

An experimental evaluation of three theories of auditory stream segregation

WENDY L. ROGERS and ALBERT S. BREGMAN
McGill University, Montreal, Quebec, Canada

Three theories of auditory stream segregation were evaluated. In two-part trials, subjects heard an induction sequence, whose effects upon an immediately subsequent test sequence were measured. The rhythm and total duration of Induction Sequence tones were varied in two experiments. The similarity between induction and test sequences aided segregation, but rhythmic predictability and longer tone durations did not. Frequency alternation during the induction sequence was not necessary to induce segregation in the test sequence. Furthermore, peripheral processes inadequately account for the segregation effects found. The data suggest that, once a distinct percept emerges from an auditory scene, properties derived from the percept (particularly changes) are fed back to control the ongoing analysis of that auditory scene. A neural adaptation to stimuli with constant properties may form part of this analysis.

When two tones of different frequencies are alternated continuously, the listener's perception changes over time. At first, an up-and-down pattern (high-low-high-low) is heard. If the sequence is presented long enough and rapidly enough, and if the tones are far enough apart in frequency, the listener hears the high notes and the low notes segregate from one another and form two apparently independent sequences of tones. This splitting phenomenon has been called *stream segregation* by Bregman and Campbell (1971), and it has been studied extensively (Bregman, 1990, chap. 2). The readiness of an alternating sequence to form two streams increases with the rate of tone presentation and the frequency separation between the high and low tones, and these factors can be traded off one against the other.

There are three theories about stream segregation. Bregman (1978a, 1990) sees streaming of alternating sequences as an illustration of the partitioning, by the auditory system, of incoming sounds into groups (called streams) likely to belong to the same sound source. He hypothesizes two main classes of grouping processes: simultaneous (combining concurrent sounds, as in hearing several co-occurring pure tones as one complex sound) and sequential (grouping sounds over time, as in linking of tones to form a melody). The streaming of alternating-frequency sequences is described, by Bregman's theory, as sequential grouping of tones according to the Gestalt

principle of proximity, with frequency proximity tending to override temporal proximity. The streaming, or segregation tendency, builds up over a period of several seconds, and the tendency to favor a two-region percept is lost slowly after the tones are stopped. These time lags are thought to be necessary (Bregman, 1978b) for the stability in the streaming process to be maintained.

Data from one experiment (Bregman, 1978b) show that the segregation of a repeating sequence becomes stronger as the number of repetitions increases. Bregman proposed that this could be a simple cumulative effect of tone duration in certain frequency regions, implying that more stimulation in a frequency region would always increase the segregation.

Van Noorden (1975) proposed another theory: that alternating-tone sequences activate pitch-motion (frequency-jump) detectors, which then become adapted and unable to follow the up-and-down pattern. Anstis and Saida (1985), who proposed a theory similar to Van Noorden's, found evidence consistent with adaptation of auditory frequency-modulation detectors: the observed adaptation is specific to the frequency region of the alternating tones, increases over time, and tends not to reverse; that is, when a stream of sounds splits, it tends not to return to the form of a coherent stream.

Anstis and Saida (1985) reviewed the evidence for the existence of neural frequency-modulation detectors, and they compared the results of their auditory experiments to those from a similar study (Anstis, Giaschi and Cogan, 1985) pertaining to breakdown of apparent motion in the visual system, which is believed to be tracked by neural motion detectors. An analogy has often been noted (e.g., by Bregman & Achim, 1973) between streaming of alternating tones and a phenomenon in vision: when two lights turn on and off in alternation at certain rates, the apparent motion first seen between them degenerates into flicker.

These experiments were performed in partial fulfillment of the requirements of W.L.R.'s PhD degree at McGill University. We thank Pierre Ahad and Rhonda Amsel for assistance, and three anonymous reviewers for suggestions and comments. The research was supported by a grant to A.S.B. by the National Sciences and Engineering Research Council of Canada and by a McGill Major Fellowship to W.L.R. W. L. Rogers is now at the Douglas Hospital, 6875 LaSalle Boulevard, Verdun, Quebec, Canada H4H 1R3. Reprints are available from A. S. Bregman, Department of Psychology, McGill University, 1205 Dr. Penfield Avenue, Montreal, Quebec, Canada H3A 1B1.

Anstis and Saida (1985) proposed that similar processes are at work in both the visual and the auditory systems when the latter attempt to link up successive stimuli over time. With respect to alternating tones, the process that tries to link tones according to frequency proximity is in conflict with the process that tries to track the up-and-down pattern of tones according to temporal proximity. If the latter process shows adaptation, the former process will come to dominate.

Anstis and Saida (1985) found that "adaptation" does not cross the ears. If a monotic sequence is presented to one ear until it segregates, and if it is then suddenly switched to the other ear, the sequence is heard as integrated again. They concluded that the adaptation was likely to be a peripheral process, occurring before the point in the nervous system at which information from the two ears comes together. They also found that adaptation was specific to the particular range of frequencies at which it had occurred. Therefore they thought of it as a process that was *local* (in the frequency domain)—that is, as a process that works within small ranges of frequencies rather than one that could be affected by any of the frequencies present in the sound mixture.

However, the theory of peripheral and local pattern detectors is contrary to certain experimental evidence: the idea that the process is a *local* adaptation conflicts with evidence that it is the *relative* spacing, rather than the absolute spacing, of tones in frequency and time domains that determines streaming (Bregman, 1990; McNally & Handel, 1977). The whole auditory context appears to be used. Also, frequency-jump detectors are not necessary to explain some cases of segregation of tone sequences. In an experiment reported in 1975, Bregman and Rudnický caused a pair of lower tones (X..X) to segregate from a pair of higher ones (AB) in a four-tone sequence (XABX) by placing other low tones (flanking tones, C) both before and after the original sequence (CCCXABXCC). The flanking tones were similar in frequency to the first and last tones of the original sequence. Although no frequency jumps were added, the extra tones "captured" tones of the same frequency range out of the overall sequence.

Jones (1976) proposed the third theory of streaming: that of rule-based predictability. According to her theory, auditory stimuli are represented psychologically on dimensions (such as pitch, loudness, and time), and the auditory system is designed to detect certain types of regularities in a sequence of sounds. These regularities are then used, by an attentional process, as rules to predict subsequent incoming sounds. Jones also proposed that changes along the other dimensions were required to be proportional to changes on the time dimension; for example, a rapid sequence of alternating tones (tones with a small time gap) would only be followed when the frequency separation between sequential tones was small—that is, in proportion with the temporal separation. Jones's theory is thus consistent with the fact that the presentation rate

and the size of the frequency jump can be traded off against each other.

According to Jones's (1976) theory, the regular spacing of the flanking tones (CCC...CC) in Bregman and Rudnický's (1975) study was what allowed streaming to occur; the regular tones set up an expectancy, and the predictability of the sequence made it easier to hold together as a stream. Jones, Kidd, and Wetzel (1981), using similar stimuli, found that when the flanking tones had a different rhythm from the other tones, the flanking tones were less effective at capturing other tones into a stream. However, French-St. George and Bregman (1989), using sequences with random variations in frequency or timing or both, showed that neither predictability of frequency nor that of timing was necessary for streaming of tone sequences into two frequency ranges. Yet it was admitted by the authors that their subjects may not have been able to predict the more regular sequences well enough to provide an adequate contrast to the "unpredictable" conditions of the experiment.

Jones's (1976) theory does not explain why repeating cycles, of tones alternating between two frequencies (very predictable sequences), are heard as one stream for several seconds before they split. Such repetitive stimuli become more predictable over time and ought to be increasingly *resistant* to streaming, according to Jones's theory.

The evidence is conflicting, or not available, for many issues related to the theories about streaming. With respect to Bregman's (1990, 1978b) theory, is there a cumulative mechanism (activated by tones in a given frequency region) that becomes a stronger influence on grouping as the cumulative tone-duration increases? Regarding the frequency-jump detector theory, are frequency-jump detectors necessary for streaming? Are streaming processes peripheral, or do they require information from both ears? If frequency-jump detectors exist, do they cause streaming by becoming adapted? Finally, concerning Jones's theory, when all other properties of a sequence are the same, do a predictable and an unpredictable sequence have different influences on the cumulative effects of streaming?

GENERAL METHOD

The same basic method was used in three experiments, to test the theories. In each trial, an *induction* tone sequence, immediately followed by a *test* sequence, was presented. The boundary between the two sequences was not signaled to the subjects. The induction sequences were varied, and their effects on a constant test sequence were measured. The subject rated the degree of segregation of the test sequence at its end, and this was used as a measure of the effectiveness of the induction sequence at building up a segregative tendency. This paradigm provided a way to vary the properties of the tones preceding the test sequence in order to determine which properties aid or resist segregation.

A test sequence with well-known properties was chosen for use with all trials: the two-frequency gallop sequence developed by Van Noorden (1975), shown in Figure 1. This pattern of three equal-length notes and a silence has a rhythm that changes noticeably when

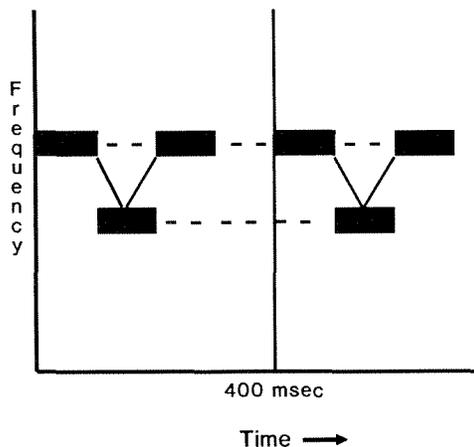


Figure 1. Two 400-msec cycles of Van Noorden's gallop sequence. Each rectangle represents one pure tone. The pattern comprises a high note, low note, high note, and silence. Connecting lines represent possible forms of perceptual grouping.

it segregates into two streams, thus aiding the subject in making a rating of segregation. When heard as a single stream, it is heard as an up-and-down pattern, as indicated by the solid lines, with a galloping rhythm. When the sequence splits into two streams, as shown by the dashed lines, the separate high- and low-pitched rhythms become isochronous.

The subjects were asked to listen for a gallop. Van Noorden (1975) found that attentional factors are important when measuring streaming effects: it is almost always easy to listen for and hear out the individual streams, but the ability to maintain perception of the up-and-down gallop pattern is highly sensitive to the tone repetition rate and the size of the frequency separation. Therefore, in order to study what makes a coherent sequence split and in order to obtain a range of ratings in these experiments, it was necessary that the subjects try to hear a gallop in the test sequence.

The experimental design was similar for all the experiments. In each, there were two experimental factors: (1) the size of the frequency separation between high and low tones of the test sequence, and (2) the type of induction sequence. The frequency separation in the test sequence was varied for two reasons. One was to check that subjects in the experiment replicated the well-known effect of greater streaming with larger frequency separations. According to Van Noorden's (1975) data, it should be easy for subjects to hear a gallop (in the control condition) when the test sequence tones are 5 semitones apart, and difficult or impossible when the gap is above 12 semitones (an octave). These two extreme values were chosen as levels of the frequency-separation factor. They served also as a reminder for the subject, during the experiment, about the extreme ends of the rating scale. Thus a statistically significant main effect was always expected for frequency separation, and this is alluded to only briefly in the results.

A second reason for varying the frequency separation of the test sequence tones was to deal with individual differences by providing at least one separation for each subject that was ambiguous enough to be affected by the induction sequence. The wide frequency separation necessary to guarantee streaming may seem surprising, given Van Noorden's (1975) data on the coherence boundary. However, his subjects were experienced at listening to alternating-frequency sequences. In our laboratory it has been found that naive listeners (the vast majority of our subjects) can hold together sequences of tones that have a large frequency gap; there appears to be a strong effect of experience on the segregation of such sequences.

The semitone (frequency ratio of 1.06:1) was used as the measure of frequency proximity. Semitones represent frequency proximity on a log frequency scale, and Miller and Heise (1950) demonstrated that proximity related to streaming is linearly related to frequency separation on a log frequency scale.

Soft white noise, of the same duration as the induction sequences, was chosen as a control for the latter because it has equal intensity over the audible frequency range, and because it has been shown to be equivalent to silence with respect to the cumulation of streaming effects (Bregman, 1978b). It is preferable to silence because it cues the subject that a trial is in progress. The purpose of the control was to provide a contrast with experimental induction sequences by inducing the minimum possible amount of segregation.

Planned comparisons were used to test the hypotheses of two experiments. For unexpected experimental results, post hoc tests were used in addition to planned comparisons. Repeated measures were used in all of the experiments. Despite the randomization of the order of presentation of stimuli, the assumptions of homogeneity of covariance were not met. Individual differences in response patterns, a common cause of nonhomogeneous covariance, have been found for many experiments in our laboratory. The Greenhouse-Geisser procedure (Greenhouse & Geisser, 1959), in which smaller degrees of freedom are used in the *F* test, is recommended when there is not homogeneity of covariance; it provides a very conservative *F* test (Winer, 1971). We used this approach for a priori hypotheses. For post hoc tests, Myers (1979) has cautioned against the use of the Tukey HSD test with repeated measures when assumptions of homogeneity of covariance are violated. Therefore, post hoc contrast effects were tested with Bonferroni-adjusted *t* tests. This technique was selected because it holds the experimentwise alpha level constant and is robust to the violation of sphericity assumptions (Maxwell, 1980).

OVERVIEW OF EXPERIMENTS

There were three goals in Experiment 1: (1) to test whether streaming can be induced by a tone sequence that has no frequency transitions, (2) to test whether predictability aids cumulative effects in streaming, and (3) to test whether the total tone duration is a factor in the cumulative effects in streaming. Thus, the experiment addressed a part of each of the major theories described. Experiment 2 was designed so that the relationship between cumulative tone duration and streaming could be examined more closely, because the first experiment did not give conclusive results. Experiment 3 was designed to test a part of the frequency-jump detector theory: whether or not peripheral information is sufficient to explain the induction of streaming.

EXPERIMENT 1

The purpose in the first experiment was (1) to test whether a coherent gallop sequence can be caused to segregate by preceding it with a sequence of tones of only a single frequency (showing that the fatiguing of jump detectors is not necessary to explain streaming); (2) to compare the tendencies of a predictable versus an unpredictable sequence, with otherwise similar properties, to induce segregation; and (3) to discover whether or not increasing the density of stimulation at one frequency results in more effective induction of streaming in test sequences containing tones of the same frequency.

Method

Subjects

The subjects were 21 adults who reported normal hearing, recruited from among students at McGill University. They were paid for participating. One subject's results were discarded because he disclosed a hearing loss only after participating in the experiment. No screening tests of hearing were conducted.

Stimuli

All stimuli were diotic. Each trial consisted of a 4.8-sec induction sequence, immediately followed by a 1.2-sec test sequence, with a subsequent 6.4-sec silence for making a response. All tones were pure tones, lasting 100 msec (with exceptions to be described below), whose stated durations included both a 10-msec quarter-sine-wave onset and decay. All induction-sequence tones, and the upper tones of the test sequences, were 524 Hz (one octave above middle C). The test sequences were three 400-msec cycles of two-frequency gallop patterns. There were four levels of frequency separation between the high and low tones of the test sequence: 5, 9, 11, and 18 semitones.

There were four induction sequence types, each containing twelve 400-msec cycles of sound. Part of a trial, showing the end of the induction sequence and the beginning of the test sequence for a sample of each of the four induction conditions, is shown in Figure 2. Condition 1 (the high-density condition) was chosen to test whether increasing the total duration of tones, and shortening the silences, in the upper frequency region would increase the induction of segregation. It contained four equal-length tones per 400-msec cycle. The density of high tones was thus twice that of the test sequence. It was expected, according to Bregman's hypothesis of cumulative frequency bias, that this induction sequence would be more effective than sequences with lower tone durations at inducing segregation of the test sequence.

Two other conditions were designed to compare the effectiveness of predictable and unpredictable sequences, with otherwise similar properties, at inducing streaming of the test sequence. Induction sequences for both these conditions contained two tones per 400-msec cycle. The tones of Induction Condition 2 (the *predictable* condition) matched the high tones of the test sequence in tempo

and length. The tones of Induction Condition 3 (the *unpredictable* condition) were unpredictable in tempo and length but had the same average tone density as did the *predictable* sequence. This is different from French-St. George and Bregman's (1989) experiment, in which the most regular stimulus was a repeating sequence of eight tones of different frequencies. The regular sequence in the current experiment was very simple (with only one frequency and an isochronous rhythm). Therefore, it was unlikely that a predictive mechanism would be unable to create the appropriate rules to describe it during the 4.8 sec of the induction sequence. Furthermore, this experiment was used to test the effect of an induction sequence upon a subsequent (different) sequence, whereas French-St. George and Bregman used the same sequences for both induction and testing.

The unpredictable sequence was constructed by combining two different types of randomization. Half of the 400-msec periods contained 100-msec tones that could appear at the start of any of the four quarter parts of a 400-msec interval (for each quarter part, a random decision was made with respect to whether a note would appear). The other half were constructed so that tones alternated with silences, but the lengths of tones and silences varied randomly from 60 to 140 msec. Cycles of the two types were alternated to make the stimulus. Overall, the sequence was designed so that the average length of a tone was 100 msec and the average number of tones per 400-msec cycle was two. There was one such sequence constructed and used for all subjects. The subjects reported that it sounded like Morse code. The stimulus was constructed with the use of the two different methods because it is not known whether the concept of irregular timing or irregular note length is more important to the concept of irregular rhythm. Therefore, both types of irregularity were included.

Condition 4, the control condition, consisted of white noise attenuated to 52 dB SPL. The level of tones was calibrated to measure 70 dB SPL (A weighting) in each ear. The frequency range of stimuli for this experiment (185–524 Hz) was chosen in a relatively flat part of the equal-loudness contours (Moore, 1982).

Procedure

Training. Each subject was first shown a diagram of the gallop pattern, and the experimenter tapped out the rhythm. An audio demonstration of the gallop pattern (three contiguous cycles of it), at five different starting frequencies and frequency separations between high and low tones, was given. The subject was then questioned to ensure his/her understanding of the gallop percept. After being shown how to make a response on the computer, having read the instruction sheet, and having been given a chance to ask further questions, the subject did five practice trials.

Testing. The subject was asked to listen to the tones of each trial and to rate whether or not a gallop was heard at the end of the trial. Rating was done on a keyboard in the intertrial silence. The rating scale went from 1 to 10, with a choice in the 1–5 range indicating a gallop percept and a choice in the 6–10 range indicating some percept other than a gallop. Choices closer to the middle of the scale indicated less confidence in the rating, whereas a 1 or a 10 indicated high confidence in the rating.

All subjects were tested under all conditions, with three replications of each condition. The conditions were randomized within blocks of trials containing one instance of each condition. Three such randomized blocks were created, and the subjects were assigned sequentially to one of the six possible presentation orders of the blocks. The experiment took place in one session.

Apparatus

Stimuli were made with the MITSYN sound processing system (Henke, 1980) running on a PDP-11/34 minicomputer with the RSX-11M operating system. The stimuli were output by a separate computer program. The sampling rate was 10 kHz, and a Rockland 851 anti-imaging low-pass filter, with a roll-off of 48 dB/octave, was

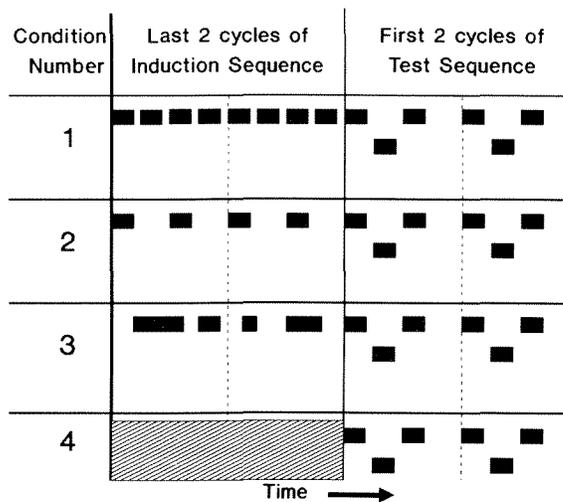


Figure 2. Part of a trial for Experiment 1. Each dark rectangle represents one pure tone, and the large striped rectangle represents white noise. A vertical solid line indicates the (unsigned) boundary between induction and test sequences, and the vertical dashed lines indicate cycle boundaries.

set to 4800 Hz. The signals were amplified with a Sony TA-F444ES amplifier and sent to an Industrial Acoustics audiometric testing room (Model 1202). The signals were played over Tektronics TDH-49P stereo headphones. The levels were calibrated with a General Radio Company sound-level meter (Type 1551-C), with a flat-plate headphone coupler (A weighting).

Results

For each subject, the results of the three trials of each condition were averaged, yielding one measure of induced segregation per condition for use as the dependent variable. Figure 3 shows the average segregation induced in the test sequence by each of the four induction conditions at each frequency separation.

A 4 (frequency separation) × 4 (induction condition) repeated measures analysis of variance (ANOVA) was conducted, using the Greenhouse-Geisser correction. Reliable main effects for both induction condition [$F(1,19) = 23.880, p < .001$] and frequency separation [$F(1,19) = 26.507, p < .001$] were found. There was no significant interaction between the two factors [$F(1,19) = 1.166, p > .25$]. The means for frequency separations of 5, 9, 11, and 18 semitones, respectively, were 2.9, 4.4, 4.6, and 6.3. The means for the high-density, predictable, unpredictable, and control conditions were (respectively) 4.9, 5.4, 5.4, and 2.5.

Two orthogonal planned comparisons were done among induction condition types. The *predictable* and the *unpredictable* conditions (with tone densities of two tones per 400-msec cycle) were compared to determine whether

regularity of rhythm had an effect. This was not significant [$F(1,19) = .004, p = .953$]. The high-density condition, with four induction tones per cycle, was compared with the average of the conditions with only two induction tones per cycle (*predictable* and *unpredictable* conditions combined). There was a reliable difference [$F(1,19) = 4.582, p < .05$], with the two-tone conditions more effective at inducing segregation.

DISCUSSION

The main effect of frequency separation confirmed that the data conform to known principles of stream segregation. The three experimental induction sequences induced segregation of the test sequence significantly better than did the control sequences, although there were no frequency changes in any of them. This demonstrates that the habituation of frequency-jump detectors is not necessary for explaining the accumulation of segregation tendencies of tones alternating between two frequencies.

We did not find that a predictable induction sequence was any more or less successful at augmenting segregation of the test sequence than an unpredictable one, when the two induction sequences were equal in frequency composition and average tone density. Because both these induction sequences had the same tone density as did the high tones of the test sequence, the tone *density* of the test sequence was predictable from the induction sequence. However, the predictability was not rule-based as in the theory of Jones (1976); that is, no rule could be used to capture tones by predicting when the next one would occur, or how long it would last, because these properties were random.

The four-tone-per-cycle induction sequence was less effective at augmenting segregation than was either of the two-tone-per-cycle induction sequences. This is an interesting finding, because it shows that more stimulation in a frequency region is not necessarily better from the point of view of capturing a stream of subsequent tones from that region.

EXPERIMENT 2

The experiment just described does not support the idea that having more tones in a frequency region will necessarily cause better capturing of tones of the same frequency. Yet with so few conditions in the experiment, the relationship between the total duration of the induction tones and the amount of induced streaming could not be determined. The following experiment was designed to elucidate this relationship. Bregman's view of a cumulative process in stream segregation predicted a frequency bias based on the cumulative *amount* of stimulation in one frequency region. Another alternative, suggested from the results of Experiment 1 above, was that streaming is based on cumulative information about the *properties* of the induction sequence, and that similarity between the induction and test sequences would be the basis for grouping. In the present experiment, all induction sequences were

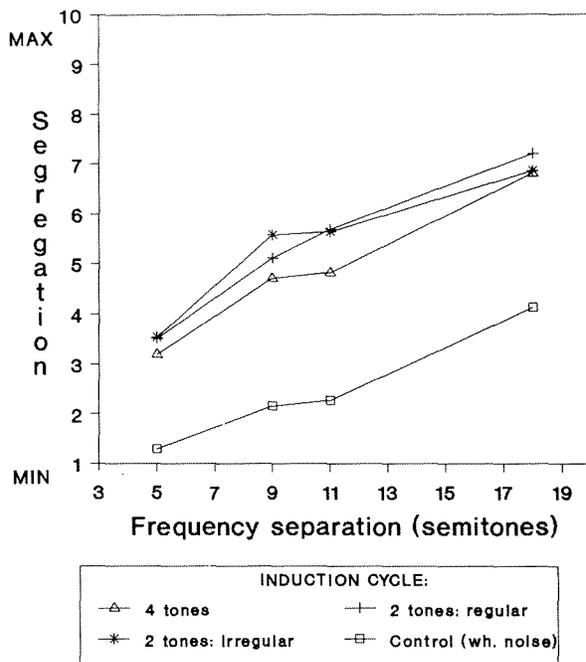


Figure 3. Results from Experiment 1. Induced segregation as a function of test sequence frequency separation, for each of the four induction conditions.

made irregular to avoid the issue of predictability. Jones's (1976) theory is therefore not relevant to this experiment.

Method

Only details that are different from the previous experiment will be described.

Subjects

The subjects were 16 graduates and undergraduates at McGill or Concordia Universities.

Stimuli

The test sequences were three repetitions of a 400-msec two-tone gallop sequence whose elements were pure tones of 80-msec duration (with linear onset and offset of 10 msec each). The difference in frequency between the high and low tones was 5, 9, 11, or 14 semitones. The high tones were always 1048 Hz.

There were six induction sequences, each 4.8 sec long (twelve 400-msec cycles), containing only tones of one frequency, the same frequency as that of the high gallop tones. These tones filled 0%, 20%, 40%, 60%, 80%, or 100% of the induction sequence. (For a diagram of one sample cycle for each of the induction sequence conditions, see Figure 4.) The 0% condition, in which there was no tone (control condition), was filled with soft white noise. The middle conditions (i.e., in which the density was not one of the extremes of 0% or 100%) contained tones that were intermittent and separated by silences. Their average length was 80 msec and their density was 1, 2, 3, or 4 tones per 400-msec period. The actual tone length and spacing within each period were random. The stimulus in the 100% condition was a long tone, 4.8 sec in duration. These stimuli were constructed by a computer program, as follows: for each 400-msec cycle of the stimulus, tone lengths were randomly chosen for the appropriate number of tones. The time remaining in the cycle was allocated to silences, whose lengths were randomly chosen, and which were interleaved with the tones. One sequence of 12 cycles was created for each of the 20%, 40%, 60%, and 80% conditions, and all subjects heard the same sequences. The subjects reported that the stimuli for these middle conditions sounded like Morse code.

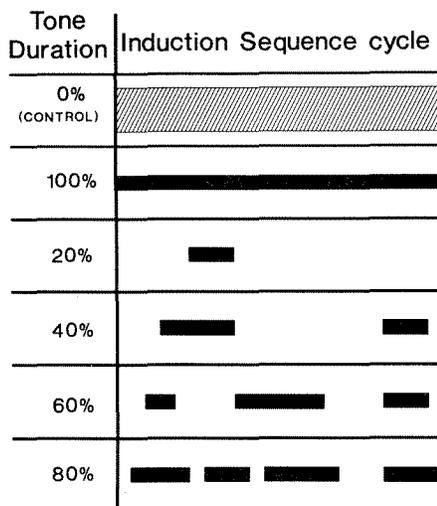


Figure 4. One 400-msec sample from the 4.8-sec induction conditions in Experiment 2. The total tone duration is expressed as a percentage of the total time of the induction sequence. The large striped rectangle indicates white noise. The black rectangles represent pure tones.

The level of the intermittent tones was 62.5 dB SPL. The continuous tone of the 100% condition was reduced to 59.5 dB because pilot studies showed it to be unpleasantly loud to subjects. The level of the white noise was 47 dB SPL.

The levels of the frequency-separation factor in the test sequence were chosen from pilot study data so that subjects were certain to experience the full range of ratings, from no segregation to complete segregation. The levels of the tone density factor were chosen to cover evenly the range between no tone and continuous tone in the induction sequence. The frequency range (466-1024 Hz) was in a relatively flat portion of the equal-loudness curves.

Procedure

Training. The subjects were presented with a menu on the computer screen, which allowed them to select a sample of a segregated two-tone sequence, a coherent gallop, a fixed-order sequence of all the induction sequences, or a sample trial. They were required to listen to each at least once, but they could choose extra repetitions. Then they were given 12 trials, in which they practiced giving ratings of segregation of the test sequence. As a result, they knew what kind of sounds to expect and were familiar with giving responses on the computer keyboard.

Task. The subjects were required to rate, on a 7-point scale, the degree to which they heard a gallop sequence clearly at the end of the trial. A rating of 1 indicated certainty that no gallop was present, and a rating of 7 indicated that a gallop was clearly and easily heard. The subjects were instructed to try to perceive a gallop. At least 5 sec of silence, the subject's response time plus 5 sec, separated the trials.

The experiment took place in one session. All subjects were tested under all conditions, with three replications of each. The conditions were randomized within blocks of trials containing one instance of each condition, and the randomization was different for each subject.

Apparatus

The sampling rate used in this experiment was 20 kHz, and the Rockland 851 low-pass filter was set to 10 kHz. Sound was presented over Sony MDR-V7 headphones. The levels were measured with A weighting, both on a continuous tone of 1048 Hz and on white noise.

Results

The subjects' scores, where low numbers indicated segregation, were recoded (by inverting the numbers of the scale) so that the y-axis on the graph of results would be comparable to that in the figure for Experiment 1. This recoding was permissible because, although the numbers on the rating scale were different, the task was identical in both experiments; the scores measured the amount of segregation when the subject was trying to hold the test sequence together as a coherent sequence. For each subject, the results of the three replications of each condition were averaged, yielding one measure of segregation of the test sequence per condition for use as the dependent variable. A 4 (frequency separation) \times 6 (total tone duration during induction) repeated measures ANOVA was conducted, using the Greenhouse-Geisser (1959) correction. Reliable main effects for both tone duration [$F(1,15) = 54.08, p < .01$] and frequency separation [$F(1,15) = 36.93, p < .01$] were found. There was no reliable interaction between the two factors [$F(1,15) = 2.56, p > .1$]. The means for frequency separations of

5, 9, 11, and 14 semitones, respectively, were 2.4, 3.3, 3.5, and 4.2. The means for the control (0%), 20%, 40%, 60%, 80%, and 100% tone-duration conditions were, respectively, 1.7, 3.1, 4.2, 4.8, 4.7, and 1.4. The results are shown in Figure 5, with segregation plotted against frequency separation for each total tone duration.

Longer tone durations did not always increase segregation. Surprisingly, the induction sequence in which the tone was always present (100% condition) induced no measurable increase in segregation; that is, it had no more effect than the control condition, in which no tone was present. Thus there was not a monotonic relationship between *tone duration* and segregation. However, as the *number of tone onsets* increased, the segregation of the Test gallop increased up to the point at which the tone-onset densities of induction and test sequences were equal (3 onsets per 400-msec cycle, for a total of 36 onsets in the induction); beyond that point there was no change.

A Bonferroni-adjusted *t* test of all pairs of means was done at the .05 level. All induction conditions except for the control condition gave reliably more segregation than did the continuous-tone condition. The 20% condition (one induction tone per cycle) was significantly different from all the other conditions: it induced less segregation than did the 40%, 60%, and 80% conditions, but more than the control or the 100% condition. The three middle-density conditions (40%, 60%, and 80%) were not reliably different from each other. There was no significant difference between the control and the continuous-tone condition.

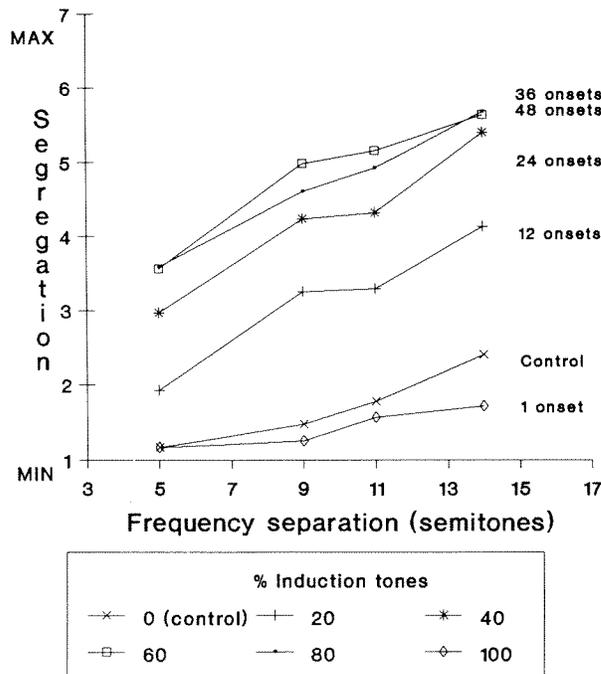


Figure 5. Results of Experiment 2. Induced segregation as a function of test sequence frequency separation, shown for each level of the tone-duration factor. The total number of tone onsets in each induction sequence is shown on the graph beside the appropriate line.

Discussion

The main effect of frequency separation confirmed that the data are responsive to factors known to affect segregation. Bregman’s hypothesis about cumulative frequency bias was disconfirmed; it is clear that the total duration of tones of one frequency does not predict the induction effect on subsequent two-frequency sequences. Instead, the second alternative for the cumulative effects of streaming was supported; the most induction occurred when there was a rough matching of properties between the induction and test sequences. The number of tone onsets (event density) appears to be important. This support for sequential grouping by similarity is an interesting finding which is discussed more fully below.

Of special interest was the failure of the continuous tone to induce segregation. The enhancement effect (Summerfield, Haggard, Foster, & Gray, 1984), which would predict enhanced perception of frequencies other than that of the continuous tone and a resulting *segregation* of the test sequence, was not operating in this case. Examination of the data for continuous-tone and control conditions, for each level of the frequency-separation factor, reveals a consistent pattern: there is less segregation induced by the continuous tone than by the control. However, the numerical difference was slight, perhaps because of a floor effect. Thus, although the continuous tone may for some reason have enhanced the *coherence* of the test sequence, rather than merely failing to segregate it, a post hoc test may not have been powerful enough to detect a difference. Several subjects mentioned the monotony (experienced as very tedious) of the continuous-tone condition, which was not present in the control because white noise is always changing. It is possible that stimuli with unchanging properties cause habituation of some comparator mechanism, which then leads auditory attention preferentially toward stimuli with other properties. The subjects in the experiment, wearing headphones and sitting in a soundproof booth, had no other sounds to turn their attention to, and they found this unnatural experience unpleasant. Their consistently lower segregation scores for the test sequence after the continuous-tone condition may have reflected a recovery from habituation once the monotonous stimulus ended.

A mechanism that keeps track of stream properties, for calculation of similarity, might deal with the sequences of this experiment as follows. We assume that the default interpretation is the perception of a single stream. After the silence that precedes the trial, tones start to arrive. Initially they are perceived as a single stream. Their properties, such as the mean frequency and the frequency range, as well as the onset rate (or tone density), are computed. During the induction sequence, there is no reason to form more than one stream, since there is only a single frequency and the statistical properties of its changes remain within the initial estimates. After the test sequence begins, two changes will be noticed: the presence of a new frequency and possibly a different tone-onset density. To the degree that such properties are changed, the

new tones will be ejected to form a stream of their own, separate from the first. A lack of change should result in the greatest maintenance of the original stream, one of high-frequency tones only. A large change should result in all the test sequence tones, even the high ones, forming a new stream. Initially this new stream will be unsegregated, until evidence for a bimodal distribution of frequencies has been accumulated. The data are consistent with this hypothesized stream-tracking mechanism, in which the recent properties of existing streams are an important input to the auditory scene analysis process.

This mechanism may have its basis in adaptation to stimuli that are not changing. Hafter and Buell (1990) have found that the auditory system shows rapid adaptation to interaural time difference information, but that the adaptation reverses quickly when there is a change in the stimulus properties. A similar process may work for stream segregation, with unchanging information given little attention once registered.

EXPERIMENT 3

The previous experiments confirmed that the adaptation of frequency-jump detectors is not necessary to explain streaming, but that does not mean that such detectors do not exist or that they do not contribute to streaming. If they exist, however, it would be surprising if they were peripheral to the point of binaural fusion. Experiments done by Steiger and Bregman (1982) indicate that the processes of sequential stream segregation and binaural fusion each influence the other—that is, that no one process makes a preemptive “decision” about what makes a stream. Smith, Hausfeld, Power, and Gorta (1982) confirmed that streaming involves competition among proximities in frequency, spatial location, loudness, and pitch. Monaural information alone is not enough to explain these results.

One finding from Anstis and Saida’s (1985) study could be interpreted as proof that stream segregation is carried out by a peripheral process: sequences of alternating tones, presented to one ear until streaming had taken place, were immediately perceived as coherent again when the presentation was suddenly switched to the other ear. This result could equally well be explained by the theory that we presented in the discussion of Experiment 2; the tones presented to the second ear had a completely different spatial location and might therefore be treated as a new stream.

In Experiment 3, we looked for a central process for the origin of streaming. Test sequences (gallop cycles) were presented to the right ear only. Induction sequences for the experimental conditions were constructed so that a gallop was also present in the right ear. However, in some conditions, the gallop was perceptually camouflaged or moved in space by means of contralateral sounds, acting according to known auditory grouping processes. A peripheral jump-detector mechanism would be expected to induce equal amounts of right-ear streaming on the test sequence in all conditions, since a gallop was always

present at that ear in both the induction and the test sequences. A central mechanism would be revealed by differences among the effects of the induction sequences.

Method

Only details that are different from those of the previous experiment will be described.

Subjects

There were 15 subjects, each of whom had participated in Experiment 2.

Stimuli

The induction sequences were 8 sec long, 20 cycles of 400 msec each. Note that these were much longer than the induction sequences of previous experiments. It was important to ensure that, if frequency-jump detectors were involved in segregation, they would have adequate time to become adapted. The test sequences were presented to the right ear only. The right ear was chosen arbitrarily, because Anstis and Saida (1985) had already shown that the effect of switching a stream of segregated alternating tones from one ear to the other is symmetrical. The upper gallop tones were 1048 Hz. All tones of both sequences were of 50-msec duration, including 10-msec linear onsets and offsets.

The frequency differences between high and low tones of the gallop were 8, 11, 13, and 16 semitones. These levels were chosen, from pilot study data, so that the subjects were certain to experience the full range of ratings from no segregation to complete segregation of the test sequence.

Five conditions for the induction sequence were chosen in the following manner: Two extreme conditions were provided such that one (the control) would have no effect on the test sequence and one (the monotonic gallop) would have a large effect. Except for the control, all conditions presented a gallop to the right ear. See Table 1.

One condition, the disguised gallop, was designed so that the right-ear gallop was perceptually hidden by tones in the left ear. The three repeating left-ear tones were chosen in light of perceptual grouping principles. The first and third tones of each cycle were presented close in frequency to the high right-ear tones and just after them in time. The middle tone of each cycle was one musical fifth below the low right-ear tone (frequency ratio of 2:3) and was presented simultaneously with the latter (see Figure 6). The high left-ear tones grouped sequentially with the high right-ear tones, and the low left-ear tones grouped simultaneously with the low right-ear tones. The stimulus was heard as an upward trill alternating with a low complex tone. That is, a gallop was not perceived even though all the required tones, and only those, were registered at the right ear.

Finally, two experimental induction conditions in which the gallop was perceptible were chosen. In one, the noise-plus-gallop condition, white noise was presented to the left ear throughout the induction sequence. In the second, the diotic gallop condition, tones identical to the right-ear tones were presented to the left ear. An incidental by-product of the design was the spatial shifting of the gallop in the latter two conditions. The white noise, by means of

Table 1
Conditions for Experiment 3

Induction Type	Left Ear	Right Ear
Control	White noise	White noise
Disguised gallop	Distractor tones	Gallop
Diotic gallop	Gallop	Gallop
Noise plus gallop	White noise	Gallop
Monotonic gallop	Silence	Gallop

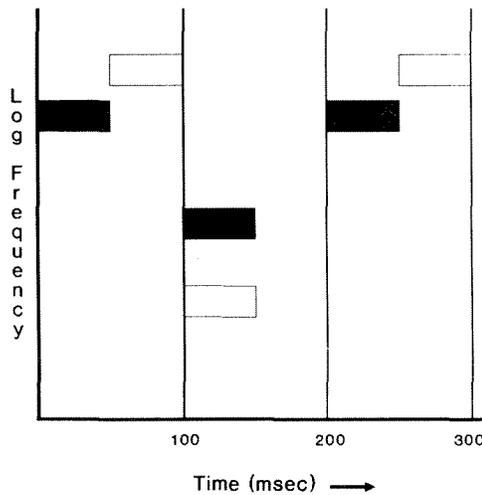


Figure 6. Experiment 3, disguised gallop stimulus. Right-ear tones (gallop) are denoted by black rectangles. Left-ear tones (distractors) are denoted by white rectangles. All tones are of 50-msec duration. The rest of the 400-msec cycle (between gallops) is filled with silence.

contralateral induction (Warren, 1982, pp. 41–45), caused a slight shift in the apparent spatial location of the right-ear gallop. The presentation of concurrent gallop tones to the left ear, in the diotic gallop condition, also shifted the gallop percept to the center. The left-ear stimuli thus altered the perceived properties of the right-ear induction gallop to varying degrees in the experimental conditions: completely (the disguised gallop), partially (the noise-plus-gallop and diotic gallop), or not at all (the monotic gallop).

Each subject was presented with all possible combinations of the two factors: induction condition (five levels) and frequency separation of the test sequence gallop (four levels). Each block of trials, with one occurrence of each condition, was randomized, and there were four blocks.

We expected that the disguised gallop condition, because its right-ear gallop did not form a separate stream during the induction sequence, would induce no more segregation than the control condition, and that the experimental conditions (i.e., noncontrol conditions) would differ from each other.

Procedure

Training. Subjects were presented with a menu on the computer screen, which allowed them to select a sample of a segregated two-frequency sequence, a coherent gallop, two cycles from each of the induction sequences (in a fixed order), or a sample trial. They were required to listen to each of those at least once, but they could choose extra repetitions. Then they were given 12 practice trials.

Task. The task was identical to that of Experiment 2, described above.

Apparatus

This was the same as that for Experiment 2, except that the loudness levels were set at 65 dB SPL. For calibration, a pure tone of 1048 Hz was used.

Results

For each subject, the results of the four trials of each condition were averaged, yielding one measure of segregation of the test sequence per condition for use as the

dependent variable. All the reported means refer to the recoded values (with numbers on the rating scale inverted, as in Experiment 2), whose high numbers indicate segregation. A 4 (frequency separation) × 5 (induction condition) repeated measures ANOVA was conducted, using the Greenhouse–Geisser (1959) correction. Reliable main effects for both frequency separation [$F(1,14) = 9.530, p < .01$] and induction condition [$F(1,14) = 47.477, p < .01$] were found. There was no significant interaction between the two factors [$F(1,14) = 1.213, p > .1$]. The means for frequency separations of 8, 11, 13, and 18 semitones, respectively, were 2.5, 2.7, 3.1, and 3.5. The means for the control, disguised gallop, diotic gallop, noise-plus-gallop, and monotic gallop were, respectively, 1.8, 1.9, 2.4, 3.1, and 5.4. The results are shown in Figure 7. The subjects' scores were recoded as in Experiment 2, so that the y-axis on this graph would be comparable to that in the figures for the other experiments.

A planned comparison between the disguised gallop and the control condition was performed. As hypothesized, there was no significant difference [$F(1,14) = 0.636, p = .439$]. A Bonferroni-adjusted *t* test of all pairs of means at the .05 level showed that the monotic (left-ear silence) condition induced reliably more segregation than did all the other conditions. Among the binaural conditions, the noise-plus-gallop condition induced significantly more segregation than did the disguised gallop and the control conditions.

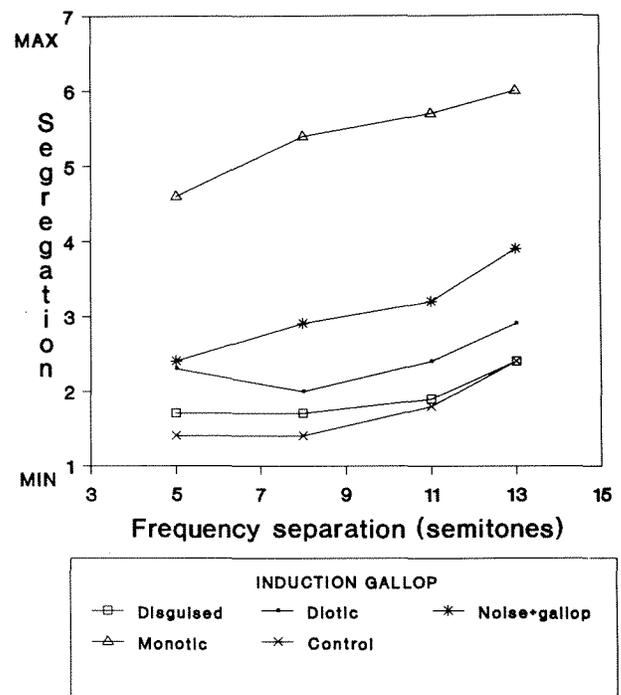


Figure 7. Results of Experiment 3: Induced segregation as a function of test sequence frequency separation, shown for each level of the induction condition factor.

Discussion

The data provide evidence for a central process in the streaming of two-frequency sequences: the experimental conditions were not equal in effect, even though all had identical right-ear stimuli. In particular, the disguised gallop did not induce any streaming on the test sequence gallop. This implies that adaptation of jump detectors, if it exists at all, must involve the tones within a perceived stream, rather than just the tones at a single ear. The main effect of frequency separation and the large difference between the monotic and the control conditions confirmed that the data are valid.

Also of interest were the results from the diotic gallop condition, which had seemed likely to induce segregation very strongly. However, it did not; the rating, averaged over all subjects and frequency-separation conditions, was 2.4. One possible explanation, which is consistent with the results of Experiment 2, is that the induction sequence had segregated but that a sudden change in stream properties was responsible for resetting the streaming mechanism. The transition from the induction to the test sequence included a change in both the intensity and spatial location of the gallop, which could be interpreted as evidence that the test sequence tones belonged to a completely different stream. This makes sense, because there is no natural object that can move through 90° instantaneously; the auditory system would be correct in treating the new tones as part of a new object.

GENERAL DISCUSSION

There is a need to reevaluate the three theories in light of the new data. Jones's (1976) theory of a predictive rule-based process in segregation is not supported by these data. In Experiment 1, the regular and irregular monotone induction sequences, with otherwise similar properties, had the same effect. According to her theory, any increased knowledge of the sequences, gained during the experiment, ought to have helped the subjects do what they were trying to do, namely to hold the test sequence together. The cumulative effect observed in our experiment, therefore, cannot have been a general increased knowledge of the stimulus. It must be viewed as something that specifically comes to favor a two-stream over a one-stream interpretation. Jones, Kidd, and Wetzel (1981) proved that *some* aspect of rhythm must be a factor in streaming. Their results can now be explained in terms of grouping by similarity (in properties such as event density or total tone duration) rather than predictive attention.

The theory of peripheral frequency-jump detectors, as proposed by Anstis and Saida (1985) and Van Noorden (1975), would need modification before it could be retained. First, frequency-jump detectors are not necessary to explain streaming, as is shown by the induction of streaming by sequences of only one frequency. Thus this theory is incomplete, especially since it also fails to deal with competition among grouping processes (Bregman,

1990). Second, even if frequency-jump detectors do contribute to streaming, they are not peripheral to the superior olivary complex, where information from the two ears first converges; or, at the very least, their output is interpreted by central processes. One might modify the theory to say that the adaptation to alternation of frequencies occurs somewhere subsequent to binaural integration but still in time to contribute to scene analysis.

Another way in which the theory would have to change is related to our finding that the adaptation recovers quickly under some conditions, such as the instantaneous change in both intensity and spatial location of the diotic gallop condition in Experiment 3. Perhaps the modified theory might propose that when properties change suddenly, the new inputs are routed to a different population of neurons, with a fresh supply of frequency-jump detectors. This might provide a mechanism similar to that discovered by Hafter and Buell (1990), which shows adaptation to unchanging information and then rapid recovery when a change does occur. That is, the results obtained by Anstis and Saida (1985) may have been caused by adaptation to the lack of change (constant repetition) in the alternating-tone sequences, rather than by adaptation to the frequency jumping.

Bregman's scene-analysis theory is largely supported by the data from the present experiments. They confirmed that the whole auditory context is used (not just local information at the periphery), and they showed that sequential grouping by similarity (of incoming sounds to existing streams) occurs. Properties involved in the calculation of similarity may include event density, total tone duration, spatial location and sound level. However, the cumulative effects of streaming are not dominated by a process that responds to cumulative tone duration in a frequency region. Furthermore, contrary to previous results (Bregman, 1978b), cumulative streaming can be reversed rapidly—that is, without a time period of several seconds. This occurred in the transition between the induction and test sequences of the diotic gallop condition of Experiment 3. The processes of hysteresis apparently do not prevent rapid changes in streaming when there is good evidence for genuine changes in the environmental sources of incoming sounds (e.g., new spatial location of newly arriving tones, and disappearance of tones in the spatial location of the established stream).

There are other implications of the new data, apart from their resolving some issues about the three theories of segregation. The mechanisms that compute the properties of existing streams may help implement grouping by similarity by providing feedback (of properties derived from previous grouping) to the scene analysis processes. It is possible that a comparator mechanism is always at work on sounds tentatively assigned to a stream to verify that any change in stream properties is gradual enough to be consistent with the new sounds' belonging to the ongoing stream. The output from the comparator may signal to the scene analyzer whether or not to continue making groupings in the same way.

Some further experiments would be useful, in the systematic investigation of the effects of changes in stream properties on cumulative stream segregation. For example, the comparison of the effects of sudden and gradual changes, in properties such as sound level or spatial location, would be interesting. This work is in progress now at our laboratory.

REFERENCES

- ANSTIS, S. M., GIASCHI, D., & COGAN, A. I. (1985). Adaptation to apparent motion. *Vision Research*, **25**, 1051-1062.
- ANSTIS, S. M., & SAIDA, S. (1985). Adaptation to auditory streaming of frequency-modulated tones. *Journal of Experimental Psychology: Human Perception & Performance*, **11**, 257-271.
- BREGMAN, A. S. (1978a). Auditory streaming: Competition among alternative organizations. *Perception & Psychophysics*, **23**, 391-398.
- BREGMAN, A. S. (1978b). Auditory streaming is cumulative. *Journal of Experimental Psychology: Human Perception & Performance*, **4**, 380-387.
- BREGMAN, A. S. (1990). *Auditory scene analysis*. Cambridge, MA: MIT Press.
- BREGMAN, A. S., & ACHIM, A. (1973). Visual stream segregation. *Perception & Psychophysics*, **13**, 451-454.
- 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.
- BREGMAN, A. S., & RUDNICKY, A. (1975). Auditory segregation: Stream or streams? *Journal of Experimental Psychology: Human Perception & Performance*, **1**, 263-267.
- FRENCH-ST. GEORGE, M., & BREGMAN, A. S. (1989). Role of predictability of sequence in auditory stream segregation. *Perception & Psychophysics*, **46**, 384-386.
- GREENHOUSE, S. W., & GEISSER, S. (1959). On methods in the analysis of profile data. *Psychometrika*, **24**, 95-112.
- HAFTER, E. R., & BUELL, T. N. (1990). Restarting the adapted binaural system. *Journal of the Acoustical Society of America*, **88**, 806-809.
- HENKE, W. L. (1980). *MITSYN: A coherent family of command-level utilities for time signal processing* [Computer program]. (Revised and extended by Daniel Solomon, Department of Psychology, McGill University)
- JONES, M. R. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review*, **83**, 323-355.
- JONES, M. R., KIDD, G., & WETZEL, R. (1981). Evidence for rhythmic attention. *Journal of Experimental Psychology: Human Perception & Performance*, **7**, 1059-1073.
- MAXWELL, S. E. (1980). Pairwise multiple comparisons in repeated measures designs. *Journal of Educational Statistics*, **5**, 269-287.
- M McNALLY, K. A., & HANDEL, S. (1977). Effect of element composition on streaming and the ordering of repeating sequences. *Journal of Experimental Psychology: Human Perception & Performance*, **3**, 451-460.
- MILLER, G. A., & HEISE, G. A. (1950). The trill threshold. *Journal of the Acoustical Society of America*, **22**, 637-638.
- MOORE, B. C. J. (1982). *An introduction to the psychology of hearing* (2nd ed.). London, New York: Academic Press.
- MYERS, J. L. (1979). *Fundamentals of experimental design* (3rd ed.). Boston: Allyn & Bacon.
- SMITH, J., HAUSFELD, S., POWER, R. P., & GORTA, A. (1982). Ambiguous musical figures and auditory streaming. *Perception & Psychophysics*, **32**, 454-464.
- STEIGER, H., & BREGMAN, A. S. (1982). Competition among auditory streaming, dichotic fusion, and diotic fusion. *Perception & Psychophysics*, **32**, 153-162.
- SUMMERFIELD, Q., HAGGARD, M., FOSTER, J., & GRAY, S. (1984). Perceiving vowels from uniform spectra: Phonetic exploration of an auditory aftereffect. *Perception & Psychophysics*, **35**, 203-213.
- VAN NOORDEN, L. P. A. (1975). *Temporal coherence in the perception of tone sequences*. Unpublished doctoral dissertation, Technische Hogeschool Eindhoven, Eindhoven, The Netherlands.
- WARREN, R. M. (1982). *Auditory perception: A new synthesis*. New York: Pergamon.
- WINER, B. J. (1971). *Statistical principles in experimental design* (2nd ed.). New York: McGraw-Hill.

(Manuscript received April 1, 1991;
revision accepted for publication August 5, 1992.)