Auditory streaming and the continuity illusion

YVES TOUGAS and ALBERT S. BREGMAN
McGill University, Montreal, Quebec, Canada

In the present experiment, auditory stream organization was investigated in the presence of perceptually restored continuity. It was found that auditory streaming processes tend to yield the same perceptual organization independently of the presence or absence of perceptual restoration. Other observations include the dominance of frequency proximity over trajectory as a perceptual organization principle, and the effect of harmonic enrichment on perceptual grouping.

Auditory stream segregation and the continuity illusion have been described in the literature as two separate phenomena. Both, however, can be considered different aspects of a perceptual process whose function is to build an accurate representation of sonic events.

Auditory stream segregation was first investigated in rapid sequences of sounds that alternated between two different frequency regions. When such sequences were played at high speeds, the sequence of tones seemed to segregate into two perceptual streams, one in the higher frequency region and one in the lower. Bregman (1981, in press) has interpreted this phenomenon as resulting from processes that act so as to group sounds that probably have arisen from the same source.

One version of the continuity illusion can be demonstrated by employing an acoustic sequence in which a long tone is repeatedly alternated with a shorter, but much louder, noise burst. The tone is heard as continuing behind the noise burst despite the fact that it actually does not. Warren (1984) has reviewed the research on this phenomenon. It is not restricted to the continuity of a tone through a noise but is more general, dealing with the restoration of any sort of sound interpreted as having been masked. This effect has been interpreted by Warren, Obusek, and Ackroff (1972) as resulting from processes that try to compensate for the masking of one sound by another in natural environments.

Both auditory stream segregation and the continuity illusion can be seen as occurring because scene analysis processes have been tricked. The expression *scene analysis* originated in the field of computer vision, where it refers to a problem in interpreting images of natural scenes: If an object is to be correctly recognized, it is necessary that all the visible surfaces that belong to that object be grouped together and treated as parts of the single object. Scene analysis consists of exploiting regulari-

ties of the visible world in order to make accurate decisions about such grouping.

In audition, there is a similar problem that can be referred to as *auditory scene analysis* (Bregman, 1984, in press). It arises from the fact that in natural listening situations, more that one sound-producing event can occur at the same time. In order to recognize one of these events, its acoustic properties must be separated from those of the others. This can be difficult, because the acoustic pressure waves of all events are mixed at the ear of the listener. Although the auditory system appears to carry out a frequency analysis of the input, energy in any one frequency region could have originated from two or more events. Furthermore, energy in different frequency regions could have come from the same event. These facts require the auditory system to carry out a process of grouping.

Both stream segregation and the continuity illusion can be understood as consequences of this scene analysis process. Stream segregation occurs when the auditory system attempts to group auditory components into streams, each stream representing a single external source of sound. The continuity illusion occurs when the system interprets a sequence of acoustic inputs as the result of a softer sound's being interrupted by, and masked by, a louder one.

An argument can be made that the illusion of continuity must depend on the processes that produce auditory streaming, as follows: The continuity illusion depends on an interpretation of a sequence of inputs as the temporary interruption of one sound by another. But what counts as acoustic evidence for "temporary interruption"? We need to have a sequence of (at least) three segments of sound, A-X-B, with certain relations between them:

- A and X should not plausibly be part of the same stream;
- X and B should not plausibly be part of the same stream;
- however, A and B should plausibly be part of the same stream.

The third requirement (concerning A and B) is another way of saying that there has to be something coherent to be interrupted. Since streams are mental representations of such "somethings," one must interpret a sequence of events as a stream before it can be heard as interrupted. These three rules give only a subset of the requirements

for hearing continuity. Continuity through interruption is a special form of stream integration that occurs when it is plausible to interpret A and B not only as part of the same stream (as the third rule states), but also as a single continuous event within that stream. There are additional requirements for this interpretation. They include the facts that there should be no audible offset of A or onset of B, that there should be auditory neural activity present during X that is acoustically appropriate to the event whose percept is being restored, and that the transitions from A to X and from X to B should not be interpretable as Event A's changing smoothly into an X or Event X's changing smoothly into a B. However, we do not wish to go into these additional requirements at this point (see Bregman, in press, chapter 3). We only wish to assert that the requirements for the continuity illusion include those for stream integration.

There are two empirical consequences to this assertion. Since stream cues are part of the "interruption" decision and not vice-versa, (1) stream cues should make a difference to continuity, and (2) continuity or discontinuity cues should not make a difference to streaming.

The first of these consequences was supported by the results of a study by Ciocca and Bregman (1987). Listeners heard a pure-tone glide continue through a noise burst, to the extent that the entering and exiting portions of the glide either were in the same frequency region or fell on a common trajectory. The second prediction is the object of the present experiment. We looked at auditory organization in a situation in which a noise burst acted to mask elements that were also sorted by stream segregation into two simultaneous perceptual streams. Our main question was whether or not the perceptual continuity process would interfere with or change the stream organization. Our stimulus consisted of variations of an "X pattern" involving two tonal glides, one ascending in frequency and one descending, such that they crossed at a midpoint (see Figure 1, part A).

In this stimulus, the organization into streams is first determined by a process that segregates simultaneous glides from one another if they are not parallel on log-
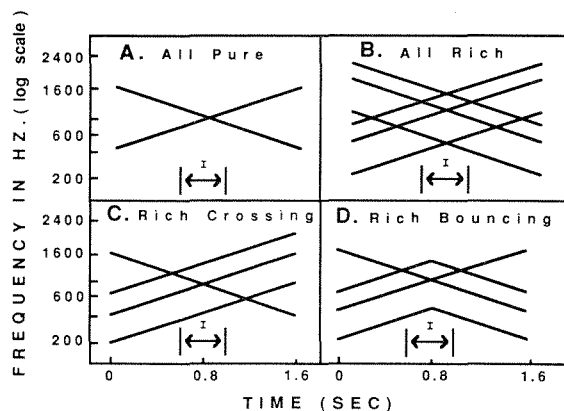
frequency coordinates and not harmonically spaced (McAdams & Bregman, 1979). Once this has been done, the sequential grouping of the segregated glides can be based on two further principles of grouping. The first, called the *proximity principle*, groups tonal elements on the basis of proximity in frequency. This has been observed in many experiments involving stream segregation. With our stimulus, it could be expected to group the two higher parts of the glides into one stream and the two lower parts into a second stream, to create what we called the *bouncing* percept (as in Figure 2A). For example, one stream would consist of the first half of the descending glide and the second half of the ascending glide. The stream would appear to descend to the midfrequency point and then bounce up again.

The second influence on grouping, called the *trajectory principle*, is thought to group tonal components if they fall into a regular trajectory on frequency × time coordinates (Ciocca & Bregman, 1987; Jones, 1976). This principle would be expected to keep the two halves of the descending glide together as a stream, and the two parts of the ascending glide together as a second stream, yielding what we called the *crossing* percept. Since the meeting of the glides at the midpoint offered two choices for perceptual grouping, the final stream organization would result from the competition of these two principles, with one possibly dominating the other in particular situations.

In the present experiment, the richness of the glides relative to one another (defined by the number of harmonics of which the glide consisted) was manipulated in order to selectively reinforce either the bouncing or the crossing organization. Richness (harmonic structure) has been shown to affect the likelihood of perceiving *crossing* (Tougas & Bregman, 1985).

Perceived continuity was studied by manipulating the stimulus properties in the temporal interval where the two glides crossed. There were three conditions: continuous glides, noise burst, or silent gap. For the noise condition, a 200-msec noise burst replaced the middle section of the X pattern (Interval I in Figure 1) in order to induce the illusion of continuity at the strategic point where the perceptual system had to decide between following the trajectories (i.e., crossing) or staying in the same frequency range (i.e., bouncing). The silent gap condition simply involved a 200-msec interval of silence replacing the glides at the cross-over point.

On the basis of our analysis of the relations between auditory streaming and perceived continuity, we expected that auditory streaming processes should yield the same organization into streams, independently of the presence or absence of perceptual restoration.

METHOD

Subjects

The listeners were 21 young adults who reported normal hearing.

Stimuli

The basic stimulus pattern was an X pattern consisting of two simultaneously gliding tones, one ascending and the other descend-

Figure 1. The four richness conditions used in this experiment. The interval I could be filled by glides (as shown), silence, or noise.

ing, both traversing the same frequency range in 1.64 sec. The richness of the glides relative to one another was manipulated in order to yield the four richness conditions depicted in Figure 1. In two of them, all the parts of the two glides were the same in richness. (By *richness*, we mean only to refer to the sound that occurs when three simultaneous partials are played.) The *all-pure* condition consisted of two pure-tone glides traversing the same two-octave frequency interval, between 400 and 1600 Hz. This was the baseline condition, in which the cues for crossing and bouncing percepts were balanced. The *all-rich* condition consisted of two glides, one ascending and the other descending, each containing a set of three harmonics, the middle harmonic of which corresponded to the sinusoidal glide used in the pure-tone condition. The components of the ascending rich glide thus had starting frequencies of 200, 400, and 600 Hz and ending frequencies of 800, 1600, and 2400 Hz, the endpoints being reversed for the descending glide. In the all-rich condition (as in the all-pure), the added harmonics did not favor either the crossing or bouncing percepts. Despite this, on the basis of previous studies (Deutsch, 1975; Tougas & Bregman, 1985; van Noorden, 1975) the all-pure and all-rich conditions were expected to produce a preference for the bouncing percept.

In addition, two conditions were intended to affect the perceptual organization, promoting either the bouncing or the crossing percept. The *rich-crossing* condition consisted of an ascending rich glide (moving from 200, 400, and 600 Hz to 800, 1600, and 2400 Hz) accompanied by a descending pure-tone glide (moving from 1600 to 400 Hz). This condition was expected to favor the crossing percept, because the two streams that would have to be formed to obtain this percept would be differentiated by number of harmonics. In previous studies (Halpern, 1977; McAdams & Bregman, 1979; Tougas & Bregman, 1985), richness differences between tones have been found to promote perceptual segregation. Finally, there was a *rich-bouncing* condition, in which the upper half of the X pattern consisted of pure-tone glides (moving from 1600 down to 800 Hz and back up again) while the lower half of the X pattern consisted of rich glides (moving from 200, 400, and 600 Hz up to 400, 800, and 1200 Hz and back down again). This condition was expected to favor the bouncing percept, because the two streams that would be created to form this percept would be different in their harmonic richness.

The contents of the 200-msec interval, centered at the temporal point at which the glides crossed, was manipulated in order to yield the following three conditions: In the *continuous* interval, the glides were left intact. In the *noise* interval, the glides were deleted and replaced with a noise burst. Finally, in the *silence* interval, the glides were replaced by silence. This last condition was meant as a control for the fact that the noise condition differed in two ways from the continuous condition, namely in the absence of glides and the presence of noise. A noise duration of 200 msec was chosen as being long enough to induce an appreciably long restoration effect without risking its loss. In previous studies, it has been found that subjects can perceive tones as being continuous over 300-msec breaks (Dannenbring, 1976; Warren, Obusek, & Ackroff, 1972). All the subjects from the present study verbally reported perceiving continuity when asked whether or not they heard the glides in the noise condition as being continuous (e.g., "with no breaks in the middle").

The glides were exponential changes in frequency, since these were expected to produce linear changes in perceived pitch. The loudness of the rich tones was subjectively adjusted (by four persons) to equal that of the pure tones.

Procedure

As is shown in Figure 2, one can basically foresee two possible perceptual responses to a set of simultaneous glides arranged in the form of an X pattern. It could be heard as a rising glide (R), ac-
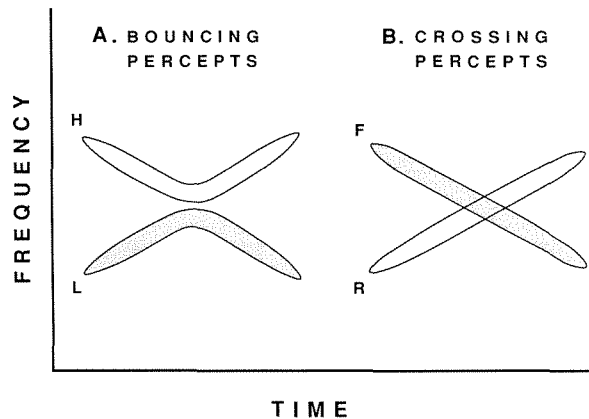

Figure 2. Perceptual contours for the bouncing percepts (H = a high glide falling and then rising; L = a low glide rising and then falling) and for the crossing percepts (R = a rising glide; F = a falling glide).

companied by a falling glide (F)—that is, as the crossing percept. Alternatively, it could be heard as the bouncing percept, consisting of a high glide falling, until it reaches the midpoint and then rising again (H), accompanied by a low glide rising to the midpoint and then falling again (L).

The subjects were shown the four contours of Figure 2 on four separate sheets of paper, and they received auditory demonstrations to familiarize them with the sound of the four contours. They were given the following instructions about the task: On each trial, they would be presented with five repetitions of a standard sound, followed 6 sec later by a comparison sound also cycled five times. They would then be given 4.9 sec to rate on a 7-point scale the clarity of isolation of the standard sound when heard as part of the comparison sound. The four X patterns were used as comparison sounds. There were four standards for each trial. Each resembled, in a plausible manner, one of the four ways of hearing the comparison that are exemplified in Figure 2. For example, the H standard for the rich-crossing condition had a pure tone descending to the temporal midpoint and a rich tone ascending from that point, all components of the rich tone starting at the temporal midpoint. Whenever a change from rich to pure or vice versa was required, it took place at the temporal midpoint; this point, of course, was audible only in the continuous condition.

Apparatus

The stimuli were synthesized digitally (at a 10-MHz sampling rate) on a PDP-11/34 computer, using the MITSYN software package (Henke, 1980), low-pass filtered at 5 kHz at output (48 dB/octave) by a Rockland Model 851 filter (Butterworth), recorded on audiotape, and played to subjects diotically over TDH-49P headphones. The intensity of glides at playback was determined by setting a pure 800-Hz steady-state reference tone, synthesized at the same intensity as a single glide component and recorded on the same tape, to 70 dBA. The digitally synthesized narrow-band noise burst was presented at about 80 dBA and had extremely sharp cut-off edges outside its 200–2400 Hz frequency range (40-dB attenuation at 100 and 2500 Hz).

RESULTS

The judgments of perceptual isolation of the four types of contours yielded scores ranging from 1 to 7, indicat-

Table 1
Mean Clarity Scores for Bouncing (B) and Crossing (C) Contours

| | Type of Richness | | | | | | | |
| | All-Pure | | All-Rich | | Rich-Crossing | | Rich-Bouncing | |
Interval Type	B	C	B	C	B	C	B	C
Continuous	5.3	3.3	4.3	3.4	2.0	6.6	6.6	2.7
Silence	5.7	3.0	5.3	3.6	3.6	6.4	6.3	2.4
Noise	5.2	2.7	5.0	3.4	2.7	6.1	6.0	3.0

ing "very clearly not isolated" and "very clearly isolated," respectively. The differences between H and L standards and between R and F standards were not of interest, since such differences do not pertain to the identity of the trajectories into which the pattern has been decomposed, but only to their salience. Therefore, the scores from the two frequency-proximity standards (H and L) were averaged to yield a bouncing score, whereas the scores from the two trajectory standards (R and F) were averaged to yield a crossing score. The mean scores for the different stimulus conditions are shown in Table 1. Generally speaking, the same pattern of results is seen for each of the three types of interval (i.e., for the three rows of the table).

The data of Table 1 were analyzed with a three-way ANOVA, with type of interval (continuous, silence, noise), type of glides (all-rich, all-pure, rich-crossing, rich-bouncing), and type of standard, or contour (crossing, bouncing), as factors. This revealed a very reliable main effect for contour [$F(1,20) = 21.35, p < .0002$], showing that the bouncing contours were heard more readily than the crossing ones. The interaction of richness with contour was significant [$F(3,60) = 112.46, p < .0001$], resulting from the fact that the bouncing contours were heard more easily than the crossing contours in all richness conditions except for the rich-crossing condition, where the reverse was true. Thus a strong tendency for the glides to cross was found only when one was rich and the other pure.

There were two significant interactions involving the type of interval, pointing to a small difference between the silent interval and the others. One of these was an interaction between interval type and contour type [$F(2,40) = 5.86, p < .006$], showing a small, but regular, enhancement of the bouncing effects in the silent interval. The other was an interaction between interval type and richness type [$F(6,120) = 2.94, p < .01$], showing that the silent interval produced slightly greater clarity of isolation for all the standards in the rich-crossing condition than in the other conditions. This small effect may have been due to the lack of complete randomization between trials in the experiment or to some unknown perceptual effect. In any case, the important results lie in the observation that, just as in the continuous and noise-filled intervals, the rich-crossing condition in the silence interval yielded crossing percepts that were still definitely more clearly isolated than the bouncing percepts; in other words, there was not much effect of interval on stream

formation. Finally, the three-way interaction between interval, richness, and contour type was marginally significant [$F(6,120) = 2.42, p < .03$], showing that under the continuous interval, the rich-crossing condition tended to yield slightly better crossing effects.

In order further to describe the behavior of the different richness conditions under the three interval types, a two-way ANOVA was performed on a bouncing-superiority score (the dominance of the bouncing percept), obtained by subtracting the mean crossing score from the mean bouncing score in each interval × richness condition for each subject. The mean scores for this measure are shown in Figure 3. They are expressed as a fraction of the maximum possible difference, +1 indicating complete dominance of the bouncing percept and −1 the opposite.

The main effect for richness, which was highly significant [$F(3,60) = 110.3, p < .0001$], accounted for over 12% of the total variance. Paired comparisons using the Tukey test revealed that the rich-crossing condition produced clearer crossing percepts than did all other richness conditions ($p < .01$), and that the rich-bouncing condition produced slightly clearer bouncing percepts than did the all-rich condition ($p < .05$). Although the main effect for type of interval was also statistically significant [$F(92,40) = 6.95, p < .003$], it accounted for only 0.02% of the total variance, and Tukey tests further revealed that none of the paired comparisons involving the means for interval type were significant at the .05 level. Finally, there was a significant interaction of richness with interval, suggesting that in the continuous interval the rich-crossing condition tended to yield slightly better crossing effects [$F(6,120) = 2.44, p < .03$]; however, it too accounted for only 0.02% of the total variance. There-
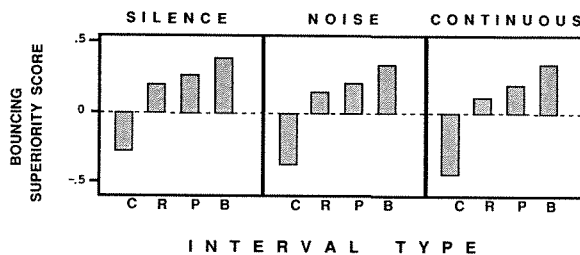

Figure 3. Mean bouncing-superiority score for each richness condition (C = rich crossing; R = all-rich; P = all-pure; B = rich-bouncing), and for each interval condition (silence, continuous, noise). The standard error of the mean is about 0.28.

ore, although the dominance of bouncing was not exactly he same in all intervals and not identically affected by he addition of richness, the differences were very small.

DISCUSSION

There are several results of interest in this experiment. First, we will present the secondary one, concerning the dominance of frequency proximity over trajectory as a principle of perceptual organization that we have observed in cases in which the stimulus pattern is composed of gliding tones of equal richness. This dominance has been observed in studies by van Noorden (1975), Deutsch (1975), and Tougas and Bregman (1985). The results of the present experiment are consistent with these studies, but they add new information, because the studies cited were done with two sequences of discrete tones alternating in time, whereas the present one was done with simultaneous continuous glides.

Some related results with glides have shown the relative weakness of the trajectory effect. Steiger and Bregman (1981) reported that when they alternated a pure-tone glide with a complex glide composed of two partials in an attempt to pull one partial out of the complex, there appeared to be no special capturing effect when the captor and target glides were aligned on a common trajectory. In another study, Dannenbring (1976) showed that when the peaks of alternately rising and falling frequency glide were deleted and replaced with noise, there was no perceptual extrapolation of the incomplete glide; instead, there seemed to be considerable rounding off of the trajectory of the glide.

The present finding that the rich-crossing condition favored crossing percepts does not contradict these two experiments. Our finding does not imply the presence of a trajectory effect per se, since the glides on the same trajectory also had the same harmonic structure; rather, it supports the results obtained by Halpern (1977), McAdams and Bregman (1979), and Tougas and Bregman (1985), which suggest that harmonic structure affects grouping.

The major finding of the present experiment is the striking similarity in the stream organizations across the three types of intervals. As can be seen in Figure 3, the three intervals yielded essentially the same stream formation and responded in the same way to the added harmonics. The finding that auditory streaming processes yielded the same perceptual organization independently of the presence or absence of perceptual restoration lends support to the idea that perceived continuity depends on the formation of streams. The results of the present experiment appear to us to be explainable in terms of a two-stage process. First, auditory streams were formed as a result of the competition between heuristics for grouping (see Bregman, 1978, in press). Then, the presence of acoustic information which suggested that part of the two streams had been masked activated a perceptual continuity mechanism, which acted to fill in the gaps, avoiding discontinu-

ities within the already formed streams. This account views perceived continuity as a process of interpolation between events that lie within an already formed auditory stream and as a process that cannot affect the formation of such streams.

To the best of our knowledge, the present experiment is the only one to have contained a condition that was expected to yield the perceptual restoration of the continuity of two streams concurrently. This was done in a situation in which the two streams themselves were created by stream-forming heuristics. Remember that the glides occurred simultaneously. Their perception as separate simultaneous sounds occurred on the basis of differences in frequency movement of the ascending and descending glides. Segregation based on frequency modulation has been demonstrated in unpublished research by Halpern (1977), cited by McAdams and Bregman (1979) and by Bregman and Doehring (1984).

The small differences between the effects obtained in the three intervals are explainable in fairly straightforward ways. For example, the slightly greater bouncing-superiority score obtained for the silent interval may be accounted for by the effect of frequency separation on auditory streaming (Bregman & Campbell, 1971; van Noorden, 1975). More specifically, the gap of silence separating adjacent glides slightly favors bouncing, since the frequency distance across which the bouncing glides must be joined is less than the distance across which the crossing glides must be joined in this condition. Similarly, the lack of a gap of silence in the continuous interval, by not separating the frequencies that have to be grouped, may intensify the effects of similarity-based grouping and account for the slightly clearer crossing percepts found for the rich-crossing condition. In any case, these differences that depend on the contents of the interval are very minimal, and they are overshadowed by the obvious similarities of the stream organizations across different types of intervals.

REFERENCES

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

BREGMAN, A. S. (1981). Asking the "what for" question in auditory perception. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 99-118). Hillsdale, NJ: Erlbaum.

BREGMAN, A. S. (1984). Auditory scene analysis. In *Proceedings of the Seventh International Conference on Pattern Recognition* (pp. 168-175). Silver Spring, MD: IEEE Computer Society Press.

BREGMAN, A. S. (in press). *Auditory scene analysis.* Cambridge, MA: MIT Press.

BREGMAN, A. S., & CAMPBELL, G. L. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology, 89,* 244-249.

BREGMAN, A. S., & DOEHRING, P. (1984). Fusion of simultaneous tonal glides: The role of parallelness and simple frequency relations. *Perception & Psychophysics, 36,* 251-256.

CIOCCA, V., & BREGMAN, A. S. (1987). Perceived continuity of gliding and steady-state tones through interrupting noise. *Perception & Psychophysics, 42,* 476-484.

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

DEUTSCH, D. (1975). Two-channel listening to musical scales. *Journal of the Acoustical Society of America*, **57**, 1156-1160.

HALPERN, L. (1977). *The effect of harmonic ratio relationships on auditory stream segregation*. Unpublished undergraduate thesis, McGill University, Psychology Department, Montreal.

HENKE, W. L. (1980). *MITSYN: A coherent family of command level utilities for time signal processing* [Computer program]. Copyright 1980 by W. L. Henke.

JONES, M. R. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review*, **83**, 323-355.

MCADAMS, S., & BREGMAN, A. S. (1979). Hearing musical streams. *Computer Music Journal*, **3**, 26-43.

STEIGER, H., & BREGMAN, A. S. (1981). Capturing frequency compo-

nents of glided tones: Frequency separation, orientation, and alignment. *Perception & Psychophysics*, **30**, 425-435.

TOUGAS, Y., & BREGMAN, A. S. (1985). The crossing of auditory streams. *Journal of Experimental Psychology: Human Perception & Performance*, **11**, 788-798.

VAN NOORDEN, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences*. Unpublished doctoral dissertation, Institute for Perception Research, Eindhoven, The Netherlands.

WARREN, R. M. (1984). Perceptual restoration of obliterated sounds. *Psychological Bulletin*, **96**, 371-383.

WARREN, R. M., OBUSEK, C. J., & ACKROFF, J. M. (1972). Auditory induction: Perceptual synthesis of absent sounds. *Science*, **176**, 1149-1151.