

Competition among auditory streaming, dichotic fusion, and diotic fusion

HOWARD STEIGER and ALBERT S. BREGMAN
McGill University, Montreal, Quebec, Canada

When a component of a complex tone is captured into a stream by other events that precede and follow it, it does not fuse with the other components of the complex tone but tends to be heard as a separate event. The current study examined the ability of elements of a stream to resist becoming fused with other synchronous events, heard either in the same ear or at the opposite ear. The general finding was that events in one ear fuse strongly with elements of an auditory stream in the other ear only when they are spectrally very similar. In this case, the fusion of simultaneous components at opposite ears is stronger than of simultaneous components heard in the same ear. However, when the spectra of the synchronous events are mismatched even slightly, components in the same ear fuse more strongly than components at opposite ears. These results are accounted for by a theory that assumes that decisions that perceptually integrate sequential events, synchronous events, and events at opposite ears are interdependent.

In natural listening situations, several sound sources are often simultaneously active. Therefore, at any instant in time, the energy present at the two ears can consist of the sum of an indefinite number of signals. This being the case, the auditory system is faced with the problem of subdividing the set of simultaneous components present at the two ears into perceptual units that correspond to separate sources.

Bregman and Pinker (1978) demonstrated a perceptual effect that might be heuristic in decomposing several cooccurring frequency components into separate perceptual units. They showed that when a pair of simultaneous pure tones, B and C, was placed in rapid alternation with a third pure tone, A, that was close in frequency to B, B and C were heard as separate pure tone elements. However, the same tones B and C fused into a single image with a richer timbre when A was far in frequency from B (see Figure 1). This effect appeared to reflect the use of a strategy that asked the question: "Was B more likely to have belonged in a unit with its synchronous counterpart C, or as part of a sequential organization with the event A which preceded and followed it?" Apparently, when A was similar (in frequency) to B, the auditory system assessed the probability of A and B both being members of a single sequen-

This research was supported by the Natural Sciences and Engineering Research Council of Canada and employed the facilities of the computer-based laboratory of the McGill University Department of Psychology. This study was performed in partial fulfillment of the requirements for Howard Steiger's PhD degree at McGill University. Requests for reprints should be sent to Albert S. Bregman, Psychology Department, McGill University, 1205 Dr. Penfield Ave., Montreal, Quebec H3A 1B1, Canada.

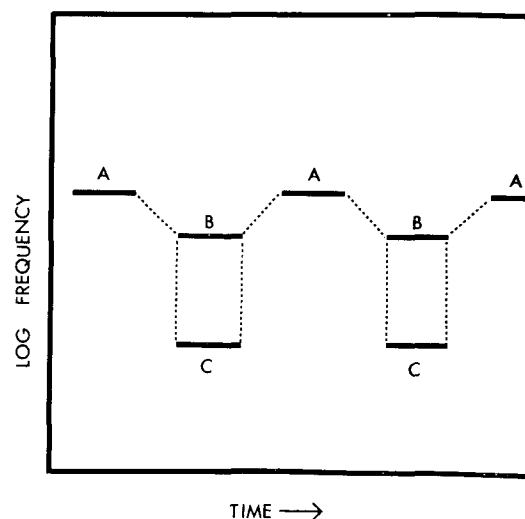


Figure 1. A representation of the stimulus employed by Bregman and Pinker (1978). The sequential attraction between the successive tones A and B (represented with diagonal dotted lines) competed with the fusion of the synchronous tones B and C.

tial organization as stronger than that of the synchronous elements B and C being parts of a single unit. As a result, the sequential A-B organization became prepotent. The sequential connections that develop between successive tones that are near each other in frequency have been called "streams" (Bregman, 1978), and the formation of streams is thought to result from heuristics for segregating individual sequential signals from acoustic mixtures. Bregman and Pinker's study suggests that sequential stream organizations compete with the tendency for synchronous events to fuse into complex tones

(Dannenbring & Bregman, 1978; Rasch, 1978). Such a competition might contribute to the assignment of cooccurring components to separate sources by allowing sequential cues to contribute to assessments of the likelihood that synchronous components arose from the same or different sources.

Recently, we (Steiger & Bregman, in press) found that sequential organization appears to compete with binaural integration in a similar manner. We placed the sequential organization effect into opposition with a binaural phenomenon called "contralateral induction" (Warren & Bashford, 1976). Contralateral induction refers to the tendency for the perceived position of a monaural pure tone to be shifted toward a simultaneous noise burst presented to the contralateral ear.

In our experiments, we presented an alternating pair of pure tones (A and B) to one ear and synchronized the presentation of contralateral noise bursts with each occurrence of B. The contralateral induction effect should have acted to produce a delateralized image of B, while streaming between A and B should have encouraged the perception of B as being part of the same signal as A and, hence, as having the same lateralized position as A. By varying the frequency proximity between A and B, we demonstrated that, although the contralateral induction effect delateralized the position of B when A was far enough in frequency from B to fail to "stream" with it, a sequential organization between A and B, created when A was close in frequency to B, inhibited the delateralization of B. Thus, sequential organization cues appeared to act to reduce the weighting assigned to binaural evidence during localization computations. In the same series of experiments, we also observed that, although the noise tended to color the timbre of tone B (when the tone and the noise were presented in synchrony), the "capturing" of B into a sequential stream with A could eliminate this effect. It appears, then, that streaming can oppose the integration of information across the ears both for the assessment of location and of timbre.

Such an influence of streaming on the use of binaural cues might improve the accuracy of perceptual computations in several ways: (1) It would add an additional cue for localization by favoring the assignment of all members of a stream to the same position in space. This extra information might help disambiguate situations in which interaural cues used to localize individual elements of a sequential signal (e.g., interaural intensity cues) were rendered unreliable, for example, by partial masking of the signal at one ear from extraneous sounds. (2) It might help the system to avoid anomalous percepts that would result when components from *different* sources haphazardly coincided in frequency at the

two ears. By interpreting such events in the light of previous events, the system might be able to determine that such components arose from different sources and, therefore, that they should not be fused and assigned a single position.

Although the results of Steiger and Bregman (in press) suggested that streaming cues influenced the use of binaural information, the converse relation also appears to hold: Binaural cues can affect streaming. For example, Judd (1979) found that when a sequence of tones was alternated between the ears, it split into streams on the basis of ear of presentation, rather than on the basis of frequency proximity between successive tones. Together, the results of Steiger and Bregman (in press) and of Judd (1979) suggest that binaural cues and sequential cues cooperate to determine a "what-is-where" decision, with neither class of cue contributing exclusively to either the "what" or the "where" aspects of the decision. Rather, the cues seem to function interdependently, each cooperating and competing with the other during perceptual decisions.

If the preceding speculation is correct, then one would expect sequential organization cues and binaural cues to interact in much the same way as did the sequential cues and the cues favoring fusion in the Bregman and Pinker (1978) study. In other words, these cues should compete during perceptual decisions that determine whether or not to fuse events occurring at opposite ears. For example, one might predict that the sequential grouping of elements at one ear would be antagonistic toward the fusion of any of those elements with synchronous events presented to the opposite ear. Conversely, when a monaural component at, say, the left ear, fuses with a synchronous element at the right ear, this should compete with the left-ear component's tendency to stream with other elements presented at the left ear. The current study attempted to verify this speculation.

Specifically, the study was designed to compare the degree to which sequential streaming between successive complex tones A and B could oppose the fusion of B with a synchronous tone C under two conditions: (1) when A, B, and C were all presented binaurally, as in the Bregman and Pinker (1978) study (we conceive of this condition as representing a competition between sequential A-B streaming and B-C fusion in the absence of interaural differences between B and C; we will refer to the fusion of different spectral components under such diotic presentation conditions as *diotic fusion*); (2) when A and B were presented monaurally, with C presented to the opposite ear (we conceive of this condition as representing a competition between sequential A-B streaming and B-C fusion, where B and C may or may not share frequency components but are presented to

opposite ears; we refer to fusion of spectral components under these presentation conditions as *dichotic fusion*). We will call the B and C tones "target" and "mask," respectively. The audibility of the target (tone B) served as an index of B-C fusion in all cases, since when B and C fused, B would cease to be audible as a discrete event. Tone A will be called "the captor," since it was designed to "capture" the target into a sequential A-B stream and, thereby, to prevent fusion of the target and mask.

In order to assess the ability of the captor-target sequence to compete with dichotic fusion of the target and mask, we wanted to manipulate interaural cues that would promote such fusion to varying degrees. Several factors are known to predispose a listener toward hearing simultaneous tones presented to the two ears as fused. A pair of pure tones, each presented to one ear, readily fuse providing they have less than a critical difference in frequency, which is proportional to the frequency of the tones themselves (Perrot & Barry, 1969; Perrot, Briggs, & Perrot, 1970). Thus, there appears to be a limited bandwidth at each frequency within which energy at the two ears becomes perceptually combined into a fused image. Other research, in examining the fusion of complex sounds presented to the two ears, has led to a more general impression that the correspondence between the global spectral features of complex events at each ear is a determining factor in fusion. For example, formants heard dichotically will fuse if they are tuned to the same fundamental frequency, but do so less readily when their fundamentals are mistuned (e.g., Broadbent & Ladefoged, 1957; Darwin, 1981). Thus, by varying the interaural correspondence in frequency of sinusoidal components of complex tones, as well as by varying the fundamentals of complex events at each ear, one should be able to promote or to discourage dichotic fusion.

In the current experiment, we manipulated both the "local" and global correspondence of dichotic targets and masks. Our target and masking tones contained from two to four partials, and the correspondence between them was varied: (1) globally, by making the fundamental frequency of the mask either the same as, or different from, that of the target, and (2) "locally" on the spectrum, by varying the number of the mask's partials that matched, or nearly matched, the frequency of partials of the target (see Figure 2a). When partials of the mask matched some of the target's partials, we could expect these "local" matches to enhance dichotic fusion. When no such corresponding partials were present, one would expect fusion to be weaker. In either case, we expected that, when tuned to the same fundamental, target-mask fusion would be superior, due to the action of a global analysis of the tones' spectra.

To compare the effects of streaming on fusion of the target and mask under these *dichotic* conditions

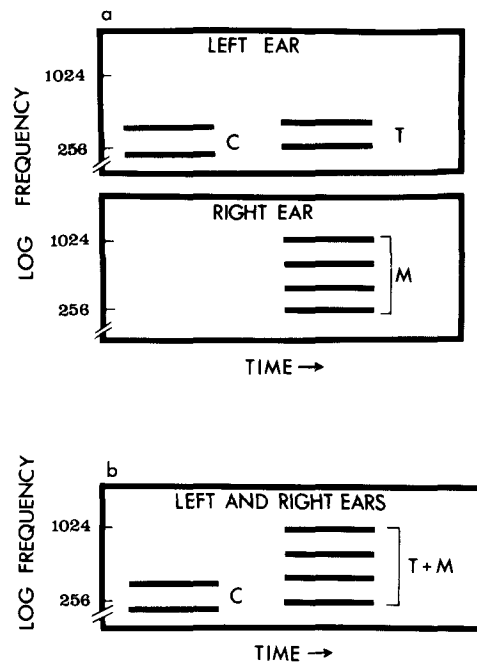


Figure 2. A representation of the frequency relations between partials in some of the stimuli used in the present experiment. A sample dichotic stimulus is shown in a above, and a sample diotic stimulus is shown in b. The partials of each tone are represented as horizontal lines. The captor, target, and masking tones are labeled C, T, and M, respectively. The dichotic target and mask shown in the figure have two partials corresponding in frequency and the same fundamental. In other conditions, one or no corresponding partials were present, and the tones' fundamentals could differ.

with its effects on *diotic* fusion of a target and mask under comparable conditions, we created diotic counterparts to each of the stimuli described above. This was achieved simply by acoustically mixing the same stimuli prepared for the dichotic condition, so that the captor, target, and mask tones, formerly presented each at one ear, were instead all presented to both ears (see Figure 2b). Thus, in this latter case, the sequential captor-target organization again competed with the fusion of the target and mask, but this time it was a competition with diotic fusion.

In summary, the experiment reported below explored the following: (1) the ability of a sequential stream to overcome diotic and dichotic fusion of various pairs of complex tones; (2) the effect of a common fundamental on the fusion (diotic or dichotic) of a pair of synchronous complex tones; (3) the effect of the number of corresponding partials in complex tones presented to the two ears on dichotic fusion.

METHOD

The Task

To measure the strength of fusion in each condition, we used the method of adjustment. Listeners initially heard only the repeating captor-target sequence and then gradually added the masking

tone to the signal by raising its level by turning a knob that they could control. They were told to increase the level of the mask to the point at which they became unsure as to whether or not the target was still audible. That point should reflect the level at which the target and mask began to fuse into a single image. We assumed that in conditions in which the target and mask fused more readily, the mask would not have to be made as intense to overcome the sequential attraction between the captor and the target, as compared with conditions in which the target and mask were less inclined to fuse.

Stimuli

All stimulus patterns were similar in structure, consisting of a target tone and a synchronous masking tone alternated with a captor tone near in frequency to the target. The target tones always consisted of two frequency components, either the first and second or the third and fourth harmonics of a 256-Hz fundamental. The target consisting of harmonics 1 and 2 will be called "low target," and that consisting of harmonics 3 and 4 will be called "high target." Different masking tones were designed so that, in each, a different number of partials would overlap in frequency with those of each target (see Table 1). When fully overlapping, the maskers consisted of harmonics 1, 2, 3, and 4 of a 256-Hz fundamental. Thus, two of the masker's frequency components matched the components of either target (either harmonics 1 and 2 in the case of the low target, or harmonics 3 and 4 in the case of the high target). When only partially overlapping, the maskers contained three harmonics of a 256-Hz fundamental, only one of which matched one component of each target. For the low target, the masker contained harmonics 2, 3, and 4, and for the high target, the masker contained harmonics 1, 2, and 3. When nonoverlapping, the low target was masked with a masker composed of harmonics 3 and 4 of a 256-Hz fundamental, and the high target was masked with harmonics 1 and 2 of a 256-Hz fundamental. Collectively, the set of maskers described above will be referred to as "same fundamental masks."

In addition, maskers corresponding to each of those described above (i.e., consisting of the same sets of partials), but tuned to fundamentals different from that of the targets, were used. These will be called "different fundamental masks." When a low target was used, the maskers were set to a higher fundamental of 263 Hz. When a high target was used, the maskers had a proportionately lower 249-Hz fundamental (see Table 1). In this manner, the frequency relations between tones in the different stimuli were kept

symmetrical. We mistuned the fundamental of different fundamental masks by 4% of an octave, since this difference fell near the threshold reported by Perrot and Barry (1969) at which pure tones at different frequencies will fuse across the ears. The partials of the targets that were near in frequency to those of the masks should thus have had some tendency to fuse dichotically, even in the different-fundamental conditions, if dichotic fusion is determined in frequency-specific binaural channels.

The captor tones for each of the high and low targets consisted of the same two harmonics as were in their respective targets. To enhance the discriminability of the captor and target tones, the fundamental of the captors was shifted slightly with respect to the fundamental of the targets. For the high target (which consisted of harmonics 3 and 4 on a 256-Hz fundamental), the captor was shifted slightly *upward* in frequency by an arbitrary value. It consisted of harmonics 3 and 4 on a 263-Hz fundamental. For reasons of symmetry, the captor for the low target (which consisted of harmonics 1 and 2 on a 256-Hz fundamental) was shifted *downward*, so that it consisted of harmonics 1 and 2 on a 249-Hz fundamental. The captors remained close enough in frequency to the targets to permit the tones to stream had they been alternated without a masker tone being present.

By crossing all levels of the factors described above (high/low target \times target-mask overlap \times fundamental of mask), 12 stimuli were produced. The 12 patterns were recorded on stereo audio tapes, with the captor and target tones always on one channel and the masking tones always on the other. In order that the identical set of 12 stimuli could be presented in both dichotic and diotic modes, the amplitudes of partials in the target and masking tones which corresponded in the two ears in the dichotic condition were halved during synthesis. Thus, the two tracks of the tape could be mixed for diotic presentation, leaving all frequency components equal in amplitude (i.e., the halved components would be added together so that the resulting intensity equaled that of the full amplitude components). Similarly, when presented dichotically (i.e., in stereo), all frequency components had equal energy when summed across the ears. To keep the conditions comparable in both same- and different-fundamental conditions, the same compensation was performed on corresponding partials of the target and masking tones (even though no partials of the masks corresponded perfectly to those of the targets in the different-fundamental condition). All noncorresponding components of the targets and masks, as well as the components of the captor, had full amplitude. In addition, during diotic presentation, the inten-

Table 1
Frequency Components of Tones in Different Stimulus Conditions (in Hertz)

	Captor	Target	Mask
Low Target: Same Fundamental			
Fully Overlapping	249,498	256,512	256,512,768,1024
Partially Overlapping	249,498	256,512	512,768,1024
Nonoverlapping	249,498	256,512	768,1024
Low Target: Different Fundamental			
Fully Overlapping	249,498	256,512	263,526,789,1052
Partially Overlapping	249,498	256,512	526,789,1052
Nonoverlapping	249,498	256,512	789,1052
High Target: Same Fundamental			
Fully Overlapping	789,1052	768,1024	256,512,768,1024
Partially Overlapping	789,1052	768,1024	256,512,768
Nonoverlapping	789,1052	768,1024	256,512
High Target: Different Fundamental			
Fully Overlapping	789,1052	768,1024	249,498,747,996
Partially Overlapping	789,1052	768,1024	249,498,747
Nonoverlapping	789,1052	768,1024	249,498

sity of the signals was reduced by 3 dB at each ear, so that the tones heard binaurally in the diotic condition matched the loudness of tones heard monaurally in the dichotic condition. (The value of 3 dB was derived during a pretest in which three experienced listeners matched the loudnesses of binaural and monaural tones.) The intensity of the signals was calibrated so that any monaural two-component complex tone was presented at 70 dB SPL. Complexes containing three or four components, or those containing components of half amplitude, were correspondingly more or less intense.

It should be noted that when the signals were mixed for diotic presentation, the spectra of the resulting target-mask complexes in the same-fundamental treatment were identical in all overlap conditions. In the different-fundamental treatment, this was not the case, since the masks contained frequency components different from those of the targets (see Table 1).

All tones were 120 msec in duration (including exponential rise and fall times of 10 msec). Each captor and target-mask pair was cycled repeatedly, with 30 msec of silence between the captors and targets and 138-msec silences embedded between consecutive cycles.

The stimuli were digitally synthesized on two channels at a 10-kHz/channel sampling rate. The digital-to-analog converter introduced a slight phase lag on the channel containing the maskers relative to the channel containing captors and targets. To compensate, the former was low-pass filtered at 3.04 kHz and the latter at 2.5 kHz, using linear phase filters during recording. Different delays of the signal, introduced by the asymmetrical filtering, canceled the undesired phase shift.

Four different tapes were prepared, each containing the 12 stimuli described above (75 cycles of each) recorded in a different random order. In addition, a practice tape containing 6 stimuli that represented a sample of the range of experimental stimuli was prepared. However, all frequencies of partials in the practice stimuli were shifted to frequencies unlike those of the experimental stimuli to avoid specific carry-over effects.

Procedure

Instructions. With written instructions and diagrams, listeners were informed that their task on each trial was to listen to an alternating pattern composed of two tones, A and B, and then to introduce and raise the level of a third tone, C, synchronous with B, until they became unsure that B was still present on every repetition of the pattern. To make this judgment, listeners were told to do the following: (1) to listen to the alternations of A and B in the absence of tone C; (2) to raise the level of C until B became inaudible; (3) to lower C until B could barely be heard again; (4) to raise C again to the point at which it became unclear as to whether B was present on every cycle of the pattern or not. At that point, they were to record the value visible on a digital meter (which registered the position of the adjustment knob) in the appropriate space on their response sheets. We felt that allowing the listeners themselves to record the values in this way could not have introduced any memory effects from trial-to-trial that might have biased the results in favor of the predicted effects. Any attempt on the part of the listener to maintain consistency across responses would, instead, have created a bias against differences between the conditions. The response sheet also provided a blank for the response "Tone B was always clearly audible." Listeners were told to tick this blank if B was clearly present at even the highest intensity of C (i.e., if it never fused with or became masked by C). The listeners were also informed that, on some sessions, A and B would initially appear to be at one ear, with C at the other, while on other sessions all three tones (A, B, and C) would seem to be centered in their heads. Listeners were told to ignore the locations of tones during their judgments of all stimuli.

Testing. Listeners were tested individually in an audiometric chamber. Prior to receiving his or her first dichotic and diotic session, each listener heard the six practice trials presented in

dichotic or diotic modes, respectively, and was required to make the desired judgment of each stimulus without corrective feedback. Four consecutive test sessions were conducted, with a short break (approximately 2 min) provided after each session. In each session, a different tape (containing all 12 stimuli) was used. On two of the sessions, the stimuli were heard in the dichotic mode (i.e., in stereo), and on the other two, in the diotic mode (i.e., with both channels mixed, and presented binaurally). Between each of the two diotic and two dichotic sessions, the position of the headphone speakers on the listener's ears was reversed, so that any ear dominance effects would be canceled out. The serial order in which trial randomizations, dichotic or diotic presentations, and headphone positions were used was random, with the restriction that each randomization, presentation mode, and headphone position occurred equally often in each serial position across the sample of listeners.

Apparatus. The stimuli were synthesized digitally on a Digital Equipment Corporation (DEC) PDP-11/34 computer, using the MITSYN software package (Henke, Note 1). Signals output by the computer's digital-to-analog converter were filtered with a Rockland 851 two-channel filter and recorded on stereo audio tapes using a Marantz SD9000 programmable cassette deck. This deck was used to present the stimuli because it permitted the listeners to advance the tape to the next trial by pressing a button once they had completed their responses.

The signals were amplified and presented over Sennheiser HD-414 headphones. The output of the playback equipment was measured at the headphones and found to be flat (within ± 1 dB) in the frequency range encompassed by our stimuli. Prior to amplification, the channel carrying the masking tones was passed through one channel of a Pioneer SA-8500 II preamplifier so that the listeners could control the mask's intensity by adjusting the preamplifier's gain control. A reference signal was fed through the other channel of the preamplifier, amplified, and then read on a digital multimeter. Since a single gain control controlled the level of both preamp channels, increasing the level of the mask caused a proportional increase in the voltage of the reference signal. Thus, the voltage of the reference signal provided an index of the sound pressure level at which the mask was presented. The levels of all signals were measured using a General Radio 1551-C sound-level meter with a flat plate coupler.

Listeners. Sixteen listeners were recruited from the McGill University graduate and undergraduate student body. All listeners reported having normal hearing.

Data analysis. The values recorded by listeners from the digital voltmeter were converted to masker sound pressure levels in decibels (B weighting), with care taken to consider the fact that different masker conditions had different numbers of partials and, hence, different intensities at the same voltage reading (which measured, in effect, the gain to which the adjustable amplifier had been set). Since it was of no interest to examine specific ear dominance effects (and the experimental design was not appropriate for this purpose), the results were averaged across headphone positions. Due to the fact that there was a maximum level to which we were willing to let listeners adjust the maskers, the intensity data were subject to a ceiling effect. As a result, there was some heterogeneity of the variances in various cells, since in some conditions listeners often adjusted the mask to the ceiling value but in others they did not. Nevertheless, since F tests are known to be robust against violations of the assumption of homogeneous variances when there is the same number of observations in all cells (Myers, 1972), and since many of the effects of interest required an examination of interactions, F tests were used in the main analysis.

As we pointed out in the Stimulus section, in the diotic conditions, when the targets and masks had the *same fundamental frequency*, the stimuli were identical, regardless of the overlap factor. This was not the case in any other condition of the experiment. Nevertheless, the same-fundamental/diotic condition was presented three times, once for each level of the overlap factor, to

preserve the symmetry of the experimental design. The reader should, therefore, bear in mind that the three values shown in Figure 3 at different levels of overlap in the same-fundamental/diotic condition reflect the responses of the listeners to identical stimuli. The fact that these stimuli were identical makes the interpretation of main effects of, or interactions with, the overlap factor in same-fundamental/diotic conditions meaningless. All effects of interest were evaluated by either planned comparisons that tested simple main effects and simple interaction effects or else by Scheffé tests, with both types of tests being performed among those means for which comparisons were valid.

RESULTS AND DISCUSSION

Figure 3 shows the mean masker sound pressure level (in decibels) required in each condition to cause the target to become inaudible. Recall that a lower level indicates that the target lost its separate identity even with low intensities of the masker. Therefore, one can think of the masker level scale as a "perceptual separability scale," with high and low values indicating high and low separability, respectively, of target from mask. It should be noted that, due to the fact that there was an upper limit on the masker intensity, the higher values probably underestimate the level that would have been required to actually obscure the presence of the target. Higher values do, however, indicate conditions in which target-mask fusion was weaker than it was in conditions that yielded lower values. To provide a measure of inter-

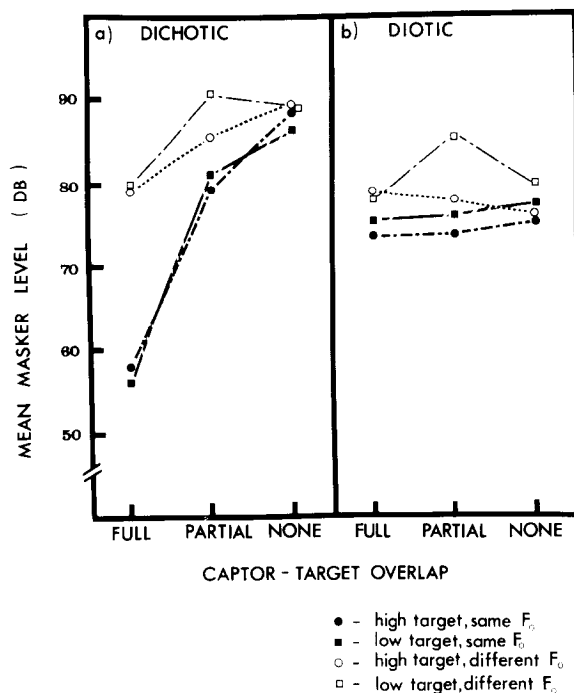


Figure 3. The mean masker level (in decibels) required to obscure the target in each condition. Means for the dichotic conditions are shown in a above, and those for the diotic conditions are shown in b.

subject variability on the current task, the standard deviations for each of the conditions are shown in Table 2. In addition, to provide a measure of within-subject reliability on the task, we computed Pearson product moment correlations, correlating the pairs of responses provided by each listener for each condition. The mean of the within-subject correlations was .56, with a range of .16 to .80. Thus, the performance of subjects was fairly consistent.

A major question in the current study was whether dichotic fusion or diotic fusion of simultaneous components opposes an antagonistic sequential attraction more strongly. A comparison of the values in the dichotic conditions with those in the diotic conditions suggests that the answer to this question is twofold: dichotic fusion was much stronger than diotic fusion for the spectrally similar targets and masks. For example, the means for the conditions in which the targets and masks were fully overlapping and on the same fundamental (i.e., the conditions containing the most spectrally matched target and masking tones) are lower (by a factor greater than 15 dB) in the dichotic than in the diotic conditions for both high and low targets. The difference was tested with a Scheffé contrast that compared the mean of the fully overlapping, same-fundamental conditions of the dichotic mode with the mean of all three overlap values in the same-fundamental conditions of the diotic mode. The difference was very significant [$F(1,30) = 230.54, p < .001$]. However, the opposite pattern of differences is found when comparing the diotic and dichotic conditions in which the targets and masks were nonoverlapping and on different fundamentals. Diotic fusion was stronger than dichotic fusion in the conditions that contained the most spectrally dissimilar target and masking tones, with both high and low targets (see Figure 3). A Scheffé test that compared the pooled high- and low-target conditions within these treatments showed the difference to be significant [$F(1,30) = 57.59, p < .01$].

These results suggest that when a complex tone results from the dichotic fusion of two *spectrally similar* elements (a target and masker having overlapping partials and the same fundamental), it strongly resists being decomposed when an antagonistic sequential attraction from a captor tone acts on the target. However, when integrated through diotic fusion, the same elements are more easily separated when a captor exerts a competitive attraction on the target. On the other hand, when the spectra of the targets and masks are *less similar* (i.e., when their partials do not overlap in frequency and when they are tuned to different fundamentals), dichotic fusion becomes somewhat weaker, losing its advantage over diotic fusion.

The data in Figure 3a show a fairly regular pattern within dichotic conditions. The target and masker tones on the same fundamental were less separable

Table 2
Standard Deviations About the Means in Each Condition

Target	F _o	Overlap					
		Dichotic			Diotic		
		Full	Partial	None	Full	Partial	None
High	Same	7.22	8.22	2.87	8.38	9.35	8.86
High	Different	10.77	7.40	1.95	12.02	11.68	8.80
Low	Same	5.33	9.78	8.06	12.37	12.68	10.23
Low	Different	9.18	1.05	2.62	14.38	5.66	10.39

than those on different fundamentals only to the extent that the target and mask contained overlapping partials. Note that in the fully overlapping, dichotic condition, when the high or low targets were tuned to the same fundamental as their masks, the tones fused much more readily than when on different fundamentals. A similar although attenuated effect of the correspondence of fundamentals is evident in the partially overlapping, dichotic conditions, again for both high and low targets. However, when the targets and masks were nonoverlapping, there appears to have been no effect of the correspondence of fundamentals. This pattern of results was examined using a test of the simple interaction between the same-/different-fundamental factor and the overlap factor (within the dichotic condition only), using Satterthwaite's degrees of freedom (Winer, 1962). The interaction was significant [$F(2,59) = 55.70$, $p < .001$]. We cannot be sure that the means in the same- and different-fundamental conditions in the nonoverlapping treatment were not forced to similar values by the ceiling effect discussed earlier. Nevertheless, even if one omits any consideration of the nonoverlapping conditions and looks only at the fully and partially overlapping, dichotic conditions, it appears that the effect of the fundamental was stronger in the fully overlapping condition. The simple main effect of the fundamental factor within the dichotic conditions was found to be significant on Satterthwaite's degrees of freedom [$F(1,35) = 120.19$, $p < .001$].

Since, in the dichotic conditions the number of overlapping frequencies at the two ears seemed to determine whether or not tuning the target and mask to same or different fundamentals had any effect, it appears that dichotic fusion depended upon the presence of corresponding partials at opposite ears, and not upon corresponding fundamental frequencies at the two ears. If fundamental frequency per se were the critical factor, the tuning of masks and target to same or different fundamentals should have made a difference even in the nonoverlapping dichotic treatment. Thus, the apparent effect of the fundamental factor seems not to have been a genuine effect of fundamental frequency on dichotic fusion, but simply a fusion

effect that was enhanced when local matches between opposite-ear partials were present. In the current experiment, we find a 4% octave mistuning of corresponding frequency components at opposite ears sufficient to drastically reduce their fusion. Cutting (1976) suggested that forms of dichotic fusion that are highly sensitive to relative frequency differences in signals at opposite ears are linked directly to a low-level mechanism which bases fusion upon interaural matching of energy in frequency-specific bands. Our assumption that fusion at such a level was active in our dichotic conditions is consistent with Cutting's proposal.

In the diotic mode, tuning the target and mask to the same or different fundamentals showed no effects as large as those evident in the dichotic condition (see Figure 3). However, a test for the simple main effect of the fundamental factor (using Satterthwaite's degrees of freedom) revealed that the targets and masks fused significantly less in the different-fundamental than in the same-fundamental conditions [$F(1,35) = 14.78$, $p < .001$]. (The interaction effect of the fundamental and overlap factors was meaningless in the diotic condition, since all the same-fundamental stimuli were identical, and the interaction was therefore not tested.) In the diotic case, the effect of the fundamental factor cannot be attributed to the matching of corresponding partials at opposite ears and, therefore, has to be different in kind from the one observed in the dichotic conditions. The differences in fusion seem to be due directly to a genuine effect of fundamental frequency on fusion decisions, events on different fundamentals appearing to bias the system against fusion.

Our analysis also showed a highly significant simple main effect of the overlap factor within the dichotic conditions, when tested on Satterthwaite's degrees of freedom [$F(2,52) = 155.97$, $p < .0001$]. There is, however, a problem in interpreting this effect. As we pointed out in the stimulus section, the frequency components in targets and masks which occupied corresponding frequency regions were reduced in intensity for the purpose of maintaining equal intensity between the different frequency regions, since energy at some frequencies was present in both the mask

Table 3
Mean Number of Trials in Each Condition in Which the Target Was Reported to Remain Audible Even at the Highest Level of the Mask

Target	F ₀	Overlap					
		Dichotic			Diotic		
		Full	Partial	None	Full	Partial	None
High	Same	.00	.50	1.25	.19	.25	.19
Low	Same	.00	.81	1.63	.31	.56	.94
High	Different	.81	1.25	1.81	.44	.44	.38
Low	Different	.75	1.88	1.88	.38	.94	1.13

Note—The means are based on two replications (one in each of the two headphone positions) of each condition.

and the target and energy at other frequencies was present only in one or the other tone. Since different numbers of these corresponding partials were present in different overlap conditions, the adjustment meant that the ratios of target-mask intensity also differed slightly at each level of the overlap factor. Such differences might be sufficient to account for the small differences evident as a function of the overlap factor in the different-fundamental conditions (see Figure 3). However, any effects of these differing target-mask ratios are certainly not enough to account for the large difference which occurred between non-overlapping and fully overlapping conditions of the same-fundamental/dichotic treatments, which is on the order of 30 dB. We, therefore, interpret this latter effect as being due to an influence of the number of corresponding partials at the two ears on fusion (as described earlier in our discussion of the interaction effect of the same-/different-fundamental and overlap factors within the dichotic condition). We should point out that, since target-mask intensity ratios were identical within each overlap condition, all of the effects involving the fundamental factor that we have discussed earlier are unconfounded by differences in intensity ratios between overlap conditions.

In Table 3, the mean number of trials (of 2) in which the target was reported to remain clearly audible at the highest masker level is shown. It can be seen that the differences among conditions on this latter measure correspond well to the differences observed among masker intensity settings. In those conditions that yielded high levels of masker intensity (see Figure 3), the mask was more often judged to be insufficient to obscure the target.

GENERAL DISCUSSION

We can summarize our findings as follows: (1) A complex captor tone is able to strip a complex component from a richer mixture when the two are alternated rapidly and heard diotically. Thus, the competition between sequential streaming and diotic

fusion observed by Bregman and Pinker (1978) using sinusoidal stimuli appears also to operate on spectrally more complex stimuli. The diotic fusion of the target and mask only becomes strong enough to resist the antagonistic effect of the captor-target stream when the mask is made fairly intense relative to the target (see Figure 3). (2) The dichotic fusion of two complex spectrally matched tones (the target and mask) is very resistant to decomposition when an antagonistic sequential organization attempts to "capture" the target. This resistance is much stronger than that provided when identical target and masking tones fuse diotically. (3) When the spectra of target and masking signals are mismatched even slightly, dichotic fusion loses its advantage over diotic fusion. (4) Dichotic fusion appears to be unaffected by fundamental frequency per se. Instead, there seems to be a strong effect that depends upon whether or not partials of the target correspond exactly in frequency to those of the contralateral mask. However, in the diotic conditions, there was some indication that a smaller effect, attributable to an analysis of the fundamental frequencies of the target and mask tones, may have caused stronger fusion of targets and masks when they were on the same fundamental.

Consider how streaming cues and fusion cues might have cooperated to yield a perceptual image of "how many" and "what" events occurred in the present stimuli. Diotic fusion is, presumably, concerned primarily with decomposing complex waveforms into smaller units that are then assigned to individual sound sources. In so doing, the system contributes to determining "what" and "how many." Since natural sound sources often emit energy spread over a wide frequency range, one would expect that the diotic fusion mechanism should have a large tolerance to differences in frequency. However, the auditory system need not, due to this fact, fuse any set of synchronous components whatever. It can trade this "willingness" to fuse widely spread components against other evidence by considering: (1) Do any of the group of simultaneous components at one instant (for example, those that compose the target in the "target plus masker" complex) seem to form the extension of events that occurred previously? (2) Is there any evidence at that one instant to suggest that some of the components *do not* belong together? To deal with the first question above, the system might allow streaming cues, which indicated that the target should be integrated with the earlier occurring captor, to influence its fusion decision. To deal with the second question above, the presence of inharmonically related components might bias the system against the fusion of a set of simultaneous components into a single unit, for such relations would often indicate that the signals from two or more simultaneously active sources had become mixed.

Another result of the current experiment was that strong dichotic fusion (which resulted when the target and mask had matching spectral components and the same fundamental) appeared to easily overcome the streaming between the captor and target tones. Thus, the target was perceptually "excised" from the captor-target stream and fused with the mask. The system had, in this context, appeared to give precedence to the evidence that suggested that the partials of the target should be assigned to the same source as the mask over the evidence that suggested that the target belonged in a stream with the captor.

This result, taken alone, suggests that dichotic fusion takes precedence over sequential organization rules during processes that organize complex signals into their component sources. One account for such an effect might be that the binaural interaction that causes dichotic fusion occurs at a level in the auditory system prior to the level involved in the streaming analysis. If this were so, then sequential organization rules might never act to "reverse" an earlier binaural fusion decision. However, there are at least two indications that sequential cues do interact with binaural localization cues, derived from the low-level integration of information across the ears: (1) The results of Bregman and Steiger (in press) discussed in the introduction suggested that sequential streaming cues competed effectively with the contralateral induction effect. Warren and Bashford (1976) pointed out that contralateral induction is closely related to auditory masking and suggested, therefore, that contralateral induction is a low-level auditory phenomenon. Streaming cues therefore do appear to influence at least this one peripheral binaural effect. (2) Furthermore, in one experiment recently completed in our laboratory, we found that sequential streaming cues could compete effectively with interaural time-of-arrival cues (see Mills, 1972) in determining the locations of tones in rapid sequences. Tones that were close enough in frequency to group sequentially were judged to be in similar positions despite time-of-arrival cues designed to cause the tones to have very different perceived positions. Tones with identical timing features but further apart in frequency were localized in positions more consonant with the time-of-arrival cues. Therefore, sequential organization cues seem to compete with the binaural interactions that occur in sound localization. Whether or not sequential cues also compete with the binaural interactions that result in events at opposite ears becoming fused into a single image remains to be determined. The current study demonstrates no such effect.

The current study does, however, demonstrate that the fusion of dichotic target and mask tones that share common partials can prevent the sequential organization of the target with a previous captor tone.

The following account explains this effect in a manner consistent with the belief that sequential cues and binaural cues have a mutual influence: Consider, for example, the likely action of binaural cues in the stimulus shown in Figure 2a. The captor is presented monaurally to the left ear. The tone should thus have been assigned a position at the left ear. The partials of the target, while also presented to the left ear, correspond in frequency to the lower pair of partials in the mask, presented to the right ear. These partials, therefore, constitute a binaural event, which may have been assigned a somewhat centered position on the basis of the interaural match at specific frequencies (Mills, 1972). The remaining partials of the mask are presented monaurally to the right ear and should have been assigned a position at the right ear. Thus, on the basis of interaural cues, the captor may have been assigned a position different from that of the partials of both the target and the mask. If, as Judd (1979) suggested, tones that are assigned different spatial positions are also less likely to group into sequential organizations, then the binaural evidence suggesting that the partials of the captor and target tones arose at different positions may have reduced the tendency for the captor and target to be assigned to the same stream. This would have had the consequence of reducing any impediment to dichotic fusion by eliminating any antagonistic sequential organization tendency.

An important point should be noted here: When the dichotic target and mask fused, a single image resulted, despite the fact that not all of the tones' partials corresponded between the ears. The target-mask fusion in this study thus consisted of the fusion of binaural events with monaural events. This result has two implications: (1) The interaural cues that we have described above could not, alone, have controlled the final percept since, based solely on interaural frequency disparities, the target should always have been differentiated from the noncorresponding partials of the mask. Therefore, it appears that, while interaural cues *contributed* to the percept (perhaps in part by acting against sequential streaming and in part by favoring fusion), the percept itself could not have resulted from the independent use of interaural cues as a source of evidence. (2) The interaural interactions involved in dichotic fusion of complex tones may not be strictly limited to frequency-specific binaural channels.

Together, the current results and those of our other studies described earlier suggest that the manner in which sequential organization cues and binaural cues are used allows for mutual influences between them. Similarly, the current study and that of Bregman and Pinker (1978) suggest that mutual influences occur between sequential streaming and dichotic fusion effects. Therefore, it seems appropriate

to characterize the action of streaming cues, diotic fusion cues, and dichotic cues as being interactive, rather than conceiving of each class of cue as making an independent contribution to auditory perceptual images.

REFERENCE NOTE

1. Henke, W. L. *MITSYN: An interactive dialogue language for time signal processing*. Cambridge: Massachusetts Institute of Technology, Research Laboratory of Electronics, 1975.

REFERENCES

- BREGMAN, A. S. The formation of auditory streams. In J. Requin (Ed.), *Attention and performance VII*. Hillsdale, N.J.: Earlbaum, 1978.
- BREGMAN, A. S., & PINKER, S. Auditory streaming and the building of timbre. *Canadian Journal of Psychology*, 1978, **32**, 19-31.
- BROADBENT, D. E., & LADEFOGED, P. On the fusion of sounds reaching different sense organs. *Journal of the Acoustical Society of America*, 1957, **29**, 708-710.
- CUTTING, J. E. Auditory and linguistic processes in speech perception: Inferences from six fusions in dichotic listening. *Psychological Review*, 1976, **83**, 114-140.
- DANNENBRING, G. L., & BREGMAN, A. S. Streaming vs. fusion of sinusoidal components of complex tones. *Perception & Psychophysics*, 1978, **24**, 369-376.
- DARWIN, C. J. Perceptual grouping of speech components differing in fundamental frequency and onset time. *Quarterly Journal of Experimental Psychology*, 1981, **33A**, 185-207.
- JUDD, T. Comments on Deutsch's musical scale illusion. *Perception & Psychophysics*, 1979, **26**, 85-92.
- MILLS, A. W. Auditory localization. In J. Tobias (Ed.), *Foundations of modern auditory theory* (Vol. 2). New York: Academic Press, 1972.
- MYERS, J. L. *Fundamentals of experimental design*. Boston: Allyn & Bacon, 1972.
- PERROT, D. R., & BARRY, S. H. Binaural fusion. *Journal of Auditory Research*, 1969, **3**, 264-269.
- PERROT, D. R., BRIGGS, R., & PERROT, S. Binaural fusion: Its limits as defined by signal duration and signal onset. *Journal of the Acoustical Society of America*, 1970, **47**, 565-568.
- RASCH, R. A. The perception of simultaneous notes such as in polyphonic music. *Acustica*, 1978, **40**, 21-33.
- STEIGER, H., & BREGMAN, A. S. Negating the effects of binaural cues: Competition between auditory streaming and contralateral induction. *Journal of Experimental Psychology: Human Perception and Performance*, 1982, in press.
- WARREN, R. D., & BASHFORD, J. A. Auditory contralateral induction: An early stage in binaural processing. *Perception & Psychophysics*, 1976, **20**, 380-386.
- WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Manuscript received October 29, 1981;
revision accepted for publication April 19, 1982.)