Effects of time intervals and tone durations on auditory stream segregation

ALBERT S. BREGMAN, PIERRE A. AHAD, POPPY A. C. CRUM, and JULIE O'REILLY
McGill University, Montreal, Quebec, Canada

Adult listeners rated the difficulty of hearing a single coherent stream in a sequence of high (H) and low (L) tones that alternated in a repetitive galloping pattern (HLH–HLH–HLH . . .). They could hear the gallop when the sequence was perceived as a single stream, but when it segregated into two substreams, they heard H–H– . . . in one stream and L—L— . . . in the other. The onset-to-onset time of the tones, their duration, the interstimulus interval (ISI) between tones of the same frequency, and the frequency separation between H and L tones were varied. Subjects' ratings on a 7-point scale showed that the well-known effect of speed's increasing stream segregation is primarily due to its effect on the ISI between tones in the same frequency region. This has implications for several theories of streaming.

When a sequence of tones, alternating between two frequency ranges, is speeded up, the tendency for the high and low tones to form separate auditory streams is increased. It has been proposed by Bregman (1990) that tones group by their proximity on a frequency-by-time surface. An increase of speed brings the tones closer together in time but does not reduce their frequency separations. This brings the consecutive tones of the same frequency closer together on the frequency-by-time surface, while leaving those of different frequencies almost as far away as they were before. This new proximity favors the grouping of a tone with the next one in the same frequency range even if the two tones are not consecutive, because the alternative grouping (with the tone that comes right after it but is of a different frequency) requires grouping across a longer distance. So we see that temporal distance is very important. But what is the best way to measure temporal distance? The effect of speed could be due to a change in any of the four types of time intervals shown in Figure 1, which all become shorter when the speed is increased. (Note: SOA means stimulus onset asynchrony—i.e., onset-to-onset time, and ISI is the label for interstimulus interval—offset-to-onset time. (1) SOA for consecutive tones in the same frequency range (SOA-within). Note that in Figure 1 there are two different intervals of this type, one for each frequency, since the low tones occur less frequently than the high ones in the galloping pattern. (2) ISI for consecutive tones in the same frequency range (ISI-within). Again, there are two different intervals of this type, since the low tones occur less

frequently than the high ones. (3) SOA for consecutive tones that cut across frequency ranges (SOA-across). (4) ISI for consecutive tones that cut across frequency ranges (ISI-across).

SOA is the interval that reflects the rhythm of the tone cycle, since onsets define a rhythm. If it is the rhythmic pulse that marks temporal separation for the auditory system, grouping should follow changes in SOA. On the other hand, if it is the silent gap between consecutive tones that defines temporal separation, ISI is the important interval.

The distinction between the *within* and the *across* intervals is also of theoretical importance. Anstis and Saida (1985) explained stream segregation as resulting from a tiring out of frequency jump detectors that register the transition from a tone of one frequency to a tone of another. The gradual tiring out of detectors explains, for them, why stream segregation (the failure of integration) takes hold only after a certain number of repetitions of an alternating high–low sequence. According to their theory, it is only when the detectors register a transition that the two tones involved can be integrated into a single stream. So, on this theory, it is the ISI-across that is important. Presumably, segregation increases as this interval gets shorter, because the detectors are less likely to register a fast transition than a slow one, especially when tired out.

Three experiments in the literature have studied whether it is the ISIs that control segregation but have obtained conflicting results. Van Noorden (1975, pp. 53–56) used himself as a listener in an experiment that used a semirandom tone sequence in which the frequency of each tone was displaced up or down from its predecessor's by an average of ΔF semitones, where ΔF ranged from 2 to 25 semitones. The tone duration (*D*) was also varied across trials from 80 to 400 msec. Van Noorden adjusted the SOA of the sequence to the highest tempo at which a single stream could be maintained. The results indicated that he had lengthened the SOA to hold the sequence together as ΔF had become
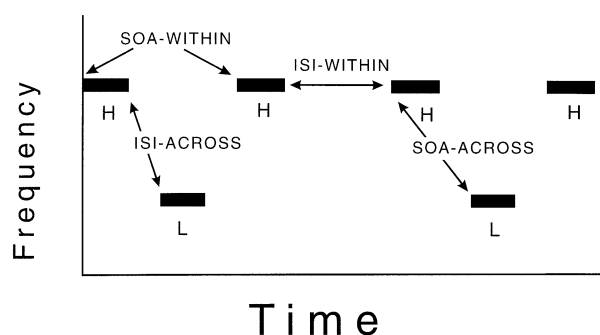
Figure 1. Two cycles of the galloping pattern, illustrating the four types of time intervals that can be distinguished: SOA-within, ISI-within, SOA-across, and ISI-across. The *within* intervals are different for high and low frequencies in a galloping pattern.

larger. The simplest account of the data was obtained when the time between tones that were contiguous in frequency (analogous to the ISI-within of the present experiment), taken from van Noorden's adjustments, was plotted against ΔF. As this time got longer, it seemed to reduce the tendency to form streams. This was reflected in the fact that a larger ΔF was required to induce streaming.

Beauvois (1998) alternated a higher and a lower tone at a fixed SOA of 130 msec at various frequency separations, while he varied the D of the higher and lower tones independently or together (*global duration*). As in our Experiment 1, which he cited as an unpublished study, his experiment showed that a greater global duration of the tone pair favored stream segregation.

However, results discrepant with those of van Noorden (1975) and of Beauvois (1998) were found in an experiment by Dannenbring and Bregman (1976). This experiment presented listeners with a fixed sequence of tones with frequencies of 2000, 614, 1600, and 400 Hz, always in that order. The ISI between adjacent tones (ISI-across) was varied across trials from 1 to 100 msec. The listeners controlled the duration of the tones (D). On each trial, they were asked, starting at a low value of D, to increase D to a value just at the threshold between one-stream and two-stream percepts, trying it above and below that value if necessary. When the adjusted duration was plotted against the gap duration, it showed a straight line between 25 and 100 msec (4 data points) with a slope of about -1. This implies that the listeners were compensating for the increasing gaps by shortening the tones by an exactly equal amount, keeping the SOA constant. There was no additional contribution from ISI per se.

Neither the cited findings of van Noorden (1975) nor those of Beauvois (1998) discriminate between ISI-within and D itself as the factor that favors stream segregation. ISI-across is probably not the main factor, because of the findings of Rogers and Bregman (1993), which will be described in the General Discussion section.

The present experiments yielded further evidence concerning the role of ISI in streaming. Owing to the limitations of their apparatus, neither van Noorden

(1975) nor Dannenbring and Bregman (1976) could allow tones to overlap in time. For some reason, Beauvois (1998) did not include such overlaps either. In the present experiments, however, we lengthened the tones past the point at which high and low tones overlapped.

EXPERIMENT 1

Method

Listeners. The listeners were 24 adults, 12 men and 12 women, who reported having normal hearing. They were recruited from a university student population and were paid for participating.

Stimuli. The stimulus was a repeating pattern of high (H) and low (L) pure tones and empty time slots (–), presented in a galloping pattern (HLH–HLH–HLH– . . .). This pattern was used instead of a simple alternating HLHL . . . pattern because, in the galloping sequence, it is easy for listeners to tell when the sequence segregates: The galloping rhythm disappears. Each tone had 10-msec onset and offset ramps that followed a quarter-sine-wave function. The tones were presented diotically. Each single tone was measured to be 60 dB in intensity (B weighting, flat-plate coupler), but the total intensity was, of course, higher whenever H and L tones overlapped in time.

The onset-to-onset time (SOA) of successive elements in the sequence was always 100 msec for the HL or LH transitions. The SOA-within for the H tones was always 200 msec. The SOA-within for the L tones was always 400 msec. The onset of each L tone was halfway between the onsets of its bracketing H tones. Although the SOAs were not the same for the H and the L tones, their durations were fixed and could not contribute to the differences across conditions.

The durations of the tones (including their onsets and offset ramps) took the following six values: 40, 60, 80, 100, 120, and 140 msec. Despite the fact that ISI-within intervals were different for the H and the L tones, any increase in tone duration (D) reduced both these intervals by the same *number of milliseconds*, although it did not decrease them by the same *proportion*. So we can equally well think of the change in D as a change in the ISI-within. Whenever the Ds were greater than 100 msec, they exceeded the SOA-across, causing the H and the L tones to overlap in time. The frequency separation between the H and the L tones was two, three, or four semitones. The H and L tones were positioned symmetrically (on a semitone scale) above and below 1000 Hz. There were 18 conditions altogether (3 frequency separations × 6 values of duration).

Apparatus. The stimuli were synthesized by using MITSYN signal-processing software (Henke, 1990) at 20,000 samples per second and were low-pass filtered on playback at 8 kHz by a passive Tchebychev filter having a 3-dB cutoff at 8000 Hz and a slope of 142 dB/octave. Testing was controlled by a program written in MAPLE (Achim, Bregman, & Ahad, 1995), controlling a PC-compatible computer equipped with a 16-bit D/A converter. The sounds were presented over Sony MDR-V7 headphones in a single-wall Industrial Acoustics test chamber.

Procedure. The listeners were told that they would hear a pattern of H and L tones in a galloping pattern. They were given an explanation, using diagrams, of the difference between hearing this as a single stream containing a galloping pattern (HLH–HLH– . . .) and as two separate streams, each with a simpler pattern, (H–H–H– . . . and L—L— . . .). They were to try to hold on to the single-stream interpretation (gallop) as much as possible and were to rate, on a 7-point scale, how easily they could do this (1 = *very easy*, 7 = *very hard*), to be entered on a keyboard.

To familiarize them with the one- and two-stream percepts, the listeners heard galloping patterns that were clear examples of integration (SOA = 140 msec, two-semitone separation) and segregation (SOA = 40 msec, four-semitone separation). Then they went through a practice run consisting of all the conditions of the exper-

iment in a random order. On each trial, the repeating HLH– pattern was presented 12 times in an unbroken sequence lasting for 4.8 sec. The experiment proper consisted of six blocks of randomized trials, in which each of the 18 conditions was presented once.

Results

The mean 1–7 segregation scores are shown in Figure 2 as a function of D and H–L frequency separation. The usual effect, in which frequency separation increases the amount of stream segregation, is seen in the separation of the three curves. Each curve shows a monotonic effect of the duration of the tone. Longer tones (which produced shorter ISIs) promoted greater segregation. A two-way repeated measures analysis of variance (ANOVA) showed significant effects of frequency separation [$F(2,46) = 120.4, p < .00001$] and of D [$F(5,115) = 86.7, p < .00001$]. The interaction of the two variables was also significant [$F(10,230) = 4.83, p < .00001$]. This seems to reflect the fact that the effects of D were a little smaller at the closest frequency separation. (Bruno Repp, a reviewer, has suggested that the reduced separation of means could have been due to a floor effect.)

Increases in D in the region above 100 msec continue to cause increases in segregation, despite the fact that, in this region, the tones temporally overlap and ISI-across becomes negative.

Discussion

In the introduction, four time intervals were mentioned as being potentially involved in the well-known effects of speed in promoting stream segregation. Two of
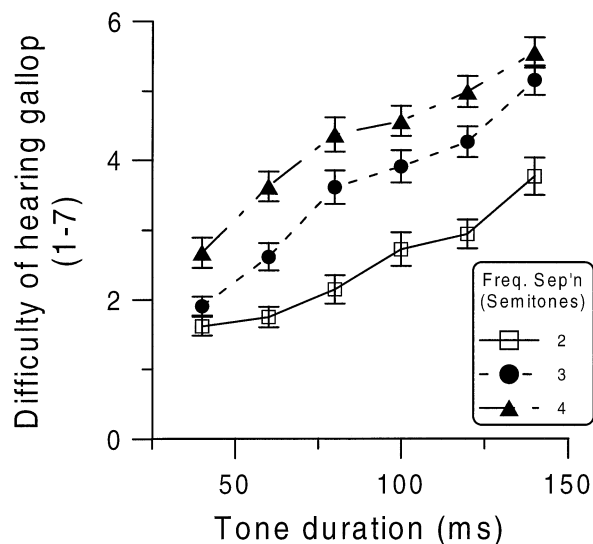

Figure 2. Experiment 1: Rated difficulty of hearing the gallop (1 = *easy*, 7 = *hard*) as a function of the tone duration in milliseconds. The parameter is the frequency separation in semitones. Vertical bars represent ±1 *SE*.

those were held constant in the present experiment: the SOA between the nearest pair of consecutive tones and the SOA between consecutive tones of the same frequency. However since the SOA was held constant, as tone D increased, ISI-within decreased by exactly the same amount. So this experiment, alone, cannot tell whether long D or short ISI-within (or both) favors segregation.

Of the experiments reviewed in the introduction, that of Dannenbring and Bregman (1976) is the only one that gave results discrepant with the present ones. In that experiment, there was an unmistakable tendency for listeners to try to hold SOA constant in order to maintain a criterion level of streaming. It may be that, in some circumstances, onsets play a stronger role than we observed in the present experiment. A factor that may have influenced the Dannenbring–Bregman experiment is learned expectations. When a listener is asked to adjust a parameter to find the threshold of streaming, there is a considerable level of ambiguity in the region of the threshold. The situation is made more ambiguous when listeners are allowed to adjust the value of the parameter up and down more than once across the threshold value. Informal observations by the first author, based on many experiments, suggest that listeners, when starting from the two-stream extreme, will hold on to the two-stream percept considerably past the threshold that they would have chosen had they been approaching the threshold from the other direction. This is a case of hysteresis of interpretations, or *set*, as the Gestalt psychologists would have labeled it. Facing such ambiguity, the listeners in the Dannenbring and Bregman experiment may simply have focused on the salient rhythm defined by the onsets and tried to achieve the value of tempo (SOA) that, on average, seemed to be present at the boundary between one and two streams.

Another possibility, of course, is that SOA has a real effect in addition to the tone duration effect observed in the present experiment. Recall that the present experiment yielded the strong ISI effect (or equivalently, the tone duration effect) by holding SOA constant. So the experiment neither supported nor ruled out an effect of SOA per se. To deal with this issue, Experiment 2 varied SOA and tone D orthogonally.

SOA-within is equal to the sum of D and ISI-within. Therefore, in Experiment 1, since SOA-within was held constant, ISI-within and D were perfectly (negatively) correlated. But how do we know that intertone silence and D do not have *separate* effects or that tone duration itself was not the causal factor? This issue was addressed in Experiment 2.

EXPERIMENT 2

SOA-within, ISI-within, and D cannot vary independently. Fixing the values of any two of these determines the third. For this reason, in Experiment 2, we ran vari-

ous combinations of conditions that allowed us to analyze the results according to *two designs* each with three orthogonal independent variables: (1) frequency × ISI-within × *D*, with SOA-within (onset-to-onset time) constrained by the values of the other variables; and (2) frequency × ISI-within × SOA-within (onset-to-onset time), with *D* constrained by the values of the other variables.

Because there are, potentially, three underlying causal factors that could affect the perception of a stimulus pattern that has only two degrees of freedom (i.e., any two of *D*, ISI-within, and SOA determines the third), the experiment cannot test one underlying factor at a time. The best we can do is to find the minimum set of assumptions that can make sense of the results of both Analyses 1 and 2, below.

Method

Listeners. As in Experiment 1, there were 24 adult listeners, 12 men and 12 women, recruited from a university student population and paid for participating.

Stimuli. The stimuli were generated as in Experiment 1 and consisted of the same galloping pattern. Individual tones were presented at 70 dBA. Tables 1 and 2 define the conditions (ISI-within, *D*, and SOA-within) that allowed the two analyses described above to be carried out. Frequency separation is not shown in the tables, but the two values used were three and five semitones.

Given that the within-stream ISI-within and SOA-within are different for the high and the low tones (since there are half as many L tones as H ones), these measures will henceforth always be specified *for the H tones only*. The values for the L tones can be calculated from these, when desired. When the ISI abbreviation is used, it will refer to the ISI-within of the H tones, except when otherwise indicated. Similarly, the SOA abbreviation will refer to the SOA-within of the H tones.

Table 1 shows the experimental conditions for Analysis 1, in which ISI and *D* were varied orthogonally, with SOA (shown in the cells) constrained by the values of the other variables, as required.

Table 1
Experimental Conditions for Analysis 1:
ISI-Within for the High Stream × Tone Duration,
Showing SOA-Within Intervals

Tone Duration	ISI-Within			
	75	100	125	150
75	150	175	200	225
100	175	200	225	250
125	200	225	250	275
150	225	250	275	300

Note—Tone duration (in milliseconds) was the same for both high and low tones. All values are in milliseconds.

Table 2
Experimental Conditions for Analysis 2:
ISI-Within for the High Tones × SOA-Within
Intervals for the High Tones, Showing Tone Duration

SOA-Within	ISI-Within		
	75	100	125
175	100	75	50
200	125	100	75
225	150	125	100
250	175	150	125

Note—Tone duration was the same for both high and low tones in the cells. All values are in milliseconds.

Table 2 shows the conditions for Analysis 2, with ISI and SOA varied orthogonally and *D* (shown in the cells) varying as required. When the same condition occurred in both designs, it was not tested twice as often as those present in only one design. Rather, the result for that condition was used in both analyses. In all, there were 36 different conditions. A block of trials consisted of one random ordering of these 36 conditions. Each listener received six such blocks.

In Analysis 1 (ISI × *D*), the values of ISI were 75, 100, 125, and 150 msec. The values for *D* were 75, 100, 125, and 150 msec. Frequency separations between H and L tones were three or five semitones.

In Analysis 2, (ISI × SOA), the values of ISI were 75, 100, and 125 msec. The values for SOA were 175, 200, 225, and 250 msec. Frequency separations between H and L tones were three or five semitones.

Procedure. As in Experiment 1, the listeners were taught the difference between hearing a single coherent stream and hearing two separate streams. They were to try to maintain the single-stream percept and, at the end of each trial, to judge how difficult it was to do so on a 7-point scale, as in Experiment 1. They then received a practice session consisting of all 36 conditions. The experiment proper was carried out in two sessions, each consisting of three blocks of 36 trials. Each trial consisted of a 7-sec presentation of the galloping sequence. Because the total trial *D* was fixed, there were fewer repetitions of the slower sequences.

Results and Discussion

Two three-way repeated measures ANOVAs were carried out, one for each design, utilizing the data from the set of conditions appropriate for that design. It is evident that the two designs are not independent but are simply different ways of looking at the data. We were particularly interested in the proportion of the systematic (nonerror) effect of all variables that would be accounted for by ISI.

Figures 3 and 4 show the results for Analysis 1 for frequency separations of three and five semitones, respectively. The mean rated difficulty of holding the gallop together is plotted as a function of ISI (i.e., ISI-within for the H tones). High values represent segregation. The parameter is the duration of the tones in milliseconds. In Figure 5, a pie chart shows, for each independent variable, the quantity $SS_E / SS_{TE} \times 100$, where SS_E is the sum of squares for a given effect and SS_{TE} is the total sum of squares for all systematic effects (ignoring noise). Since the sums of squares are additive, the percentage of the total gives an index of the relative effect size; it is like η^2, a common measure of effect size, except that it ignores error sums of squares. The main results were the following.

1. *Frequency separation*. A comparison of Figures 3 and 4 shows that an increase in frequency separation of the H and the L tones strengthened the segregation [mean of 2.9 and 4.6 for the three- and five-semitone separations, respectively; $F(1,23) = 92.2, p < .00001$]. Figure 5 shows that the frequency separation accounted for a major proportion of the observed effects, as is typical in stream segregation experiments.

2. *ISI*. There was a strong effect of ISI, seen in both Figures 3 and 4. As it got larger (longer intervals between H tones), segregation became weaker [$F(3,69) = 83.1, p < .00001$]. ISI accounted for almost as large a proportion of observed effects as did frequency separation.
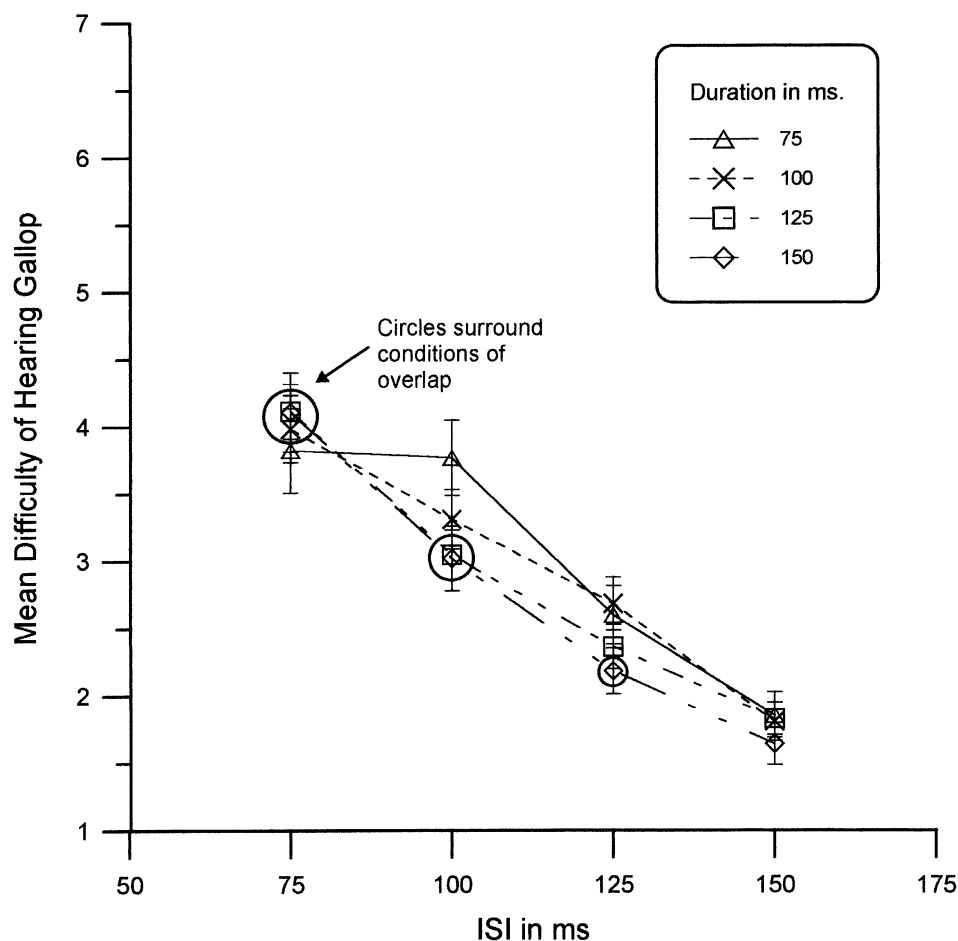
Figure 3. Experiment 2, Analysis 1: Mean difficulty of hearing the gallop (1 = *easy*, 7 = *hard*), as a function of ISI-within in milliseconds at a *three-semitone separation* of high and low tones. The parameter is the duration of the tones in milliseconds. Circles completely surround data points for which the high and the low tones overlapped temporally. Vertical bars represent ±1 *SE*.

3. Duration. As D got shorter, there was a tendency for segregation to become stronger. Although it was weak and inconsistent, accounting for only 2% of the systematic effects, the effect was statistically significant [$F(3,69) = 4.1, p < 01$].

4. Duration × ISI. There was a significant interaction of D and ISI [$F(9,207) = 6.2, p < .00001$]. The effects of ISI became stronger as D increased—that is, as the sum of ISI and D became larger. Since SOA is the sum of ISI and D, this interaction suggested that SOA itself might have been having an effect. However, this interaction accounted for only 2% of the systematic effects, as is shown in Figure 5.

5. Frequency separation × D. There was also a significant interaction between the frequency separation and D, so that a more pronounced effect of ISI was found at the higher frequency separation [$F(3,69) = 3.2, p < .03$]. Although significant, it accounted for less than 1% of the systematic effects. It may have resulted from a floor effect, since the results for the longer ISIs at the three-semitone separation were near the bottom of the scale.

6. Temporal overlap. The circles in Figures 3 and 4 surround those data points for which the H and the L tones were temporally overlapped. These overlaps occurred for the 175-, 200-, and 225-msec durations at the shortest ISI (75 msec), for the 225- and 250-msec durations at the next longer ISI (100 msec), and for the 150-msec duration for the next ISI (125 msec). It is important to note that the segregation value for a data point for which this overlap occurred was always greater (harder to hear the gallop) than for the point to the right of it on the same curve and less than the point to the left of it (when there was one). That is, the trend for segregation to be greater with shorter ISIs continued through the point at which the short ISI caused the H and the L tones to overlap temporally.

Since ISI and D add up to SOA, if it was SOA rather than D that was important, then at any fixed value of ISI,
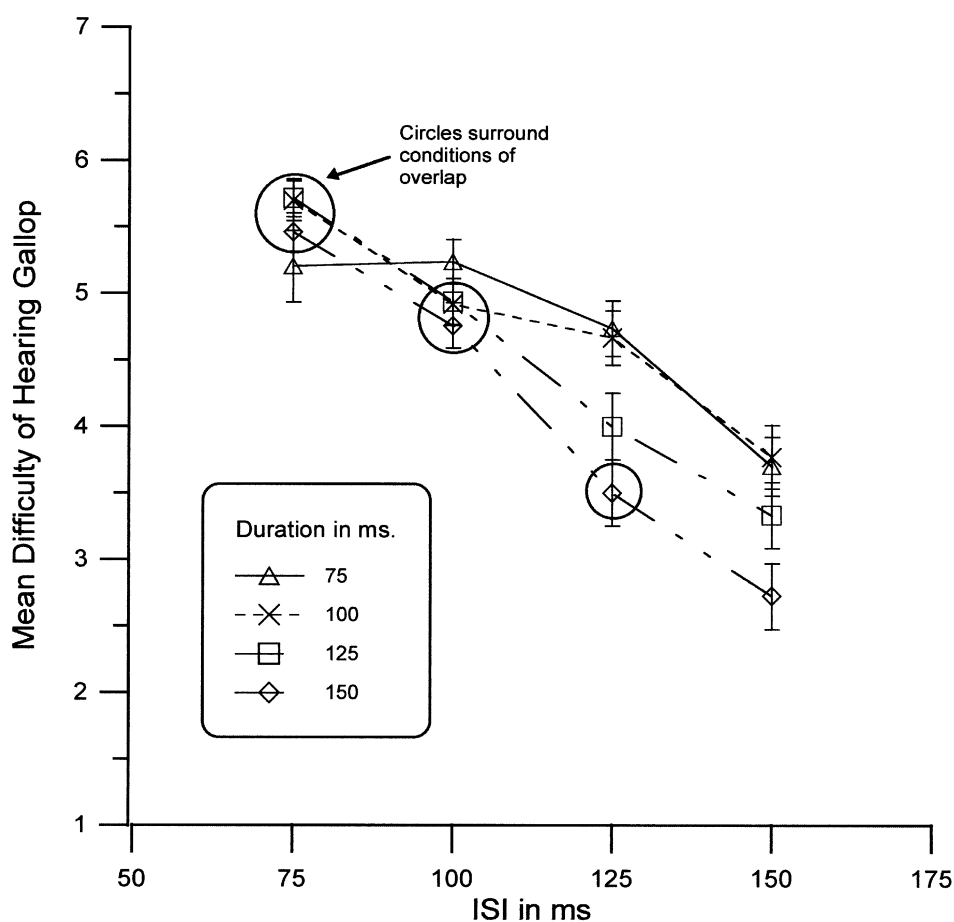
Figure 4. Experiment 2, Analysis 1: Mean difficulty of hearing the gallop (1 = *easy*, 7 = *hard*), as a function of ISI-within in milliseconds at *five-semitone separation* of high and low tones. The parameter is the duration of the tones in milliseconds. Circles completely surround data points for which the high and the low tones overlapped temporally. Vertical bars represent ±1 *SE*.

an increase in SOA would look as if *D* was influencing segregation. However, there was only a very small effect attributed to *D*. If there was an effect of SOA, it must also have been very small.

Figure 6 shows the results for Analysis 2 for the three-semitone separation of H and L tones. The mean rated difficulty of holding the gallop together is plotted as a function of ISI (i.e., ISI-within for the H tones). The parameter in this analysis is SOA in milliseconds. Figure 7 shows the plot for the five-semitone separation. Figure 8 shows a pie chart display of the proportion of the total of all the systematic effects that is accounted for by each of the independent variables. The main effects were the following:

1. *Frequency separation*. A comparison of Figures 6 and 7 shows again that an increase in frequency separation of the H and the L tones strengthened the segregation [mean rating of 3.3 and 5.0 for the three- and five-semitone separations, respectively; $F(1,23) = 100.7, p < .00001$]. Figure 8 shows that the frequency separation accounted

for an even greater proportion of the systematic effects than it did in Analysis 1. This is at least partly due to the fact that the ISI effect was somewhat weaker than in Analysis 1 (since the proportion of the effects of all the variables accounted for by one variable depends on the relative strength of the others). The reduction in the ISI effect was caused, at least in part, by the fact that it had a smaller range of values than it did in Analysis 1 (compare Tables 1 and 2). The range was restricted by the constraint that $D = SOA - ISI$. If ISI had been allowed to become too large, D would have become shorter than 50 msec, our shortest permitted D.

2. *ISI*. The effect of ISI, although not as strong as in Analysis 1, was still reliable, as is shown in Figure 5. As the ISI got longer (greater intervals between H tones), segregation became weaker [$F(2,46) = 42.1, p < .00001$]. Figure 6 shows that it accounted for 25% of the systematic effects.

3. *SOA*. There seemed to be a trend for smaller values of SOA to lead to more segregation, but this was not sig-
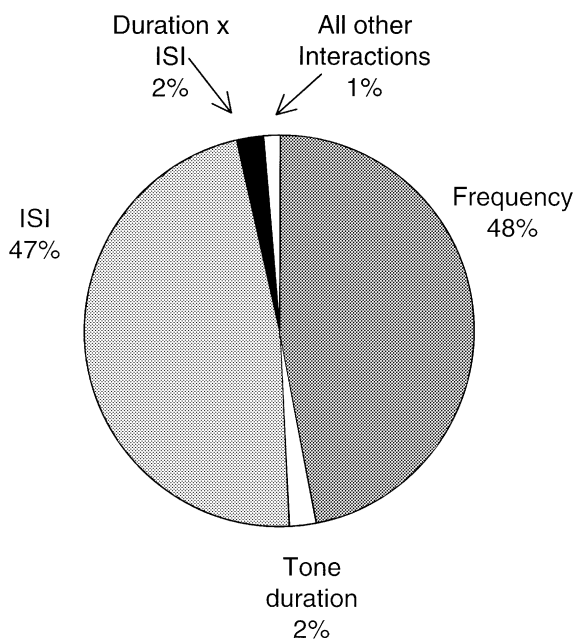
Figure 5. Experiment 2, Analysis 1: Proportions of the total systematic effect of all variables accounted for by frequency separation. ISI-within, duration (*D*), and interactions. ISI and *D* varied orthogonally.

nificant at the .05 level ($.05 < p < .10$). It accounted for 3% of the systematic effects.

4. *Frequency separation $\times$ ISI*. There was a significant interaction between frequency separation and ISI [$F(2,46) = 3.4, p < .05$], but it accounted for only about 0.5% of the systematic effects and is not shown separately in the pie chart of Figure 8. The facilitation of segregation by shorter SOAs may have been a bit stronger at the three-semitone separation.

5. *SOA $\times$ ISI*. The interaction of SOA with ISI accounted for only 2% of the systematic effect of all the variables, but it was significant [$F(6,138) = 2.7, p < .02$]. There may be a slightly more consistent effect of SOA at an ISI of 100 msec.

6. *Temporal overlap*. Again, the data points at which the H and the L tones were temporally overlapped do not seem to deviate from the general trends observed in Figures 6 and 7.

The results of Analysis 2 confirm the results of Analysis 1 in showing a strong effect of ISI. Although we speculated that the *D* effect in Analysis 1 could have been an indirect effect of SOA, we find in Analysis 2, where SOA is varied orthogonally with ISI, that SOA has a weak effect, accounting for only 3% of the systematic effect of all the variables. Indeed, in this analysis, the SOA effect could be an indirect effect of *D*. Since SOA = ISI + *D*, when ISI is fixed, SOA can only increase through an increase in *D*. That is, they are positively correlated. If we think of *D*/SOA as a single variable that was varied orthogonally with ISI, in neither analysis did it account for

more than 3% of the systematic effect of all the variables, whereas ISI accounted for from 25% to 47% (most of the remainder being accounted for by frequency separation). It is meaningless to make exact comparisons of the effects of ISI and frequency separation on segregation, since we cannot equate them on the range over which they varied in the experiment (how much frequency separation is equal to how much temporal separation a priori?).

GENERAL DISCUSSION

The results support the view that the effects of ISI are direct, rather than being an indirect effect of its consequences for *D* or SOA in a particular experimental design. If the latter two variables have an effect, it seems to be small. We can address a number of theories, using our data.

The first is the idea proposed by Anstis and Saida (1985) that rapid alternation of frequencies serves to habituate frequency transition detectors. Only when all the frequency transitions are registered by such detectors can the sequence be integrated as a single perceptual stream. This theory makes ISI-across the important measure of rate.

This theory, however, fails to explain the results of the experiments of Rogers and Bregman (1993), which studied the accumulation of the streaming tendency. They examined how the early part of a sequence of tones induced a streaming tendency in the later part. They found that, even if the earlier part had only high-frequency tones (i.e., no frequency jumps at all), it was able to induce a streaming tendency in the later part (which did contain frequency jumps between H and L tones), as compared with control stimuli, which had unrelated sounds or only silence in the earlier part. Hence, the cumulative property of the streaming tendency does not depend—or does not depend exclusively—on the tiring of jump detectors. Therefore the ISI-across time interval is not a crucial one.

With regard to the present data, if ISI-across were the important interval, we would expect to find some change in the trend of the results at the point where ISI-across became negative. Consider, for example, what might happen if frequency jump detectors were responsible for sequential integration. One can imagine three ways in which they could work. (1) If such detectors respond to the tone whose onset follows the *offset* of a given tone, A, they might miss a tone, B, whose onset precedes the offset of A, or they will have to consult a memory of recent events to detect the onset of B. It is hard to imagine how such a mechanism could work at all if sounds were overlapped in time by more than a few milliseconds, as they often are in natural environments. (2) On the other hand, if the detectors are sensitive *only to onsets* and not to the steady states of the tones, the time interval *D* (or ISI) should have *no* effect. Only SOA should matter. The present results do not support this view. (3) A third possibility is that the frequency transition detectors, rather than waiting until the end of tone *n*, will switch to tone *n* + 1 as soon as they detect its onset. This would mean that any addi-
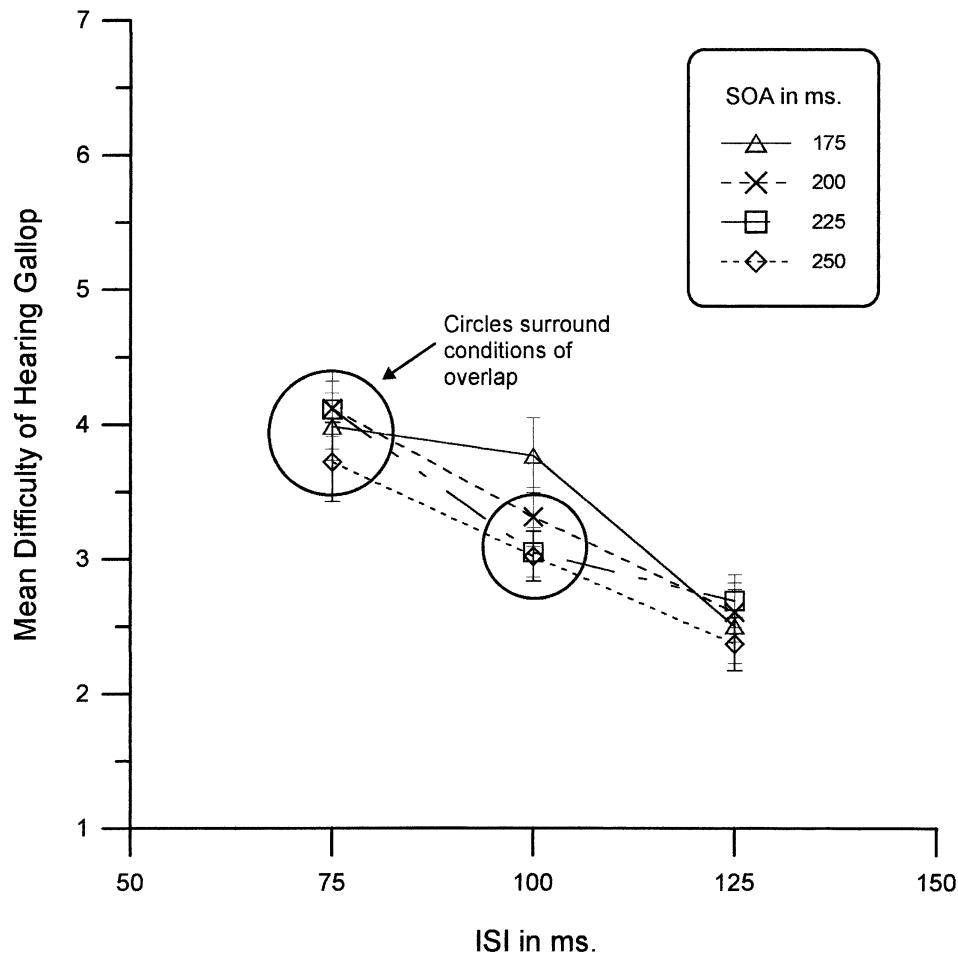
Figure 6. Experiment 2, Analysis 2: Mean difficulty of hearing the gallop (1 = *easy*, 7 = *hard*), as a function of ISI-within (for the high tones) in milliseconds at a *three-semitone separation* of H and L tones. The parameter is the SOA-within in milliseconds (for the high tones). Circles completely surround data points for which the high and the low tones overlapped temporally. Vertical bars represent ±1 *SE*.

tional duration of tone *n*, past the onset of tone *n* + 1, will have no further effect on streaming. However, decreases in ISI past the point of overlap did increase segregation in the present experiment. Therefore, all three of the proposed modes of operation for frequency transition detectors are incompatible with our data.

Van Noorden (1975) proposed a similar theory of stream segregation, although he phrased it in terms of a band-pass filter. His notion was that the auditory system is equipped with a filter that can be tuned to different frequencies. When a sequence of alternating H and L tones is received, only if these tones are all able to pass through this filter in succession will a single integrated stream be heard. He assumed that the band-pass filter will change its center frequency to "catch" the incoming tones. In order to account for the well-known fact that both the frequency difference between H and L tones and their rate of presentation favor stream segregation, he assumed that the filter can only change its center frequency at a certain

maximum rate. We have two possibilities for the mode of operation of this filter mechanism. (1) The filter might switch whenever there was an *onset* at some frequency other than the one it was tracking. If this were true, the persistence of tone *n* past the onset of tone *n* + 1 would have no effect on segregation. However, the present data tell us that it does have an effect. (2) The filter could change its tuned frequency only at the *offset* of the tone it was currently tracking. If tones were overlapped in time, this would make it harder to catch the following tone, or at least all of it. Therefore, overlap should increase the amount of streaming (as it does). So the latter mode of operation would explain the present data. However, using the maximum slew rate (rate of change of tuned frequency) of a filter to explain stream segregation does not explain the role of competition of frequency proximities on stream segregation. Competition of frequency proximities can be illustrated in a situation containing four tones (A, B, C, and D) in a rapid cycle (Bregman,
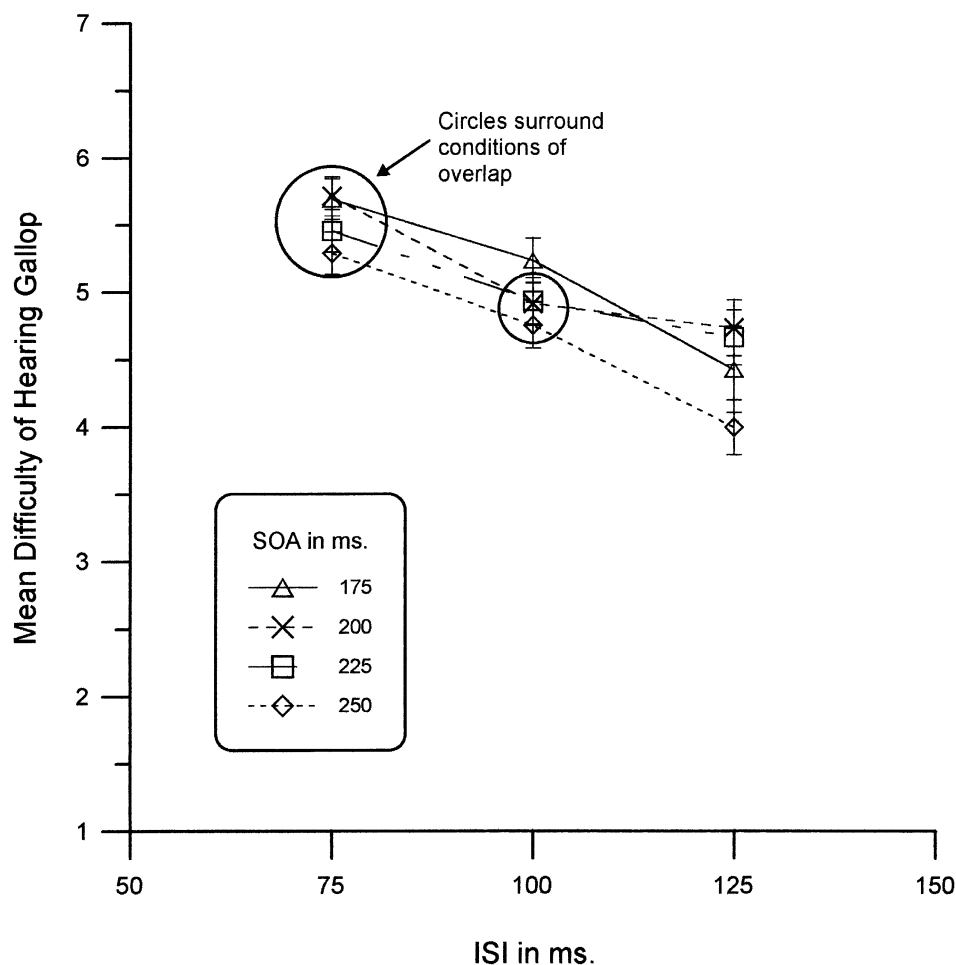
Figure 7. Experiment 2, Analysis 2: Mean difficulty of hearing the gallop (1 = *easy*, 7 = *hard*), as a function of ISI-within (for the high tones) in milliseconds at a *five-semitone separation* of high and low tones. The parameter is the SOA-within in milliseconds (for the high tones). Circles completely surround data points at which the high and the low tones overlapped temporally. Vertical bars represent ±1 *SE*.

1978a). Suppose that A and B are in a higher frequency range and C and D in a lower range. Let us set the frequency ratio (*R*) between A and B to be the same as the ratio between C and D (also *R*). We can set *R* to a small enough value that A groups with B to form a high-frequency stream and C groups with D to form a low-frequency stream. Let us examine the high-frequency stream. Under the filter theory described above, the fact that A and B have formed a stream means that the filter has managed to change its tuned frequency fast enough to first allow A to pass and then B. Similarly, a lower frequency filter must have passed both C and D. Now, without changing *R*, let us transpose the C and D tones upward in frequency so that C is very close to A in frequency and D close to B. We find that a new perceptual grouping takes place. A and C now form one stream, and B and D a second stream. This cannot be explained by a rate-of-change limit on the filter that passes A. We did not change

the frequencies of A and B or their temporal separation; so the same filter should again pass both A and B. How could A and B end up in different streams? The filter explanation has no answer for this. A Gestalt explanation would be that groupings are determined, at least in part, by *relative* proximities in the frequency-by-time space. Tone A will group with the nearest tone(s). A particular tone having a fixed distance from it is not always among the nearest.

Furthermore, a theory that explains the streaming phenomenon by appealing to limits on the slew rate of a filter cannot explain why segregation increases with more repetitions of a sequence of alternating H and L tones (Bregman, 1978b).

Beauvois (1998) offered a computer model to explain stream segregation. It is too complex to be described here. Suffice it to say that, in this model, the effects of tone lengthening (as in our Experiment 1 and in Beauvois's own
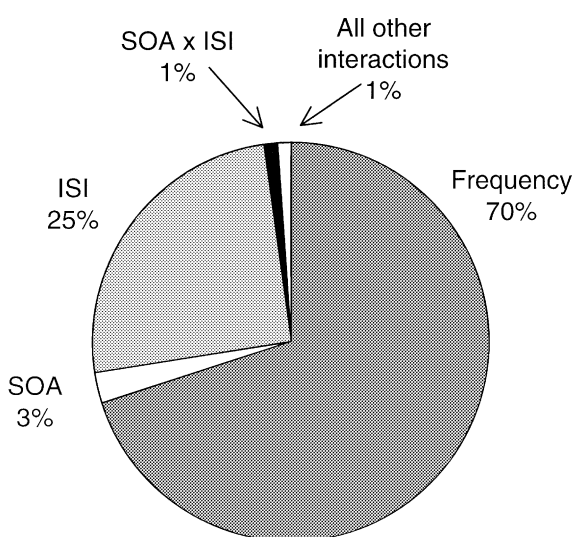
Figure 8. Experiment 2, Analysis 2: Proportions of the total systematic effect of all the variables accounted for by frequency separation, ISI-within, SOA, and interactions. ISI and SOA varied orthogonally.

experiment) are not seen as the result of changes in ISI-within, but as the direct result of a process that increased its strength as the tones became longer. So the model unambiguously attributes the temporal effects to the duration of the tones. However, the data of the present Experiment 2 suggest that it is the shortening of ISI-within, not the lengthening of the tones per se, that promotes stream segregation. Therefore, the present results are not explainable by his model. While Beauvois *did* draw the conclusion that it was ISI-within, rather than *D*, that was actually responsible for the "speed effect," he failed to note that this was incompatible with the computational model he offered as an explanation of the data. Furthermore, his experiment could not experimentally distinguish the effects of *D* from those of ISI-within, since he kept SOA constant and varied only *D*, as in our Experiment 1. Nonetheless, one important finding of Beauvois's study is unaffected by the present conclusions. Replicating a finding by Hartmann and Johnson (1991), Beauvois found that as the *difference* in tone durations of H and L tones increased, segregation did likewise. Apparently, the property of *duration* is a feature of a tone that affects its similarity or contrast with other tones. However, we have found that *D* has little effect on grouping when *all* the tones in the sequence are made longer or shorter, given a fixed ISI.

Our proposal for a more plausible stream segregation mechanism is this. Activity in a neural circuit that has just registered a frequency gradually dies out. Only if stimulation from a subsequently arriving tone, similar in frequency to the earlier one, is registered before the activity dies out do the two tones develop a link between them—a link whose strength is inversely proportional to

the elapsed time and the frequency separation. Links between all the frequency components detected in a certain-sized temporal window (say, on the order of a second) then enter into a competition that determines the set of streams that will be perceived. In other words, links are set up between tones that are close together in frequency and in time (where *close together in time* refers to ISI), and competition of these proximities determines the perceptual grouping (cf. Bregman, 1978a).

In order to accommodate the findings of Bregman (1978b) on the cumulative effects of repetition on segregation, we would also have to assume that the presence of recent repeated exposures to a narrow range of frequencies changes the link-assigning process. We can imagine two possible ways this could happen: (1) Repetition could increase the strength of the links assigned in that frequency range, and the termination of exposures could lower this strength, or (2) receiving a series of sounds having a narrow range of frequencies could *increase the weight* assigned to frequency separation (relative to temporal separation) in weakening the strength of the links, and the termination of exposures might restore the default weighting of frequency separation relative to temporal separation. Either of these versions could account for the gradual build-up of bias favoring a two-stream interpretation as an unbroken sequence gets longer and the subsequent dissipation of the two-stream interpretation during a silence. According to either version of this explanation, ISI-within is the important time interval, because this is the interval during which the link assignment process returns gradually to its unbiased form.

We have framed this argument in terms of frequency, but we intend it to apply to such variables as spatial location and timbre as well.

There are phenomena other than the segregation of alternating H and L tones that also suggest that it is ISI-within that is the important interval. The first of these is the sequential capturing of a part of a complex sound so that the part is heard as a separate sound. Suppose we play a repeating cycle in which a sound, A, alternates with a simultaneous pair of sounds, B and C, and in which A and B are identical sounds. The shorter the silent interval between A and B (the ISI), the better A is able to capture B into a sequential stream (A–B–A–B– . . .), leaving behind the repetitions of the C sound (C—C— . . .) as a residual sound in a different stream. Rappold, Mendoza, and Collins (1993) found this with narrow-band noise bursts serving as A, B, and C, but we have informally observed the same effect in our laboratory, using tones. If we assume that the sequential grouping between tones in the same frequency range results from the same underlying mechanism in this phenomenon and in the streaming phenomenon, this argues that it is the ISI-within that is the relevant timing variable in the streaming phenomenon.

Another phenomenon that displays a similar ISI effect is apparent continuity. An example of this can be created by alternating a soft pure tone with a much louder noise burst, leaving no gaps between them. Under some con-

ditions, the separate occurrences of the pure tone are integrated perceptually and heard as a long unbroken tone, continuing behind the noise. The ISI effect in the apparent continuity experiment is this: If one shortens the noise burst that separates the tones (analogous to shortening the *silence* between tones of the same frequency in the present experiment), the continuity of the tone behind the noise improves (Dannenbring, 1976; Warren, 1982, p. 152). This shortening of the noise reduces the interval between the end of the tone that precedes the noise and the beginning of the one that follows it and, therefore, is an example of a change in ISI.

These phenomena suggest that the time between the offset of an earlier tone and the onset of a later one of the same frequency (the ISI-within of Figure 1) is the important one in determining sequential integration or segregation. If one conceives of stream segregation as a Gestalt-like grouping by proximity (in time, in frequency, and in some other descriptive dimensions), silence seems to be what defines distance for sequential grouping. This conclusion, of course, may be limited by the fact that the tones typically used in experiments on streaming come on quickly, have a steady state at a constant amplitude, then go off quickly. The effect of having tones come on abruptly and then decay exponentially in amplitude, as many plucked and struck instruments do, is still unknown. Presumably, the energy present in the decay would have some effect on grouping.

REFERENCES

ACHIM, A., BREGMAN, A. S., & AHAD, P. A. (1995). *Manager of Auditory Perception and Linguistics Experiments (MAPLE)*. Montreal: McGill University, Department of Psychology, Auditory Perception Laboratory.

ANSTIS, S., & SAIDA, S. (1985). Adaptation to auditory streaming of frequency-modulated tones. *Journal of Experimental Psychology: Human Perception & Performance*, **11**, 257-271.

BEAUVOIS, M. W. (1998). The effect of tone duration on auditory stream formation. *Perception & Psychophysics*, **60**, 852-861.

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

BREGMAN, A. S. (1978b). Auditory streaming is cumulative. *Journal of Experimental Psychology: Human Perception & Performance*, **4**, 380-387.

BREGMAN, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA.: MIT Press.

DANNENBRING, G. L. (1976). Perceived auditory continuity with alternate rising and falling frequency transitions. *Canadian Journal of Psychology*, **30**, 99-114.

DANNENBRING, G. L., & BREGMAN, A. S. (1976). The effect of silence on auditory stream segregation. *Journal of the Acoustical Society of America*, **59**, 987-989.

HARTMANN, W. M., & JOHNSON, D. (1991). Stream segregation and peripheral channeling. *Music Perception*, **9**, 155-184.

HENKE, W. L. (1990). *MITSYN: A synergistic family of high-level languages for time signal processing. Version 8.1*. Belmont, MA: Author.

RAPPOLD, P. W., MENDOZA, L., & COLLINS, M. J. (1993). Measuring the strength of auditory fusion for synchronous and non-synchronous amplitude-fluctuating, spectrally disparate narrow-bands of noise. *Journal of the Acoustical Society of America*, **93**, 1196-1199.

ROGERS, W. L., & BREGMAN, A. S. (1993). An experimental evaluation of three theories of auditory stream segregation. *Perception & Psychophysics*, **53**, 179-189.

VAN NOORDEN, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences*. Unpublished doctoral dissertation, Eindhoven University of Technology.

WARREN, R. M. (1982). *Auditory perception: A new synthesis*. New York: Pergamon.