Journal of Experimental Psychology:
Human Perception and Performance
1978, Vol. 4, No. 3, 380-387

Auditory Streaming Is Cumulative

Albert S. Bregman
McGill University, Montreal, Canada

The auditory system appears to begin listening to an input with a bias toward hearing the input as a single stream, but it gradually accumulates evidence over a period of seconds which may lead to the input's being split into substreams. Several seconds of silence or of unpatterned noise slowly remove the bias of the mechanism in favor of these streams. These effects were demonstrated in experiments in which young adult listeners sped up sequences of tones until they split. The sequences varied in the number of tones packaged between recurrent "separators" (periods of silence or of white noise) and in the lengths of these separators.

If a sequence of tones of different pitches is played rapidly enough, it seems to split perceptually into two or more concurrent substreams. Subgroups of tones closely related in frequency, or following a smooth trajectory in frequency, will form part of the same stream. The splitting increases when the subgroups are farther away in frequency or when the sequence is played faster (Bregman & Campbell, 1971; Heise & Miller, 1951; Miller & Heise, 1950; Van Noorden, Note 1). The splitting phenomenon may also be observed with repeating short cycles of speech sounds (Cole & Scott, 1973; Dorman, Cutting, & Raphael, 1975; Lackner & Goldstein, 1974).

These effects are seen by the author (Bregman, 1978b; Bregman & Dannenbring, in press) as the product of an auditory "parsing" mechanism. In natural environments, the sounds emitted by different sources reach our ears, mixed together. The auditory system must group the acoustic components so as to recover the original individual sources. Only after this is done can pattern recognition processes operate with any chance of success. It would be hopeless to try to recognize a speaker's words if these were mixed with the phonemes of another concurrent speaker. We can view stream segregation, then, as a preprocessing mechanism creating auditory objects (streams) upon which pattern recognition processes can operate. One piece of evidence for the importance of streaming in pattern recognition is the following fact: When subjects hear a repeating cycle of three high (H) tones and three low (L) tones alternating in the order H L H L and so forth and the cycle splits into two perceptual streams, a high one and a low one, listeners can identify only patterns that occur in a single stream. They are unable to recognize a three-tone pattern if it crosses the streams (Bregman & Campbell, 1971). Furthermore, when a stream splits into substreams, the series of tones and silences in each substream forms a new rhythm located in that substream. This change of rhythm can be used as an index of segregation (Bregman, 1978a; Van Noorden, Note 1).

To perform an analysis into streams, the auditory system must contain rules for decomposition of the input. One such rule may be the rule of frequency proximity: If successive moments of the signal are similar

Acknowledgment is made of funding from the National Research Council of Canada and the Quebec Ministry of Education. I am grateful to Jean Beninger for research assistance. Experiment 3 was carried out as an undergraduate research project by Norman Finkelberg.

Requests for reprints should be sent to Albert S. Bregman, Department of Psychology, McGill University, 1205 McGregor Avenue, Montreal, Quebec, Canada H3A 1B1.

with respect to frequency spectrum, they should be considered part of the same source. A second may be the rule of continuity: If changes are continuous, then only one source exists. This rule has been shown to exist (Bregman & Dannenbring, 1973). Both the discontinuity and the dissimilarity of successive moments of sound can be viewed as evidence that there is more than one source involved in producing the acoustic input.

There is reason to believe that stream segregation should increase slowly with continuous listening. Human perceptual systems seem to be biased toward simple perceptions; therefore, evidence may need to be built up before the auditory system is willing to interpret an input as a product of two sources rather than one. A good piece of evidence for two streams in stream segregation experiments using a series of tones would be a bimodal distribution of frequencies of successive tones. This would indicate the existence of two sources, each one undergoing small changes in frequency. There would have to be a mechanism, therefore, that would respond, increasing strongly over time, to any accumulation of signals that were restricted to a limited region of frequencies. This mechanism would form a stream in the heavily populated frequency region and would more readily accept new inputs in that region as part of the ongoing stream. In this way, the strength of a stream could be built up over time. If there were two heavily populated frequency regions, two streams would build up strength. Any neural mechanism with these properties would have the effect of accumulating evidence about the number of streams and their positions in the frequency spectrum.

The purpose of the present experiments was to show the existence of evidence-accumulating mechanisms in stream formation, by showing that with a fixed distribution of tone frequencies, stream segregation increased over time. This was done by a process of titration: If two factors, A and B, favor stream splitting, then A's contribution can be measured by seeing how much of B must be added to split the stream. In this experiment, the splitting force of the tone distribution (A) was titrated against the splitting effects of speed (B). Speed is known to increase stream segregation; if a tone sequence moves out of a given frequency range, then the sooner it moves back into that range, the more likely the auditory system is to conclude that a separate stream exists in that range. Putting it another way, if two tone ensembles (high pitch and low pitch) alternate, then the faster they alternate, the greater the perceptual splitting (Van Noorden, Note 1). For this reason, speed could be used to measure stream-splitting effects. The lower the speed necessary to split the sequence into substreams, the greater the splitting force from some other cause. Using the method of titration, the experiments measured the increasing tendency of a pattern of tones to split into substreams as the total number of tones contained between two silent periods was increased.

One other factor had to be considered. There must be some time interval large enough so that evidence is not accumulated across that interval. Pilot experiments showed that a 4-sec interval had this property and that if a 4-sec silence was inserted into a tone sequence, this would serve to reset the stream-forming mechanism. Thus only the information packaged between 4-sec silences was expected to effect stream formation.

Experiment 1

Method

The first experiment was designed in the following way. Two different frequencies were used, two high (784 and 831 Hz) and one low (330 Hz), presented in a sequence H1, L, H2, L, H1, L, and so forth. When this sequence splits into two streams, the upper stream involves an alternation of the two high tones, and the lower stream involves the steady repetition of a single tone. Thus the two streams are easily distinguishable on a qualitative basis (one is frequency modulated and the other is not).

Four conditions were created by varying the number of tones packaged between 4-sec silences. There were 4, 8, or 16 tones in the packages, or else no silences at all. The longer packages were simply formed out of repetitions of the four-tone package. The sequence (package, silence, package, silence, etc.) repeated indefinitely while a subject adjusted the speed. The sequence started at a slow speed (600 msec per tone, onset-to-onset time), and the listener could turn a knob to speed it up until the point of splitting was determined. The subject then informed the experimenter who recorded the speed and began the next trial. The highest speed possible was 30-msec onset-to-onset time. The listener was allowed to turn the knob back and forth during the trial to zero in on the speed of splitting.

The tone sequences were generated by a Wavetek Model 136 tone generator controlled by a PDP-11 computer. Each tone consisted of a preliminary 6-msec silence, a 12-msec linear rise in amplitude, a variable length steady-state portion, and a 12-msec linear decay in amplitude. (It was actually tone duration, not silence between tones, that altered the onset-to-onset speed of the sequence.) The subjects sat in a room that was quiet but not sound-treated and heard the signal at 80-dB (SPL) through Sennheiser Model HD 414 headphones. They communicated with the experimenter through an intercom unit.

Each of the four conditions occurred five times for each subject in randomized blocks. Twelve students ranging in age from 19 to 32 years served as paid subjects, but one subject was discarded because of extremely erratic and anomalous performance. These judges received an explanation and demonstration of stream segregation and kept in view a visual illustration of one- and two-stream percepts.

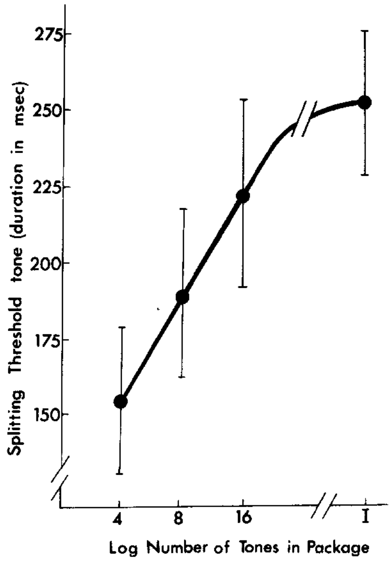

Figure 1. Speed thresholds in Experiment 1 as a function of package size (4, 8, 16, or indefinite repetition). (High values represent a greater tendency for streams to split independently of speed. Vertical bars are ± 1 *SE* of the mean.)

Results and Discussion

There were five splitting speeds given as judgments in each condition by each subject. The median of these was selected, and a mean of these medians was calculated across subjects. Figure 1 shows the results. As package size increased, the speed required for segregation decreased, or, as the figure shows, the splitting threshold measured in terms of tone duration increased. The difference across conditions was highly significant statistically by an analysis of variance, $F(3, 30) = 20.6$, $p < .001$. Furthermore, 9 of 11 subjects showed the same monotonic trend in their data, based on only five observations per condition. The other two subjects deviated from the monotonic trend in only one of their four median judgments. Thus the effect was very robust in this experiment. It should be noted that there is no artifactual "constancy" underlying the inverse relation of package size to time per tone. Neither the overall number of tones per unit of time nor the tone-to-silence ratio is held constant by the relation shown in Figure 1.

Experiment 2

In designing the previous experiment, an arbitrary decision was made that a 4-sec silence would be adequate to isolate successive packages of tones from one another. Although this was successful in producing a significant effect of package size, I did not know how critical this choice was. Accordingly, the second experiment held the number of tones in a package constant at four and varied the length of the silence between tones.

An attempt was also made to use the experiment to study how the distribution of tones can affect streaming. We know that if, in a rapid sequence of tones, there are two subsets of tones (each in a restricted range of frequency), these subsets will segregate from one another increasingly as they are moved away from each other in frequency. This can be called the *between-subsets* effect. One might also expect a *within-subsets* effect in which a narrower range of frequencies within each subset would facilitate segregation. The auditory system might be carrying out some sort of cluster analysis in which a more compact range of frequencies within a subset would allow an earlier decision that a cluster was present leading to earlier stream segregation. The present experiment, therefore, held constant the between-subset frequency separation and varied the within-subset range.

Method

Stimuli. Three patterns of tones were used (see Figure 2). All patterns consisted of a repeating cycle of four tones labeled A, B, C, and D, respectively. This corresponded roughly with the four-tone package of Experiment 1. Between repetitions of the cycle, there was a silence whose duration was 4.0 sec (as in Experiment 1), .7 sec, 1.5 sec, or 0 sec (i.e., no silence, corresponding to the indefinite repetition condition of Experiment 1).

The basic pattern, shown at the left of Figure 2, employed four frequencies, 802, 337, 901, and 318 Hz, which appeared in that order and were labeled A, B, C, and D, respectively. In this pattern, A and C are called the high set and B and D are the low set. The within-set frequency range is the separation between A and C or between B and D. In the basic pattern, A and C are separated by two semitones, and B and D are separated by one semitone. The two other patterns are variations of the basic pattern in which the within-set frequency range is increased: A and C are now separated by five semitones, and B and D are separated by three semitones.

The intention in designing this experiment was to vary the within-set frequency range while holding the separation between sets constant. But it is not clear how to define the separation between the high and low sets. Should it be the separation between the nearest two tones, A and B, or between the centers of the high and low ranges? Rather than deciding this arbitrarily, the experimenter used both definitions. In Figure 2, the frequency separation between the nearer elements of the two sets in the basic pattern is labeled 1, and the separation between the (logarithmic) centers of the high and low sets is labeled 2. The second pattern (Match-1) is matched with the basic pattern on Separation 1 (15 semitones), and the third pattern (Match-2) is matched with the basic pattern on Separation 2 (16.5 semitones). In Figure 2, the tones that were used are represented by large dots, and the logarithmic centers of the high and low ranges are represented by small dots.

Procedure and subjects. The procedure was identical to that of Experiment 1 except that the stimuli were presented through Micromonitor MX-1 electrostatic headphones. The subjects were 16 young adults ranging from 15 to 28 years of age, many of whom were familiar with stream segregation phenomena.

Results and Discussion

The means for all condition are shown in Figure 3. The length of the silence showed a highly significant effect, $F(3, 42) = 22.4$, $p < .001$. No other effects were statistically significant. The choice of 4 sec of silence to separate packages in Experiment 1 appears to have been close to optimal, since the rate of decline in the curve has slowed down considerably by this point.

It is noteworthy that the continuous tone sequence (0-sec time interval) is significantly different from the .7-sec interval by the Tukey test ($p < .05$). By the time .7 sec of silence has elapsed, the stream-forming

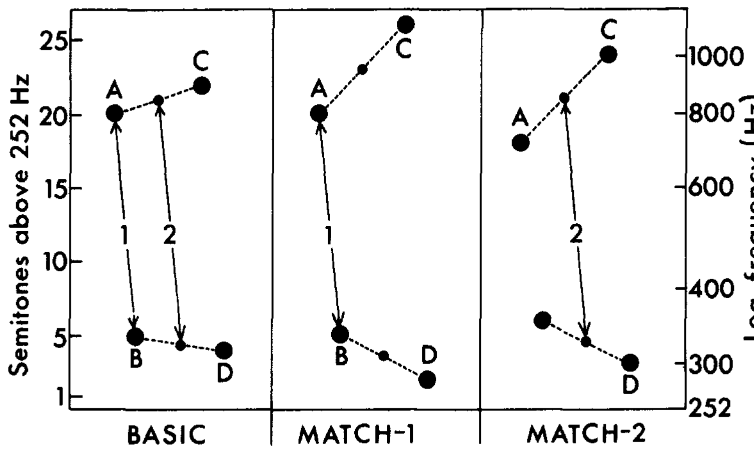

Figure 2. The 4-tone patterns used in Experiment 2. (The Match-1 and Match-2 patterns are matched with the Basic pattern on Separations 1 and 2, respectively.)

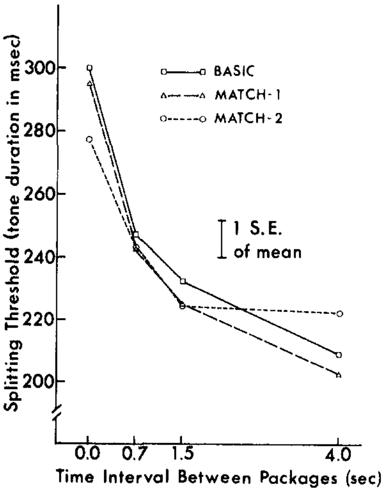

Figure 3. Speed thresholds in Experiment 2 as a function of the silent interval between 4-tone packages. (The broken lines represent the conditions with the wider within-subset range. The vertical bar is 1 *SE* of the mean.)

mechanism has reset itself by a substantial amount. But the .7-sec silence is also significantly different from the 4-sec silence by the Tukey test ($p < .05$). This means that the bias has not totally dissipated in .7 sec.

We can think in terms of the biasing and recovery of the stream-forming mechanism over time. Figure 1 shows that some number of tones, greater than 16, at about 250 msec per tone, were required to fully bias the mechanism in favor of stream splitting. By multiplying the time-per-tone threshold by the number of tones in the package, we can determine that this took something over 4 sec to occur. Figure 2, on the other hand, shows the dissipation of the bias caused by a four-tone burst which had lasted from 800 msec to a little over 1 sec, depending on the condition. This recovery continued for at least 700 msec. It is likely, however, that the course of recovery would have continued to be observable for an even longer period of time if Experiment 2 had produced stronger biasing by using longer packages of tones. I would guess that neither biasing nor recovery is completed in less than 6 sec.

The attempt to affect the rate of biasing by manipulating the within-subset frequency range was unsuccessful. The basic, Match-1, and Match-2 conditions did not differ significantly from one another, $F(2, 28) = .65$. No interactions were significant.

Experiment 3

The earlier experiments showed that a sequence of tones, located in two frequency regions, caused a streaming mechanism to cumulatively build up a bias for splitting the input into two substreams. It was found also that silence acted to unbias the mechanism. The present experiment addressed the question whether it is silence itself or merely the absence of a bimodal distribution of tones which resets the mechanism.

Method

The method was similar to that of Experiment 1 except that a package of 74-dB tones alternated with silence on some trials and with 88-dB white noise on other trials. Although white noise has quite different auditory effects than silence, it is similar to it in not having a bimodal distribution in the frequency domain.

Package sizes were 4, 8, 16, and indefinitely repeating, as in Experiment 1. The interval between packages was 1 sec in duration and consisted of white noise on half of the trials and silence on the other half. Each of the eight conditions occurred three times for each subject in randomized blocks.

Eleven university students ranging in age from 18 to 25 years performed this experiment as volunteers. The data from one subject were discarded when it was discovered that the subject had responded to the wrong criterion.

Results and Discussion

The speed thresholds (in milliseconds per tone, onset-to-onset time) are shown in Figure 4. Values for both noise and silence as separators are shown for all package sizes except the indefinite repetition (I) condition in which there were no separators at all. The value for I is based on twice as many trials as the value for each of the other conditions is. The noise and silence conditions did not differ significantly. The only statistically significant effect was that of package size, $F(3, 30) = 5.44$, $p < .01$. In Figure 4, it is evident from high thresholds in the 16-tone-package condition that bias was built up as much by the noise-bracketed packages as by the silence-bracketed ones. Yet we can see, from the low thresholds in the 4-tone-package condition that the recovery of the streaming mechanism after the noise was not much different than after silences. (Otherwise the 4-tone-package condition with the noise as separator would have built up a splitting tendency across successive packages and would have shown a threshold more like that shown by the longer packages having silences as separators.) The observed variation as a function of package size (i.e., as a function of how often 1-sec separators occurred) was actually greater with noise-filled separators. Taking sampling error into account, we can conclude that tone-free intervals have similar effects whether filled with silence or with 88-dB noise.

Let us compare the results in Figures 1, 3, and 4 for the splitting threshold for the indefinite repetition condition (0-sec silence in Figure 3), a stimulus condition that was more or less the same in the three experiments. This threshold shows a large variation in magnitude across experiments and conditions from a high of 300 msec in Figure 3 to a low of 165 msec in Figure 4. This should caution us against taking the absolute numerical values too seriously. The variation probably arises from slight differences in stimulus pattern, instructions to subjects, and signal-to-noise ratio across experiments. We should attend only to the pattern of results within each experiment and the rough order of magnitude of the numbers.

Finally, there is an issue about interpretation that has to be addressed. This article takes the view that subjects' thresholds were direct indexes of streaming and that the accumulation of evidence had a direct effect on pre-attentive organization. Another alternative was pointed out by Posner (Note 2). The accumulation of evidence may not affect streaming at all. Streaming may reach its maximum immediately. However, a second discriminative activity may have to examine the output of the streaming mechanism and to decide that two streams exist before the subject can perform the judgment task. It may be the second postulated process that requires the build-up of evidence before it

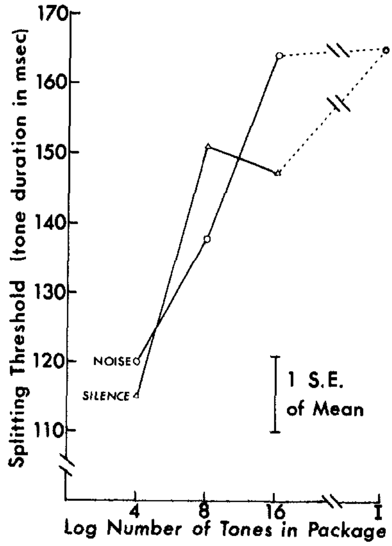

Figure 4. Speed thresholds in Experiment 3 as a function of package size with silence or white noise as separators. (The vertical bar is 1 *SE* of the mean.)

judges that streams are present (i.e., it has to "see" the already formed streams for a longer period of time). To summarize this view, the accumulation of evidence does not affect streaming itself but affects the judgments about streaming.

There is one fact in these experiments that argues against this hypothesis. Notice that when subjects judged that streaming was not occurring with short packages, they turned the speed *up* to make the sequence split. In doing so, they made the package even shorter; that is, they undoubtedly reduced the discriminability of the stimuli. Yet this produced a stronger impression of streaming. Longer packages produce a better impression of streaming only when they have been lengthened by increasing the *number* of tones, not when they have been lengthened by lengthening each tone. This seems to occur because the "information for two sources" in the signal is conveyed by a rapid transition to another frequency range and back again. As more such transitions occur, the impression of streaming builds up.

Another argument against the interpretation of the cumulative effect as arising from factors that occur after the streams are organized (i.e., factors involved in judgment) is that the results of the present experiments can be seen to fit in consistently with other ones. Existing studies show effects of speed of alternation of high and low tones on pre-attentive stream segregation (e.g., Van Noorden, Note 1). A faster alternation causes greater substream formation. This fact has been indexed by tasks other than the direct judgment of streaming (e.g., by judgments of same vs. different orders in research by Bregman & Dannenbring, 1973). The result can be exactly restated as follows: A longer elapsed (silent) time in a particular region of frequencies (e.g., high) decreases the formation of a substream in that region. This suggests that the separate high substream loses its ability to incorporate new elements as more time passes since the last appropriate element was encountered. The present Experiment 2 can simply be viewed as creating an extension of this effect by inserting a silence all the way across the frequency spectrum. Experiment 1 simply adds the idea that the tendency of the most recent element to be incorporated into a substream (thereby strengthening it) depends not only on the temporal separation of that element from its immediate predecessor but also from more distant predecessors in the same frequency range.

The concept of the build-up of strength of a stream is also consistent with the findings of Bregman and Rudnicky (1975). In that experiment, the perception of the order of two target tones (2,200 and 2,400 Hz) was made difficult by bracketing them temporally between two lower distractor tones, both at 1460 Hz. The difficulty arose from the fact that the four tones formed a single perceptual event. When this four-tone sequence was, in turn, bracketed by a captor sequence at 1,460 Hz, the captors absorbed the distractors into a separate stream, thereby isolating the targets. The order of the targets was now easily detectable. We can make contact with the present experiments by asking the following: Why, in the experiment just cited, was a captor stream required to strip the distractor tones off the target tones? Why did the distractor tones not strip themselves off to form a separate stream from the targets, regardless of the presence of the captors? (After all, the only thing done by the captor tones was to add more tones at the frequency of the distractor tones.) The answer is that in the four-tone pattern (distractors and targets alone), there was no cumulative build-up of a stream in the lower frequency range; hence the distractors and targets formed one coherent stream. Adding the captors allowed evidence to build up for a stream in the 1,460-Hz frequency range and to thereby affect the streaming of the distractors. Thus the streaming of the distractors is affected not only by their relation to the targets but by their relation to prior and later tones as well. Research in progress in this laboratory is now showing that a longer sequence of captor tones leads to better stripping off of the distractors, a fact that argues for a cumulative build-up of streams. Note that this effect is being demonstrated in a situation in which direct judgments of streaming are not required. The present experiments are simply a more direct way of studying the parameters involved. Taken alone, they do not localize the effect of evidence accumulation in the pre-attentive realm.

Conclusions

The continued occurrence of tones in a particular restricted frequency region increasingly biases a stream-forming mechanism to consider there to be an independent stream in that region. The biasing continues for more than 4 sec. Either a period of silence or of wide-band noise can gradually remove this bias. The recovery process probably also continues for more than 4 sec.

Why should this sluggishness exist when other processes in the auditory system are carried out in milliseconds if not in microseconds? This relatively slow biasing and unbiasing of the streaming process is valuable because it acts as a conservative evidence-accumulating process. Streams are not split until evidence for substreams continues for a few seconds. Similarly, our auditory systems do not assume that a substream has ceased to exist simply because it has not been heard from for one or two seconds.

This conservatism prevents the system from oscillating wildly among perceptions of various numbers of streams in a complex environment.

Reference Notes

1. Van Noorden, L. P. A. S. *Temporal coherence in the perception of tone sequences.* Doctoral dissertation, Technische Hogeschool, Eindoven. (Published by the Institute for Perception Research, P.O. Box 513, Eindhoven, The Netherlands, 1975.)
2. Posner, M. Personal communication, November 1977.

References

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

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

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.

Cole, R. A., & Scott, B. Perception of temporal order in speech: The role of vowel transitions. *Canadian Journal of Psychology,* 1973, *27,* 441–449.

Dorman, M. F., Cutting, J. E., & Raphael, L. J. Perception of temporal order in vowel sequences with and without formant transitions. *Journal of Experimental Psychology: Human Perception and Performance,* 1975, *1,* 121–129.

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

Lackner, J. R., & Goldstein, L. M. Primary auditory stream segregation of repeated consonant-vowel sequences. *Journal of the Acoustical Society of America,* 1974, *56,* 1651–1652.

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

Received September 30, 1977 ■

Errata to Curtis and Rule

In the article "Binocular Processing of Brightness Information: A Vector-Sum Model" by Dwight W. Curtis and Stanley J. Rule (*Journal of Experimental Psychology: Human Perception and Performance,* 1978, Vol. 4, No. 1, pp. 132–143), the final exponent in Equation 5 (p. 138) should be $1/2t$ rather than $(1/2)t$. The corrected Equation 5 should read as follows:

$$\Psi_{ij} = [\Psi_{Li}{}^{2t} + \Psi_{Rj}{}^{2t} + b(\Psi_{Li}\Psi_{Rj})^{t}]^{1/2t}.$$

Also, the table titles of Tables 1 and 2 (pp. 139 and 140, respectively) refer to the wrong equation numbers: Table 1 presents the parameter estimates of Equation 9 and Equation 10 (not Equations 6 and 7, as indicated), and Table 2 lists the parameter estimates for Equation 9 (not Equation 6).