Auditory Streaming and the Building of Timbre*

ALBERT S. BREGMAN AND
STEVEN PINKER
McGill University

ABSTRACT

In a natural environment, the auditory system must analyse an incoming wave train to determine two things: (a) which series of frequency components arose over time from the same source and should be integrated into a sequential stream, and (b) which set of simultaneous components arose from one source and should be fused into a timbre structure. A set of experiments was performed in which subject judged the stream organization and the timbre of a repeating cycle formed by a pair of more or less synchronous pure tones, B and C, and a preceding pure tone, A, whose frequency was varied in its proximity to that of the upper tone of the BC pair. These experiments demonstrated that fusion and sequential organization of streams are carried out using two sorts of information which compete to determine the best perceptual description of the input. Proximal frequencies between sequential components promotes a sequential organization and the simultaneity of onset of frequency components promotes perceptual fusion.

It is useful to create a distinction between two concepts in audition: 'source' and 'stream.' A source is a physical event which gives rise to a train of acoustic pressure waves; an example is the playing of a violin. A stream, on the other hand, is a psychological organization whose function it is to mentally represent the acoustic activity of a single source over time; in our example it is the experience of a violin playing. The transformation of sources into streams would be a simple matter if each source gave off a pure tone whose frequency were distinct from that of other sources. A bank of filters could then distinguish one source from another. However, there are two ways in which natural listening environments differ from this ideal. First of all, each source gives off a waveform which consists of a set of many individual frequencies that change from moment to moment. Secondly, many sources are active at once, and the pressure waves arising from them are mixed at the ear of the listener. At a particular moment, this complex waveform contains frequencies arising from all the sources active at that moment. A filter bank could not determine which frequencies came from which sources. The auditory system must use temporally extended information in this wave train to recover separate descriptions of the several sources which gave rise to it. These separate descriptions are the 'streams' referred to earlier. We see this process of parsing the acoustic information to form coherent streams as analogous to the process of parsing the retinal information to form 'objects' in vision, a process which is currently being studied in artificial intelligence research under the label of 'scene analysis' (e.g., Winston, 1975). This acoustic factoring process has two aspects: the grouping together of all the simultaneous frequency components that emanate from a single source at a given moment, and the connecting over time of the changing frequencies that a single source produces from one moment to the next.

If the perceptual division into streams

*We are grateful to Alexander Rudnicky who did much of the pilot research, to Stephen McAdams who helped us in data collection and analysis, and to the computer-based laboratory at the Psychology Department, McGill University, supported by the National Research Council of Canada and the Quebec Ministry of Education. Reprints can be obtained from A.S. Bregman, Psychology Department, 1205 McGregor Ave., Montreal, Quebec H3A 1B1. Steven Pinker is now at the Department of Psychology and Social Relations, Harvard University.

corresponds exactly to the set of individual acoustic sources, we hear veridically. If not, we experience auditory illusions. By studying some of these illusions, it has been possible to determine a number of the rules by which parts of the incoming pressure wave are assigned to streams. One such illusion, which we can call the illusion of multiple streams, occurs when a sequence of tones, of alternating high and low pitches, is played very rapidly. At slow speeds, only one stream, containing all the tones, is heard. At a high speed, however, the high and low tones are heard as forming separate streams. It has been found that a pure tone tends to be included in a particular stream when its frequency is close to those already in the stream and when there is little temporal lag between the new element and the last similar element (Bregman & Campbell, 1971; Dowling, 1973; Miller & Heise, 1950; Noorden, 1975). Tonal elements tend to segregate from noise-like elements (Dannenbring & Bregman, 1976). Smooth transitions from the previous parts of the stream to a current element favor its inclusion in the stream (Bregman & Dannenbring, 1973; Heise & Miller, 1951). Tones localized in one spatial position tend to form a single stream (Judd, 1977). Furthermore, the illusion of multiple streams builds up cumulatively as the tones alternate (Bregman, 1976). The consequences of stream segregation, aside from the subjective experience of two streams of sound as opposed to one, include the inability to perceive the temporal relationships among elements in different streams or to perceive patterns that cut across streams (Bregman & Campbell, 1971; Dannenbring & Bregman, 1976). It has also been shown that multiple streams are formed in parallel and compete for new incoming elements (Bregman & Rudnicky, 1975). Many of these findings have been reviewed by Bregman (1978) and by Bregman and Dannenbring (in press).

These experiments have been concerned only with how successive portions of an acoustic input are connected up with earlier ones. The experiments to be reported here are concerned also with the other aspect of stream formation. How are simultaneously present frequency components sorted into appropriate groups that each represent a single source?

At this point we must introduce the notions of perceptual fusion and timbre. Any moment of sound is a complex structure which may have arisen from one acoustic source or may be a mixture of sounds from two or more sources. If two sets of acoustic features (say sets of pure tone components) have arisen from a single source they should be fused in perception (i.e., experienced as features of a single stream of sound); if they arise from two separate sources, they should participate in two separate fusions. A subset of the features assigned to each stream will determine the acoustic quality of that stream – its timbre.

Let us imagine a simple example. Suppose two frequencies are played at the same time. If the two are perceived as separate streams, we will hear two pure tones. If they are grouped as one stream, we will hear a single rich tone. What determines whether or not they will be grouped? Helmholtz (1878) suggested one answer.

> ... when one musical tone is heard for some time before being joined by the second, and then the second continues after the first has ceased, the separation in sound is facilitated by the succession of time. We have already heard the first musical tone by itself, and hence know immediately that we have to deduct from the compound effect for the effect of this first tone ... When a compound tone commences to sound, all its partial tones commence with the same comparative strength; when it swells, all of them generally swell uniformly; when it ceases, all cease simultaneously. Hence no opportunity is generally given for hearing them separately and independently (pp. 59–60).

There are two principles embedded in Helmholtz's speculation; they could be referred to as 'continuation' and 'common fate.' Both of these are principles that have been studied by the Gestalt psychologists (e.g., Koffka, 1935). The first principle states that, if an element of a perceptual array forms a simple continuation of another series, it should be treated as part of that series. In audition, if a set of frequency components, not yet assigned to a stream A, are a good continuation of components of a stream, they should be assigned to that stream.

The second principle is that of 'common fate': sets of sensory elements that change in parallel ways will be perceived as whole and distinct entities. An example of this principle in vision is Johansson's demonstration (1964) that in an array of randomly moving dots, a correlation in the trajectories of any two dots will cause them to be perceived as parts of a single rigid object. In audition, the principle might apply to the movement of pure tone components over time. If two of them change in amplitude together (e.g., come on at the same time) or are frequency-modulated together, they should fuse into a single stream with a timbre determined by the set of fused components.

Both principles really represent valid generalizations about the world. Generally speaking, parts of the same event resemble one another in some way. The principle of good continuation relates to resemblances in the relations between elements in a series, and the principle of common fate relates to resemblances in the behaviour of simultaneous elements.

There are, however, kinds of resemblance other than simultaneous changes that pure tone components could have and that might influence their grouping. One is the harmonic relationship among them. When a mass has a dominant frequency of vibration, harmonics (integral multiples) of this frequency tend also to be present. Hence, it might be useful to perceptually group all harmonically related frequencies into one stream.

Each principle of grouping has some general validity in exploiting a regularity in the world of acoustic sources. However, none of them is completely reliable. We may expect, therefore, that in the human auditory system these principles would operate in parallel, competing with or reinforcing one another. The experiments presented below try to demonstrate the activities of principles of stream formation by putting them into competition with one another.

If a pure tone of frequency f alternates rapidly with a complex tone containing frequency f, we might expect that this component of the complex tone would be stripped out and placed in a stream with the pure tone. Noorden (1975), for example, has shown that, in a rapidly repeating cycle, a preceding pure tone will make audible an otherwise inaudible component of a complex sound. We might also expect that the complex sound which had lost one of its components would have a purer timbre. This would indicate that a pure tone component could not play two roles at the same time; it could not both be a continuation of a preceding stream and also contribute to the timbre of a complex tone. This latter aspect has not yet been established experimentally.

The four experiments reported below are all based on the same arrangement of stimuli (see Fig. 1). A pure tone, A, repeatedly alternates with a more or less simultaneous pair of pure tones B and C. B and C are always of equal duration, whether their onsets are synchronous or asynchronous. The frequency of tone A may be varied from just above to considerably higher than that of B. The frequency of C may also be varied. The question is always whether B joins into a sequential stream with A, or is fused with C so as to make a contribution to the perceived timbre of a compound tone (BC) formed out of B and C.

In Experiments I and II, two factors were varied: the frequency difference between A and B, and the relative synchrony of onset and offset of B and C. Earlier work has shown that pure tones that are close to-

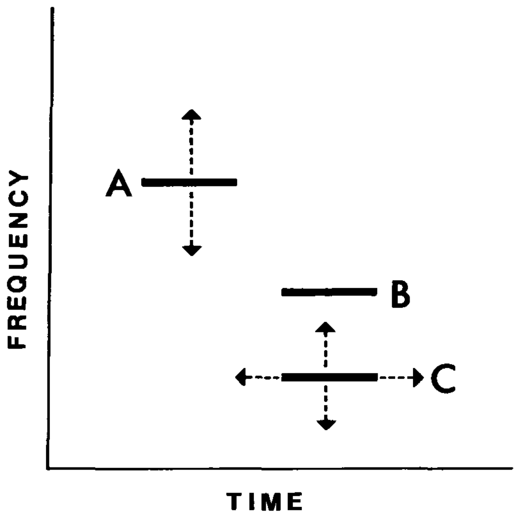

FIGURE 1 Schematic diagram of the test stimuli.

gether in frequency should become part of the same stream. Hence as A approaches B in frequency, it should capture B into a sequential stream, isolating C which would be heard as a pure tone. Conversely as C's onsets and offsets are made more synchronous with B's, the tendency of C to capture B into a unified timbre structure should increase. This structure should have the lower tone C as its predominant pitch but should possess a complex timbre because of the presence of B. In this case, A should be heard as isolated. In Experiment I we approached the question by asking listeners to directly judge the richness of the tone having C's pitch; in Experiment II we asked them to judge whether or not A and B were in the same stream.

EXPERIMENT I

METHOD

Procedure

Subjects, tested individually, were informed that the experiment was to investigate the factors that contribute to the 'perceptual richness' or 'timbre' of a sound, defined as 'the quality that distinguishes, for example, the sounds of different musical instruments producing notes of the same pitch.' They were told that they would be required, on each trial, to judge the richness of the note with the lowest tone, relative to two standard tones 'of the same pitch' which would be presented separately. They were also told that the to-be-judged tone was the second member of each rhythmically defined pair of sounds. They were to judge the test tones on a 7-point rating scale, with the two endpoints, 1 and 7, anchored to the 'pure' and 'rich' standard tones respectively. They were asked to use the full range of the rating scales in making their judgments.

Trials

There were 45 trials, comprising three presentations of each of the 15 tonal patterns, all in a random order. Each trial began with a warning buzz (a 100 Hz square wave tone lasting 68 msec), followed one second later by the 'pure' standard tone, one second of silence, the 'rich' standard tone, and two seconds of silence. Then came the pattern of test stimuli for that trial, repeated 40 times with no delay between repetitions, and then ten more seconds of silence during which the subject was expected to indicate his judgment on a sheet of paper. A computer automatically sequenced the 45 trials of the experimental session, as well as the warning buzz, standard tones, test stimuli, and silences within each trial.

Stimuli

In each experimental condition a different pattern of test stimuli (schematized in Fig. 1) was presented. The pattern always consisted of repetitions of the following cycle: 117 msec of silence, a single pure tone (A) also lasting for 117 msec, 47 msec of silence, and a compound tone consisting of 527 hz (B) and 300 hz (C) components of equal loudness and of equal duration (147 msec). The BC compound is somewhat dissonant and is describable as 'rough' or 'complex.'

The test stimuli for the different conditions were created by orthogonally varying the frequency of the captor tone A (559, 978, or 1713 Hz), i.e., from a semitone to just about 1.7 octaves above B, and the asynchrony of the onsets (and hence the offsets) of the two equal-length components, B and C, of the compound tone. The lower frequency component C could either begin (and hence, end) in synchrony with B, or else lead or lag it by 29 or 58 msec (25 or 50 per cent of the length of one tone). B always began 47 msec after the offset of A. Thus there were three 'captor' frequencies and five 'synchrony' conditions to yield 15 stimulus patterns. Each condition occurred three times for each listener.

In order to eliminate the clicks produced by instantaneous tone onsets and offsets, each 117 msec tone was actually composed of a 12 msec linear rise in amplitude to full value, a 93 msec steady state, and a 12 msec linear decay in amplitude to zero. The amplitudes of the captor tone A and of each of the two components, B and C, of the compound tone were adjusted before synthesis so that the three would sound equally loud when played back through the headphones.

The tone to be used as a 'pure' standard was of the same frequency as the to-be-judged note in the test pattern (300 Hz) but had twice the amplitude, had smoother onset and offset glides (25 instead of 12 msec), and lasted about three times as long (350 msec). The tone to be used as a 'rich' standard had the same temporal properties as the 'pure' standard, but possessed both of the frequency components (300 and 527 Hz) present in the compound tone. The two components were of equal amplitude, and each was one-half the amplitude of the pure standard.

Apparatus

All stimuli were generated by a PDP-11 digital computer, which numerically synthesized each experimental stimulus and stored it as a sequence of numbers on a disk file. At the time of tape recording, these numbers were converted to voltages by a digital-to-analog converter and output in real time at 31,000 samples per second. This signal was low-pass filtered at 9000 Hz by a Rockland Model 852 Dual Hi/Lo filter to remove the digitization noise.

The filtered signal was amplified by a Sony TA-1055 integrated amplifier, and the entire experimental session was recorded at 7½ inches per second (19 cm/sec) by an Akai GX-400D SS tape recorder. To remove hiss, the sound was filtered during playback either by the low-pass filter set at 12,000 Hz, or by the 'high filter' feature of the amplifier which attenuated the upper portion of the sound spectrum by 6 db per octave above 5000 Hz. Subjects listened to the sound binaurally through Koss Pro/4AA stereo headphones while seated in an Industrial Acoustics Company 1202 audiometric testing room. The loudness of the test stimuli, as measured in the testing room by a sound level meter equipped with a headphone coupler, was approximately 92 db SPL for each speaker of the headphones.

Subjects

Twenty-five university students volunteered to participate in the experiment. Two of the subjects accidentally failed to respond in one trial, and one misunderstood the task instructions, so data were analysed for only 22 of the subjects.

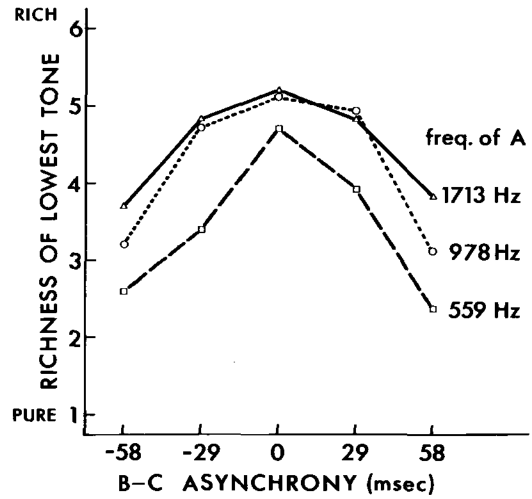

FIGURE 2 Mean richness ratings for "the lowest tone" in Experiment 1, as a function of BC asynchrony. The parameter is the frequency of A. Negative asynchronies indicate that the onset of C preceded that of B.

RESULTS

Mean richness ratings of 'the lowest tone' are shown in Figure 2 as a function of BC asynchrony. The timbre of C was judged as richest when C was exactly synchronous with B, the richness dropping off monotonically with increasing asynchrony. This effect was statistically significant, $F(4, 84) = 21.14, p < .001$. Whether B led or trailed C made no obvious difference.

The effect of A's frequency upon the richness judgments was also as predicted. As A's frequency became higher (i.e., as its frequency moved away from that of B), the 'lowest tone' was judged as richer, indicating a greater contribution of B to the BC compound tone. This effect of the frequency of A was statistically significant, $F(2, 42) = 10.79, p < .001$. The interaction of the two experimental variables was not statistically significant.

DISCUSSION

We take the results as showing that the perceived complexity of a moment of sound is not purely a consequence of its frequency composition. Instead, sequential and simultaneous processes of organization act to assign to the frequency components of that moment a role in one or more concurrent streams. This organization determines what timbres are heard. Timbre seems to be a perceptual description of a stream, not of an acoustic waveform.

We have interpreted the results as due to stream organization. There is one possible criticism that has been suggested to us by L.P.A.S. van Noorden. Perhaps the effects of the asynchrony of B and C on the 'richness' judgments came purely as a result of the shorter period of time that B and C overlapped. Shorter periods of time simply produce less of the tonal interaction phenomena which the subject was calling richness. While this criticism can be entertained for the effects of asynchrony, it cannot be responsible for the effects of the frequency difference between A and B. Here the judged richness of the BC compound tone was affected by the frequency of a tone preceding the compound tone. Judgments of richness are clearly not the best way to measure segregation phenomena. However, we did not study timbre judgments as a means but as an end. It is striking to find that timbre, an apparently basic quality of sound, is directly affected by stream organization. In Experiment II, we abandon richness judgments and employ direct judgments of stream segregation to ascertain whether the synchrony and frequency relations manipulated in Experiment I were actually affecting stream segregation.

EXPERIMENT II

We suggested earlier that, under certain circumstances, the alternation of A with the BC compound would yield a stream consisting of A and B together, and cause a rejection of C into a separate stream. Experiment I did not look at this directly, but inferred such an organization from its effects on the perception of the BC compound tone. Experiment II looked more directly at the relation between A and B. We asked listeners, this time, to directly judge whether the A and B tones were heard in the same stream or in separate streams. We varied the same two variables, the frequency of A and the BC asynchrony.

We expected a direct effect of the frequency separation of A and B; other research on stream segregation indicates that the closer the frequencies of two successive tones are, the more likely they are to form a unified stream. In addition, we expected an indirect effect on the streaming of A and B to arise from the competing effects of C. When C is synchronous with B, it should act to pull B into a timbre structure with itself and thereby to make B unavailable as a pure tone to group with A in an AB stream. In this case, A ought to be heard as isolated. The strongest effects of degree of BC synchronization were expected at intermediate frequency differences between A and B. At the highest AB frequency difference (just about 1.7), A and B should rarely stream together, and at the smallest AB frequency difference (one semitone) they should always stream together. At the intermediate frequency difference (10.7 semitones) the AB streaming should be unstable and very much affected by the BC asynchrony.

METHOD

Stimuli

The test stimuli that the subjects were to judge were identical with those of Experiment I. However, instead of being provided with pure and rich comparison tones before each trial, the subjects were presented before the experiment with a single sequence of tones designed to demonstrate the difference between one and two auditory streams. The taped sequence, which has been used in this laboratory prior to a number of experiments that require subjects to discriminate one from two streams, begins with a repeating cycle of four tones of nearby frequencies. As the tape progresses, the pitches of the second and fourth tones decrease until two distinct streams of sound, a repeating pair of high tones, and a repeating pair of low tones, are heard. Then the pitches of the lower pair increase until the two streams unite again into a single stream; the pitches of the two pairs diverge a second time and then converge a second time. The entire tape lasts 50 seconds.

Trials

The order of trials and distribution of stimulus patterns among trials were identical with those of Experiment 1. However, the internal structure of each trial was altered. Instead of a warning buzz, a pure tone, and a rich tone, each trial began with two orienting tones, corresponding to the frequencies of the two successive tones (A and B) that might or might not be heard as being in the same stream in the ensuing trial. The first tone A (either 1713, 978, or 559 Hz, appropriate to the particular trial), lasted .5 sec, followed by a .5 sec silence, and then the second tone B also lasting .5 sec. Five seconds of silence (enough to preclude any stream integration) intervened between the orienting tones and the test stimuli. Next there followed forty cycles of the test stimulus pattern, and then a ten second silence during which the subject was to indicate his judgment.

Apparatus

The apparatus was identical with that of Experiment 1, with the following exceptions. The testing was done with a Revox 77A tape recorder in a quiet (although not sound-treated) room. The sound was not filtered during playback, and the sound level was 87 db for each headphone speaker.

Procedure

Subjects were tested individually. After the subject had read diagrammed instructions explaining what is meant by auditory streams, he listened to the demonstration tape. The experimenter asked the subject whether he understood the distinction between one and two streams and whether he felt that he could discriminate between the two percepts. If the subject indicated that he could not, the tape was played a second time. No subject failed to understand after the second playing. The subject then read a set of instructions concerning the experiment itself. The instructions described each trial as beginning with two isolated tones, followed a few seconds later by the same two tones repeating in quick succession. The subjects were instructed to listen to those tones and decide whether they heard them in the same stream or in two different streams; they were told to ignore any tones other than the pair that they heard at the trial onset. They were then to mark their judgment on a seven-point scale, one end of which corresponded to hearing the tones in one stream, the other to hearing the tones in two different streams; the intermediate positions corresponded to the relative *ease* and *frequency* of hearing one pattern or the other when the subject's perception fluctuated. A similar technique was used successfully by Bregman and Dannanbring (1973). The subjects were also asked to wait until the silence that followed the stimuli before marking their judgments.

Subjects

Nineteen graduate and undergraduate students of McGill University volunteered to participate in the experiment. Two of these accidentally omitted a trial, and one confessed to misunderstanding the task instructions after he had completed the experiment. Of the remaining sixteen usable subjects, eight had participated in Experiment 1 four to six weeks earlier.

RESULTS

The mean stream segregation ratings (averaged across subjects and replications) are plotted in Figure 3. Low values indicate that A and B were perceived in the same stream. Hence high values indicate effective competition from C. As expected, A and B were less likely to be heard in the same stream the further apart their frequencies (i.e., the higher A's frequency was). This effect of frequency separation was statistically significant ($F(2, 30) = 71.27, p < .001$).

Asynchrony of B and C had mixed effects, depending on the AB frequency separation. The effect of degree of BC asynchrony was statistically significant taken over all frequency separations ($F(4, 60) = 7.26$, $p < .001$); however, there was a significant interaction between frequency separation and degree of asynchrony ($F(8, 120) = 3.76$, $p < .001$). For the smallest frequency separation, asynchrony between the overlapping tones produced no noticeable effect on the ratings, whose average was 1.73 (a rating of 4 indicates an equal likelihood of splitting and not splitting). For the largest frequency separation, the two conditions in which the lower tone lagged behind its upper companion (+25% and +50%) asynchrony produced a smaller

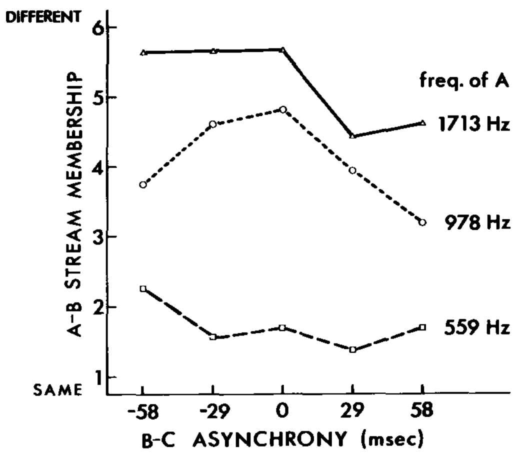

FIGURE 3 Mean judgments of whether A and B were perceived in the same stream in Experiment 2. Low values indicate that they were.

tendency toward segregation than the other three synchrony levels, although the combined mean (5.24) and all the individual means were on the 'two streams' portion of the 7-point scale. For the intermediate frequency separation, however, the greater the asynchrony in either direction, the less the stream segregation. Stimuli with synchronous tone pairs were judged to be more likely to split (mean rating of 4.88), whereas stimuli with very asynchronous tone pairs were judged to be more likely *not* to split (3.23 and 3.81 mean ratings for the −50% and +50% conditions, respectively).

For all frequency separations, there was an asymmetry in the ratings, such that when the lower tone C preceded the upper one B, greater segregation between A and B was obtained than when B followed C. This difference was tested statistically in an analysis of variance by collapsing the asynchrony dimension to two values, negative and positive (ignoring the synchronous conditions). This has the effect of contrasting the six means representing asynchrony in one direction with the six representing asynchrony in the other direction; the difference was statistically significant, $F(1, 15) = 11.57, p < .005$.

DISCUSSION

The results of the experiment support the view that sequential factors of stream organization compete with other processes which act to fuse co-occurring frequencies into complex sounds. While Experiment I looked at the effects of the competition upon the fusion process, Experiment II looked at the effects upon sequential streaming. The effects were expected to be strictly compensatory. When fusion was strong, sequential integration was expected to be weak, and vice-versa. The results were not so cleancut. In particular, we saw no evidence for an asymmetrical effect of the direction of BC asynchrony on richness judgments, but there was such an effect on stream judgments. The latter effect is not hard to explain. When B led C, the frequency movement between the onsets of A, B, and C was unidirectional (showed 'good continuation'). This situation has been shown by other studies to favor sequential integration (Divenyi & Hirsh, 1974; Heise & Miller, 1951; Noorden, 1975). Our results agreed with this principle. However, no such asymmetrical effects were found in Experiment I. We can only speculate that somehow the subjects were listening differently under the different instructions used in the two experiments and that the stream organization was not identical in the two cases. However, we did demonstrate, in Experiment II, a clear overall effect of BC asynchrony upon the stream segregation judgments, with the clearest, most symmetrical effect seen when A and B were at an intermediate frequency separation. It was expected that the strongest effects of synchrony would be seen at this separation since the intermediate separation tends to produce unstable sequential organizations at the speed we used.

EXPERIMENT III

We suggested earlier that frequencies should tend to fuse when there is some evidence that they have arisen from the same source, and we hypothesized that simple harmonic relationships (e.g., ratios of 2:1 or 3:1 or 3:2) among the frequencies might serve as evidence for a common source. We therefore manipulated the simplicity of the frequency relations between B and C in Experiment III.

Quite apart from considerations of how the auditory system succeeds in sorting frequencies correctly into streams, there is evidence that frequency relations influence the perception of tones close together in time. In both the phenomenon of stream segregation and that of masking, closer frequencies tend to interact more. We therefore varied the gross separation of the frequencies of B and C by a factor of two.

The frequency separation of two simultaneous tones is known to affect the perceived quality of the compound – some combinations are more dissonant than others. Therefore we could not use richness judgments to study the effects of the BC frequency relation. Instead we again used direct judgments of stream organization as in Experiment II; i.e., we did not actually study the fusion of B with C but, instead, the freedom of B to group sequentially with A.

METHOD

Procedure

The same instruction of subjects and the same procedure were used as in Experiment II.

Stimuli

An ABC pattern was employed similar to those in the earlier experiments. A, B, and C were sinusoids with 11.2 msec rise/fall times. The rises and falls in amplitude were S-shaped to avoid clicks. The amplitudes of the tones were adjusted so that all tones would sound equally loud through the headphones.

B was always 600 Hz. One factor that was varied was the frequency separations between A and B. A was either two semitones above B (672 Hz) or 10.7 semitones above B (1113 Hz). The synchrony of B and C was also varied; C was either synchronous with B or preceded B by 33.9 msec (30% of the length of the tones). In addition, there were two variations of the frequency relationship between B and C: the gross interval size and the harmonicity (relative consonance or dissonance) of the exact interval used in that range. In the consonant condition three intervals were used. If the frequency of C is designated f, then B was $1.5f$, $2f$, or $3f$. Note that B was held constant and C was varied; i.e., C was 200, 300, or 400 Hz. In the dissonant condition all of the consonant intervals were diminished by 1 semitone yielding $1.414f$, $1.888f$, and $2.828f$ relationships; i.e., C was 212, 318, or 424 Hz. The waveforms of the 'mistuned' compounds, when viewed on an oscilloscope, were observed to have a periodic change in the shape of the amplitude envelope, or 'second-order beats.'

Trials

The structure of a trial was identical with those in Experiment II, except that the recycling pattern of A followed by BC had slightly different time parameters: 112.8 msec of silence, 112.8 msec of tone A, 45.2 msec of silence (in the BC synchronous condition), and then 112.8 msec of the BC complex (in the synchronous condition). In the asynchronous conditions, C always led B; C started 11.3 msec after the termination of A, and B began 33.9 msec after the onset of C. (The A-B interval was always 45.2 msec.)

There were six blocks of 16 trials each. For all trials in one block, C had the same frequency, the six possible frequencies of C varying across blocks. Within each block, four different patterns (two frequency separations between A and B combined with two synchrony conditions) were presented four times each. The trials were distributed within the block so that each of these patterns appeared once in each group of four successive trials. Otherwise the order of trials was random. The first four trials in each block, unknown to the subject, were treated as practice trials and were excluded from the analysis. The order of presentation of the blocks was counterbalanced over subjects using a Latin square design, so that there were six orders of the six blocks.

Each trial began with one to five short 100 Hz square-wave tones which corresponded to an equal number of slashes printed next to each trial on the score sheet. This was to remind the subject which trial was being presented. Two seconds later the standards were presented. This was followed by 5 sec of silence to rule out any temporal continuity between presentation of the standards and the pattern itself. Forty cycles of the stimulus pattern were then presented with no gaps between cycles, and this was followed by 10 sec of silence during which time the listener was to mark the judgment on the rating scale.

Apparatus

The stimuli were prepared as in Experiment I and played to subjects in an Industrial Acoustics Company 1212 audiometric testing room

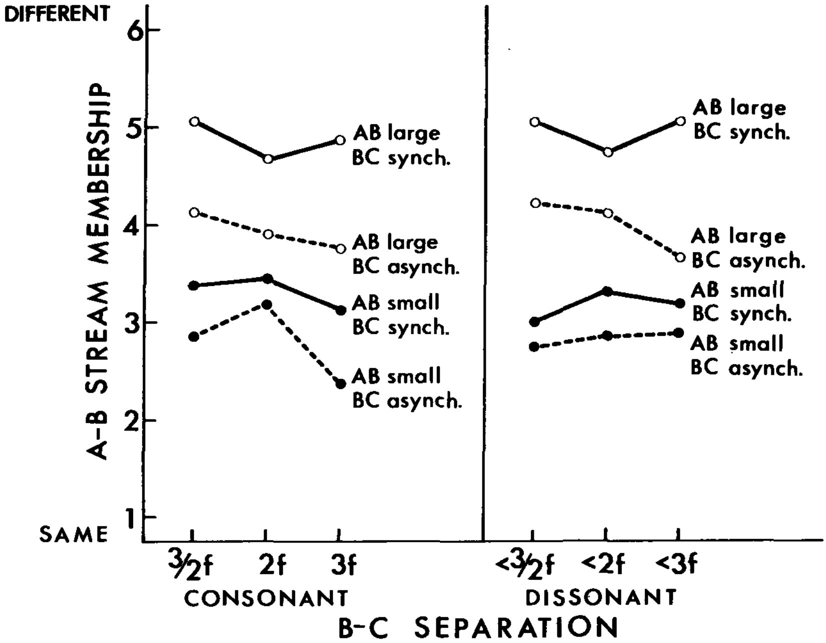

FIGURE 4 Mean judgments of whether A and B were perceived in the same stream in Experiment 3. Low values indicate that they were.

through Micromonitor MX-1 electrostatic headphones. The output level for each pure tone component was approximately 80 db SPL.

Subjects

Twenty-four students and researchers at McGill University volunteered to participate in the experiment without pay. Twelve of these volunteers participated in order to fulfil a requirement for a psychology course.

RESULTS

The mean segregation ratings are shown in Figure 4. High values indicate that A and B were judged as being in different streams, hence that C offered effective competition. The scores in the figure can therefore be viewed as 'effective competition (by C) scores.' These judgments are plotted as a function of frequency separation between B and C. Consonant separations (simple frequency ratios) are shown on the left and dissonant separations are shown on the right. The parameters are the size of the frequency drop between A and B (large or small) and the degree of synchrony between B and C (synchronous or asynchronous).

For all conditions when B and C were synchronous, A and B were less likely to be judged as being in the same stream, $F(1, 18) = 9.01, p < .01$. This confirms the results of Experiment II. The size of frequency drop between A and B also shows a consistent effect. With larger drops, A and B are less likely to be in the same stream, $F(1, 18) = 19.37, p < .001$. There were no other statistically significant effects. The consonance of B and C made no difference at all (the mean for consonant intervals was 3.73 and for dissonant ones was 3.74). The gross magnitude of the frequency difference between B and C also had no significant effect; the values of AB streaming for the three gross ranges used were 3.81, 3.78, and 3.62 (for approximately 1.5*f*, 2*f*, and 3*f* respectively). While being statistically nonsignificant, these latter results go in an expected direction; when B is further from C it is more likely to go with A (C offers less 'effective competition'). However, this result was not consistent across the variations in other conditions; if the effect is real, it is relatively weak.

It is particularly surprising that the octave relation between B and C did not keep B away from A better ('$2f$ consonant' in Figure 4). The mean value for octaves was 3.81 and for all other conditions combined was 3.72. While this difference is in the right direction, the 'effective competition' value for the octave was not the highest; the value for $1.5f$ was higher (3.86).

We conclude that there were only two clear effects: BC asynchrony and AB frequency separation. Both of these were found in both Experiments I and II.

EXPERIMENT IV

In Experiment IV we simply took all the conditions of Experiment III and asked subjects to make richness judgments on 'the tone with the lowest pitch.' Since different BC intervals will have different amounts of richness for reasons other than the degree of BC integration, the effect of BC frequency separation is meaningless; however, the effects of AB frequency separation and BC asynchrony are meaningful and would constitute a replication of previous findings.

METHOD

Stimuli

The portion of the stimuli involving the pattern of tones A, B, and C were those used in Experiment III. However, instead of being preceded by the A and B target tones as was the case for stream judgments, they were preceded by rich and pure standard tones as in Experiment I.

Procedure

The instruction of subjects and the procedure were carried out as in Experiment I. However the trials were blocked, with a particular value of BC frequency separation (i.e., frequency of C) occurring only in one block. In that block, the 'pure' and 'rich' standards contained the appropriate frequency of C. Since the listeners were told to judge the stimuli only in relation to the standards used in that block, it was hoped that intrinsic richness differences across blocks would not swamp all other effects. Within each block, BC asynchrony and AB frequency separation were varied in a randomized order, and the order of blocks was counterbalanced across subjects. The first four trials in each block, unknown to the subjects, were treated as practice trials and excluded from the analysis of results.

Subjects

Twenty-eight young adults, students and assistants in the psychology department, served originally as subjects. Two of these were discarded for not conforming with instructions and new subjects replaced them in the twenty-six subject design.

RESULTS

The mean richness scores are plotted in Figure 5. Although the BC gross frequency separation and degree of consonance both showed statistically significant effects, we will not comment on them since we cannot tell whether the BC intervals that were used varied intrinsically in perceived purity, or whether the variations in purity came from variations in the streaming of B with A.[1]

However, there were, as in Experiment I, clearcut effects of BC asynchrony, $F(1, 18) = 23.1, P < .001$, and of the frequency drop between A and B, $F(1, 18) = 27.1, p < .001$. In addition there was a significant interaction between these two factors, $F(1, 18) = 6.93$, $p < .05$. The greatest effects of BC asynchrony were found when A and B were closer in frequency. This may have arisen because the sequential (AB) and the simultaneous (BC) grouping tendencies were more balanced at the smaller AB frequency separation; this is suggested by the fact that the richness values for the smaller AB separation fall nearer to the middle of the richness scale. Hence variations in BC asynchrony could show their strongest effects.

GENERAL DISCUSSION

We can summarize the most important effects from the four experiments as follows. (1) The frequency separation of A and B

[1]The following F-ratios are included only for the sake of completeness. For gross frequency separation, $F(2, 36) = 13.5\ p < .01$. For degree of consonance, $F(1, 18) = 12.8, p < .01$.

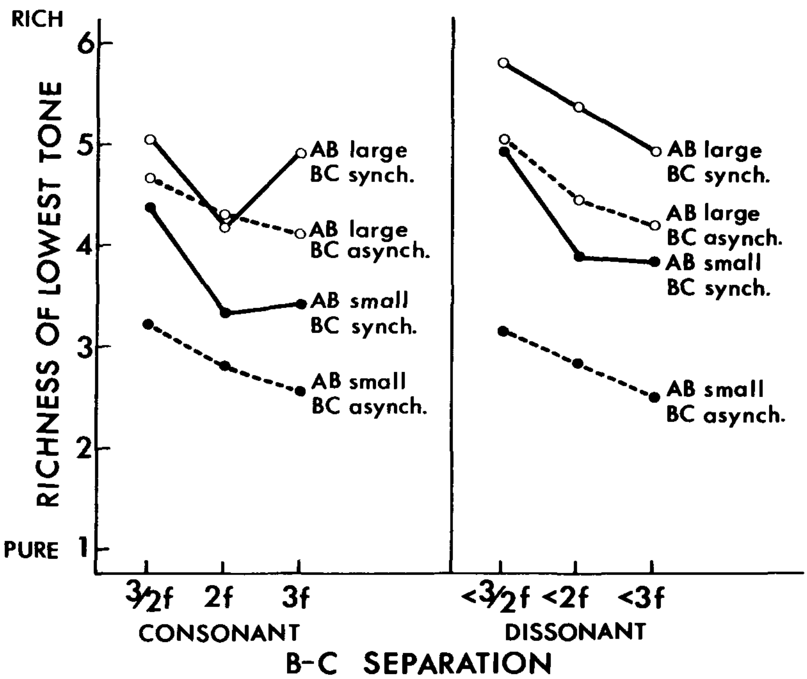

FIGURE 5 Mean richness ratings for "the lowest tone" in Experiment 4, for consonant BC intervals (left-hand graph) and dissonant BC intervals (right-hand graph).

influenced the degree to which the B and C frequencies were heard as a unified compound. (2) The degree of synchrony of C with B influenced whether B was heard as being in the same stream as A. A finding of lesser significance was the fact that when A was closer to B in frequency, A and B were more likely to be heard in the same stream. This result was highly predictable in the light of prior research on stream segregation (e.g., Noorden, 1975). There was another relevant finding which, although consistent, is subject to the alternative interpretation mentioned in the Discussion of Experiment 1: the judged richness of the BC compound was greatest when B and C were synchronous in onset and offset.

The first two effects mentioned above clearly show that sequential and simultaneous grouping effects can compete, that frequency separation is a strong determinant of the sequential effect and that onset/offset synchrony is an important determinant of the simultaneous effect (perceptual fusion). The distinction between the sequential and the simultaneous applies itself naturally to music where it expresses itself as melody and timbre respectively. There is much that psychologists must do to clarify how these two interact perceptually.[2] However, the sequential and simultaneous organizing processes also play their role in ordinary listening situations, the two together acting to parse the input waveform so as to find the meaningful sources. The sequential effect hooks frequency components into streams. The simultaneous effect, by selecting which harmonics to fuse, determines the acoustic quality (timbre) of each moment of each stream.

[2] It has been pointed out by a reviewer that a technique in the performance of music, for emphasizing a particular note or voice, is to desynchronize it slightly from its expected time. A medial voice may be brought out of a harmonic structure by advancing or delaying its onsets. It should not be surprising that musicians long ago discovered these effects empirically.

RÉSUMÉ

Dans le cours normal des choses, le système auditif doit faire l'analyse du flot des ondes afférentes pour reconnaître deux réalités: (a) la suite de fréquences venant successivement d'une même source, qui doivent conséquemment être rattachées à un flot séquentiel et (b) les composantes provenant simultanément d'une même source et dont la fusion doit alors aboutir à la structure de timbre. Dans une série d'expériences, les sujets ont été amenés à juger de l'organisation du flot des ondes et du timbre dans une stimulation cyclique formée d'une paire B et C de tons purs plus ou moins synchronisés et d'un troisième ton pur A, précédant les deux autres et dont la fréquence se rapprochait plus ou moins, selon les cas, de celle du ton le plus élevé de la paire BC. Ces expériences démontrent que la fusion et l'organisation séquentielle des flots d'ondes se font à partir de deux types d'informations qui se font concurrence dans l'obtention de la meilleure description perceptive de la stimulation. Quand les composantes sont successives, la proximité des fréquences favorise une acquisition séquentielle et quand les débuts des composantes sont simultanées, le fusion perceptuelle se trouve favorisée.

REFERENCES

BREGMAN, A.S. Stream segregation: an evidence processing system. Paper presented at the meeting of the XXIst International Congress of Psychology, Paris, France, July 1976

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

BREGMAN, A.S., & CAMPBELL, J. Primary auditory stream segregation and the perception of order in rapid sequences of tones. *J. exp. Psychol.*, 1971, **89**, 244–249

BREGMAN, A.S., & DANNENBRING, G.L. The effect of continuity on auditory stream segregation. *Percept. Psychophys.*, 1973, **13**, 308–312

BREGMAN, A.S., & DANNENBRING, G.L. Auditory Streaming. *Scientific American*. In press

BREGMAN, A.S., & RUDNICKY, A.I. Auditory segregation: Stream or streams? *J. exp. Psychol.: Human Percept. Perform.*, 1975, **1**, 263–267

DANNENBRING, G.L., & BREGMAN, A.S. Stream segregation and the illusion of overlap. *J. exp. Psychol.: Human Percept. Perform.*, 1976, **2**, 544–555

DIVENYI, P.L., & HIRSH, I.J. Identification of temporal order in three-tone sequences. *J. Acoust. Soc. Am.*, 1974, **56**, 144–151

DOWLING, W.J. The perception of interleaved melodies. *Cognit. Psychol.*, 1973, **5**, 322–337

HEISE, G.A., & MILLER, G.A. An experimental study of auditory patterns. *Am. J. Psychol.*, 1951, **64**, 68–77

HELMHOLTZ, H.L. *On the sensations of tone as a physiological basis for the theory of music* (1878). Second English edition, New York: Dover, 1954

JOHANSSON, G. Perception of motion and changing form. *Scand. J. Psychol.*, 1964, **5**, 181–208

JUDD, T. *An explanation of Deutch's scale illusion*. Manuscript submitted for publication, 1977

KOFFKA, K. *Principles of Gestalt psychology*. New York: Harcourt and Brace, 1935

MILLER, G.A., & HEISE, G.A. The trill threshold. *J. Acoust. Soc. Am.*, 1950, **22**, 637–638

NOORDEN, L.P.A.S. VAN. Temporal coherence in the perception of tone sequences. Unpublished doctoral dissertation, Technische Hogeschool Eindhoven, Eindhoven, the Netherlands, 1975

WINSTON, P.H. (Ed.) *The psychology of computer vision*. New York: McGraw-Hill Computer Science Series, 1975

(*First received 18 October 1977*)
(*Date accepted 30 January 1978*)