Perception & Psychophysics
1973, Vol. 13, No. 2, 308-312

The effect of continuity on auditory stream segregation*

ALBERT S. BREGMAN† and GARY L. DANNENBRING††
McGill University, Montreal, Quebec, Canada

A rapid, repeating cycle of alternating high and low tones was presented under three conditions. In the "discrete" condition, transitions between tones were abrupt; in the "ramped" condition, successive tones were connected by frequency glides. In the "semiramped" condition, there were partial glides in frequency (as in speech). "Discrete" sequences were most likely to split perceptually into high and low streams, making order discriminations difficult. The "ramped" condition was least likely to split, and order perception was easiest. Results for the "semiramped" condition were intermediate. The discussion relates these findings to the acoustic properties of speech and to the process of auditory stream formation

Not long ago, Warren, Obusek, Farmer, and Warren (1969) reported what they perceived as a remarkable inability of unpracticed human Ss to make a judgment of order. The stimulus was a repeating cycle of four sounds: hiss, buzz, sine tone, and vowel, each lasting 200 msec. Few Ss could name the order of the sounds. This struck them as puzzling in view of the already known capacity of the auditory system to discriminate the order of sounds at a much higher rate. For example, in normal English speech, phonemes occur more quickly—80-100 msec per phoneme (Efron, 1963). Speech can be reported correctly at rates as fast as 30 msec per phoneme (Foulke & Sticht, 1969). Winckel (1967) reports that temporal order of musical notes is resolvable down to about 50 msec per note. Warren et al (1969) found that the order of short repeating sequences of spoken digits was reported more accurately than unrelated sounds presented at the same rate, and concluded that verbal sounds were related in some fashion that permitted more rapid perceptual following. There is some confirmation for this in the work of Thomas, Hill, Carroll, and Garcia (1970), who found that the order of a repeating cycle of four vowel segments spliced together could be correctly identified at 125 msec per segment (but not at 100 msec per segment). Yet, even this rate is not as fast as we can discriminate the order of sounds in speech.

Bregman and Campbell (1971) have proposed that the difficulty experienced by Ss in experiments using repeating cycles of sounds is due to the fact that subsets of the sounds group into separate perceptual streams and that the judgment of fine order relations across streams is impossible. Using sine tones, they created two subsets based on frequency differences, using a high set (2,500, 2,000, and 1,600 Hz) and a low set (550, 430, and 350 Hz). In a repeating cycle of mixed high and low tones, Ss could discriminate the order of the high tones relative to one another or of the low tones among themselves, but could not order the high tones relative to the low ones. They proposed that the auditory system creates substreams of sound whenever sounds of different types occur rapidly in a mixed sequence. This process was referred to as "primary auditory stream segregation" (PASS). They further proposed that continuity of the acoustic properties of successive moments of sound might be a factor tying the successive sounds of speech into a single stream. They pointed out that speech is not composed of discrete segments of sound as in the experiments with repeated cycles. As Liberman (1970) points out, the individual phonemes connect, overlap, and dissolve into one another.

The present experiments were conducted to assess the role of acoustic continuity in PASS. It was hypothesized that stream segregation would be reduced when, in a sequence of alternating high and low tones, there was a frequency glide joining successive tones. This hypothesis was tested in two experiments. In the first, Ss were required to make a judgment of same or different orders for two sequences, each containing high and low tones. In the second, Ss judged directly whether a sequence of two high and two low tones did or did not split into substreams.

EXPERIMENT I

On each trial, Ss were presented with two auditory sequences, a "standard" and a "comparison," each consisting of a repeating pattern of four tones: two high tones and two low tones. They were required to decide whether the order of the tones in the standard was the same as or different from the order of the tones in the comparison. The patterns varied according to the

*This research was supported by Grant 9401-40 from the Defence Research Board of Canada to A. S. Bregman, and is based in part on a master's thesis submitted to McGill University by G. Dannenbring. We would like to thank the Montreal Neurological Institute for the use of the PDP-12 computer, and Esther Benezra for help in experimentation.

†Requests for reprints should be sent to A. S. Bregman, Psychology Department, McGill University, Montreal, Canada.

††The authorship of this article is to be considered as equal. The order of names was determined by a coin toss.

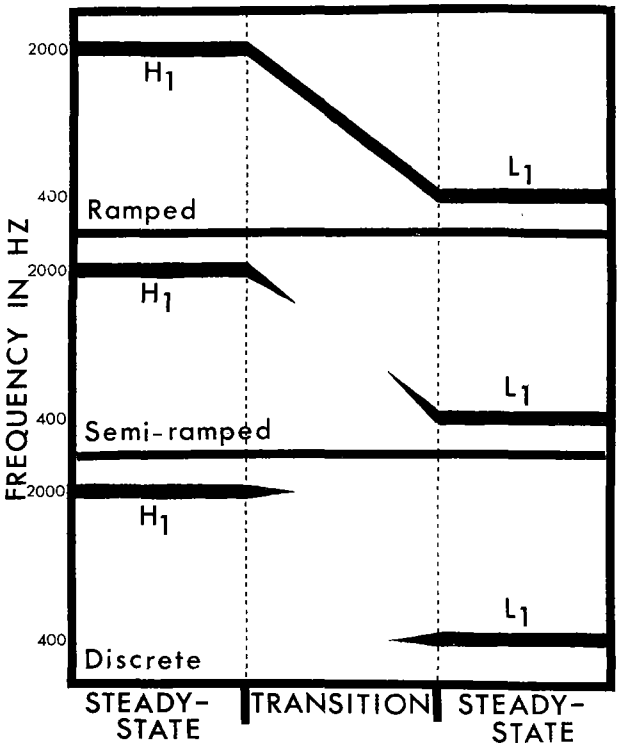

Fig. 1. Examples of the transitions between H_1 (2,000 Hz) and L_1 (614 Hz), showing the three types of ramping conditions. (The width of the tones illustrates amplitude, with a narrowing of the width indicating a drop in amplitude. The total duration of the transition between the steady-state tones is 40 msec.)

duration of the tones and the type of transition between the tones. It was expected that when the tones were discrete, with spaces between them, and presented at a rapid rate of speed, Ss would be unable to follow the pattern; i.e., the auditory stream would perceptually split, with the high tones being heard as a pattern separate from the low tones. When the tones were connected by a gliding, continuous frequency change, it was expected that Ss would be able to follow a rapidly presented pattern better than when the tones were separate and discrete. Performance in a third condition, in which the frequency glided part of the way between the tones, was expected to fall somewhere between the other two conditions. It was thought that Ss would perform well in all conditions when the tones were presented at a slow rate of speed; thus, an interaction was expected between rate of presentation and the type of transition between the tones.

Method

Apparatus

Stimulus materials were generated by means of a PDP-12 computer (Digital Equipment Corp.) operating a Wavetek (Model 136 VCG/VCA) signal generator, which produced the sine tones used in the experiment. Four parameters were used to specify tones for the computer: frequency, amplitude, transition (or ramp) time, and steady-state time. The tones were adjusted to a subjective equal loudness by E. Sequences were specified as a series of tones, which the computer repeated until instructed to stop. Tonal sequences were recorded with a Sony (TC-200) tape recorder, with noise above 2,000 Hz and below 400 Hz being filtered out with a Krohn-Hite (Model 310-C) bandpass filter. These sequences were then presented to Ss in a small room by means of the Sony tape-recorder loudspeakers (75-85 dB at the S's ears).

Auditory Sequences and Design

Two auditory sequences, a "standard" and a "comparison" were presented twice each on each trial as follows: standard, comparison, standard, comparison. Each was a 5-sec repeating cycle of two high (H) and two low (L) tones, e.g., H_1 L_1 H_2 L_2 H_1 L_1 H_2 L_2 ... (5 sec). The tones of each pattern always alternated between high and low. Since the pattern within each sequence was repetitive, there were only two possible cyclic orders in which the tones could be presented: Order 1, H_1 L_1 H_2 L_2, and Order 2, H_1 L_2 H_2 L_1. All of the patterns consisted of the same four "steady-state" tones: two high tones (2,000 and 1,600 Hz) and two low ones (614 and 400 Hz). Their frequencies were selected so that the separation between H_1 and H_2 and between L_1 and L_2 was 200 mels (Pedersen, 1965). In one half of the trials, the tones of the comparison sequence were presented in the same order as the standard; in the other half of the trials, they were presented in the opposite order.

If the auditory sequences segregated into two streams based on frequency, both the standard and comparison sequences would sound the same, with one stream heard as the repeating pattern H_1-H_2- and the other as the repeating pattern L_1-L_2-, despite the actual order of the tones. Hence, auditory stream segregation would perceptually destroy the actual difference between Order 1 and Order 2, resulting in an inability on the part of S to discriminate between the two.

The patterns varied according to two conditions: ramping condition, in which the transition between the tones was varied, and steady-state time, in which the length of the tones was varied. Three different ramping conditions were used, and these are shown in Fig. 1. In the following descriptions, all ramped changes in frequency or amplitude are linear. In the *ramped* condition, both frequency and amplitude changed from one tone to the next over the 40-msec break between the steady-state time of the tones. In the *semiramped* condition, at the end of the steady-state part of a tone, there was a 10-msec 45-dB fall in amplitude, along with a change of frequency part of the way toward the next tone; for the following tone (after a 20-msec "silence"), the first 10 msec consisted of a resumption of the final part of the glide transition, together with a 45-dB rise in amplitude. There was a 20-msec space between the tones, which was attenuated 45 dB and subjectively silent because of tape and background noise. In the *discrete* condition, there was a 10-msec change in *amplitude only* at the beginning and end of a tone, with a 20-msec silence between the tones. There were three different steady-state times of the tones used in this experiment: 100, 150, and 225 msec.

Thirty-six different tests were constructed. Each appeared twice in a total series of 72 trials. The 36 were generated as follows: Nine classes of tests were generated by factorial combinations of three ramping conditions and three steady-state times. The standard and the comparison sequence on any test were always the same with respect to these two variables. Two other variables were added factorially: two levels of similarity of standard and comparison (same or different) and two orders for the standard (Order 1 or 2). The tests were blocked into four sets of 18 trials by a combination of randomization and counterbalancing. All Ss then received all 72 tests in the same order, with a short break following Trial 36.

All Ss were instructed to indicate on a response sheet whether the order of the tones in the comparison sequence sounded the same as or different from the order of the tones in the standard sequence. They were also asked to indicate, by putting a mark on a 100-mm line with the extremes labeled "not at all confident" and "very confident," how confident they were that they had made the correct response.

Pretest

Ss also took a pretest prior to the actual experiment to eliminate those Ss who had extreme difficulty in determining the order of pairs of tones. Each trial of the pretest consisted of four presentations of pairs of tones, with 110 msec between the pairs. Each trial was presented in the same manner as the experimental trials, with the first sequence being the standard, the second the comparison, and the two then presented again. Ss had to decide whether the order of the tones in the comparison sequence was the same as or different from the standard.

Each tone in the pretest lasted 25 msec. The transitions between the tones were constructed in the same manner as the discrete condition of the experiment, with a 10-msec drop in amplitude only at the end of each tone, a 10-msec rise in amplitude at the beginning of a tone, and a 20-msec silence between the two tones of a pair. There were eight trials in the pretest. The frequencies of the tone pairs used, in the order presented, were (in hertz): 2,760-1,525, 3,400-1,860, 373-200, 200-455, 1,525-3,400, 2,760-1,860, 246-373, and 3,400-1,860. In Trials 1, 4, 5, and 8, the tones were presented in the same order in the comparison pattern as in the standard; for the remaining trials, they were presented in reverse order. Those persons who failed to meet a criterion of six out of eight correct in the pretest were not used as Ss in the experiment.

Subjects

The Ss were 36 graduate and undergraduate students at McGill University who were naive as to the purposes of the experiment. Sixteen Ss who failed to meet the criterion of two errors or less in the pretest were eliminated from the experiment.

Results

The confidence rating given for each trial by Ss was treated as a measure of similarity when S checked the "same" box and as a measure of difference when he checked the "different" box. These two scales were then treated as a single, continuous scale, the two scales being placed end to end with "not at all confident" being the midpoint. Thus, this continuous scale ranged from "very confident different," through "not at all confident" (indicating that S made a guess between "same" and "different"), to "very confident same." The raw measure for each trial was the distance along this scale in millimeters. This was the "rated similarity" (RS), which thus took a value of 0 for "very confident different," 100 for "not at all confident," and 200 for "very confident same," with points in between depending upon the confidence rating for that trial. A dependent variable, D, developed by Bregman and Campbell (1971) for scoring responses of this type, was calculated for each condition on each S. This is a nonmetric measure of the degree to which Ss could discriminate "same" from "different" stimulus pairs. To obtain D, the RS for both the physically same and physically different trials within a cell were ranked together. These rankings were then used in the following equation: $D = [2(M_d - M_s)]/N$, where M_d = the mean of the ranks of the RS for the physically different trials, M_s = the mean of the ranks of the RS for the physically same pairs, and N = the total number of judgments being ranked (eight in this case).

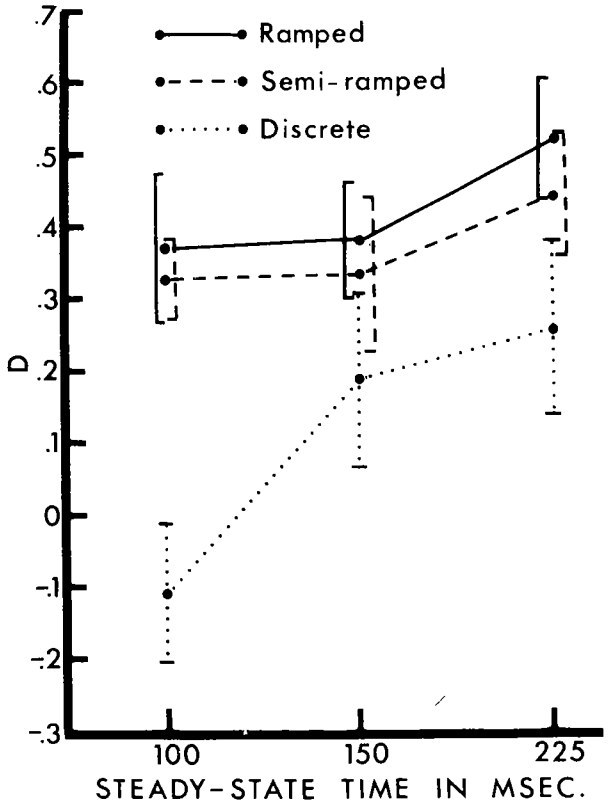

Fig. 2. The effect of ramping condition at each steady-state time for Experiment I. Brackets indicate ±1 standard error of the mean.

If all the ranks of the physically same trials are above (i.e., numerically smaller than) the ranks of the physically different trials, the D value would be 1.00, indicating no overlap of the distributions of physically same and physically different ranks, and thus perfect discrimination between the two. D equals zero when the judgments are random, indicating a lack of discrimination between physically same and physically different. A D value of −1.00 indicates systematic incorrect judgments for that cell on the part of S. Any response bias that does not affect the ordering of the judgments along the 200-mm scale, such as an overall shift in judgment towards "same" or a change in the variance of ratings, is eliminated by this procedure.

The mean D values, with the standard error of the mean for each cell, are shown graphically in Fig. 2. These results show that overall performance in the ramped and semiramped conditions was superior to that of the discrete condition at each steady-state time. Performance also improved in each condition as the

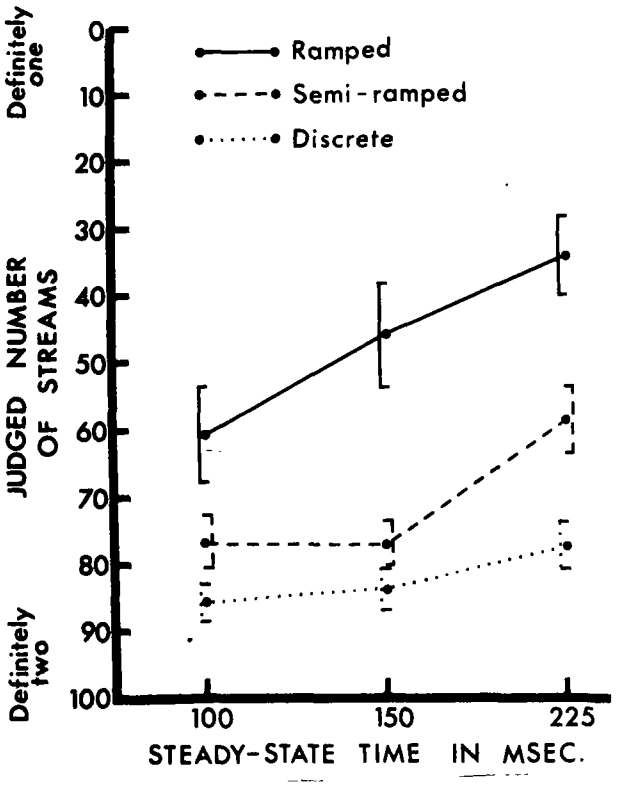

Fig. 3. The effect of ramping condition at each steady-state time for Experiment II. The Y axis corresponds to the position in millimeters of the marks on the response sheet. Brackets indicate ±1 standard error of the mean.

length of the tones increased. An analysis of variance revealed a significant difference between ramping conditions, $F(2,38) = 14.47$, $p > .001$, and between steady-state times, $F(2,38) = 4.45$, $p < .025$. There was no significant interaction between the two.

EXPERIMENT II

Although Experiment I confirmed the prediction that Ss would experience greater difficulty in determining the order of tones in the discrete condition as compared to the ramped and semiramped conditions, it did not show directly that this difficulty was due to auditory stream segregation. For this reason, a parallel Experiment II was conducted in which Ss were asked to directly judge auditory stream segregation. It was expected that the auditory stream would show a greater tendency to split into two streams in the discrete condition as compared to the ramped condition, with the semiramped condition falling between the two. It was also expected that the auditory stream would seem to be split to a greater degree for all three conditions at faster rates of presentation.

Method

Subjects

The Ss were 15 naive students at McGill University, who were paid for their services.

Apparatus and Procedure

The apparatus and stimuli were identical to those used in Experiment I. There were, however, several changes in the procedure. The pretest was eliminated, since ability to make order judgments was not relevant to this task. Ss were asked to listen to the stimuli and to decide, for each trial, whether the sequence sounded like one or two auditory streams. This decision was indicated by placing a mark along a 100-mm continuum with the extremes labeled "definitely one stream" and "definitely two streams." The mark was to indicate the degree to which the sequence sounded like one or two streams.

Results

Raw scores were obtained by measuring in millimeters the position of the marks made by the Ss on the response sheets. Means of these raw scores are shown graphically in Fig. 3.

An analysis of variance performed on the data showed a highly significant difference between ramping conditions, $F(2,28) = 30.76$, $p < .001$, between steady-state times, $F(2,28) = 14.48$, $p < .001$, and for the interaction between the two, $F(4,56) = 9.99$, $p < .001$. These results indicate that the discrete stimulus sequences sounded like two streams to a much greater degree than the ramped condition did, with the semiramped condition falling between the two. In addition, there was a tendency for the auditory streams to sound more like one stream as the steady-state time increased.

DISCUSSION

The first conclusion one can draw from these results is that continuity between tones presented at a rapid rate causes an increased ability to follow the pattern, as was demonstrated in Experiment I. Data from the semiramped condition in Experiment I indicate that this condition was functionally quite similar to the ramped condition. This demonstrates that complete continuity between the tones is not required to reduce primary auditory stream segregation (PASS); rather, it seems that a ramped frequency change "pointing" toward the next tone, allows S to follow the pattern more easily.

The notion that the difficulty experienced by Ss in the discrete condition of Experiment I is due to PASS is supported by Experiment II. The auditory stream tended to be most segregated (i.e., sound like two streams rather than one) in the discrete condition, where Ss had experienced difficulty in making order judgments. It was least segregated in the ramping condition, and here Ss had been best able to make order

judgments. The amount of perceived stream segregation in Experiment II decreased as the duration of the stimuli increased, and this was reflected in Experiment I by the enhanced ability of Ss to make order judgments at longer tonal durations. The lack of the expected interaction between steady-state time and ramping condition in Experiment I is readily explained: the 225-msec rate was not slow enough for all conditions to converge on the perfect performance you would expect at very slow speeds (as evident in Fig. 2); nor, correspondingly, was it slow enough to yield the uniform perception of a single stream (see Fig. 3).

An interesting point is the resemblance of the sounds in the semiramped condition (see Fig. 1) to the spectrographic patterns produced by individual syllables in speech production (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). The partial gliding, or "pointing," in the acoustic components of speech may be a strong factor holding it together as a unified stream.

The present two experiments, together with the others reviewed in the introduction, lead us to the following statements: (1) correct judgments of order in a recycling sequence depend upon the stream's not splitting; and (2) splitting increases when subsets of sounds occupy different frequency regions, when the tone rate is higher, and when the transitions in frequency are discrete.

Saying that the stream splits more easily at high rates may be equivalent to saying that it splits when very acoustically similar segments follow one another closely in time. In normal speech, while the event rate is high, subsets of sounds that are very similar acoustically may not occur at short enough intervals to overcome the unifying effects of glided transitions; so separate streams do not form. However, one might expect that if a single consonant-vowel syllable were recycled quickly enough, the rapid succession of identical sounds would cause PASS to occur and the consonant would split away from the vowel and form a separate stream. This has, in fact, been found by Cole and Scott (1972). Another interesting finding by these investigators was that when the syllable "sha" split up, the glided formant transition on the front end of the "a" remained united with the "a" to form the perceived syllable "da." Thus, again, we have evidence that glided transitions resist PASS.

The capacity to make judgments of order may be related to the PASS phenomenon as follows. In order to make a judgment about a pair of sounds in a stream, some perceptual process must *isolate* them, i.e., it must code the pair as a structural unit that is distinct from its environment. However, if the sensory system has a strong bias to encode the two parts of the pair as components of *separate* coded units (because of strong general rules of coding), then seeing the pair *as a pair* will be impossible or difficult. Thus, the ease of making a perceptual judgment will be determined by the correspondence of the grouping required by the judgment with the grouping patterns determined by general rules of the sensory system.

The process that encodes a sequence of auditory events into organized streams seems to have several describable properties. First, it incorporates an input into a stream if it closely resembles inputs previously assigned to the stream (in terms of frequency, loudness, overtone structure, duration, etc.). Secondly, it responds to continuities and discontinuities in a property, preferring to assign inputs to the same stream if there are no *sudden* changes; this is why the ramped condition is superior to the discrete condition in the present experiment. Thirdly, the coding mechanism is describable as a "predictive tracking device." It is a "tracking device" in that it modifies its criteria for inclusion of an input into a stream as a function of very recent properties of the stream. It is also "predictive" because if a change in a signal is preceded by a "pointer" in the direction of the change, the coding process incorporates the new input more easily into the stream.

REFERENCES

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.

Cole, R. A., & Scott, B. Phoneme feature detectors. Paper presented to the Eastern Psychological Association, Boston, April 1972.

Efron, R. Temporal perception, aphasia, and déjà vu. Brain, 1963, 86, 403-424.

Foulke, E., & Sticht, T. G. Review of research on the intelligibility and comprehension of accelerated speech. Psychological Bulletin, 1969, 72, 50-62.

Liberman, A. M. The grammars of speech and language. Cognitive Psychology, 1970, 1, 301-323.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. Perception of the speech code. Psychological Review, 1967, 74, 431-461.

Pedersen, P. The mel scale. Journal of Music Theory, 1965, 9, 295-308.

Thomas, I. B., Hill, P. B., Carroll, F. S., & Garcia, B. Temporal order in the perception of vowels. Journal of the Acoustical Society of America, 1970, 48, 1010-1013.

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.

Winckel, F. *Music, sound, and sensation.* New York: Dover, 1967.

(Received for publication August 14, 1972; revision received December 6, 1972.)