Perception & Psychophysics
1978, Vol. 23 (5), 391-398

Auditory streaming: Competition among alternative organizations

ALBERT S. BREGMAN
McGill University, Montreal, Quebec, Canada

It has been proposed that auditory stream splitting in rapid tone sequences occurs whenever a tone falls outside some critical region surrounding its predecessor and some tracking mechanism cannot shift its frequency setting fast enough. If this were true, a certain pair of tones would split apart or not, depending on their separation in time and frequency. Actually their splitting apart depends on the context of other tones. Alternative groupings compete for tonal elements. This was demonstrated using adult subjects who listened to a rapid repeating four-tone cycle and made three types of judgments: (1) discriminating the order of two of the tones, (2) saying whether two of the tones could be heard as a separate pair, and (3) judging the rhythmic pattern. It is proposed that stream formation is a pattern-factoring mechanism, sensitive to pattern properties.

If a sequence of discrete tones is presented rapidly, it seems to "split" perceptually into two or more parallel sequences as if two or more different sources of sound, each restricted to a certain range of frequencies, were emitting different, but interwoven, sounds (Dowling, 1973; Miller & Heise, 1950; Van Noorden, 1975). This phenomenon has been referred to as "stream segregation" or "streaming" (Bregman, 1978; Bregman & Campbell, 1971; Bregman & Dannenbring, 1973, in press; Bregman & Rudnicky, 1975). Each of the separate "parts" or "sources" is referred to as a stream. It has been shown that it is hard to hear patterns that include elements of different streams (Bregman & Campbell, 1971) or to temporally locate elements of one stream with respect to members of the other stream (Dannenbring & Bregman, 1976; Norman, 1967; Van Noorden, 1975).

There are two possible approaches to this phenomenon. One approach views it as the *breakdown* of a mechanism that normally follows the true succession of stimuli. The other approach sees it as an *accomplishment* of the nervous system: a process of taking a complex stream apart into its probably meaningful components. The approach that views it as a breakdown is encouraged by the finding of Miller and Heise (1950) that when there is a rapid alternation of tones, the splitting into two streams depends on the frequency separation of the tones. It seems furthermore to depend on the ratios between the tones. Splitting occurs when the difference is about 15% for pairs of tones located in various frequency regions. One proposal says that at high speeds a tracking mechanism cannot follow a tone when it is outside the critical band of its predecessor (Norman, Note 1). This is suggested by the fact that the width of the critical band around a particular frequency is proportional to that frequency, as is the splitting threshold studied by Miller and Heise (1950). Van Noorden (1975) offered two preliminary hypotheses relating streaming to auditory physiology. One related a listener's ability to integrate successive tones to the degree to which these tones stimulate overlapping populations of hair cells in the cochlea (pp. 21, 24). Secondly, Van Noorden offered the hypothesis that there were "pitch motion" detectors which required a longer time interval to register "motion" between successive tone bursts the further apart they were in frequency (pp. 48-51). He related this to Körte's laws concerning apparent motion in vision. (In this regard, see also Bregman & Achim, 1973.) Both proposed mechanisms are, in effect, filters. Prior tones define the setting of the filter; and subsequently, "the tones which fall inside the passband of this filter are perceived better than those which fall outside." Furthermore, "the filter can only follow the tones with a limited velocity" (Van Noorden, 1975, p. 40). This accounts for increased segregation at higher speeds.[1] All these hypotheses have in common the idea that the auditory system is built to integrate successive sounds, and that streaming arises when stimulus factors push the integration mechanisms beyond their limits. They

---

Acknowledgment is made of research support from the Defense Research Board and the National Research Council of Canada, and the FCAC program of the Quebec Ministry of Education. The author wishes to acknowledge the research assistance of Jeff Selig, Graham Reynolds, and Jean Beninger. Requests for reprints should be addressed to the author, Department of Psychology, McGill University, 1205 McGregor Avenue, Montreal, Quebec, Canada H3A 1B1.

also try to predict streaming from the relationship between temporally adjacent tones.

An alternative view is that the auditory system is not built to integrate successive sounds willy-nilly, but to integrate those that probably arose from the same source (Bregman, 1978; Bregman & Dannenbring, in press). Since the pressure waves arising from different sources of sound are mixed at the ear, the auditory system must have mechanisms to sort out the contributions of different sources. In doing so, it will make use of similarities in the frequency spectrum from moment to moment (Dannenbring & Bregman, 1976), of continuities in the direction of frequency changes (Bregman & Dannenbring, 1973; Heise & Miller, 1951), of synchrony of onset and modulation of frequency components, of abruptness of change and so on. This approach views stream formation as an accomplishment, not a breakdown; it factors an input into streams that probably arose from different sources.[2]

A sensible stream-forming mechanism would create more than one stream at a time. Then, as each new portion of the acoustic input arrived, it could be assigned to the stream that it fit best. There are experimental data which support the hypothesis that two or more streams, in the process of being constructed, compete for new inputs (Bregman & Rudnicky, 1975).

If an input is assigned to the best-fitting stream, then any explanation of streaming which relies only on relations between each moment of sound and the next will fail. Depending on the set of active streams, an adjacent pair of sounds may or may not enter the same stream, regardless of their frequency separation from one another. Thus, stream membership arises out of the competition of alternate possibilities for the grouping of sounds and not from the specific frequency separation between an adjacent pair.

The present experiments set up conditions in which a successive pair of tones, A and B, with a fixed frequency and temporal separation, would either be integrated into the same stream or segregated into separate streams, depending on the context of other tones. There were always four tones, ABXY, in a rapid repeating cycle. The streaming of A and B was influenced only by manipulating the frequencies of X and Y. Stream segregation was detected by its influence on three types of responses. Experiment 1 studied the ability of listeners to tell the order of A and B. Experiment 2 asked listeners to judge whether A and B could be easily perceived as a separate pair; the same experiment also studied stream segregation by its effects on the rhythm of the sequence.

EXPERIMENT 1

Two conditions were set up in which a pattern of two tones, A and B, was to be recognized when combined with two other distractor tones, X and Y, in a four-tone repeating sequence, ABXY. Subjects were required to discriminate whether the tones A and B appeared in the order AB or BA in this sequence. In one condition, X and Y were chosen so as to segregate away from the pair AB and form a separate stream, XY. Thus A and B would be isolated and left together in the same stream; hence their order would be perceptible. This is called the "isolated" condition. In the other condition, X and Y were chosen so that A would be grouped with X forming a stream AX and B would be grouped with Y in the stream BY. Here A and B would be absorbed into separate streams and their order would be hard to perceive. This is called the "absorbed" condition. The tones A and B were identical in the two conditions. Hence, if B were outside some critical region around A in one condition, it would be in both. Furthermore, A and B were temporally adjacent in both conditions.

Method

Procedure. On each trial, a rapid repeating pattern, consisting of two tones (A and B) and two silences, was presented as a "standard." Then a repeating pattern of four tones (A, B, and the two distractor tones) was presented as a "comparison." The listener judged whether A and B were in the same order in the standard and comparison patterns and gave an estimate of the difficulty of the decision.

The exact sequence of events was as follows: After a 2-sec high-pitched warning tone and a 4-sec silence, the subject heard the standard, repeating for 5 sec, then a 1-sec silence, followed by the comparison pattern, repeating for 5 sec. After a 1-sec silence, he again heard the standard and comparison presented, as before. Then the subject recorded his judgment during the 11-sec intertrial interval. Each tone (or silence in the standard) was 65 msec in duration. In order to prevent the subject from using the first or last tone that he heard as an anchor point, the standard and comparison sequences were brought on gradually in amplitude over a 1-sec interval and went off gradually in the same way. (Note: This onset/offset fading is essential when streaming is to be studied via judgments of order.)

Experimental design. Eight tones, labeled 1 to 8, were selected so as to be grouped by proximity on a logarithmic scale of frequency into two major clusters (Tones 1 to 4 vs. Tones 5 to 8) with a large frequency gap between the clusters (see Figure 1). The tones of each major cluster were, in turn, grouped by proximity into two minor clusters (e.g., Tones 1 and 2 vs. Tones 3 and 4) again separated by a gap in frequency. The frequencies of the tones were as follows: 200, 246, 373, 455, 1,525, 1,860, 2,760, and 3,400 Hz. The target tones, A and B, were always selected so as to be in the same major cluster but from two different minor clusters.

To create the "isolated" condition, the two distractor tones, X and Y, present in the comparison sequence, were chosen from the *other* major cluster, one from each of its minor clusters. Hence X and Y were grouped together by proximity in frequency, as A and B were. To create the "absorbed" condition, A and B were selected as before, but this time the two distractors were chosen from the same major cluster as A and B, one of the distractors, X, being adjacent to A in frequency (from the same minor cluster) and the other, Y, adjacent to B. Hence, A and B would be captured by X and Y, respectively, into separate streams. A diagram showing the spacing of tones, an "isolated" comparison sequence, and an "absorbed" comparison sequence are shown in Figure 1. The arrows indicate the nature of the intended perceptual grouping. There was no difficulty in recognizing the pitches of the different tones, since the nearest tones were about 3.5 semitones apart.

Each of the eight tones was selected once as A and once as B, yielding 16 standard sequences (i.e., A, B, silence, silence). For each of these, there were "isolated" comparison sequences and two "absorbed" comparison sequences. In one-half, the comparison sequences A and B were in the same order as in the standard, and in the other half, they were in the reverse order. Thus there were 64 different standard-comparison arrangements. These were split into two counterbalanced blocks of 32 trials. All subjects had all conditions in the same order.

Stimuli. The tones were sinusoidal and were generated by a Wavetek (Model 136) voltage-controlled oscillator, controlled by a PDP-12 computer. Each tone consisted of a 10-msec linear rise from zero to full amplitude (about a 45-dB change in S/N), a 25-msec steady state at maximum amplitude. There was a 20-msec silence between tones. Thus, a new tone occurred every 65 msec. The eight tones were attenuated differentially by trial and error to eliminate subjective loudness differences caused at various places in the overall electrical-acoustical-perceptual system. The stimuli were played to the subjects in a small room via the speakers of a Revox 77A tape recorder. Listening volume was set at a comfortable level (75 to 85 dB SPL at the subject's ear).

Response scale. The response scale for each trial consisted, first, of two boxes, one marked "same" and one marked "different." The subjects were told to check the "same" box only if they could detect the two tones of the standard within the comparison sequence in the same *order* as in the standard. They were also asked to rate each judgment on a 7-point "difficulty" scale ranging from "very easy" to "very difficult," and were encouraged to use the entire scale.

Subjects. The subjects were 28 McGill University students. Each was given a pretest of eight trials with two-tone standard sequences and two-tone comparison sequences; i.e., they were asked to discriminate "A, B, silence, silence," from "B, A, silence, silence." It was felt that this level of auditory perceptual skill was necessary as a prerequisite for entering the main experiment. Pretest trials were presented in the same manner as in the main experiment. Any subject with more than three errors on the pretest was not permitted to go on to the main experiment. Hence, only 21 persons participated in the main experiment.

Results

The response protocols consisted of judgments of "same" or "different" for the standard and comparison sequences of each trial, with subjective ratings of "difficulty" along a 7-point scale. These two measures were combined into a single "rated similarity" score by multiplying the difficulty measure (1 to 7) by +1 if the judgment "same" had been made, and by −1 if the judgment "different" had been made. This gives a 14-point scale, ranging from −7 (easily made judgment of difference) to +7 (easily made judgment of similarity), with values near the center of the scale representing difficult judgments or ambiguous cases.

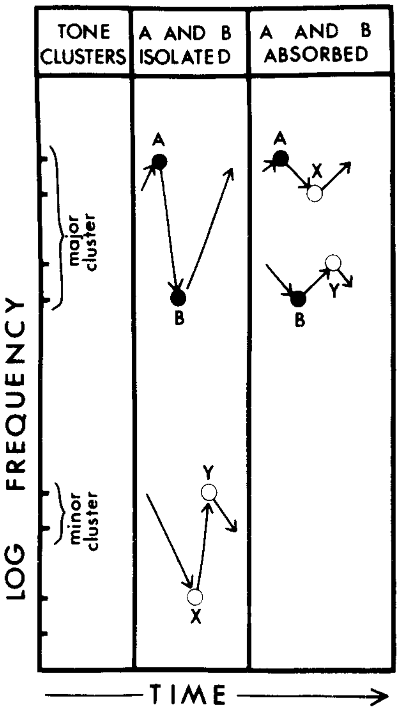

Figure 1. Illustration of: (1) the separation of tone clusters, (2) a sequence in which A and B are isolated, and hence grouped, and (3) one in which they are grouped into separate streams. (Arrows show perceptual streams.)

These rated similarity (RS) scores then became the raw measure for the calculation of a dependent variable D (Bregman & Campbell, 1971) for each subject in each condition. D is an easily calculated nonmetric measure representing the degree to which subjects could discriminate cases where the standard and comparison were the same from cases where they were different in each condition. D compares physical similarity and rated similarity, assigning high scores when these correspond. First, all RS scores for a given experimental condition are ranked. (This means that the gap in the scale between +1 and −1 is meaningless.) Then the ranks assigned to physically same and physically different pairs of stimuli are compared, and the overlap of ranks assessed. A D value of +1 represents complete discrimination (perfect separation of ranks), zero represents random judgments, and −1 shows systematic reversal of judgments. The equation for D is given below:

$$D = \frac{2(M_d - M_s)}{N},$$

where M_d is the mean of the ranks of the RS scores for physically different pairs; M_s is the mean of the ranks of the RS scores for physically same pairs; and N is the total number of judgments being ranked.

Because D is based on rank information only and is assessed separately for each subject in each condition, it is insensitive to individual differences in the use of the underlying scale or to the response biasing effects of different conditions.

The RS scores of each subject were grouped into four conditions, first vs. second block of trials in the experiment and isolated vs. absorbed conditions. The results are shown in Table 1.

Table 1
Mean D Scores

Stream Condition	Block of Trials	
	First	Second
Isolated	.541	– 008
Absorbed	000	– 051

This table was analyzed using ANOVA. There was a significant effect of trial block [$F(1,20) = 16.5$, $p < .001$], of stream condition [$F(1,20) = 18.6$, $p < .001$], and of the Condition by Block interaction [$F(1,20) = 12.9$, $p < .005$]. It is apparent that performance fell off to a random level in both conditions in the second half of the experiment. Apparently the attentional demands of the experiment were quite exhausting. However, performance was systematically different in the two stream conditions in the first half, with moderately good performance in the isolated condition and random performance in the absorbed condition.

Discussion

The ability to judge the order of two temporally adjacent tones, A and B, in a four-tone recycling series has been shown to depend on the choice of the two other tones. If the overall distribution of tones causes A and B to be perceptually grouped and isolated from context, their order may be judged with some accuracy. If the distribution of tones is such as to absorb A and B into separate streams, their order is perceptually indeterminate. This seems to support a view of auditory stream segregation in which there are strong whole-pattern (Gestalt) effects, against any of the theories which imply that a sequence-following mechanism breaks down when the frequency jump between a successive pair of tones is too fast.

Patterns of grouping, however, are not by themselves a sufficient explanation of segregation. The speed of the tonal sequence plays a strong role, too. There seems to be a law operating such as the following: the closer two subpatterns are to one another in frequency, the higher the speed necessary to segregate them into separate streams. Thus, in the present experiment the 65-msec/tone event rate was not chosen at random, but by trial and error. If we had chosen a slower event rate (say 120 msec/tone), then, when A and X were chosen from one minor cluster and B and Y chosen from the adjacent one (absorbed condition), there would have been no grouping of AX and BY, with the corresponding segregation of the two streams. Instead, a single stream, ABXY, would have been heard, because all four tones are relatively close in frequency. However, at that same tone rate, the upper and lower major clusters *would* have segregated from one another, because they are farther apart in frequency. On the other hand, faster rates would have segregated every tone from every other one, producing the effect of four separate unrelated streams in comparison sequences. We therefore chose to use the slowest rate that would segregate adjacent ensembles.

The arbitrariness of the choice of the 65-msec tone duration does not restrict the conclusions of the present study. It merely illustrates the potent interactions of tone rate with the distribution of frequencies in inducing the formation of streams.

EXPERIMENT 2

We know that if two tones are segregated into separate streams, this will cause difficulty in judging their order (Bregman & Campbell, 1971). However, the converse is not always true: difficulty in judging the order of two tones does not always imply that they were in separate streams. In particular, an alternative interpretation, not involving stream segregation, has been suggested by an anonymous reviewer for Experiment 1: In the absorbed conditions of Experiment 1, since A and X were near in frequency, the subject might have confused X for A and thereby made an incorrect judgment of the order of A and B; similar confusions could have occurred between B and Y with a similar harmful influence on correct judgments. It seemed to the experimenter in listening to the stimuli in the absorbed conditions that this was not the problem; rather, the stimulus pattern in the comparison sequence seemed to bear no relation to the standard. The pitch interval A-B was simply missing perceptually and replaced by the intervals A-X and B-Y. Furthermore, the rhythmic pattern of the isolated and absorbed conditions were different. If elements of the two streams are represented by the digits 1 and 2, the isolated condition has the rhythm, 11221122 . . . , etc., and the absorbed condition has the rhythm 12121212 . . . , etc. However, since different listeners often give different phenomenological descriptions of the same stimuli, it seemed desirable to devise an experimental framework in which such judgments could be collected systematically. The present experiment, therefore, gathered direct judgments of the rhythmic pattern and of the perceptual grouping of tones A and B and used these to verify that the stream membership of tones A and B had indeed been influenced by the context (i.e., by tones X and Y).

An important consideration in designing Experiment 2 was to rule out any explanation in terms of the confusability of the target tones with the distractor tones. This explanation was described above. To gather evidence against it, a task involving a direct judgment of the perceptual grouping of A and B was employed. This seemed to be a task where an explanation of the absorbed condition based on a hypothesized confusion between A and X (e.g., identification of X as A) would predict different effects than would an interpretation in terms of stream segregation. The subject is asked, "Can A and B be heard as a separate pair?" in a situation where A is near in frequency to X and B is near Y. If he confuses A for X, or B for Y, he would be more likely to answer "yes," since A and B, A and X, or B and Y would all sound like the required pair. An explanation in terms of stream segregation, on the other hand, would predict that the subject would say "no," since A and B are absorbed into separate streams and all that the subject can hear is AXAX . . . , etc., or BYBY . . . , etc.

For these reasons, listeners were asked both to judge how easily A and B could be heard "as a separate pair" and to make judgments of rhythm. The isolated and absorbed conditions used a number of ABXY four-tone patterns, centered at different frequencies and extending over different ranges.

Method

Task 1: A-B isolation judgment. In this task, the subject first heard a warning knock, then, after 2 sec, as a standard, the pair AB in isolation (i.e., A, B, silence, silence, repeated 12 times). After another gap of 2 sec, he heard the four-tone ABXY sequence repeated 12 times. All of this was repeated a second time, and then a short buzz signaled to the subject that he had 8 sec to make a judgment, before the onset of the next trial.

The subject was required "to judge whether the standard pair is easily perceived as a separate pair" and to rate the ease of doing so.

Task 2: Rhythm judgment. First, two kinds of rhythm were described to the subjects and illustrated using, as elements, two kinds of hisses, J and K, one rhythm where they alternated by twos (JJKKJJKK, etc.), and one where they alternated singly (JKJKJKJK, etc.). They were told, "Do not worry about hearing pairs. Just get the feeling of the rhythm." The first type was named "uneven" and the second type, "even."

On each trial, after a warning knock and 2 sec of silence, the subject heard two rhythmic standards, the hiss patterns described above. First the "even" rhythm JKJK was repeated 12 times. Then, after a 2-sec silence, the "uneven" rhythm JJKK was repeated 12 times. Then, after a 2-sec gap, the comparison sequence of tones ABXY was repeated 12 times. (These were the same ABXY stimuli as were used in the stream judgment task.) All of this was repeated a second time, and then a short buzz signaled the subject that he had 8 sec to make a judgment before the onset of the next trial.

The listener was asked to judge "whether the pattern resembles the even or the uneven rhythm" and to rate the ease of doing so.

Rating scales. The rating scale for each trial consisted of two printed boxes, one of which was to be checked by the subject. In the stream judgment task, the boxes were labeled "yes" and "no" (the standard pair was easy or not easy to hear as a separate pair in the comparison sequence). In the rhythm task, they were labeled "even" and "uneven." To the right of the boxes there was a 7-point scale ranging from "very easy to decide" to "very hard to decide."

Stimuli. There were seven conditions involving different patternings of four tones. Two of them, A and B, were the targets for the stream judgment task, and the other two, X and Y, were distractor tones. They were presented in the order ABXYABXY..., etc., in a repeating cycle in which the four-tone pattern ABXY was repeated 12 times on each trial. At the onset, the cycle was faded in over a 1-sec interval; at the offset, it faded out over 1 sec; this helped to prevent subjects from using the first or last tones as anchor points.

The frequencies chosen for A, B, X, and Y in the seven different stimulus conditions are shown in Table 2. Each tone was a sine tone, lasting 100 msec, including a 10-msec S-shaped rise in amplitude at the onset and a 10-msec S-shaped fall in amplitude at the offset to eliminate clicks. There were 10-msec silences between tones. Hence, the onset-to-onset period was 110 msec. The two silences that replaced X and Y in the standard were each 110 msec in duration. The amplitudes of different frequencies were adjusted by trial and error to produce equal loudnesses for three judges. (The 1,000-Hz tone was measured as 70 dB SPL out of the headphones.)

The hisses used to illustrate the rhythms were each 110 msec in duration, including rise/fall times of 30 msec. One type of hiss was unfiltered white noise. The other was white noise high-pass filtered at 1,000 Hz. The unfiltered noise hiss was presented at 77 dB and the filtered noise at 62 dB SPL.

Table 2
Stimulus Conditions and Results of Experiment 2

	Target*		Distractor*		Isolated/ Embedded	Separation		Mean for Task**	
Condition	A	B	X	Y		AB†	AX††	1	2
1	2,800	1,556	600	333	I	10.2	26.7	12.03	9.86
2	600	333	2,800	1,556	I	10.2	26.7	12.69	9.96
3	2,800	2,642	1,556	1,468	I	1.0	10.2	13.16	11.69
4	333	314	600	566	I	1.0	10.2	13.76	11.72
5	2,800	1,556	2,642	1,468	E	10.2	1.0	3.71	2.96
6	600	333	566	314	E	10.2	1.0	3.84	2.99
7	2,800	600	1,468	314	E	26.7	11.2	5.21	3.96

*Frequency of tones in hertz.
**High scores indicate AB streaming.
†Separation in semitones between A and B and between X and Y.
††Separation in semitones between A and X and between B and Y.

Apparatus. The sine tone stimuli were generated by a Wavetek Model 136 VCA-VCG function generator, controlled by a PDP-11 computer. The hisses were provided by a Lafayette Instrument Co. white-noise generator, Model 15011, and filtered by a Multimetrics Model AF-520A filter. The switching of the noise signals was done by the computer via an MMC Model VCAM-4A voltage-controlled amplifier/mixer. Signals were tape-recorded on Sony 208 recording tape by an Akai GX 400 DSS tape recorder and played back to the subjects in an Industrial Acoustics Company audiometric testing room, model 1202, through Koss Pro-4AA headphones.

Design. Each subject had both tasks (stream judgment and rhythm judgment). Seven of the subjects had the stream judgment first and nine had the rhythm task first. Each task presented five blocks of trials with no separation between blocks. Each block presented the seven ABXY frequency conditions in a random order.

Subjects. Twenty-one young adults with some musical training volunteered as subjects; the data from five were discarded because they got out of step with trials on the rating scales or because they complained of fatigue.

Results

For the task in which the listeners rated whether A and B could be heard as a separate pair, the two categories YES and NO and the 7-point rating scale for ease were combined to produce a 14-point scale. This scale went from 1 for "very easy NO" to 14 for "very easy YES"; middle values corresponded to hard decisions. The average results for the seven stimulus conditions are shown in Table 2. High scores indicate AB streaming.

Similarly, the responses on the rhythm judgment task were converted to a 14-point scale, with 1 representing "very easily heard UNEVEN rhythm," 14 representing "very easily heard EVEN rhythm," and middle values corresponding to hard decisions. The average results for the seven stimulus conditions are shown in Table 2. High scores indicate AB streaming.

Both tasks showed the same pattern of results. The first four conditions, which were called "isolated" because the frequency difference between the target tones was low compared to the difference between the targets and the distractors, all produced high scores. These scores indicated that A and B were grouped into one stream and X and Y in the other. In the first task, this led to the judgment that A and B were an isolated pair. In the second task, it led to the judgment that the rhythm of ABXY was uneven (i.e., grouped by twos).

Analyses of variance were performed on the mean scores for each task taken separately. In both cases, the isolated conditions (1 to 4) were contrasted with the embedded conditions (5 to 7). The contrast was very highly significant for the AB isolation judgment [$F(1,90) = 575.5$, $p < .001$], and also for the rhythm judgment [$F(1,90) = 269.9$, $p < .001$].

Individual pairs of conditions were compared using Scheffé's method. In each task, every isolated condition was very significantly different from every embedded condition, every $F > 48.0$ and every $p < .001$. There were no other significant differences.

Discussion

Task 1. This task requires subjects to directly judge whether A and B are easily heard as a separate pair in ABXY. The significance of every post hoc comparison of isolated vs. absorbed conditions by the Scheffé test allows us to look closely at the individual conditions.

The ability to hear A and B as a pair could not be affected by their temporal proximity; they were adjacent on all conditions. Furthermore, their frequency proximity was not the sole determinant, as we can see by comparing Condition 1 with Condition 5 or comparing Condition 2 with Condition 6. In each of these comparisons, the frequency separation of A from B is held constant.

What of the hypothesis, offered as a criticism of the results from the absorbed conditions of Experiment 1, that when X was near A in frequency, the listener confused A with X and made errors of identification? In the absorbed conditions of the present task, errors of identification (i.e., accepting X as A or B as Y) would have led the listener to say that he *could* hear a pair AB when he was, in fact, hearing XB (or AY or XY). Thus, an increase in the number of "yeses" should have occurred most strongly in Conditions 5 and 6, where the interval between A and X (B and Y) was only a semitone. Yet our subjects said they could *not* hear A and B as a pair in these conditions. Hence, the problem, even at this low separation of frequencies, was not one of confusing X for A or B for Y, but arose directly from the inability to isolate the pair as a pair because of their membership in separate streams. This argues that, in Experiment 1, the results of the absorbed conditions arose from stream organization and not from confusion of frequencies. The subjects simply could not bring A and B together into a perceptual unit because they were in different streams. This accords with the experience of the experimenter who heard the two separate streams AXAX, etc., and BYBY, etc., and could not switch attention from one stream to the other fast enough to relate A to B.

Conditions 1 and 2 cast doubt on theories of stream formation which claim that segregation becomes compelling at these speeds with rather low-frequency separations. Stream splitting in Miller and Heise's experiment (1950) occurred when two tones alternating at 100 msec/tone were separated by a ratio of about 1.15 (about 2.4 semitones). The results of Van Noorden (1975) show that tones, at the rates we used (110 msec/tone), should split into separate streams at about 6 or 7 semitones' separation (p. 13, Figure 2.7) or perhaps 3 semitones (p. 15, Figure 2.9). Yet, in the present experiment, in Conditions 1 and 2, A and B were heard in one stream even though they were 10.2 semitones apart. The difference between the present experiment and those cited above is that the cited experiments alternated only two tones, A and B. When A "split away from" B, it was actually grouping with itself on its own successive repetitions. There was implicit competition between the prior As and Bs for the "privilege" of grouping with subsequent As and Bs. Since A is identical in frequency to subsequent As, it acts as very strong competition to B, and splitting (the linking of consecutive As into a stream) occurs at relatively low-frequency separations. In the present experiment, four tones were involved and the competition of groupings was present explicitly. In Conditions 1 and 2, A grouped with B at a separation of 10.2 semitones because any other grouping with A or B would have involved tones having much larger frequency separations from A and B. In Conditions 5 and 6, where A and B were also 10.2 semitones apart but where alternative groupings involved lower frequency separations, A and B split into separate streams. (Parenthetically, the reader might ask why A did not simply group with itself on its subsequent occurrences, rather than with B. The answer relates to the temporal separation of repetitions of A. The onset-to-onset time for As alone is 440 msec, too slow for A to group itself in preference to other tones.)

Discussion

Task 2. This task did not ask listeners to focus on any particular tones defined by their frequencies, but only to describe the rhythmic structure. The tones were ordered in such a way (ABXY) that if the groupings are AB and XY, the rhythm should be describable as 1122..., etc. ("uneven"); if the groupings are AX and BY, the rhythm should be 1212..., etc. ("even"). Since no recognition of definite tones is required, this task is perhaps the purest measure of perceptual grouping. In the isolated conditions, the grouping was always AB and XY; in the absorbed conditions, it was always AX and BY.

Just as in the other tasks, grouping in this task depended not on simple frequency proximity but on competing proximities. In condition 7, for example, the grouping was AX despite the fact that A and X were 11.2 semitones apart. However, in Condition 3, where A and X are only 10.2 semitones apart, A does not stream with X; it prefers to stream with B, which is much closer in frequency.

There is one final observation, which, while not significant by the Scheffé test in either task, is consistent across tasks. This is the tendency to find a greater AB grouping in Conditions 3 and 4 than in Conditions 1 and 2 and a greater AB grouping in Condition 7 than in Conditions 5 or 6. While the very conservative Scheffé test does not find these differences to be significant, ordinary F tests show 9 out of the 14 possible relevant comparisons in the two tasks to be significant at the 5% level or better. Furthermore, these differences make sense. In Conditions 3 and 4, the separation of the two target tones was lower than it was in Conditions 1 and 2, and correspondingly, these tones seemed to show a greater tendency to group. This was true despite the fact that in Conditions 3 and 4 the distractor tones were placed closer to the target tones by an even greater amount measured in semitones, i.e., on a logarithmic scale (see Table 2). Hence, it appears that moving A and B closer together by n semitones more powerfully improves the AB grouping than does moving the distractors n semitones further away. Apparently, the distances AB and AX do not compete in an additive way. Expressing the competition as a ratio between the two distances in semitones (i.e., AX/AB) predicts the correct rank order of the conditions, but predicts greater differences between Conditions 1 and 2 vs. Conditions 3 and 4, or between Conditions 5 and 6 vs. Condition 7, than were actually obtained. Such numerical predictions are doubtful, in any case, because of the arbitrary nature of the response scale. We are left with a general qualitative hypothesis that the effects of frequency proximity upon the "attraction" between tones falls off in a nonlinear way, with the addition of an n-semitone distance having a decreasing effect when the original separation is larger.

CONCLUSIONS

The two experiments, taken together, support the idea that there is a competition of alternative groupings in the formation of auditory streams. Thus, the frequency separation between a consecutive pair of tones does not directly influence grouping. All theories which rest on this assumption (probably because it can be given a simple physiological implementation), e.g., the "critical band" hypothesis of Norman (Note 1), are in conflict with the data presented above. The data are more consonant with a theory of stream segregation as the effect of unknown, but complex, physiological mechanisms which have been evolved to factor an input acoustic wavetrain so as to group those sounds which probably arose from the same source, and which take as "evidence" a variety of relationships in the acoustic waveform (Bregman, 1978; Bregman & Dannenbring, in press). These mechanisms probably have much in common with the "scene analysis" processes studied in research on computer vision (Winston, 1975).

REFERENCE NOTE

1. Norman, D. A. *Rhythmic fission: Observations on attention, temporal judgments and the critical band.* Unpublished duplicated manuscript, Harvard University, 1966.

REFERENCES

Bregman, A. S. The formation of auditory streams. In J. Requin (Ed.), *Attention and performance VII.* Hillsdale, N.J: Erlbaum, 1978.

Bregman, A. S., & **Achim**, A. Visual stream segregation. *Perception & Psychophysics*, 1973, **13**, 451-454.

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

Bregman, A. S., & **Dannenbring**, G. L. The effect of continuity on auditory stream segregation. *Perception & Psychophysics*, 1973, **13**, 308-312.

Bregman, A. S., & **Dannenbring**, G. L. Auditory streaming. *Scientific American*, in press.

Bregman, A. S., & **Rudnicky**, A. Auditory segregation: Stream or streams? *Journal of Experimental Psychology. Human Perception and Performance*, 1975, **1**, 263-267.

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

Dowling, W. J. The perception of interleaved melodies. *Cognitive Psychology*, 1973, **5**, 322-327.

Heise, G. A., & **Miller**, G. A. An experimental study of auditory patterns. *American Journal of Psychology*, 1951, **64**, 68-77.

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

Norman, D. A. Temporal confusions and limited capacity processors. *Acta Psychologica*, 1967, **27**, 293-297.

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

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

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

NOTES

1. These hypotheses were viewed by Van Noorden as working hypotheses only and he subsequently modified them (Van Noorden, 1975, Chapter 4; see also Van Noorden, 1977).

2. These streams are susceptible to further analysis. Even when small frequency separations cause a familiar tune to be in the same stream as another one, a listener, by active search, can find and hear the familiar one (Dowling, 1973, pp. 331ff). This is, in effect, the "fission boundary" phenomenon of Van Noorden (1975). Streaming does not, however, permit streams which it has segregated to be put *back together* by subsequent processes.

(Received for publication September 12, 1977;
revision accepted January 15, 1978.)