Journal of Experimental Psychology
1971, Vol. 89, No. 2, 244–249

PRIMARY AUDITORY STREAM SEGREGATION AND PERCEPTION OF ORDER IN RAPID SEQUENCES OF TONES[1] †

ALBERT S. BREGMAN[2] AND JEFFREY CAMPBELL[3]

McGill University

A recent finding of the inability of listeners to judge the order of three or four nonspeech sounds presented in a repetitive cycle is explained by the concept of stream segregation. Two experiments showed that at high presention rates of a short cycle of six tones (three high and three low), *S*s invariably segregated the tone sequences into streams based on frequency and could perceive only those patterns relating elements of the same subjective stream.

Recently, Warren, Obusek, Farmer, and Warren (1969) reported a remarkable inability of listeners to judge the order of three or four nonspeech sounds (e.g., high tone, hiss, low tone, buzz) presented repetitively in a loop. Warren et al. reported they had to lengthen each sound to about 700 msec. in duration before half of the *S*s could judge the order of sounds correctly. The difficulty *S*s encountered in their experiment may be related to another phenomenon, that of auditory stream segregation.

In the course of investigating organizational processes in the perception of rapid sequences of sounds, we have encountered a phenomenon in which a single rapid sequence of tones seems to "break up" perceptually into two or more parallel sequences, as if two or more different instruments, each restricted to a certain class of sounds or range of frequencies, were playing different but interwoven parts. We call this phenomenon primary auditory stream formation. A stream may be defined as a sequence of auditory events whose elements are related perceptually to one another, the stream being segregated perceptually from other co-occurring auditory events. We assume that attention cannot be paid to more than one such stream at a time, i.e., that the *apparent* simultaneous streams produced by this process have the same properties as *actual* simultaneous streams set to separate ears. The assumption that we pay attention to only one "ear channel" at a time has been extensively investigated and has been summarized by Neisser (1967).

Musicians are familiar with stream formation under the names of "implied polyphony" or "compound melodic line" (Bukofzer, 1947, p. 289; Piston, 1947, p. 23), where a single instrument, by alternating high and low tones, gives the effect of two instruments playing. Plentiful examples are found in the sonatas for violin and cello solo, in the violin concerti, and other works of Bach.

Miller and Heise (1950) have investigated auditory stream segregation under the name "trill threshold," which they studied using two repeatedly alternating, 100-msec. sine-wave tones. Miller and Heise noted that the splitting of the signal into two streams depends upon the rate of alternation and upon the difference in frequency between the two tones and found that stream splitting could be obtained with as little as a 15% difference in frequency ($\Delta F/F \times 100$) and could be obtained throughout the frequency range from about 150 Hz. to 7,000 Hz. Our own observations in experiments not reported here have confirmed the effects of frequency difference and presentation rate and also suggested an interaction between the two. The higher the presentation rate, the less the frequency difference required for stream splitting.

The two following experiments relating order perception to stream segregation bear

[1] This research was supported in part by Defense Research Board of Canada Grant 9401-40 and by National Research Council of Canada Grant APA-127. Thanks are due to M. C. Corballis for statistical advice and to Geoffrey Selig for assistance in data gathering and analysis.

[2] Requests for reprints should be sent to Albert S. Bregman, Department of Psychology, McGill University, Montreal 110, Quebec, Canada.

[3] Now at Loyola College, Montreal, Canada.

a direct relevance to the studies of Warren et al. (1969).

Experiment I

In this experiment, we hypothesized that the perception of order of very rapid events would be restricted to events in the same apparent stream. Therefore, we made up tape loops containing a sequence of six sine-wave tones, three from a high-frequency range and three from a low range. It was hypothesized that *S* would be able to judge order relationships only among the three high tones or among the three low tones, but would not be able to relate the temporal positions of high tones to those of low tones.

Method

Six different sine-wave tones appeared once each on a tape loop. Their frequencies were 2,500, 2,000, 1,600, 550, 430, and 350 Hz., and they were labeled A, B, C, D, E, and F, respectively. Each tone was of 100-msec. duration. The first three were considered to be members of a high-tone ensemble (H) and the second three to be from a low-tone ensemble (L). Two kinds of tapes were made, varying the arrangements of high and low tones. In Cond. 1, the arrangement was HLHLHL and in Cond. 2 it was HHLLHL. Cond. 2 was included so that effects of stream segregation on order perception might be seen to be independent of how the tones from the same frequency subset were spaced on the tape. The assignment of Tones A, B, and C to different H positions and of D, E, and F to different L positions was counterbalanced by creating six different tapes for each condition. These were produced by splicing together 100-msec. sections of magnetic tape, square cut for accurate timing.

There were 16 *S*s per condition, tested individually. Each was given practice in listening to the tones and identifying them with the letters A to F. A practice tape loop was constructed with the stimuli in the order A B C D E F and each tone at 300-msec. duration. The *S*s listened to this until they were able to write down the correct order of tones. The letters A to F alongside an arrow indicating the descent of the tones from highest to lowest were written on their answer sheets as a further guide. The *S*s were instructed to listen to each tape for as long as they wished and then to write down the order of the six tones. They were encouraged to sing, beat time, write, or in any other way aid themselves in discovering the order of the tones. Order of presentation of the six tapes per condition was counterbalanced across *S*s.

Subjects.—The *S*s were 32 male and female volunteers from a student population at McGill University.

Results

Since the tones had been divided into triplet subsets, the preservation of order relationships in the judgments was scored on the basis of triplets of tones. On each tape, taking combinations of three tones out of six, there are 20 distinct triplet combinations. Of these, only 2, ABC and DEF, are within-stream triplets. The remaining 18 triplets relate items from both streams; these are called across-stream triplets. Each *S*, therefore, could be scored on a total of 12 within-stream triplets and 108 across-stream triplets on the six tapes which he judged. If, in his written recall, the three tones of a triplet appeared in the same order as on the tape loop, in any of its three rotational equivalents (e.g., ABF, FAB, or BFA), it was scored as correct. The chance probability for any one triplet of tones being correct if scored in this manner is .50, but the 20 triplets for each tape are not independent. However, by a principle of indifference, the chance expected value of within-stream triplets should be the same as for across-stream triplets, and the truth of this hypothesis can be evaluated using rank-order statistics. Each *S* received two scores, W, the percent correct of within-stream triplets, and A, the percent correct of across-stream triplets. In Cond. 1, the mean value of W was 73.4 (SD = 17.3), while the mean for A was only 60.1 (SD = 11.1). In Cond. 2, the mean value of W was 66.1 (SD = 13.4) and of A was 55.2 (SD = 4.7). The difference between the within- and across-stream scores for each condition was statistically significant at beyond the .001 level by the Wilcoxon test. The differences in percent correct between Cond. 1 and 2 were not significant either for within-stream triplets, for between-stream triplets, or total triplets, by White's modification of the Mann-Whitney test (White, 1952). More important, it can be seen that the magnitude of the difference between the within- and across-stream scores is about the same in the two conditions. There is a consistent superiority of within-stream judgments independently of how stimuli from the two classes are distributed in the loop.

While the statistical analysis was unbiased as to favoring within- or across-stream triplets, the observed percents correct cannot be directly related to the relative difficulty of the two types of judgments. This arises because the percents correct for within- and across-stream triplets are not statistically independent. As one goes up, so does the expected chance value of the other, although not in the same degree.

There is an additional important observation in this experiment, however, suggesting that the success in within-stream judgments was primary and that the observed success in obtaining greater than the .50 chance value on the across-stream triplets is either a statistical consequence of the within-stream judgments or is a different type of judgment entirely. This observation is that every *S* reported the items in a stream-by-stream order. That is, the listener first wrote down the items of one stream (H or L) and then filled in the items of the other stream (sometimes inserting them between items of the first stream). In addition, 59% of all judgments actually claimed that the items were in the orders HHHLLL or LLLHHH on the tape. These orders never occurred on the tapes and would be expected from random guessing only 30% of the time. Such a segregation of the items suggests a complete inability to relate items in the two streams. It appears, then, that the listeners organized the material into two subjective substreams, made order judgments within each one relatively successfully (about 70% correct), and then tried to relate the two, achieving a lower degree of success (about 55%), perhaps due largely to chance.

Experiment II

This experiment was done for three reasons: (*a*) to eliminate the statistical dependency between the within- and across-stream judgments, (*b*) to eliminate the necessity for *S*s to remember labels for the tones, and (*c*) to demonstrate a complete inability to relate elements across streams, even with an extremely sensitive recognition measure.

Method

Each *S* listened to two tape loops on each trial. One loop was a standard (ST) containing three tones and three silent gaps. The other was a comparison loop (CO) in which the former silent gaps were filled by the three tones not used in the standard tape. This comparison tape always contained six tones, three H and three L. The *S*s judged whether the three tones of the ST occurred in the same order and temporal spacing in the CO. The ST loops contained either within-stream triplets (three tones from the same frequency range) or across-stream triplets (two tones from one range, H or L, and the third tone from the other range). The high-frequency (H) tones were 2,500, 2,000, and 1,600 Hz.; the low-frequency (L) tones were 550, 430, and 350 Hz. Tone durations (and silence durations on the ST tapes) were 100 msec.

On each trial, *S* heard a warning tone, then 5 sec. of ST, 5 sec. of CO, 5 sec. of ST again, and 5 sec. of CO again. Then he made his judgment. The CO always contained the three tones of the ST in it, but not necessarily in the same order. The *S* registered his judgment of sameness on a continuous 100-mm. rating scale marked at the two ends with the labels "same" and "different." It was explained to *S* that a rating near the center of the scale expressed a lack of confidence in the judgment, whereas the two extremes reflected complete confidence.

Stimulus sequences for both ST and CO loops were constructed by splicing together 100-msec. segments of magnetic tape, square cut for accurate timing.

Four conditions were constructed by generating combinations of two variables: (*a*) type of triplet—within-stream (WIT) or across-stream (AC_1) and (*b*) spacing on the tape loop—balanced (BAL) or unbalanced (UNB).

The BAL standards were constructed with a symmetrical arrangement of tones, i.e., tone–silence–tone–silence–tone–silence, and UNB standards were constructed with an asymmetrical arrangement of tones, i.e., tone–tone–silence–tone–silence–silence. As in Exp. I, the BAL versus UNB conditions were included in order to establish whether the spacing of high and low elements in the cycle would affect stream segregation. Hence, there were four conditions: WIT BAL, WIT UNB, AC_1 BAL, and AC_1 UNB. Within each condition, half of the stimuli required the response "same," and half, the response "different." A CO that was the same as the ST contained the three tones of the ST in the same order and spacing. A CO that was different from the ST had the order of the three tones reversed but retained the original spacing. Occurrences of particular tones in particular positions and conditions were randomly arranged. Each of the four conditions was tested 24 times with each *S*.

In discussing the design of the AC_1 ST triplets, when constructed as outlined above, with U. Neisser (Personal communication, April 1970), it was pointed out to us that there exists a strategy

whereby *S*s could tell that an AC_1 triplet had been changed in the CO even if he could not relate tones from the two ensembles. To see why, let us assign the names A and B to the two same-ensemble (H or L) tones and give the name X to the single tone from the other ensemble. Suppose that *S* can listen only to the order of A and B. Now the original ST may be A–silence–X–silence–B–silence (repetitively). Notice that in the forward direction, B's onset follows A's offset by 300 msec., while A's onset follows B's offset by only 100 msec. This will lead to a particular rhythmic relation between A and B which is changed when the AXB triplet is reversed to make a different CO. Hence, if *S*, despite instructions that he is to listen for the order of three events, listens to the rhythm of two of them, he can detect a different CO. We call this the two-tone strategy.

To check whether *S*s were employing this strategy and to eliminate its effects if they were, we constructed another type of across-stream triplet where the two-tone strategy could not work. In this type, called AC_2, the ST triplet was of the form A–silence–X–B–silence–silence. The CO sequences, when "different," were of the form: A–tone–tone–B–X–tone. In this case, the CO keeps the same rhythmic relation between A and B but shifts X from a position inside the A-B interval to a position outside it. Hence this condition was labeled AC_2 IN. A second similar condition shifted X from a position outside the A-B interval in the ST to inside it in the CO. This was referred to as AC_2 OUT. Each of these two conditions appeared 24 times for each *S*, 12 same and 12 different.

The experiment consisted of 144 trials, split into blocks of 24; each of the six conditions appeared four times in each block. The *S*s were given rests between blocks.

Subjects.—The *S*s were 21 summer course students at McGill University, paid for their participation.

Results

The *S*'s protocols were scored by measuring the distance in millimeters along the response scale in the direction of sameness. This is the raw measure, rated similarity (RS). A dependent variable, D, was calculated for each *S* in each condition, e.g., in WIT BAL. The measure D reflects the degree to which *S*s could discriminate same from different stimulus pairs (CO and ST) in that condition. That is, it compares the physical similarity to the rated similarity, assigning high scores when these two correspond. To obtain D, first the RS ratings for both physically same and physically different stimulus pairs are ranked together. If there were complete discrimination, the ranks of RS scores for all physically same pairs should be ahead of (i.e., have a lower numerical value than) the ranks of the RS scores for all physically different pairs; in other words, there should be no overlap in the two distributions. The D measure shows the degree of overlap between similarity ratings for physically same and physically different pairs. The following formula defines D:

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

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.

This statistic takes on the value + 1.00 when all physically same pairs of stimuli are higher on RS than are all the physically different pairs. It takes the value zero when judgments are random. When judgments are systematically reversed (i.e., all physically same comparison are lower on RS than physically different ones), it takes the value − 1.00. The calculation of ranks separately for each *S* in each condition prevents certain kinds of response bias from creating differences between *S*s or conditions. The procedure eliminates any source of response bias that does not affect the ordering of the judgments on the 100-mm. sameness scale. Such response factors as a general shift toward same or a shift in the overall range of judgments should be eliminated. The measure simply assumes that the overlap of ranks of judgments for same and different stimuli is a distribution-free measure related monotonically to discrimination.

Its advantage over the nontheoretical use of the signal detection d′ measure is that it makes none of the assumptions inherent in the latter measure and yet would appear to be equally insensitive to response bias. In addition, D is easy to calculate and is usable when only a small number of responses are obtained in each condition.

The mean D scores across *S*s for each of the six experimental conditions and for halves of the experiment are given in Table 1. The WIT conditions produced a highly

TABLE 1

MEAN D VALUES, EXPERIMENT II

Cond.	First half	Second half	All trials
WIT BAL	.72	.73	.73
WIT UNB	.82	.76	.79
AC_1 BAL	.25	.13	.19
AC_1 UNB	.08	.16	.12
AC_2 IN	−.01	−.06	−.04
AC_2 OUT	.08	−.10	−.01

skewed distribution (with many D scores of 1.00), and, therefore, in comparing these conditions with others, parametric statistics were inappropriate. Instead, the Wilcoxon test was used. Conditions WIT BAL and WIT UNB, combined, were compared with AC_1 BAL and AC_1 UNB combined. These were significantly different at beyond the .001 level. The comparison of all WIT conditions combined, with all AC conditions (AC_1 and AC_2) combined, also reveals a highly significant difference ($p < .001$). Every *S* attained a higher mean D score on the WIT conditions combined than on the AC conditions combined. The WIT BAL and WIT UNB means were not significantly different from one another, and first-half WIT performance showed no significant difference from second-half performance. Combining all six experimental conditions, there was no significant difference between performance in the first and second halves of the experiment.

The AC conditions, AC_1 BAL, AC_1 UNB, AC_2 IN, and AC_2 OUT for the two halves of the experiment were compared by analysis of variance. Only the effects of type of AC (AC_1 versus AC_2) were significant, $F(1, 20) = 12.33$, $p < .005$, indicating that *S*s were able to use the "two-tone strategy" in the AC_1 conditions.

In separate analyses, it was ascertained that there were no significant differences between AC_1 BAL and AC_1 UNB or between AC_2 IN and AC_2 OUT. This result and the earlier comparison of WIT BAL and WIT UNB shows that the order of tones, per se, in a sequence, has no significant effect on performance.

DISCUSSION

The main results are quite striking. The within-stream judgments showed a high degree of accuracy, while the AC_2 condition (in which the two-tone strategy was impossible) showed chance level performance. The mean value of D across all AC_2 conditions was −.023, which is extremely close to the expected chance value of zero. Thus, we have been able to show that at the rates used, there is essentially no ability to relate material from different streams. We conclude that the apparent success *S*s had in Exp. I with across-stream triplets arose from the statistical interdependence of within- and across-stream triplets.

The comparison of AC_1 and AC_2 conditions shows that some *S*s, at least, were able, to some small degree, to detect changes in the temporal pattern of two tones providing that they were in the same subjective stream. This capability, plus the relatively high performance on WIT comparisons shows that the speeds involved were not too high for accurate order judgments provided that the comparison restricted itself to elements of a single stream. Thus, the shifting of attention from stream to stream, rather than the comparison process itself, constituted the time-limited process in the present experiment.

Returning now to consider the experiment of Warren et al. (1969), we propose to explain the low performance of their *S*s as a stream segregation effect. When three unrelated sounds (e.g., high tone, low tone, hiss) are presented repetitively in a loop, this generates three streams. Each of the sounds groups with its own prior and subsequent repetitions, rather than with the other two sounds. Listeners cannot switch their attention from stream to stream fast enough to make the necessary order judgments.

The greater ability to make temporal order judgments in loops containing four spoken digits, as reported by Warren et al. (1969), simply implies that a sequence of speech sounds constitutes a unitary stream for the auditory system. Why should this be? There are indeed subsets of vocal sounds that might form similarity groups, e.g., fricatives, vowels, stops, etc. The vocal sound stream may not split into substreams because splitting depends not only on similarities in the component sounds but also on the nature of the transition from sound to sound. In our laboratory, we have noticed that when the frequency glides gradually (though quickly)

from tone to tone, there is less tendency for the sequence to split than with instantaneous transitions. The transitions in speech are not instantaneous.

Finally, we would like to attempt to relate our work to the work on sensory "channels" stimulated by Broadbent's (1958) theorizing and summarized by Neisser (1967). The distinction between a stream and a sensory channel is that a stream is an organizational entity and is not definable by any single physical property. We created streams in this experiment by segregating two subsets of tones by frequency. This was only for the sake of convenience. We believe that we could cause streams to cross one another in frequency if we maintained the integrity of each one by smooth transitions as in Fig. 1. It would be expected, in such a case, that *S* might fail to relate Tone X and Tone Y, which are close together in frequency but in different streams. If such is the case, no theory employing filters that operate so as to block out signals with some specified attribute (other than the trivially defined attribute of being in Stream A) could account for such an effect.

Another difference between the present research and that on dichotic channels, for example, is the compelling degree of segregation of streams in the present case. While *S*s in dichotic experiments prefer to group stimuli by channels and can shut out material from an unattended channel, there has been no demonstration of a complete inability to relate material in the two ear channels such as we see in the present experiments.

A third difference is that in the typical study of divided attention, stimuli from different ear channels co-occur in time; and hence some competition of attention is forced on *S*. When material in the two channels alternates rather than coincides in time (Moray, 1960), *S*s prefer to report material in true sequential order. In the present study, material alternates in two frequency ranges. Yet *S*s report in a stream-by-stream order, cannot identify sequences that cross the organizational streams, and have no clear idea of temporal relations between the two streams.

A fourth difference is in the scale of time. The present stimuli are coming 5 to 10 times faster than those in divided attention studies.

The present authors propose that the organizational process that creates auditory streams is distinct from the limited-capacity channel proposed by Broadbent (1958). The role of the former is to preorganize the material

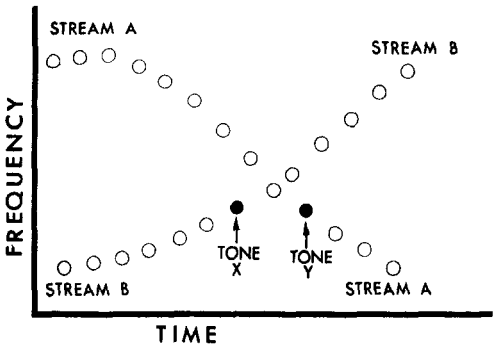

FIG. 1. Two streams, segregated by frequency, crossing each other.

and extract higher order perceptual attributes. Then the limited-capacity channel, or processor, can select and process material from one such stream at a time. The stream-forming processes described in this paper probably fall into the category of "preattentive processes" discussed by Neisser (1967). It is expected that their effects will be of the sort described by Gestalt psychologists (e.g., Köhler, 1947). While such concepts as "figure-ground segregation," "common fate," and "good form" are not yet reducible to elegant mathematical rules, they will undoubtedly serve heuristic functions in the study of stream segregation and other aspects of sequence perception.

REFERENCES

BROADBENT, D. E. *Perception and communication.* New York: Pergamon Press, 1958.

BUKOFZER, M. *Music in the Baroque era.* New York: W. W. Norton, 1947.

KÖHLER, W. *Gestalt psychology.* New York: Liveright, 1947.

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

MORAY, N. Broadbent's filter theory: Postulate H and the problem of switching time. *Quarterly Journal of Experimental Psychology,* 1960, **12**, 214–220.

NEISSER, U. *Cognitive psychology.* New York: Appleton-Century-Crofts, 1967.

PISTON, W. *Counterpoint.* New York: W. W. Norton, 1947.

WARREN, R. M., OBUSEK, C. J., FARMER, R. M., & WARREN, R. P. Auditory sequence: Confusion of patterns other than speech or music. *Science,* 1969, **164**, 586–587.

WHITE, C. The use of ranks in a test of significance for comparing two treatments. *Biometrics,* 1952, **8**, 33–41.

(Received October 24, 1970)