals, there is again no significant reduction in the number of trials on which at least one of the categories reported is appropriate to the combination of the two formants.

The same conclusion can be drawn from the results of the more complex conditions that presented four formants simultaneously. When four formants are presented binaurally, with each pair of F_1 and F_2 on a different F_0, there is no suggestion that subjects are more likely to report the categories formed by the pairs that have the same F_0, indeed the proportion of trials on which this happens (43·1 %) is less than would be expected by chance. Additionally, in the split formant condition, there is no evidence that a common F_0 can override a common ear in determining which pairs of formants will be categorised together. In fact there is only weak evidence (an increase from 57·6 to 67·0%) that a pair of formants presented to one ear will be categorised together more often if it differs in F_0 from the pair on the other ear than if it has the same F_0.

The detailed pattern of responses to the various combinations of formants and fundamentals gives no support to the notion that grouping is determined by a common fundamental, but suggests rather that certain formant/fundamental combinations are dominant and tend to be incorporated into a perceived category with other formants whatever the latters' fundamental. Others, especially the low F_1 on a high F_0 contour, are particularly weak.

Discussion

The results of this experiment support Cutting's conclusion that a common F_0 provides little justification for grouping two formants together for perceptual categorisation. But both this experiment and Cutting's used short sounds (mine lasted 140 ms and Cutting's formant transitions 70 ms). I have made a number of observations myself on longer sounds, similar to those used here, and find that for sounds lasting longer than about 0·5 s grouping by F_0 *does* become significant dichotically, but not binaurally for pairs of formants. In the split-formant condition longer sounds were no more likely to group by F_0 than were the short original sounds. The former observations were made on the F_{1b} plus F_{2b} combination which yields /uə/ when fused, but a predominantly /iə/ percept when not fused. The original 140 ms formants were digitally spliced onto their temporal mirror images to give a basic 280 ms sound—/uəu/, which was repeated ten times. The initial percept of /uə/ gave way under dichotic presentation of formants on different fundamentals to /iəi . . ./ heard on the ear receiving the first formant after a couple of cycles. This change in percept did not happen for the dichotic condition when the fundamentals were the same on both ears or in the binaural condition with the two formants on different fundamentals. Neither could I find any evidence from my own listening that repetition of the sound in this way gave any increased grouping by F_0 in the split-formant condition, although my observations on this were limited. Obviously more observations and experiments are needed to confirm these initial observations, but they do form an interesting parallel to streaming phenomena for sounds with alternating frequencies (Bregman and Campbell, 1971; Darwin and Bethell-Fox, 1977), where the auditory system requires some time to overcome the initial hypothesis that only a single source of sound is present (Bregman, 1978).

So far, our experiments have proved remarkably unsuccessful in demonstrating any effect of grouping by fundamental on the perception of timbre. The results of the first experiment (on vowel identification) and the two-formant conditions of the second and third experiments indicate that the mechanism responsible for assembling vowel quality is able to group together formants on different fundamentals at least when the formants so grouped make a plausible speech sound. Except in the four-formant conditions of the last experiment there has not been any alternative grouping of the formants (apart from them standing in isolation) that a common fundamental could have provided. So the two-formant results should be interpreted as showing that vowel categories *can* be achieved across different fundamentals, rather than as indicating that F_0 plays no role in formant grouping. Speech constraints can override a difference in periodicity, at least until streaming effects exert a more powerful influence on fundamental than in the short sounds used in these experiments. In the four-formant conditions of the last experiment no such explanation can cover the failure to find significant grouping by fundamental, but here the formants that had to be disentangled were rather close together in frequency (especially F_{1a} and F_{1b}); this may have made the task of sorting out which formants were on which fundamentals more difficult. Scheffers (1979) found only weak evidence for subjects being able to use fundamental frquency to separate out two simultaneous vowels.

Experiment IV

In order to test whether grouping by fundamental occurs under more favourable conditions than those in the previous experiment, a different strategy is adopted. Two syllables are used which have two of their three formants in common; moreover the second formant trajectory of one of the syllables (/li/) is equivalent to the third formant of the other (/ru/). Speech constraints, in carefully synthesised versions of the syllables, might then be sufficiently ambivalent about which is the preferred grouping for any grouping by fundamental to have a chance to appear. Moreover, the formants in the following experiment are more widely spaced in frequency than those of Experiments II and III, perhaps making easier the job of assigning a fundamental to individual formants. It should be noted, however, that this restriction considerably weakens the utility of F_0 grouping with more than one speaker, since the formant trajectories of two speakers would normally overlap more than do the formants we have used here.

Method

Stimuli

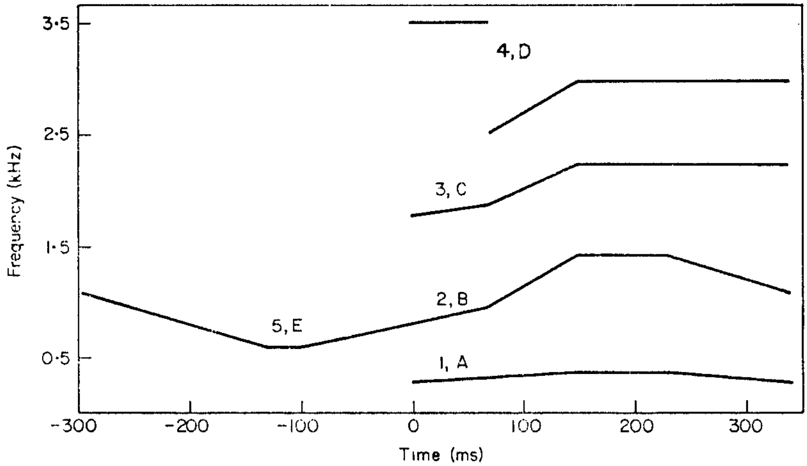

FIGURE 6. The right-hand side of the figure shows the formants used to synthesise the syllables /li/ and /ru/ in Experiment IV. The left-hand side shows the addition to the second formant used in the second part of Experiment IV.

The sounds for this experiment were based on the two syllables /li/ and /ru/. Considerable care was taken in synthesising good versions of the two syllables, taking formant values from analyses by linear predictive coding of natural utterances, since pilot experiments with inadequate syntheses had failed to find any convincing effect of F_0 grouping, except with very experienced listeners. Good synthetic versions of these syllables can be produced using for each syllable three formant trajectories from a total of four. The right-hand side of Figure 6 shows the four basic formants, which lasted 330 ms, /ru/ is given by the first three formants, while /li/ is heard if the fourth formant is substituted for the second. The upper three formants were 11, 23 and 27 dB weaker respectively than the first formant throughout the syllable. The frequency discontinuity in /li/'s highest formant is clearly visible on spectrograms and is thought to be a consequence of a spectral zero in the third-formant region, created by the formation of a tube behind the tongue tip, disappearing as the /l/ is released (Fant, 1960, p. 162). The falling third-formant transition conventionally used in synthesising an /l/-vowel syllable is not, in my experience, evident on spectrograms. These four formants were synthesised on two fundamentals, 110 and 170 Hz. I will refer to those on the lower F_0 with the digits 1–4, and those on the higher F_0 with the letters A to D. Their addition is denoted by concatenation, so that the canonical form of /ru/ is 123 or ABC.

The following combinations of formants were used in the main experiment: (1) the canonical forms of the two syllables on the low or high F_0 (123, ABC for /ru/; 134, ACD for /li/); (2) the pair of formants that are common to both syllables (13, AC); (3) the single formants that are added to the common pair (2, B; 4, D); (4) all four formants on a common F_0 (1234, ABCD); (5) all four formants with those representing a syllable on one F_0 and the additional formant on a different F_0 (123D, ABC4 for /ru/; 1B34, A2CD for /li/). As an adjunct to the main experiment a second set of conditions was run which included some of the above sounds, namely (1) (4) and the /li/ versions of (5) but which had additional sounds derived from the latter two categories. Their second formant had a 300-ms precursor whose final slope was the same as the initial slope of the original. The prefixed formant is referred to as 5 or E depending on its F_0. so that a sound consisting of all four basic formants on the low F_0, prefixed by this trajectory for the isolated second formant, but on the high F_0, is denoted E1234. All combinations of 1234, ABCD, 1B34 and A2CD with 5 and E were used.

The reasons for including this extended precursor are as follows. The previous experiments have used only 100-ms differences between the onset times of the different formants and have shown no effect of this variable on categorisation; so a longer onset time was used in this experiment to try to find some effect of onset time. The particular form of the precursor trajectory allows a cyclical stimulus to be constructed, similar to those used by Bregman in auditory streaming experiments, which changes smoothly in both the frequency and the frequency slope of the second formant. Had no effect of the precursor on categorisation been found in this experiment, then its effect could have been tested with the presumably more potent cyclical stimulus. In the event the precursor did show an effect, so it was not necessary to use the cyclical stimulus.

Procedure

Subjects were run individually in a sound-attenuating booth, on-line to a PDP-12 computer that played stored waveforms in response to button presses. The appropriate sound started 500 ms after a button press. Subjects could repeat any sound by pressing one of the buttons, or register their categorisation of it to the computer by pressing others.

The 12 subjects (staff and students of the laboratory) were given extensive practice in identifying the canonical forms of the two syllables and the isolated additional formants. Two buttons were labelled "ru" and "li" and two others were left blank, the subject providing his own labels for the isolated second and fourth formants. We will refer to these responses as "2B" for sounds 2 and B, and "4D" for sounds 4 and D; typical labels actually given were "ua" and "gli" respectively. Subjects initially heard the eight sounds as often as they wished, being able to repeat each sound or the series indefinitely. When they thought they had learned to label each sound they were given a test of three tokens of each sound in random order. If they had a perfect score they were given a further identification test of five tokens of each sound. If they failed at either test they returned to the previous stage.

Throughout this training and the experiment proper, subjects were always able to listen as often as they wished to any sound and were entirely self-paced.

TABLE V

Percentage of trials on which a given category was heard in first part of Experiment IV, and mean number of categories reported for each stimulus. Letters and digits refer to the formants of Figure 6

	/ru/	/li/	"2B"	"4D"	Mean number of categories
13		100			1·0
AC		100			1·0
2			100		1·0
B			100		1·0
4		15		88	1·03
D		5		95	1·0
123	98		5		1·03
ABC	100	3	10	15	1·28
134		100			1·0
ACD		97		3	1·0
1234	98	12	3	10	1·18
ABCD	100	10	2	33	1·45
123D	*100*	13	8	*42*	1·63
ABC4	*100*	23	2	*37*	1·62
1B34	63	*42*	*90*	7	2·02
A2CD	48	*65*	*77*	5	1·95

Categories expected if grouping is by pitch are in italics, underlined.

When they had reached 100% correct identification in the five-token test, they were then given another five tokens each of the 16 different basic stimulus combinations, which are listed in Table V. Subjects were told that they would hear the sounds that they had heard before plus some that might sound like compounds. If they heard more than one of the basic sounds they were to press more than one button. No provision was made for them recording the double occurrence of the same category. They were also told that they could listen again to the demonstration of the eight basic sounds whenever they wished, by contacting the experimenter.

After they had taken this experiment they passed on to the second part which used the additional prefixed stimuli as described earlier, and listed in Table VI. This part was done after they had again attained perfect identification of the eight basic sounds. For this part, subjects were told that they would hear again some of the sounds from the previous main experiment together with some in which the sound was preceded by another sound which they were to ignore. Otherwise their task was identical.

They were given a practice run of two tokens of each of the 16 sounds to ensure that they understood the task. This second part again had five tokens of each sound for the subjects to identify and again they were free to ask for a demonstration of the basic sounds whenever they liked, and to listen to each trial as often as necessary.

Results

In the first part of the experiment there was a clear effect of grouping according to F_0. On the null hypothesis that there is no F_0 grouping we would expect responses to all stimuli with all four formants present to be substantially the same. As

Table VI

Percentage of trials on which a given category was heard in second part of Experiment IV. Letters and digits refer to the formants of Figure 6

	/ru/	/li/	"2B"	"4D"
123	100	2		
ABC	90	17	5	7
134	3	98		
ACD		88		10
1234	85	33	5	3
ABCD	62	43	18	20
1B34	33	77	47	
A2CD	8	85	58	5
51234	55	45	2	
EABCD	37	67		10
E1234	38	65	2	
5ABCD	33	57	3	12
E1B34	7	93	10	
5A2CD	5	90	2	8
51B34	15	88	3	
EA2CD		90	10	10

Table V shows, this was not the case. When all four formants were on the same F_0 all subjects heard predominantly /ru/ (on 98% of trials for 1234 and 100% for ABCD), with "4D" being given in addition on 33% of trials to ABCD. When the fourth formant was played on a different F_0 from the other three, there was an increase in "4D" responses for 123D (41·7%) over 1234 (10 subjects for, none against), but not for ABC4 (36·7%) over ABCD. However, when the *second* formant was on a different F_0 from the other three there was a clear increase in /li/ responses for 1B34 (41·7%) over 1234 (6·7%; $T=5{\cdot}5$, $n=10$, $P<0{\cdot}025$) and for A2CD (65·0%) over ABCD (10·0%; all 12 subjects for), and in "2B" responses for 1B34 (90·0%) over 1234 (6·7%; all subjects for), and for A2CD (65·0%) over ABCD (1·7%; all subjects for). All the canonical syllables were identified at better than 95% correct and the two two-formant sounds (13, AC) were unanimously heard as /li/.

Fewer categories were reported when all four formants were on the same F_0 (1·34) than when one was on a different one (1·89; all subjects for). But there were also fewer sounds reported when all four formants were on the lower fundamental (1·18) than on the higher (1·45; six for, none against), and fewer when the odd formant was the fourth (1·66) than when it was the second (2·0; 11 for, one against).

In the second part of the experiment there were generally more /li/ responses when there was precursor second formant (74·4% overall) than when there was not (59·6%; 11 subjects for, none against). But neither the F_0 of the precursor nor its relation to the F_0 of the second formant in the main syllable had any significant effect on the number of /li/ responses, although the effect of the second formant's F_0 within the main syllable replicated the first part of the experiment. The data in Table VI also show that subjects are biased towards /li/ in the second compared to the first part of the experiment. It is impossible to tell whether this is due to the greater experience the subjects had had with the sounds, or to the absence of /ru/-like sounds in the second half. The number of sounds reported is not analysed in this part of the experiment since subjects were told to ignore the precursor sound.

Discussion

There is a clear effect of grouping by F_0 in the first part of this experiment. Of the responses given to stimuli which had formants on different fundamentals, 78% corresponded to categories expected from grouping by F_0. There is also a small effect of formant asynchrony in the second part of the experiment, with the number of /li/ responses (the response indicating that the second formant has not been integrated into the category) increasing by 15% when the second formant is extended forward to start 300 ms earlier. But this increase did not depend on the relation between the F_0 of the precursor and that of the second formant in the main sound. An obvious difference between the precursor used here and that from Experiment II is that the present one is three times as long. That may explain its success in influencing the perceived speech category.

Even for the main sound in the first part of the experiment the grouping effect of F_0 is not overwhelming. For stimulus 1B34, which has the formants for /li/ on the low fundamental and the second formant of /ru/ on a higher F_0, there is still a majority of /ru/ responses. This sound illustrates a further point. Although the second formant was heard as a separate category ("2B") on 90% of trials, it also contributed to the perception of /ru/ on 63% of trials (note that the first and third formants by themselves are heard as /li/). On some trials at least, the same formant is thus being heard in two ways: as itself, and also as part of a more complex whole. This same phenomenon has been noted when isolated second formants or just their transitions are played to one ear and the remainder of a syllable to the other ear. The formant transition is heard as a non-speech sound, but it also gives a distinctive place of articulation to the syllable heard in the other ear (Cutting, 1976; Liberman and Isenberg, 1980; Rand, 1974).

A third point from Experiment IV is that having a group of simultaneous formants on the same fundamental does not prevent the individual formants being heard as separate categories. This is particularly clear in sound ABCD, which has all four formants on the high fundamental. Nine out of 12 subjects gave at least one "4D" response to this sound in addition to always hearing /ru/. The observation that a formant that violates phonetic constraints stands out from the speech background, despite being on the same fundamental, is a common one with inadequately synthesised speech, where the offending formant often detaches from the speech as a separate sound source. A common fundamental provides neither a necessary nor a sufficient condition for formants, to be grouped together, although, as this experiment has shown it can exert some influence.

General discussion

The following points have emerged from these experiments.

(1) With binaural presentation, the identification of a three-formant vowel's (Experiment I) or a two-formant diphthong's (Experiment II and III) category is not

significantly impaired by exciting its formants with different fundamentals, or by starting formants at different times, even though both these variables increase the number of categories reported.

(2) When the constituent formants of a two-formant diphthong are presented dichotically (Experiment III), identification of its category is not impaired by ringing the formants on different fundamentals or by starting one formant earlier, though again the number of sounds heard increases. Informal observations with longer, cyclic sounds indicated that grouping by F_0 might occur for dichotic sounds lasting more than half a second or so.

(3) When all four formants of two two-formant diphthongs are played binaurally or dichotically (Experiment III), there is no evidence that the categories heard are determined by what formants have common fundamentals.

(4) With binaural presentation of four widely-spaced formants (Experiment IV), from which two different three-formant syllables could be formed, the relative F_0 of the formants and their relative onset-times influenced, but did not absolutely determine, which syllable was heard. The relative and the absolute fundamentals also influenced the number of categories reported.

It has proved surprisingly difficult to demonstrate a grouping effect of fundamental on speech categories. A clear effect was eventually found in the last experiment, but only under conditions rather remote from those encountered at the now well-known cocktail party. The conservative conclusion from these experiments on single syllables is that F_0 provides at best a weak constraint on what formants the speech categorising mechanism groups together. Other factors, such as phonetic constraints or the relative strength of different formants on different fundamentals, may override a common fundamental in determining which formants form a category. The relative onset-time of formants also exercised a surprisingly weak effect.

Informal observations after Experiment III indicated that strong grouping by F_0 was possible with longer, cyclic sounds but only with dichotically presented pairs of formants. Some recent experiments by Brokx, Nooteboom and Cohen (1979) have looked for effects of grouping by fundamental binaurally in extended speech, and some of their results do indicate such an effect. They looked at the number of content words correctly identified in semantically anomalous seven-word sentences heard against the competing background of a read passage of prose. Both the target sentences and the prose had been processed by linear pediction analysis and re-synthesis (Atal and Hanauer, 1971) so that they had either a common fundamental (though presumably different phases of exciting pulse-train) or fundamentals that differed by various numbers of semitones. A difference of three semitones between the two fundamentals produced an increase of 20% in the number of content words correctly reported, an advantage that was abolished again by making the difference in fundamental an octave. Again though, their effect was not overwhelmingly strong since even with identical fundamentals, their subjects were getting almost half of the target words correct. They could presumably use timing differences and instantaneously favourable signal-to-mask ratios to help disentangle the message from the mask.

Two factors perhaps help to explain why fundamental frequency is not a particularly strong grouping principle for speech sounds, the first is psycho-acoustic, the second phonetic. If the individual harmonics within a formant must be resolved out in order for the fundamental to be used for grouping, then the spacing between harmonics must be more than the critical band. The limits of grouping by fundamental are then equivalent to those of pitch perception. For a single male voice with a fundamental averaging around 100 Hz, the fourth and fifth harmonics are dominant in the perception of pitch, while for higher fundamentals the dominant harmonics rest around 1 kHz or so (see Plomp, 1976. Ch. 7 for a review). But when more than one voice is present, the task of resolving harmonics within overlapping formants is compounded by their increased number within a critical band. If the fundamental were extracted by temporal mechanisms (Moore, 1977, Ch. 4), similar problems would also arise with overlapping formants. It may be that the wider spacing of the four formants in Experiment IV allowed better use to be made of their harmonic spacing (or periodicity) than was possible for the overlapping formants of Experiment II.

The second, phonetic point concerns the nature of the excitation in natural speech. A purely periodic excitation is found only in ideal voiced speech, and corresponds to the vocal cords closing completely and regularly each cycle. In the rest of speech, there is some random excitation produced either at the glottis by incomplete closure of the folds leading to turbulent air flow with consequent noise, or at a constriction elsewhere. Whisper and voiceless sounds are entirely noise excited, but it is less well known that much of conventionally voiced speech has some noise excitation particularly in the higher formants. This noise becomes particularly noticeable in "breathy" voice, but the quality of synthetic speech is generally improved if formant excitation can be a variable mixture of low frequency buzz and high-frequency hiss (Holmes, 1973; Makhoul *et al.*, 1978). If grouping is not to be appreciably worse for breathy voice and whisper than it is for voiced speech (and we do not know whether it is), the grouping mechanisms must at least be able to take together, using phonetic or other auditory constraints, formants that do not have the same excitation.

A similar point can also be made concerning onset-time. Although formants often start and stop at similar times there are many occasions in speech when they do not. Stop bursts have little low-frequency energy, while nasal murmur and voice-bar lack high frequencies. Formants that start and stop at different times must be integrable into a common phonetic category.

Nevertheless, given suitable conditions, we have been able to show some grouping effect due to fundamental and to onset-time. If we assume that these processes are prior to those involved with phonetic categorisation rather than an integral part of them, and we have no evidence to the contrary, then we could describe our results in terms of a cafeteria analogy. Let us assume that low-level grouping processes are analogous to the rules that the staff of a cafeteria use to assemble basic elements into a dish that is displayed for the customer to take. These rules provide a generally useful grouping that the staff hope reflects the common tastes of their customers. The individual customer (in our instance, the speech categorisation process) is free to take more than one dish for his meal according to his particular tastes. So formants on different fundamentals etc. can be gathered up if they correspond to a phonetic category, but it is relatively difficult to reject a component of a dish.

Since the grouping effects we have found have been weak, it is possible that a better analogy (and one that does not assume that auditory grouping processes precede phonetic categorisation) is a soup-kitchen rather than a cafeteria, with each customer receiving a bowl that contains all the elements, and extracting from it what he will. Such a model is compatible with observations by Liberman and Studdert-Kennedy (1978) on the intelligibility of speech masked by additional formants that are on the same fundamental and which start and stop at the same time as the speech formants. They report that despite the inability of the eye to group from a spectrogram the subset of sounds constituting the speech signal, the ear has no difficulty in disentangling speech from mask. My own observations on similar stimuli indicate that although there is a great deal of variability between subjects in how easily they hear the speech, there is *not* a striking increase in intelligibility when the speech and the masking formants are on different fundamentals. Quite what the relationship will be between observations on this type of material and those from streaming phenomena with repeating sounds is not clear.

In summary, the subjects in Experiment IV did show some tendency to hear as a single phonetic category formants that started at the same time and were on a common fundamental. These effects were not overwhelming and failed to emerge at all in the earlier experiments. It seems unlikely that a common fundamental or onset-time exercises a very strong constraint in grouping together frequency components of normal speech. Whether there are other general auditory constraints that can do this, or whether all such grouping must be left to phonetic rather than auditory processes is not yet clear.

I would like to thank Mo Gooding and Jenny Mason, who made up and ran the experiments, supported by a Science Research Council grant; Al Bregman, with whom I had interesting discussions about auditory grouping while he was at Sussex on a Guggenheim Fellowship; Christopher Longuet-Higgins for use of his LPC algorithms; the SRC for computing facilities and two reviewers for helpful comments on the typescript. Preliminary versions of Experiment III have appeared in Darwin (1979*a*,*b*).

References

Atal, B. S. and Hanauer, S. L. (1971). Speech analysis and synthesis by linear prediction of the speech wave. *Journal of the Acoustical Society of America*, **50**, 637–55.

Bregman, A. S. (1978). The formation of auditory streams. In Requin, J. (Ed.), *Attention and Performance VII.* Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Bregman, A. S. and Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology*, **89**, 244–9.

Bregman, A. S. and Pinker, S. (1978). Auditory streaming and the building of timbre. *Canadian Journal of Psychology*, **32**, 19–31.

Broadbent, D. E. (1952). Failures of attention in selective listening. *Journal of Experimental Psychology*, **44**, 428–33.

Broadbent, D. E. (1958). *Perception and Communication.* Oxford: Pergamon.

Broadbent, D. E. and Ladefoged, P. (1957). On the fusion of sounds reaching different sense organs. *Journal of the Acoustical Society of America*, **29**, 708–10.

BROKX, J. P. L., NOOTEBOOM, S. G. and COHEN, A. (1979). Pitch differences and the integrity of speech masked by speech. *IPO Annual Progress Report*, **14**, 55–60. Eindhoven: Netherlands.

CHERRY, E. C. (1953). Some experiments on the recognition of speech with one and with two ears. *Journal of the Acoustical Society of America*, **25**, 975–9.

CUTTING, J. E. (1976). Auditory and linguistic processes in speech perception: inferences from six fusions in dichotic listening. *Psychological Review*, **83**, 114–140.

DANNENBRING, G. L. and BREGMAN, A. S. (1978). Streaming versus fusion of sinusoidal components of complex tones. *Perception and Psychophysics*, **24**, 369–76.

DARWIN, C. J. (1976). The Perception of Speech. In CARTERETTE, E. C. and FRIEDMAN, M. P. (Eds), *Handbook of Perception VII: Language and Speech*, pp. 175–226. New York: Academic Press.

DARWIN, C. J. (1979*a*). Perceptual grouping of speech components. In CREUTZFELDT, O., SCHEICH, H. and SCHREINER, C. (Eds), *Hearing Mechanisms and Speech*, Experimental Brain Research Supplementum 2, pp. 333–40. Berlin: Springer-Verlag.

DARWIN, C. J. (1979*b*). Perceptual grouping of speech sounds. *Proceedings of the Institute of Acoustics, Autumn Conference*, pp. 89–92.

DARWIN, C. J. and BETHELL-FOX, C. E. (1977). Pitch continuity and speech source attribution. *Journal of Experimental Psychology: Human Perception and Performance*, **3**, 665–72.

DEUTSCH, D. A. and ROLL, P. L. (1976). Separate "what" and "where" decision mechanisms in processing dichotic tonal sequences. *Journal of Experimental Psychology: Human Perception and Performance*, **2**, 23–9.

EGAN, J. P., CARTERETTE, E. C. and THWING, E. J. (1954). Some factors affecting multichannel listening. *Journal of the Acoustical Society of America*. 26, 774–82.

FANT, G. (1960). *Acoustic Theory of Speech Production.* The Hague: Mouton.

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

HEISE, G. A. and MILLER, G. A. (1951). An experimental study of auditory patterns. *American Journal of Psychology*, **64**, 68–77.

HELMHOLTZ, H. (1954). *On the Sensations of Tone.* New York: Dover.

HOLMES, J. (1973). The influence of glottal waveform on the naturalness of speech from a parallel formant synthesizer. *IEEE Transactions*, AU-**21**, 298–305.

KLATT, D. H. (1980). Speech perception: a model of acoustic–phonetic analysis and lexical access. In COLE, R. (Ed.), *Perception and Production of Fluent Speech.* Hillsdale, New Jersey: Laurence Erlbaum Associates.

LIBERMAN, A. M. and ISENBERG, D. (In press). Duplex perception of acoustic patterns as speech and non-speech. *Haskins Laboratories Status Report on Speech Perception*, SR-**62**, 47–58.

LIBERMAN, A. M. and STUDDERT-KENNEDY, M. G. (1978). Phonetic Perception. In HELD, R., LEIBOWITZ, H. and TEUBER, H.-L. (Eds), *Handbook of Sensory Physiology VIII: Perception.* Heidelberg: Springer-Verlag.

MAKHOUL, J., VISWANATHAN, R., SCHWARTZ, R. and HUGGINS, A. W. F. (1978). A mixed-source model for speech compression and synthesis. *Journal of the Acoustical Society of America*, **64**, 1577–81.

MILLER, G. A. and HEISE, G. A. (1950). The trill threshold. *Journal of the Acoustical Society of America*, **22**, 637–8.

MITCHELL, O. M. M., ROSS, C. A. and YATES, G. H. (1971). Signal processing for a cocktail party effect. *Journal of the Acoustical Society of America*, **50**, 656–60.

MOORE, B. C. J. (1977). *Introduction to the Psychology of Hearing.* London: Macmillan.

NOOTEBOOM, S. G., BROKX, J. P. L. and de ROOIJ, J. J. (1978). Contributions of prosody to speech perception. In LEVELT, W. J. M. and FLORES-D'ARCAIS, G. B. (Eds), *Studies in the Perception of Language*, pp. 75–109. New York: Wiley.

PARSONS, T. W. (1976). Separation of speech from interfering speech by means of harmonic selection. *Journal of the Acoustical Society of America*, **60**, 911–8.

PLOMP, R. (1976). *Aspects of Tone Sensation.* London: Academic Press.

RAND, T. C. (1974). Dichotic release from masking for speech. *Journal of the Acoustical Society of America*, **55**, 678–80.

RASCH, R. A. (1978). Perception of simultaneous notes such as in polyphonic music. *Acustica*, **40**, 21–33.

SCHEFFERS, M. T. M. (1979). The role of pitch in perceptual separation of simultaneous vowels. *IPO Annual Progress Report*, **14**, 51–54. Eindhoven, Netherlands.

TREISMAN, A. M. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology*, **12**, 242–8.

Manuscript received 30 *September* 1980