Canadian Journal of Psychology, 1990, 44(3), 400-413

Auditory Grouping Based on Fundamental Frequency and Formant Peak Frequency

Albert S. Bregman, Christine Liao, and Robert Levitan
McGill University

ABSTRACT The perceptual grouping of a four-tone cycle was studied as a function of differences in fundamental frequencies and the frequencies of spectral peaks. Each tone had a single formant and at least 13 harmonics. In Experiment 1 the formant was created by filtering a flat spectrum and in Experiment 2 by adding harmonics. Fundamental frequency was found to be capable of controlling grouping even when the spectra spanned exactly the same frequency range. Formant peak separation became more effective as the sharpness (amplitude of the peak relative to a spectral pedestal) increased. The effect of each type of acoustic difference depended on the task. Listeners could group the tones by either sort of difference but were also capable of resisting the disruptive effect of the other one. This was taken as evidence for the presence of a schema-based process of perceptual grouping and the relative weakness of primitive segregation.

RÉSUMÉ Nous avons étudié le regroupement perceptuel d'un cycle de quatre sons en fonction des différences dans les fréquences fondamentales et les fréquences de pics spectraux. Chaque son avait un seul formant et au moins 13 harmoniques. Dans l'expérience 1 le formant était créé par le filtrage d'un spectre plat et dans l'expérience 2 par l'addition d'harmoniques. Nous avons trouvé que la fréquence fondamentale était capable de contrôler le regroupement même quand les spectres traversaient exactement la même étendue de fréquences. La séparation entre les pics de formant devenait plus efficace à mesure que l'acuité (l'amplitude du pic relatif au piédestal spectral) augmentait. L'effet de chaque type de différence acoustique dépendait de la tâche. Les sujets pouvaient regrouper les sons par l'une ou l'autre des différences et étaient aussi capables de résister à l'effet perturbateur de l'autre différence. Ceci a été considéré comme une évidence de la présence d'un processus schéma-dépendant du regroupement perceptuel et de la faiblesse relative de la ségrégation primitive.

When a repeating cycle of tones, some in a higher frequency region and some in a lower region, is repeated rapidly, the listener hears two streams of sound, one composed of the higher tones and the other of the lower tones (Bozzi & Vicario, 1960; Miller & Heise, 1950). This phenomenon has been referred to as *auditory stream segregation* (Bregman & Campbell, 1971). Stream segregation makes it hard to hear patterns that cross the two streams, be they melodic patterns (Bregman & Campbell, 1971; Dowling, 1973) or rhythmic patterns (Bregman, 1978b; Handel, 1984; van Noorden, 1975). The time interval between events in two different streams is also hard to judge (e.g., Divenyi, 1971; Fitzgibbons, Pollatsek, & Thomas, 1974; Kinney, 1961; Neff, Jesteadt, & Brown, 1982; van Noorden, 1975).

The segregation of the two subsets of tones becomes greater with increasing frequency separation and with increasing speed (van Noorden, 1975, 1977, 1982).

Author's address: Dr. A.S. Bregman, Department of Psychology, McGill University, 1205 Docteur Penfield Avenue, Montreal, PQ, Canada H3A 1B1.

It also increases with repetitions of the cycle (Anstis & Saida, 1985; Bregman, 1978a). While the segregation in most of these studies was produced by frequency separation, other kinds of differences between sounds can also cause them to form separate perceptual streams. For example, streams can be formed on the basis of differences in spatial location (Axelrod & Guzy, 1968; Judd, 1979; Massaro, 1976) or differences in the direction of the gliding of tones (Steiger & Bregman, 1981).

Much of the research on stream segregation has been done with pure tones. Such tones are unique in that they confound two height dimensions that occur in more complex periodic sounds. These dimensions are the height of the fundamental frequency and the height of the mean of the spectral energy. With pure tones, these are always the same. In complex tones, however, the former is thought to give rise to our sense of musical pitch while the latter is responsible for a timbral dimension, the *brightness* of the tone (Wessel, 1979).

Differences in timbre can contribute to segregation. Research on this basis for segregation has defined timbre in many different ways: as the contrasts between different instruments playing the same pitch (Erickson, 1975; Smith, Hausfeld, Power, & Gorta, 1982), between arbitrary qualitative classes such as buzzes, pitches, and noise bursts (Dannenbring & Bregman, 1976; McNally & Handel, 1977), between tones with and without extra partials added to a fundamental (McAdams & Bregman, 1979), between tones differing on brightness (Wessel, 1979), or in terms of tones formed from different subsets of harmonics of the same fundamental, the subsets occupying different frequency bands (Singh, 1987; van Noorden, 1975).

The question of whether fundamental frequency (F_0) alone, independently of the region of the spectrum occupied by the harmonics, can affect grouping has not been clearly answered in the literature. Positive evidence has been found. For example, sudden F_0 changes have been shown to dissociate the parts of a synthesized speech syllable (Darwin & Bethell-Fox, 1977; Green, Stevens, & Kuhl, 1989). Furthermore, vowels synthesized on different fundamentals and formed into a repeating loop tend to segregate into streams grouped by fundamental frequency (Noteboom, Brokx, & de Rooij, 1978). However, these experiments employed speech stimuli exclusively.

The present experiments used complex tones to study the way in which their stream segregation was affected by their fundamental frequencies and their timbres by varying the two independently. A similar manipulation was performed by Singh (1987). She found a contribution of both the (missing) F_0 and the spectral region of the harmonics. To vary timbre, Singh used tones formed of four consecutive harmonics and varied the harmonic number of the lowest harmonic. Tones composed of higher harmonics had the same (missing) F_0 but sounded brighter.

The present research differed from Singh (1987) in that we studied the perceptual grouping of tones with richer spectra. Rather than differing in the selection of harmonics, the spectra of different tones in the present experiments differed in the location of a peak (formant). Like Singh, we studied how the spectral shape traded off against the fundamental frequency in controlling the grouping. The use of richer tones allowed us to avoid one of the problems faced by Singh. Her stimuli had only a small number of harmonics. She noted that the pitch of such stimuli becomes less pronounced when higher harmonics are used. Therefore they confound the saturation of the pitch with the brightness. Our tones consisted of at least 13 harmonics, with the lowest harmonic never higher than the fourth.

Experiment 2 investigated how the *sharpness* of the spectral peaks (difference between peak and valley intensities) affected their ability to control perceptual organization. In speech recognition research, it is generally accepted that the sharpness of the formants has little effect on phonetic recognition (e.g., Klatt, 1982; Roberts, 1988, chap. 3). Yet sharpness could still make it easier to follow a formant over time or to group segments of a formant across interruptions by extraneous sounds.

In Experiment 1, the spectral peaks were created by filtering a flat spectrum. In Experiment 2, in order to specify sharpness more exactly, peaks were produced by controlling the intensity of each harmonic explicitly through additive synthesis. The latter procedure also made it possible to hold the highest and lowest frequencies constant and to ensure that a harmonic always fell at the nominal spectral peak.

EXPERIMENT 1

The stimuli in this experiment were similar to those used by Singh (1987). It was a cycle of four tones in which two types of similarity were present, one based on the fundamental frequency (F_0) of the tones and another based on the frequency of a formant peak. Henceforth, these two factors will simply be referred to as F_0 and peak, respectively. The main purpose of the experiment was to study how each of these two factors affected stream segregation, especially when the two were brought into competition with one another.

Method

Task: On each trial four tones, each 100 ms in duration and differing on both peak and F_0, were formed into a 400-ms repeating loop, and the subject performed a judgement that was designed to reveal how strongly each of the two factors were promoting the formation of similarity based streams. First the subjects heard 20 repetitions of a 400-ms cycle consisting of two tones (symbolized as A and B), selected from the four chosen tones, and two silent slots (symbolized as hyphens), each 100 ms in duration, in the order AB--AB-- .... This was the *standard cycle*. Then, with no break in the AB--AB--... rhythm, this was followed by 20 repetitions of a sequence of four tones in which the two remaining tones (CD) were inserted into the two silent slots that had occurred in the standard cycle to form a *comparison cycle* (ABCDABCD...). The subjects were asked to judge how easily they could hear the standard pair (AB) continue with the same quality and rhythm when the additional sounds were added in. The instruction about rhythm was used because when sequences of sounds are absorbed into a larger stream, they tend to lose their original rhythms. The reference to quality in the instructions was used to help guarantee that the subjects were identifying the correct sounds within the cycle (since in certain cases, only the timbre distinguished the standard tones from the other tones in the comparison cycle). The judgement was made on a numerical scale running from 1 (hard) to 5 (easy). The subjects had 2.5 s to make each judgement.

The choice of the tones in the comparison cycles was varied on different trials to manipulate independently the frequency separation of the peaks and the F_0s. There were two kinds of tests created by different selections of the two tones to be used in the standard cycle. On some trials the two tones of the standard were alike in F_0; on such trials F_0 was the tested factor and peak was the interfering factor (except in certain cases where there was no interfering factor). On other trials the standard pair were alike in peak frequency; here the roles of tested and interfering factors were reversed. It was expected that if the two tones of the

TABLE 1
Design of Stimuli for Experiment 1

	Spectral Peak Position				
Fundamental	1000	1212	1470	1782	2161
128	1	2	3	4	5
155	6	7	8	9	10
188	11	12	13	14	15
228	16	17	18	19	20
277	21	22	23	24	25

Note. The numbers in the table are the identification numbers for the tones (to be read as Tone 1, Tone 2, etc.).

standard (A and B) were alike on the tested factor and different enough from C and D, they should form a separate perceptual stream in the comparison cycle and thus be heard clearly. Therefore, a high clarity score for the tested factor was taken as evidence that the particular choice of separations of F_0 and formant frequency had created a stimulus in which the tested factor was dominant in controlling the formation of streams and was resisting the disruptive influence of the interfering factor.

Design: Table 1 can be used to explain the design of the sequences. Twenty-five harmonically rich tones were created, each having a duration of 100 ms. They are shown numbered from 1 to 25 in Table 1. The F_0 of each is shown at its left and the frequency of the formant peak is shown above it. On each trial four tones were selected from this table and used for the comparison sequence. Tone 1 in the table can be called the *pivotal* tone. It appeared in every four-tone comparison cycle. The choice of the other three tones defined the similarities of F_0 and peak in that set of four tones. One of these, which we can call the *contrasting* tone, always differed from the pivotal tone both on F_0 and formant frequency; an example might be Tone 19. It was expected to be segregated from Tone 1 via both of these factors. The third tone (e.g., Tone 4) would be the same as Tone 1 on F_0 but the same as the contrasting tone on formant frequency. The fourth tone (e.g., Tone 16) was chosen to be the same as Tone 1 on formant frequency but the same as the contrasting tone on F_0. Therefore, the four tones always occupied the vertices of a rectangle in Table 1, defined by two formant frequencies and two F_0s. It can be seen that, in our example, Tones 1 and 4 both have an F_0 of 128 Hz while Tones 16 and 19 have F_0s of 228 Hz. If F_0 is a strong basis for stream segregation and formant frequency has negligible effects, two streams of tones should be heard, one with a 128-Hz F_0 (1 & 4) and the second with a 228-Hz F_0 (16 & 19). However, if F_0 has a negligible effect upon segregation and the formant frequency has a strong influence, again two streams should be heard, one consisting of the two tones whose formant peak is at 1000 Hz (1 & 16) and the other consisting of the two whose peak is at 1782 Hz (4 & 19). Since Tone 1 was always present, the selection of the contrasting tone prescribed the separations among the four tones on F_0 and formant freqency, the choice of the other two tones following automatically.

Across all the trials of the experiment, all the tones in Table 1, except for Tone 1 itself, were chosen as the contrasting tone. Notice that tones in the top row or the leftmost column were not excluded from serving as the contrasting tone despite the fact that in these cases a degenerate case of similarity would be obtained in which only one basis of similarity would exist in the group of four tones. For example, because Tone 1 was always selected, then when Tone 16 was chosen as the contrasting tone, the rule of selection of the other two tones led us to select Tones 1 and 16 again as the third and fourth tones. Hence only the factor of F_0 could differentiate the four tones (1, 1, 16, & 16). The trials that used the degenerate cases allowed us to study the stream-forming strength of either F_0 or formant frequency without competition from the other factor.

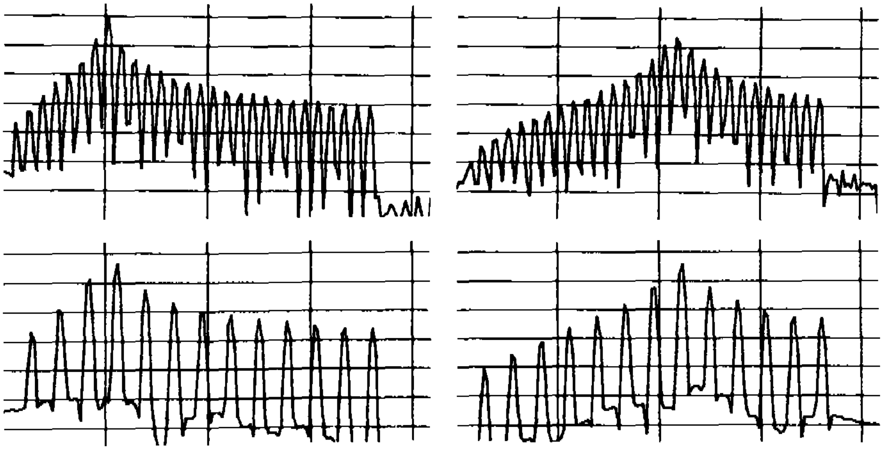

Figure 1. Spectra of four tones used in Experiment 1. Horizontal grid lines are spaced by 10 dB, vertical ones by 1000 Hz. F_0s and peaks are as follows: upper left = 128, 1000; upper right = 128, 2161; lower left = 277, 1000; lower right = 277, 2161. Note that the frequency scale is linear.

After each selection of four tones was made from Table 1, it was tested on four different trials that differed in the order of the tones in the comparison loop (e.g., CBADCBAD...). This arrangement in a particular order also determined which two tones would serve as the tones in the standard loop (e.g., CB--CB--...) since the standard loop always used the first two tones from the comparison loop. This procedure automatically produced the combinations of tested factor and interfering factor mentioned earlier.

The experiment consisted of 100 trials. Four tapes were made, each with a different random order of the 100 trials. Four subjects were tested on each tape.

Stimuli: Complex tones of 100-ms duration were used. They were digitally synthesized using the MITSYN signal processing software (Henke, 1980) running on a PDP-11 computer. Sequences were synthesized at 15,360 samples/s output on a 12-bit D/A converter, low-pass filtered at 4 kHz, and recorded on an analog tape recorder. At testing time, the playback from the tape recorder was again low-pass filtered at 4 kHz to reduce tape noise and played to listeners over TDH-49P headphones in an Industrial Acoustics 1202 test chamber.

There were 25 different tones: five different F_0s combined with five different frequencies of the formant peak. Table 1 shows that both the frequencies of the fundamentals and those of the spectral peaks were spaced by ratios of approximately 1.21. Each tone consisted of harmonics 1 to N. Since we used only frequencies up to about 3600 Hz, the tones with higher fundamentals had fewer harmonics (but never less than 13). The digital synthesis that generated the spectrum of these tones first passed a flat spectrum (all components in sine phase) through a formant filter (second-order digital resonator) and then through a first-order difference filter to give a spectrum with a similar roll-off both above and below the tuned frequency. The bandwidth of the formant filter was set at 0.1 times the formant peak frequency. Four examples of the resulting spectra are shown in Figure 1. Notice that there is not always a spectral component at the nominal formant peak. The tones were gated on and off in 10 ms by a quarter sine wave function. Tone 1 (of Table 1) was presented to the subjects binaurally at 79 dBA as measured from the headphones with a flat-plate coupler. The other tones were matched to it in loudness according to the subjective judgement of one of the experimenters (RL).

Subjects: There were 16 subjects, 15 university students between the ages of 21 and 27, paid for participating, and one 46-year-old male (ASB). They had about 15 minutes of introduction and pretraining on the task.

TABLE 2
Experiment 1: Mean Clarity Scores for the F_0 Tests

Interfering Factor (formant peak)	Tested Factor (Fundamental) 128	155	188	228	277	*M*
1000	2.7	4.1	7.4	8.9	8.9	6.4
1212	2.4	4.1	7.8	8.4	8.8	6.3
1470	2.7	3.9	5.3	7.9	8.6	5.7
1782	2.9	3.6	5.2	6.6	8.3	5.3
2161	2.9	3.2	4.9	6.6	7.3	5.0
M	2.7	3.8	6.1	7.7	8.4	

Note. Range of scores is 2 to 10.

Results

For each combination of F_0 and formant frequency, the subjects gave four scores running from 1 to 5. Table 2 shows the sum of the two scores for the same F_0 standards. The scores run from 2 to 10 with a higher score indicating that there was more segregation by F_0.

Recall that one of the tones in the comparison loop always had an F_0 of 128 and a formant frequency of 1000. The cell entries tell how much segregation there was by F_0 when the other tones had the F_0 and formant frequency shown in the margins. We observe that the F_0-based segregation becomes stronger (scores become larger) as F_0 separation increases (running across the table from left to right). There is a monotonic increase in all rows. Comparing each score to the one just to its left, the number of increases is significant at $p < .002$ by the sign test. There is a small tendency for a decrease in segregation as the formant frequency separation (the interfering factor here) became larger (running down the table). It is significant at $p < .01$ by the sign test. These trends are displayed most clearly by the row and column means. The cells that violate the top-to-bottom weakening trend are always in Row 1 or Column 1, which contain the conditions in which only one factor was varied. We also note that the *decrease* in segregation as the interfering factor increased in strength was not as large as the *increase* when the tested factor increased in strength by the same amount.

The leftmost column in Table 2 shows what happens when all four tones had the same F_0. Segregation was very low and was not made any worse by larger separations on the interfering factor (formant frequency separation). On the other hand, the top row shows what happens when all tones had the same formant frequency. We see a strong increase in segregation as F_0 separation increased.

Table 3 shows the results for the tests of segregation by formant (peak) frequency. High numbers here indicate that the frequency of the formant peak was effective in controlling segregation. Therefore, the strong observed increase in scores from left to right indicates an increase in segregation as the separation on the tested factor of formant peak increased. This increase is significant at $p < .002$ by the sign test. In the results shown in this table, F_0 was the interfering factor. Therefore we should expect to see a decrease in segregation scores as F_0 increased (running down the table). The trend taken over the whole table, if it exists at all, is very weak ($p > .22$ by the sign test). However, we can see an interesting pattern if we ignore Row 1

TABLE 3
Experiment 1: Mean Clarity Scores for the Formant Peak Frequency Tests

Interfering factor (Fundamental)	Tested Factor (Formant peak position)					
	1000	1212	1470	1782	2161	*M*
128	2.7	3.1	8.1	8.8	9.3	6.4
155	2.9	5.8	8.7	9.3	9.6	7.3
188	2.8	4.8	8.3	9.1	9.3	6.9
228	3.6	3.5	6.4	8.2	9.1	6.2
277	2.9	3.1	4.6	7.1	8.4	5.2
M	3.0	4.1	7.2	8.5	9.1	

Note. Range of scores is 2 to 10.

TABLE 4
Experiment 1: Relative Strength of F_0 and Peak

N Steps of Difference on Interfering Factor	*N* Steps of Difference on Tested Factor					
	0	1	2	3	4	*M*
0	0	10	−7	10	−4	0
1	5	17	9	−9	8	−10
2	−1	−9	−30	−12	7	−12
3	7	1	−12	−16	−5	−8
4	0	1	3	−8	−11	−3
M	−3	−3	−11	−9	−7	−6

Note. Entries are Table 2 entry minus Table 3 entry times 10. Positive numbers indicate that F_0 is a stronger influence.

and Column 1, which contain the degenerate tests in which only F_0 or formant frequency was varied. In each of the four remaining columns, where the two factors competed, the score declined monotonically as the F_0 separation increased over four steps. The conditions in which only one factor was varied seem always to give lower scores than would be predicted by extrapolating the trends from the other 16 cells. Perhaps this is due to some difference in strategy when the listener hears variation on only one factor.

Overall we see the same pattern in both Tables 2 and 3. Increases in separation on the tested factor strongly strengthened segregation, and increases in separation on the interfering factor produced an interfering effect that was substantially weaker.

A visual comparison of the relative strengths of segregation by F_0 and formant frequency can be obtained by subtracting each cell in Table 3 from the corresponding cell in Table 2. Table 4 shows the results of this subtraction. In order to eliminate decimal points and make the table more readable, each result has been multiplied by 10. The first thing to notice is that the mean for the whole table is −6 and that only 5 out of the 25 subtractions give positive numbers. This means that, on the whole, the formant frequency separations were stronger than the F_0 separations in controlling segregation when the two types of separations were equated in terms of the ratios of the frequencies involved.

The mean of zero taken across the first row reflects the fact that when the interfering factor was inoperative (zero steps of difference) the two factors were, on average, equally strong with some variation from condition to condition. These strong effects are the ones that appear in Row 1 of Tables 1 and 2. However, it appears that the formant frequency tended to dominate when the two factors were acting at cross purposes. This interpretation is consistent with the observation that the strongest superiority of formant separation (largest negative numbers in Table 4) appears on the diagonal that runs from top left to bottom right. On this diagonal, the number of steps of separation on the two factors was the same, and the superiority of formant separation was due to some sort of competitive advantage. Another apparent pattern is that cells toward the middle of the table have higher values. This suggests that the greatest domination by the stronger factor occurs when both factors have a medium strength.

EXPERIMENT 2

Experiment 1 showed that separations in peak frequencies affected the grouping of tones. Experiment 2 looked at the role played by the sharpness of the peaks in affecting this grouping. Since we already knew that both F_0 and peak affected segregation, we decided not to explore a range of variation on both factors again but simply to set up a conflict between the two and look at how the sharpness of the peak enhanced the ability of the frequency separation of the peaks to compete effectively with an F_0 separation in controlling the grouping.

Method

The listener's task and the timing of the tones were the same as in Experiment 1, but the tones themselves were different. Because the formant peaks in Experiment 1 had been created by filtering, not all the peaks were exactly the same shape, and there was no guarantee that there was a harmonic at the nominal spectral peak defined by the centre frequency of the filter. These factors might have accounted for the irregularities in the superiority of peak over F_0 (Table 4). In order to control more exactly the shape of the spectrum in Experiment 2, the spectra of the stimuli were created by adding pure tones together rather than by filtering a complex tone.

The use of additive synthesis in this experiment also allowed us to rule out a factor that might have influenced the segregation in Experiment 1. The stimuli of Experiment 1 were created by filtering harmonics 1 to N. Since the first harmonic was always present, as the fundamental increased so did the lower edge of the spectral band occupied by the signal. Therefore, F_0 was not completely independent of spectral range.

All the tones in Experiment 2, no matter what their pitch or formant frequency, spanned exactly the same spectral range, sharing the same lowest and highest harmonics (500 Hz & 4000 Hz). This ruled out spectral edges or spectral balance as bases for grouping. This design was made possible because the fundamental was omitted in all cases, the lowest harmonic always being the one at 500 Hz. Thus the tones having an F_0 of 125 Hz had 29 harmonics, those with an F_0 of 166.7 Hz had 22, and those with an F_0 of 250 had 15. These should have been sufficient to produce highly saturated pitches. Furthermore, there was always a harmonic exactly at the nominal spectral peak. All harmonics started in sine phase. They were synthesized at 19,200 samples/s, low-pass filtered at 4.5 KHz, and recorded on an analog tape recorder.

Variation of F_0 and Peak: In contrast to Experiment 1, there were only three F_0s (125, 167, & 250 Hz) and three peak frequencies (1000, 1500, & 2000 Hz). Because of constraints imposed by the other design goals, they could not be spaced by equal ratios. Furthermore, not all combinations of these values were used in synthesizing the tones. The combinations that were

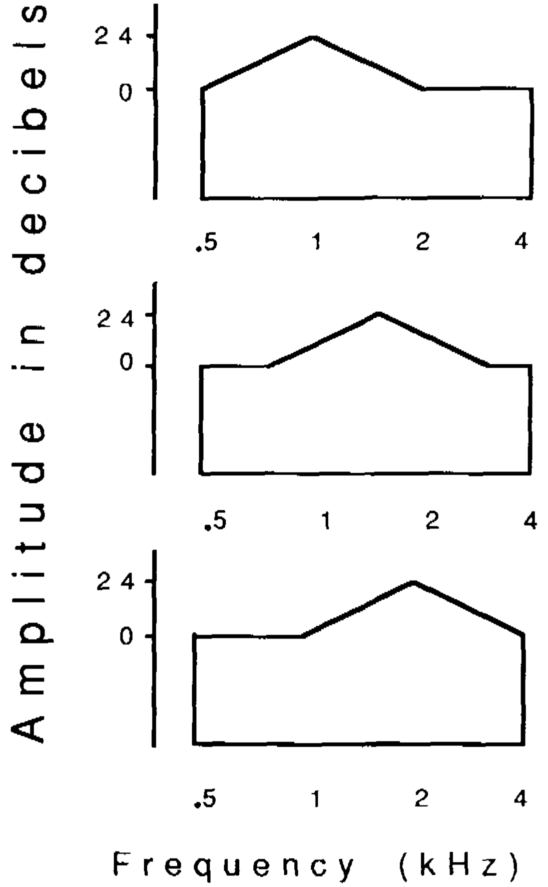

Figure 2. Experiment 2: Outlines of the spectra (frequency and amplitude boundaries) for the tones having the three different spectral peak positions. The frequency scale is logarithmic.

used allowed us to create sequences of tones bearing either a 2:1 or 3:2 relationship to one another in terms of both their F_0s and their formant peak positions. In every condition, the separation on F_0 was the same as the separation of peaks when both were measured as ratios. Thus for the 2:1 ratio, the four tones in the cycle, expressed in the form, [F_0, peak], were [250, 2000], [250, 1000], [125, 2000], and [125, 1000]. For the 3:2 ratio, the values were [250, 2000], [250, 1500], [167, 2000], and [167, 1500]. In other words, the factors always competed on an equal footing (defined in terms of frequency ratio) with one qualification: the sharpness of the spectral peak was sometimes too small to allow separation of the peak frequencies to be a meaningful perceptual difference. In these cases, we observed pure effects of F_0.

Peak Sharpness: Each combination of F_0 and formant frequency was used to synthesize five different tones, each with a different value of formant peak sharpness. The spectral envelope of each tone can be plotted on log-frequency-by-decibel co-ordinates as a triangle standing on a pedestal. The sharpness was defined as the height of the triangle, the triangle always having the same width. Outlines of the spectra (frequency and amplitude boundaries) for tones with the three different peak positions are shown in Figure 2. Note that the dB scale in these diagrams is a purely relative one, expressed in terms of the amplitude of the components in the pedestal; the actual playback intensities were adjusted to achieve equal subjective loudness

TABLE 5
Experiment 2: Mean Clarity Scores

Rate of F_0s and Peaks	Peak Sharpness (in dB)					
	0	6	12	18	24	*M*
			F_0 Tests			
3:2	6.7	7.1	7.9	8.1	8.0	7.6
2:1	8.1	8.4	7.8	8.0	7.6	8.0
M	7.4	7.8	7.8	8.0	7.8	—
			Peak Tests			
3:2	2.8	2.9	3.3	3.7	4.9	3.5
2:1	2.7	2.9	3.5	5.5	7.3	4.4
M	2.8	2.9	3.4	4.6	6.1	—

Note. Range of scores is 2 to 10.

of all tones. The different peak sharpnesses were obtained by setting the amplitude of the harmonic at the formant peak to be 0, 6, 12, 18, or 24 dB above the amplitude of the equal intensity harmonics of the pedestal. The stimuli having 0 dB peaks had, of course, flat spectra. For all the other stimuli, the sides of the triangle dropped to the level of the pedestal in one octave in both the high-frequency and low-frequency directions.

The 125, 1000 tone was presented binaurally at 65 dBA and the other tones were matched to this one for subjective loudness by the experimenter (CL). The experiment was presented as counterbalanced halves where only one ratio of F_0 and formant frequency (2:1 or 3:2) appeared in each half. The other variables were randomized within halves. The experiment lasted for about 1 hour for each subject, including instruction and practice. The subjects were 20 young adults, ranging from 20 to 28 years of age.

Results

The results for the F_0 tests are shown in the upper half of Table 5 on the same 2–10 scale as the results of Experiment 1. High numbers represent a better ability to hear the grouping when F_0 was the tested factor. The trends that will be mentioned here are all supported by an ANOVA described later. Generally speaking, the scores were always very high, indicating that the subjects always thought that the F_0-based grouping was strong despite the fact that it was being opposed, in all cases, by a formant peak separation of equal size. Increasing the sharpness of the peaks did not seem to lead, in any systematic way, to greater interference (lower scores).

The results for the formant frequency tests are given in the lower half of Table 5. Here the scores were generally low, indicating that either the force exercised by peak separation on grouping was weak at most of the peak sharpnesses that were employed or else that it was overpowered by a disruptive effect of the interfering factor (F_0). We also see that the effects of the frequency separation of the peaks became greater as the peaks themselves became sharper, this trend being stronger for the 2:1 separation. Remember that the 2:1 separation applied all the F_0s and formant peak separations in a cycle. Therefore, sharpening the peaks helped a formant separation of 2:1 to overcome more strongly the disruptive effects of an F_0 separation of 2:1. Overall, the scores were higher for the 2:1 ratio than for the 3:2 ratio.

The results given in Table 5 were subjected to an overall ANOVA. It confirmed the finding that the F_0 tests were very much easier than the formant frequency tests, $F(1, 19) = 351.8, p < .0001$. There was a significant overall effect of peak sharpness, $F(4, 76) = 52.8, p < .0001$, and a significant sharpness by ratio interaction,

TABLE 6
Experiment 2: Relative Strength of F_0 and Peak Frequency

Rate of F_0s and Peaks	Peak Sharpness (in dB)					
	0	6	12	18	24	*M*
3:2	39	42	46	44	31	40
2:1	54	55	43	25	3	36
M	47	49	45	35	17	38

Note. Entries are Table 5 F_0 test values minus peak test values times 10. Positive numbers indicate that F_0 was a stronger influence.

$F(4, 76) - 4.24$, $p < .01$. There was also a significant sharpness by type of test interaction, $F(4, 76) = 28.4$, $p < .0001$, indicating that peak sharpness had a stronger effect, understandably, when the tests were focussed on the effects of peak separation. Finally, there was a highly significant three-way interaction between the ratio, the test, and the peak sharpness, $F(4, 76) = 29.5$, $p < .0001$. This occurred because the sharpness had the strongest effect in the peak test at the 2:1 ratio.

The competition of the two factors is best displayed by the entries in Table 6. These entries are analogous to some of the entries in Table 4, namely, the ones in which the F_0 and the formant frequency separations are the same (along the major diagonal) and where these separations are similar to the 2:1 and 3:2 ratios used in the present experiment. Specifically, the four-tone cycle represented by the third row, third column (value of -30) in Table 4 embodies an approximately 3:2 ratio and the one for the fifth row, fifth column (value of 11) embodies an approximately 2:1 ratio. It seems that the relative dominance of F_0 separation and formant peak separation were reversed in the two experiments, with F_0 dominating in the second one. However, looking back at the lower half of Table 5, we see that the peak separation began to resist the disruptive effects of the F_0 separation at the higher peak sharpnesses. We must recall that the peaks were created in different ways in the two experiments.

GENERAL DISCUSSION

The experiments confirm that F_0 alone can be used to segregate sequences of sounds, even when the upper and lower limits of the frequency bands are identical. Therefore, they confirm the results of Singh (1987) showing a tradeoff between pitch (our F_0) and timbre (analogous to our spectral peak) in controlling the formation of streams. They also show that F_0 itself, and not some secondary spectral effect, can control segregation. In addition, they show that the formant peak frequency controls grouping more strongly when the peak is made to stand out more from the spectrum. This effect continues up to a sharpness of 24 dB, the highest sharpness studied. Therefore, while formant sharpness may not play a role in phoneme identification when a voice is presented in isolation, it may play a role in segregating voices from interfering sounds.

The differences in the relative strength of the effects of F_0 and peak in the two experiments should not be treated as having any importance, since the formant peaks were created in a different way in each experiment. In Experiment 2, the maximum peak sharpness (height) was 24 dB. Figure 1 shows that in Experiment 1, because of the first-order difference filter, the lowest harmonics were attenuated by much more than 24 dB relative to the most intense harmonic (up to 40 dB in some cases). As we know from Experiment 2, this would have strengthened the effects of peak separation.

It is possible that the effect that has been described as due to the separation of spectral peaks was not due to that at all, but to a difference in the relative intensity of higher versus lower harmonics, a property that gives rise to the perception of brightness (Wessel, 1979). This would have been affected automaticaly whenever the frequency of the spectral peak was changed. In Experiment 2, the effect of peak sharpness might have been due to the fact that the effects of the peak on the balance of high versus low harmonics was greater as the peaks became more intense relative to the pedestal. When the spectrum has only a single spectral peak, the brightness is automatically correlated with the position of this peak. One way to decide whether spectral peaks themselves controlled grouping, independent of brightness, would be to use tones containing at least two spectral peaks. Then the frequencies of the two peaks could be varied in opposite directions so as to keep brightness constant. A second way to vary the separation of peaks, in a single-formant spectrum, without affecting brightness would be to change the shape of the peak as it was moved upward, compensating for a higher peak position by a sharper roll-off of the harmonics above the peak.

Although it appears that both F_0 and formant frequency (or perhaps brightness) were used by listeners to achieve segregation, we must ask what kind of segregation it was. We observed, particularly in Experiment 1, that each factor was relatively weak in interfering with the segregation based on the other. Even in Experiment 2, where the effects of peak separation were not as strong, when the peaks were 24 dB high the listeners were able to group the tones by their peak separation and to resist the disruption by an equally large separation in F_0 (giving a value of 7.3 out of 10 in the lower half of Table 5).

What does this resistance to interference mean? Bregman (1990) has proposed that there are two kinds of perceptual processes that lead to perceptual grouping. They are illustrated by the results of van Noorden (1977) who found two boundaries relating stream segregation to the frequency separation of the tones at different tone rates. The *temporal coherence boundary* (TCB) was found when the listener was asked to hold all the tones together as one stream. It seems to represent a limit on integration. This integration was found to get harder as the tones were separated further in frequency or were more rapidly alternated. The *fission boundary* (FB), on the other hand, was found when listeners were asked to focus on one of the streams that remained in a restricted frequency range. Success in this task required only a small frequency separation (1–3 semitones) and was unaffected by the tone rate.

Bregman (1990) sees these two forms of segregation as resulting from two kinds of perceptual process. The TCB derives from *primitive segregation*. One of the marks of primitive segregation is that it makes it hard for the listener to perceive relationships between sounds that have been placed into different streams by this process. Primitive segregation is viewed as a preattentive grouping process. The second kind of process, called *schema-based segregation*, is responsible for the FB. It occurs when the listener uses a schema or criterion to extract the parts of a signal that fit it.

In the present experiment, the listeners profited by differences in between the two target tones and their neighbours to extract them from the sequence. However, as long as the feature that they were using was defined clearly enough (e.g., when the peaks were sharp enough in Exp. 2), they were not prevented from perceiving the pair *as a pair* by the fact that they were different from each other on the interfering quality. These two facts argue that the task was strongly affected by a schema-based process of segregation, capable of grouping the pair by their shared property, and that the primitive segregation was weak. A strong primitive segregation should have disrupted the integration of the two target tones when they were different on the interfering property even when they were grouped by the tested property.

The importance of the schema-based process in this task may have resulted from the 20 repetitions of the two-tone pattern in the standard cycle (AB—AB—...). There was plenty of time for the listener to build up a mental description (schema) of the target pair. Indeed, it seems that any time a listener tries to hear a pattern inside another one, he or she requires a mental description of the pattern being sought. Any such task, therefore, must have a heavy loading on the schema-based segregative process.

Our results can be compared to those reported by van Noorden (1975). Just as we did in Experiment 2, he varied the fundamental frequencies in a rapid sequence of tones while keeping the frequency components of all the tones in a fixed band. Contrary to what he had expected, when he acted as his own subject he was able to hear perceptual coherence up to a separation of an octave. This result can be compared to the results of Experiment 1 for the peak test in which there was no peak frequency difference but only an F_0 difference (Table 3, Column 1). We observed little interference from F_0 even when differences on this factor exceeded an octave (lower left cell value of 2.9). Van Noorden's results are therefore consistent with our own and with the hypothesis that schema-based integration was heavily involved in his task. As a highly trained subject he would have developed strong schemas concerning the sequence of tones and would therefore have been able to overcome the disruptive effects of a relatively weak primitive segregation based on F_0.

References

Anstis, S., & Saida, S. (1985). Adaptation to auditory streaming of frequency-modulated tones. *Perception and Psychophysics, 11*, 257 271.

Axelrod, S., & Guzy, L.T. (1968). Underestimation of dichotic click rates: Results using methods of absolute estimation and constant stimuli. *Psychonomic Science, 12*, 133–134.

Bozzi, P., & Vicario, G. (1960). Due fattori di unificazione fra note musicali: La vicinanza temporale e la vicinanza tonale. *Rivista de Psycologia, 54*, 235–258.

Bregman, A.S. (1978a). Auditory streaming is cumulative. *Journal of Experimental Psychology: Human Perception and Performance, 4*, 380–387.

Bregman, A.S. (1978b) Auditory streaming: Competition among alternative organizations. *Perception and Psychophysics, 23*, 391–398.

Bregman, A.S. (1990). *Auditory scene analysis: The perceptual organization of sound.* Cambridge, Massachusetts: The MIT Press.

Bregman, A.S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology, 89*, 244–249.

Dannenbring, G.L., & Bregman, A.S. (1976). Stream segregation and the illusion of overlap. *Journal of Experimental Psychology: Human Perception and Performance*, 2, 544–555.

Darwin, C.J., & Bethell-Fox, C.E. (1977). Pitch continuity and speech source attribution. *Journal of Experimental Psychology: Human Perception and Performance, 3*, 665–672.

Divenyi, P.L. (1971). The rhythmic perception of micromelodies: Detectability by human observers of a time increment between sinusoidal pulses of two different, successive frequencies. In E. Gordon (Ed.), *University of Iowa studies in the psychology of music* (Vol. 7). Iowa City: University of Iowa Press.

Dowling, W.J. (1973). Rhythmic groups and subjective chunks in memory for melodies. *Perception and Psychophysics, 14*, 37–40.

Erickson, R. (1975). *Sound structure in music*. Berkely, CA: University of California Press.

Fitzgibbons, P.J., Pollatsek, A., & Thomas, I.B. (1974). Detection of temporal gaps within and between perceptual tonal groups. *Perception & Psychophysics, 16*, 522–528.

Green, K.P., Stevens, E.B., & Kuhl, P.K. (1989). The use of rate information during phonetic perception depends on pitch continuity. *Journal of the Acoustical Society of America, 85*, (Suppl. 1), S52.

Handel, S. (1984). Using polyrhythms to study rhythm. *Music Perception, 1*, 465–484.

Henke, W.L. (1980). *MITSYN: A coherent family of command-level utilities for time signal processing*. Cambridge, MA: Copyright by W.L. Henke. (Revised and extended by D. Salomon, McGill University)

Judd, T. (1979). Comments on Deutch's musical scale illusion. *Perception and Psychophysics, 26*, 85–92.

Kinney, J.A.S. (1961). Discrimination in auditory and visual patterns. *American Journal of Psychology, 74*, 529–541.

Klatt, D.H. (1982). Prediction of perceived phonetic distance from critical-band spectra. *Proceedings of the IEEE International Conference on Speech, Acoustics, and Signal Processing*, 1278–1281.

Massaro, D.W. (1976). Perceiving and counting sounds. *Journal of Experimental Psychology: Human Perception and Performance, 2*, 337–346.

McAdams, S., & Bregman, A.S. (1979). Hearing musical streams. *Computer Music Journal, 3*, 26–43.

McNally, K.A., & Handel, S. (1977). Effect of element composition on streaming and the ordering of repeating sequences. *Journal of Experimental Psychology: Human Perception and Performance, 3*, 451–460.

Miller, G.A., & Heise, G.A. (1950). The trill threshold. *Journal of the Acoustical Society of America, 22*, 637–638.

Neff, D.L., Jesteadt, W., & Brown, E.L. (1982). The relation between gap discrimination and auditory stream segregation. *Perception and Psychophysics, 31*, 493–501.

Noteboom, S.J., Brokx, J.P.L., & De Rooij, J.J. (1978). Contributions of prosody to speech perception. In W.J.M. Levelt & G.B. Flores d'Arcais (Eds.), *Studies in the perception of language*. New York: Wiley.

Roberts, B. (1988). *Vowel identification in the presence of extraneous sounds*. Unpublished Ph.D. dissertation, Cambridge University, United Kingdom.

Singh, P. (1987). Perceptual organization of complex-tone sequences: A tradeoff between pitch and timbre? *Journal of the Acoustical Society of America, 82*, 886–899.

Smith, J., Hausfeld, S., Power, R.P., & Gorta, A. (1982). Ambiguous musical figures and auditory streaming. *Perception and Psychophysics, 32*, 454–464.

Steiger, H., & Bregman, A.S. (1981). Capturing frequency components of glided tones: Frequency separation, orientation and alignment. *Perception and Psychophysics, 30*, 425–435.

van Noorden, L.P.A.S. (1975). *Temporal coherence in the perception of tone sequences*. Unpublished doctoral dissertation, Technische Hogeschool Eindhoven, Eindhoven, The Netherlands.

van Noorden, L.P.A.S. (1977). Minimum differences of level and frequency for perceptual fission of tone sequences ABAB. *Journal of the Acoustical Society of America, 61*, 1041–1045.

van Noorden, L.P.A.S. (1982). Two channel pitch perception. In M. Clynes (Ed.), *Music, mind and brain*. New York: Plenum Press.

Wessel, D.L. (1979). Timbre space as a musical control structure. *Computer Music Journal, 3*(2), 45–52.

Received 11 July 1989
Revision received 26 February 1990
Accepted 26 February 1990