Journal of Experimental Psychology: Human Perception and Performance
1975, Vol. 1, No. 3, 263–267

Auditory Segregation: Stream or Streams?

Albert S. Bregman and Alexander I. Rudnicky
McGill University, Montreal, Canada

When auditory material segregates into "streams," is the unattended stream actually organized as an entity? An affirmative answer is suggested by the observation that the organizational structure of the unattended material interacts with the structure of material to which the subject is trying to attend. Specifically, a to-be-rejected stream can, because of its structure, capture from a to-be-judged stream elements that would otherwise be acceptable members of the to-be-judged stream.

The present experiment concerns a phenomenon labeled "primary auditory stream segregation" by Bregman and Campbell (1971). When the auditory system is presented with a rapid sequence of tones varying widely in frequency, it tends to split the input into two or more perceptual streams, each incorporating tones in a narrower range of frequencies. Both the speed of the sequence and the frequency separation of the tones contribute to splitting (Van Noorden, 1971). Splitting is also affected by the smoothness of changes in pitch (Bregman & Dannenbring, 1973; Heise & Miller, 1951).

Furthermore, this splitting appears to affect pattern recognition. Patterns cutting across the two streams are not easily perceived (Bregman & Campbell, 1971). The present experiment tried to further specify the relations between stream segregation and pattern recognition. Apparently, a listener "follows" only one stream at a time; but does this mean that only one stream has been structured by processes of organization? An alternative is that a preliminary organizational process operates to decompose the input into several auditory streams and that a later process then allocates attention (pattern recognition processes) to one of these streams (Kahneman, 1973, chap. 5). In this view, the organizational process would operate automatically, using proximity relations in the pitch-time "field" to form the streams, in ways described by Gestalt psychology (e.g., Koffka, 1935).

The reasoning involved in the present experiment is as follows: We wished to show that two auditory streams were being organized at the same time. We were obliged, therefore, to employ a task in which the ability to attend to the elements of one stream is affected by the perceptual relations in (as opposed to the mere presence of) unattended material. However, this effect could not be one of interference. If elements of the unattended stream simply interfered with the attended stream, one could say that these elements were not kept out of attention but, rather, entered the single existing stream against the efforts of the subject. No organization of the unattended material would need to be postulated. We proposed, instead, to have the structure of the material that the subject was trying to ignore enable him to ignore this material more effectively. Hence if potentially distracting elements were absorbed into a coherent stream that was easy to "keep separate" from the task-related sounds, the listener's performance on the attended material should improve.

Method

The material chosen for the judgment task was a rapidly presented pair of tones, A and B, whose order was to be judged by a listener. If a person is asked to judge the order of an isolated pair of tones, he is able to do so even at very high presentation rates. Apparently, he is able to note the change in frequency between the onset and termination of the tone burst and can judge the order

This research was made possible by grants from the National Research Council of Canada and the Quebec Department of Education. Requests for reprints should be sent to A. S. Bregman, Psychology Department, McGill University, P.O. Box 6070, Montreal, Quebec, Canada H3C 3G1.

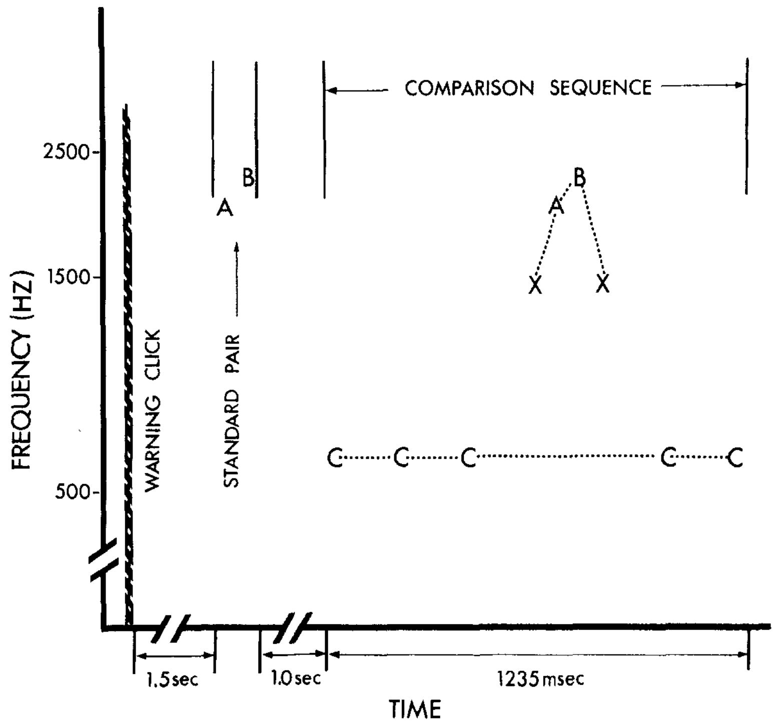

FIGURE 1. Diagram of test trial in the 590-Hz captor condition. (A and B = target tones, X = distractor tone, C = captor tone.)

on this basis. However, if we create a burst of four tones by adding two distractor tones of identical frequency (X) to the beginning and end of the tone burst, discriminating the order of the A and B elements in the sequences XABX and XBAX becomes very difficult. It is easy to see why. The four-tone pattern begins and ends with tones of the same frequency and thus the cues derived from onset and termination are no longer of any use. This effect is used in the present experiment to create a situation where a potentially easy task, the order judgment involving A and B, is made difficult by the existence of other material that is perceptually grouped with A and B. However, if we could create a "captor" stream which, by reason of frequency proximity, would capture the two Xs and strip them away, the AB judgment should become easy again. Thus with the addition of more tones to be ignored in a judgment task, that task might paradoxically be made easier.

Task

The stimuli used in the experiment are shown in Figure 1, on a logarithmic scale of frequency. Tones A and B, the target tones, were always at 2,200 Hz and 2,400 Hz respectively. The distractor tones (Xs) always occurred immediately before and after the target tone pair and had a fixed frequency of 1,460 Hz. This frequency was chosen empirically as one which would group with A and B if no other tones were present but which was far enough from the AB pair that it could be absorbed into other streams.

On experimental trials the listener first heard a warning click, followed by a pair of tones (AB), played in isolation as a standard. Then, shortly afterward, he heard a sequence of tones containing A and B. He was to judge whether they had occurred in the same order as in the standard.

There were four conditions. In one there was no captor stream. In the other three, the captor tones (indicated by Cs in Figure 1) were at 590, 1,030, or 1,460 Hz. In any one condition the captor tones were all at the same frequency. Figure 1 shows the 590-Hz captor condition. The elements are connected by dotted lines which indicate the expected perceptual organization. In the 590-Hz condition, the captor stream lay well below the XABX pattern and was expected to form an unrelated stream, leaving the XABX pattern as a separate stream. At the other extreme, the 1,460 Hz captor stream was at the same frequency as the Xs and was expected to absorb them, isolating A and B perceptually. A 1,030-Hz captor condition was also included to see whether captors at an intermediate frequency would capture the X elements. All captor conditions contrasted with the no-captor condition, which provided a baseline difficulty level for the task. In this condition the

Cs were replaced by silences, leaving all the temporal relations as in the captor conditions.

The onset-to-onset time for the tones was 65 msec. Tones in the captor stream occurred at regular intervals, separated by 130-msec silences. The distractors (Xs) fell temporally exactly where two captors would have fallen if the captor stream were completed. The target tones exactly filled the 130-msec silent interval between the two distractor tones. The exact sequence CCCXABXCC is shown in Figure 1.

Instructions

Before beginning the actual experiment, each listener was given a detailed explanation of the task. A drawing of the tones (similar to Figure 1 except for the dotted lines) was also given to him and remained in view throughout the experiment. After preliminary explanations, a sample tape was played to familiarize the listener with the nature of the stimuli. He was exposed to several repetitions of ascending pairs, descending pairs, and alternations of the two. The sample pairs were physically identical with those he would hear later. Samples of a "block mark," (a signal for the end of each page in the answer booklet) were also played at this time.

In addition to stating whether the two AB pairs in a trial were in the same or different order, the listener was asked to mark, on a 7-point scale, the degree of confidence he had in his judgment on a particular trial. This scale ranged from "not certain at all" to "very certain."

Experimental Details

The stimuli were sine tones produced by a Wavetek Model 136 tone generator, with a PDP-11 computer controlling the temporal course of frequency and amplitude. Each "tone" consisted of a preliminary 9-msec silence, a 7-msec linear growth of amplitude from zero to the maximum value for that tone, 45 msec at full loudness, and a 5-msec linear decay to zero amplitude. Thus the onset-to-onset period for successive tones was 65 msec. The maximum loudness for each tone was adjusted to equate all tones on subjective loudness as judged by the experimenter.

The signal was filtered to remove noise above 2,600 Hz and recorded on a tape recorder. The tape recorder playback during the experiment was filtered to remove frequencies under 200 Hz, mainly motor rumble and 60-Hz hum.

The listener sat in a small room for the experiment, hearing the signal binaurally through Sennheiser HD-414 headphones.

The sequence of trials was generated by combining three variables orthogonally: (a) four captor conditions, (b) same versus different AB orders in the standard and comparison sequences, and (c) AB versus BA order in the standard. This yielded 16 conditions which were randomized to produce a 16-trial block. Six such blocks were created and the order in which these blocks were presented was varied rotationally from listener to listener (e.g., 123456, 234561).

All trials and the markers separating blocks were spaced 15 sec apart on the stimulus tape. Each trial was preceded by a warning click 1.5 sec before the standard AB pair, which was then followed by the comparison sequence after a 1.0-sec silence.

Subjects

The listeners were obtained through a subject pool in the psychology department at McGill University. They ranged in age from 16 to 26 years. All were paid for their participation.

Pilot studies found that a sizable number of listeners performed at an almost perfect level, while another sizable group performed at about chance. Since neither group could give us information regarding differences among conditions, we decided in advance of the experiment to eliminate data from these groups. Accordingly, if a listener's overall performance level showed more than 90% or less than 60% correct judgments (chance is 50%), his data were put into the rejected category. Out of a total of 31 people tested, 14 were rejected because of poor performance and 4 because of near-perfect performance. Thus, the data of only 13 listeners in the midrange of performance were considered acceptable.

Results

For each trial the listener's same–different judgment and confidence rating were combined into a single rated similarity score. This was done by multiplying his confidence rating (+1 to +7) by −1 if his judgment was "different." Thus we have a scale ranging from −7 (very certain different) to +7 (very certain same).

From these scores a *D* score was calculated for each listener in each condition (Bregman & Campbell, 1971). The *D* statistic is a nonmetric measure which indicates the degree to which two conditions can be discriminated (in this case, AB vs. BA). A score of 0 indicates random judgments and a score of +1 represents perfect order discrimination. The measure is insensitive to individual differences in the use of the underlying scale or to the response-biasing effects of different conditions.

For each captor condition, the mean percent correct and the mean *D* value for the accepted subjects are presented in the upper part of Table 1. The data show that the

TABLE 1
MEAN PERFORMANCE SCORES FOR FOUR CAPTOR CONDITIONS

Subjects	Captor condition: None	590 Hz	1,030 Hz	1,460 Hz
Accepted[a]				
D	.36	.38	.60	.80
% correct	65	65	76	82
All[b]				
D	.30	.33	.41	.50
% correct	63	63	68	70

[a] $n = 13$.
[b] $n = 31$.

task is most difficult with no captor stream or when that stream is removed in frequency from the XABX pattern. It is easiest when the captors are at the same frequency as the distractors, and is intermediate in difficulty when the captors are near the XABX pattern.

An analysis of variance for a repeated-measures design was performed on the D scores for accepted subjects. A significant effect of captor condition was found, $F(3, 36) = 14.92$, $p < .001$. Comparison of the means of the different captor conditions was done using the Newman-Keuls procedure. All differences, except the one between the no-captor and 590-Hz conditions, reached significance at the .01 level or better. Thus the pattern of results visible in the upper part of Table 1 can be accepted as it appears to the eye.

To give an indication of the results with an unselected subject population, the results including the rejected subjects are shown in the lower part of Figure 1. Even with these subjects included, the same trend across conditions is seen. The main effect is significant for D scores ($p < .01$). The Newman-Keuls procedure shows differences between the 1,460-Hz condition and both the no-captor and 590-Hz conditions ($p < .01$). The differences among conditions are thus less reliable with an unselected population which includes subjects performing at lower than 60% correct (chance is 50%) or at above 90% correct overall. Some of the subjects rejected by the 60% correct criterion had actually performed at a level much lower than chance, probably due to some confusion leading to a systematic reversal of judgments. This would, no doubt, have attenuated the results more than mere noise would have done. Nonetheless, the effects of the conditions (especially on D scores) are robust enough to be visible even with the unscreened data.

Discussion

The ability to make the AB order judgment in our task clearly involves the ability to accept the AB tone pair and reject everything else for purposes of this judgment. Ignoring the notion of auditory streams for the moment, how much complexity must we postulate to exist in this process in order to explain our results?

A simple acceptance process, which tunes itself on the AB standard pair and then reacts when the pair occurs again, is not sufficient. It should operate equally effectively in all conditions. We have to postulate some process which operates to reject some of the material and which effectively isolates the AB pair in some conditions and not in others. However, a simple rejection process that simply tunes itself to the Xs which are at the same frequency on all trials) and also to material lower in frequency than X is again not adequate to explain the results. It, too, should work under all our conditions.

The differences among the conditions must depend on what happens before the first X occurs. Thus the rejection process in the captor conditions must tune itself on the early elements of the sequence; then, if the Xs are within the rejection region, but the AB pair is not, an effective rejection can take place. This hypothesis can explain some of the results. In the 590-Hz condition the rejection process is tuned too far away from the frequency of the Xs and thus misses them. In the 1,030-Hz condition the rejection range sometimes includes the Xs and in the 1,460-Hz condition it always includes them. However, this idea alone cannot explain why the no-captor condition is so difficult. Why could the rejection process not tune itself on the first X element and thus reject the Xs? The problem must lie in whatever is required to tune the rejection process. We propose the following hypothesis to deal with this problem.

Suppose that listeners have an adaptive rejection process which alters its frequency range over time. It starts with a very wide range of application; then if it encounters only tones in a narrow frequency range it narrows its range of operation. Furthermore, suppose that if it "catches" both to-be-rejected tones and target tones, some later process can ignore its output. Our results can now be explained. In the captor conditions there is enough time for the rejection process to narrow in on the captor tones. Sometimes, as a result, it rejects the Xs as well. However, in the no-captor condition it cannot narrow in on the frequency of the Xs fast enough and therefore must reject the AB pair as well. As a consequence, while some later process can ignore its output, that process still has no means of segregating the AB tones from the Xs.

Now that we have described an adequate rejection process, we must examine its relation to the acceptance process which selects the AB pair as a unit for further analysis. We propose that these are, in fact, identical processes, namely, the stream formation processes referred to in the introduction of this article. These processes have the self-tuning properties required for the proposed rejection process. By following sequential patterns of frequencies, these processes structure the auditory input into concurrent streams which other processes, which we call *attention,* can either select or reject. This implies that whenever a sequence of tones forms a unified perceptual stream, it is both easy to select for pattern recognition purposes and easier to reject as a whole without its elements intruding on another concurrent stream which is being accepted. This latter effect arises from a "mutual exclusion" property of streams. When a sound is incorporated into one stream, it tends to be unavailable to a second stream. This is the property of "belongingness" referred to by Gestalt psychology: An element cannot be both part of a figure and part of a ground at the same time. The present account differs from the Gestalt position only in suggesting that ground is not really ground; it is unattended figures. The mutual exclusion property in audition means that if a tonal element, X, fits the criteria for some rejected stream much better than those for the accepted stream, it will not intrude into the latter. Yet if no stream other than the accepted one had existed, the tonal element X might very well have been incorporated into the accepted stream (e.g., the Xs of this experiment).

If the foregoing account of the formation of parallel streams is correct, it has considerable significance for pattern recognition. Auditory stream segregation is the nervous system's attempt to decompose a complex input into the simple, separate sources which give rise to it. If successful, it permits the pattern recognition process to select or reject sources rather than particular frequencies or frequency regions for puposes of identifying familiar patterns.

REFERENCES

Bregman, A. S., & Campbell, J. Primary auditory stream segregation and perception of order in rapid sequence 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.

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

Kahneman, D. *Attention and effort.* Englewood Cliffs, N.J.: Prentice-Hall, 1973.

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

Van Noorden, L. P. A. S. Rhythmic fission as a function of tone rate. In *IPO Annual Progress Report* (No. 6). Eindhoven, Netherlands: Institute for Perception Research, 1971.

(Received September 26, 1974)