Capturing frequency components of glided tones: Frequency separation, orientation, and alignment

HOWARD STEIGER and ALBERT S. BREGMAN
McGill University, Montreal, Quebec H3A 1B1, Canada

If two sinusoidal glides, Y and Z, synchronous in onset and offset, glide in parallel on a log frequency scale, they fuse and sound like a single rich glide. However, it is possible to capture Y (the target) into a sequential stream by using as a captor a pure tone glide, X, that precedes the YZ glide in a repeating cycle. The cycle then breaks perceptually into two streams, an XY stream and a Z stream. The strength of capturing depends on the similarity of the captor and target glides with respect to both frequency range and orientation. There appears to be no special capturing effect when the captor and target glides are aligned on a common trajectory.

In many normal listening situations, a listener's ear receives a complex acoustic mixture composed of the combined effects of many simultaneously active sources of sound. Before the recognition of individual sounds can take place, this complex pattern must be "parsed" by the auditory system so that separate descriptions of each source of sound within the mixture can be developed. It appears that the auditory system uses several heuristics to parse such complex signals (Bregman, 1978a). Two sorts of processes, which we can refer to as sequential and simultaneous grouping, have been studied. In sequential grouping, the auditory system decides which of a series of non-simultaneous sounds to assign to the same perceptual stream. The effects of this type of grouping can be observed in the illusion of "multiple streams." Several researchers have reported that when a sequence of sine tones in two different frequency ranges is presented rapidly, listeners perceptually group the tones that are close in frequency so as to form two separate streams, reporting the experience of two separate sources of sound (e.g., Bregman & Campbell, 1971; Miller & Heise, 1950; Norman, 1967; van Noorden, 1975).

The second process, the grouping of simultaneous sounds, can also be referred to as fusion. When two persons, A and B, speak at the same time, the mixture of their voices at any given instant of time

gives rise to a large set of spectral components. Only by grouping (or fusing) the subset arising from A's voice alone (or B's) could the auditory system compute the correct timbre for that person's voice at that instant, and hence, what speech sound was spoken. The fusion of co-occurring spectral components is strongly influenced by the synchrony of their onsets and offsets. For example, while a pair of sinusoids with synchronous onsets and offsets fuse perceptually to form a complex tone, sinusoids with onsets or offsets that are asynchronous can be heard as separate pure-tone components (Dannenbring & Bregman, 1978; Rasch, 1978).

It has also been determined that the sequential and simultaneous grouping processes compete with one another. Bregman and Pinker (1978) reported some studies which presented subjects with a rapid alternation of a pure tone, X, and a complex tone having two pure-tone components, Y and Z. If the frequency of X was far in frequency from Y and Z, the event YZ sounded like a complex tone, but when X was near in frequency to Y, the listener heard a sequential pattern, X,Y,X,Y, etc., of pure tones accompanied by a sequence Z-Z-Z, etc., also heard as pure tones. This demonstrated the fact that proximity in frequency causes pure-tone components to group sequentially. In the same experiment, if Y and Z were asynchronous in onset and offset, their tendency to segregate was increased. The overall results could be summarized by the statement that the sequential grouping of X and Y was antagonistic to the simultaneous fusion of Y with Z, and vice versa.

While these experiments have suggested several factors which may be involved during the decomposition of complex waveforms into their component sources, the generality of the studies is limited by the fact that they have examined only the per-

425

ceptual grouping of steady state sine tones. Since a great many naturally occurring sounds (particularly in speech) consist of glided changes in frequency, we considered it important to examine the factors which induce sequences containing frequency glides to group perceptually into streams.

The present experiments used signals that resembled those of Bregman and Pinker (1978). However, in the present experiments, X, Y, and Z were frequency-gliding sinusoids, as shown in Figure 1. This figure shows two cycles of one of the patterns used in the experiment, a pure tone glide, X, descending in frequency, followed by a synchronous pair of upward gliding sinusoids, Y and Z. In our discussions, X is referred to as the captor tone, and Y as the target, since X can capture Y and prevent it from fusing with Z.

One factor that could be expected on logical grounds to affect the sequential grouping of glides is whether or not they line up on a common trajectory. This principle has been used in a computer system that is capable, to some degree, of separating two voices (Parsons, 1976). Furthermore, Bregman and Dannenbring (1973) found that a rapid alternation of high and low tones hung together better as a unified stream when adjacent tones were connected by frequency glides or even by partial glides. The authors explained the benefit of partial glides as arising from a tendency of the auditory system to extrapolate glides so as to anticipate the frequency region in which subsequent sounds would fall.

We might also expect frequency proximity to influence sequential grouping, as it did in the Bregman and Pinker studies (1978). As X approaches Y in frequency, we might expect it to better capture Y into a sequential grouping. However, we do not intuitively know how to measure the proximity in frequency of glided tones. We might consider their

average frequencies, or their starting or ending points. Alternatively, the relevant distance could be the frequency separation between the terminal frequency of X and the initial frequency of Y.

Finally, the factor of similarity of slope between X and Y might be expected to have an influence. Psychophysical data (e.g., Pollack, 1968) and selective adaption studies (Gardner & Wilson, 1979) demonstrate that humans are sensitive to the orientation of glided events, even at very brief presentation durations. In vision, patterns composed of identical elements, half with one slope and half with another slope, have been shown to segregate (Beck, 1972). Similarly, one might expect that if glides X and Y in our repeating patterns had different slopes, they might not group as well, and that Y would remain fused with Z.

In summary, the experiments we report below had three objectives: (1) to see whether there was any evidence for the existence of an auditory mechanism capable of extrapolating a trajectory from glide X to glide Y, (2) to see whether such a trajectory-following tendency was as powerful as the similarity in frequency of X and Y, and (3) to determine whether the correspondence in the orientations of glides X and Y would cause them to enter a common stream.

EXPERIMENT 1

The first experiment compared the roles of common trajectory, common frequency range, continuity, and common orientation between successive glides in causing the glides to form a sequential stream.

Method

Stimulus patterns. Figure 2 shows the combinations of sine-tone glides that were used to produce all conditions. The pairs of synchronous tones (Y and Z) that were used to form complex glides were always a pair of ascending or descending tones gliding through the ranges labeled "2" and "4" in Figure 2. The tones glided in parallel at a separation of 1 octave. Glides were always exponential changes in frequency, since these were expected to produce linear changes in perceived pitch (Békésy, 1960). In different conditions, either the upper or lower of the synchronous tones acted as "target." In Figure 2, the target is always labeled "Y" and the other tone, "Z." Several captor tone options, designed to capture each target, are shown to the left of each YZ complex. One type of captor was designed to test whether the auditory system could track and extrapolate trajectories. These captors are referred to as "trajectory" (T) captors. They are positioned so that they begin a frequency trajectory which is continued by the target tone, Y.

A second group of captors was designed to test whether the auditory system tends to group gliding tones according to their simple proximities in frequency. Thus, these captors were in the same frequency range as their target. These are referred to as "same-frequency-range" captors. In order to determine whether similar orientations in successive glides strengthen their perceptual grouping, the same frequency-range captors were subdivided into two types. One type occupied the same frequency ranges as their targets and also had the same slope of frequency change over
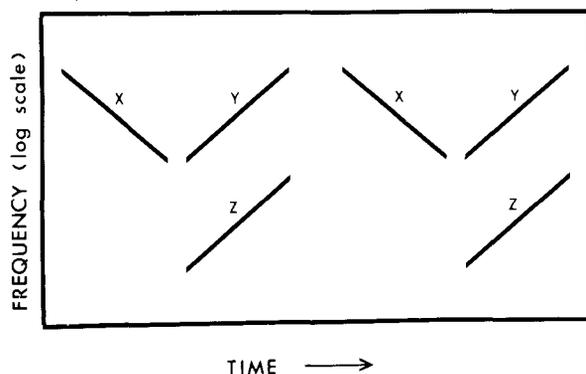

Figure 1. A graphic representation of the general stimulus pattern used in the current study. The figure shows repeated alternations between a pure-tone-captor glide (X) and a pair of synchronous glides, Y and Z.
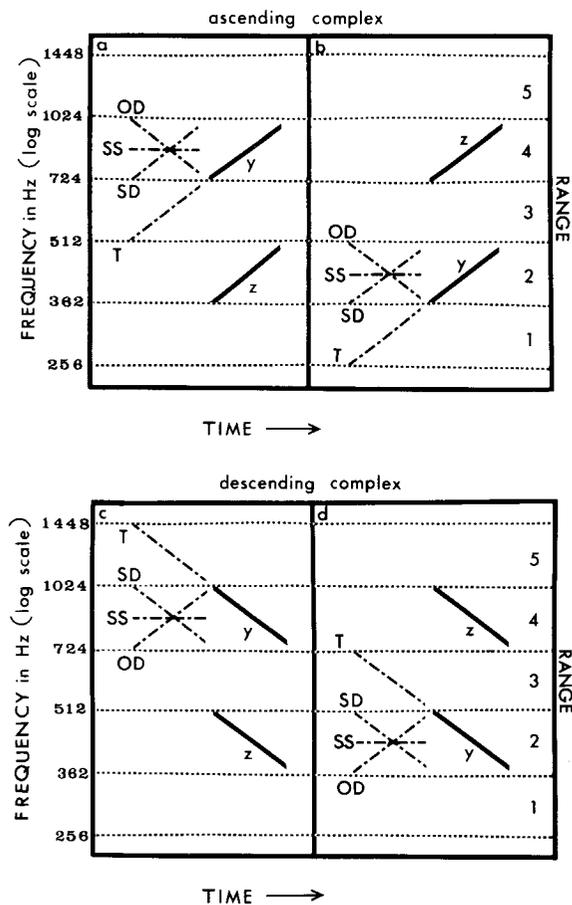
Figure 2. Stimulus patterns used in Experiment 1. The trajectory (T), same direction (SD), steady state (SS), and opposite direction (OD) captors are shown in dotted lines. Captors, when the upper and lower components of the ascending complex acted as target (Y), are shown in A and B, respectively. Corresponding captors for the descending complex are shown in C and D, respectively.

time; these were called same-direction (SD) captors. The other type, while also occupying the same frequency range as their targets, had a slope that was opposite to that of their targets (e.g., if their targets glided upward, they glided downward); these were called opposite-direction (OD) captors. Any superiority of SD over OD captors in capturing power would support the existence of an orientation effect in the grouping of glides. Glides were always exponential in frequency, since these were expected to produce linear changes in perceived pitch (Békésy, 1960).

The design of these sets of stimuli also permitted us to examine whether or not continuity between successive tones promotes their sequential grouping. The separation between the terminal frequency of the captor and the onset frequency of the target was identical in the T and OD conditions. However, the corresponding separations in the SD condition were much larger (see Figure 2). Thus, if continuity (which can be conceived of as "local" frequency proximity) between events has a very strong effect upon sequential grouping, then T and OD captors should attract their targets more strongly than SD captors. Finally, we used steady state tones as captors to determine whether or not the presence of a nonglided captor with energy inside the frequency range encompassed by Y was sufficient to induce sequential captor-target streaming (see Figure 2). This condition will be referred to as steady state control (SS).

In each condition, the appropriate captor tone and complex (YZ) were presented in a repeatedly alternating cycle. Each trial consisted of a warning beep, followed 2 sec later by 30 repetitions of the appropriate pure-tone/complex-tone sequence. A 6-sec silence separated successive trials. In addition, to examine the influence of the speed of alternation between events on perceptual grouping, the captor and complex tones were alternated at two rates. (The "alternation rate" factor will be described more fully below.)

Four different randomizations of 32 trials each were prepared (consisting of the 16 conditions shown in Figure 2, each presented at two alternation rates). To produce practice trials, four stimuli comprising a representative sample of the range of experimental stimuli were also prepared. However, all frequencies of tones were shifted somewhat relative to those in the main experiments to avoid specific carry-over effects.

Subjects. Twenty-four volunteers, all of whom were university students ranging in age from 20 to 28 years, were tested. All listeners reported having normal hearing.

Procedure. Listeners were tested individually in an audiometric chamber. The stimuli and possible percepts they could produce were described with written instructions and diagrams. The subjects were instructed to call a stimulus "fused" when it sounded like a rhythmically even alternation of a pure and a rich tone, and "decomposed" when it appeared to contain (1) two streams (one with a rhythm twice as rapid as that of the other) and (2) three pure tones on each cycle of the pattern. The instructions also warned that the sequences would be presented at two speeds, and that care should be taken to avoid confusing the changes in speed for rhythmic differences between the possible percepts that could result. Once the instructions had been read, the four practice trials were presented. These trials were repeated a second time for listeners who had difficulty in making the required judgments of the stimuli. Two listeners who could hear no such differences between the sequences after the second presentation of the practice trials were rejected from the experiment. Listeners rated the strength of the percept produced by each pattern on a 7-point scale labeled FUSED at the low end and DECOMPOSED at the high end. The stimuli were presented in one uninterrupted session.

Acoustic parameters of tones. All tones were sinusoidal. The frequency glides consisted of exponential sweeps in frequency. The amplitude envelopes of all tones had approximately exponential attacks and decays of 10 msec., which (in the case of glides) rose and fell as the tones glided in frequency. Sets of tones with two durations were used to enable the construction of sequences with different alternation rates. Slow sequences contained 230-msec tones (including rise and fall times) separated by 20-msec silences. Rapid sequences contained 130-msec tones (including rise and fall times), again separated by 20-msec silences. All frequency glides consisted of ascending or descending sweeps through one of the five ½-octave intervals between 256, 362, 512, 724, 1,024, and 1,448 Hz. To produce glides of the appropriate durations, 250- or 150-msec ½-octave sweeps were gated off after 230 or 130 msec, respectively. This procedure (1) ensured that T condition glides were perfectly collinear when separated by 20 msec on log frequency-time coordinates, and (2) permitted the separation between the terminal frequencies of captors and onset frequencies of targets in the T and OD conditions to be equalized.

The level of each sinusoidal component in the stimulus was approximately 78 dB.

Apparatus. The tones were synthesized digitally on a Digital Equipment Corporation (DEC) PDP-11/34 computer, using the MITSYN software package (Henke, Note 1). Stimuli were output by the computer's digital-to-analog converter, filtered to remove frequencies above 4,000 Hz, and recorded on audiotape. The resulting monophonic tapes were played back binaurally over Sennheiser HD-414 stereo headphones, in an Industrial Acoustics 1202 audiometer chamber. To measure the level of the stimuli,

a flat-plate coupler was used to connect the headphones to a General Radio Type 1551-C sound-level meter.

Results

The data were analyzed in a four-way ANOVA with repeated measures, testing "alternation rate," "glide complex orientation," "captor frequency range," and "captor types." The data points for two omitted responses, one in each of two cells, were replaced with cell means, and, in compensation, 2 df were subtracted from all error terms.

Figure 3 shows mean ratings of "decomposition" for each of the "captor-type" conditions when presented in both high- and low-frequency ranges, as well as the overall means (shown in dotted lines) at each level of "captor type." While the ratings on each pair of T, SD, and SS conditions were similar regardless of whether the captor was in the high or low range, the OD condition showed a clear tendency to produce higher ratings when the captor was high rather than low range. This difference was probably responsible for an observed significant interaction of the "range" and "captor-type" factors $[F(3,67) = 2.84, p < .05]$. Despite the interaction, the values for the "captor-type" factor at each level of the "range" factor and overall "captor-type" values suggest that decomposition ratings were generally lower in the T and OD conditions relative to those

on SD and SS conditions. The main effect of "captor type" was very reliable $[F(3,67) = 9.73, p < .0001]$.

To make sure that the observed interaction between "captor type" and "frequency range" did not restrict the generality of the main effect of "captor type," we looked at the effect of "captor type" in each range taken alone. Sattersthwaite's df and pooled error terms were used (Winer, 1962). The effect of "captor type" was significant in both high- and low-range conditions $[F(3,125) = 3.78, p < .05,$ and $F(3,125) = 11.25, p < .01,$ respectively].

In addition, Tukey tests were used to compare the means for different captor conditions. Each of the two best conditions (SD and SS) produced better capturing than each of the two worst conditions (OD and T) at the .05 level or better. No other differences were significant.

A significant effect of rate of alternation was also found $[F(1,21) = 5.14, p < .05]$. The rapid and slow rates yielded means of 4.54 and 4.03, respectively, the rapid rate showing the greater decomposition. The effects of rate did not interact with any other factor.

Discussion

The overall pattern of results can be accounted for by assuming that there are two factors affecting the capturing of Y by X: (1) whether or not X and Y are in the same frequency range—this would explain why the T captor was less effective than the SD captor and why the SS captor was reasonably good—and (2) whether or not X and Y have the same orientation—this would explain the rank ordering, in our data, of the three conditions, SD, SS, and OD, as resulting from an increasing difference in orientation between captor and target. These two factors were taken up individually in Experiments 2 and 3.

It is clear that having X and Y fall on a common trajectory was not a particularly effective way of promoting a sequential grouping. Both the SD and the SS captors were more powerful. There is no evidence, therefore, for a trajectory-following mechanism in this study. However, one problem with the evidence in this study is that we are only able to compare trajectory-based capturing with capturing based on other factors (e.g., frequency proximity). Therefore, we cannot say that the effect is nonexistent, but only that it is weaker than capturing based on other dimensions. Experiment 3 tried to avoid this problem.

One interesting result in the present experiment was the absence of a strong effect of continuity of frequency, that is, the frequency proximity of the end of X with the beginning of Y. Indeed, the two conditions that had the greatest continuity in this sense (i.e., T and OD) showed the worst capturing.

The "alternation-rate" factor yielded a significant main effect in which the higher rate produced
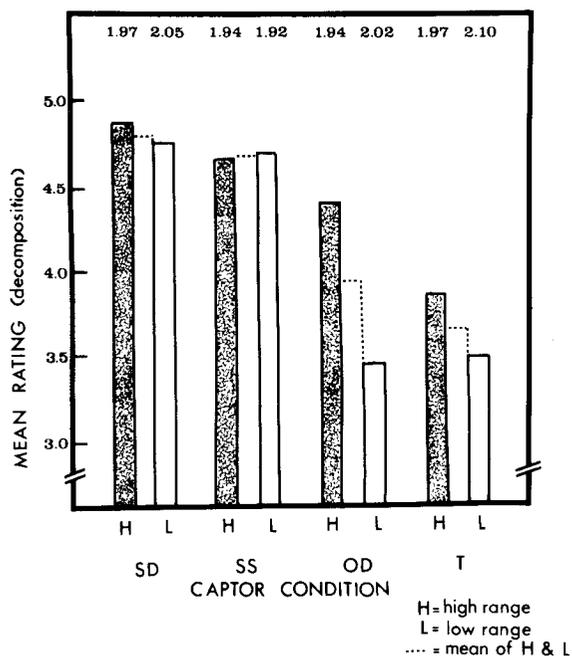

Figure 3. Mean ratings (decomposition) at each level of the "captor-type" factor for both high- and low-range captors. The overall "captor-type" means are shown in dotted lines. The pooled standard deviation for each condition is shown at the top of the figure.

more decomposition than did the lower rate. This difference can probably be attributed to the well-known tendency for the streaming effects yielded by high alternation rates to be stronger than those yielded by slow rates (e.g., Thomas & Fitzgibbons, 1971; van Noorden, 1975).

EXPERIMENT 2

This experiment again employed patterns formed from an alternation of a pure-tone captor and a simultaneous pair of tones. The captors in all conditions stayed within the range of frequencies containing the target tone. The conditions are shown in Figure 4. A subset of the captors labeled A to E had the same midfrequency as the target (on a log frequency scale), but varied in five steps from an orientation identical to that of the target to an opposite orientation. The captor in the middle condition, C, was a steady state sine tone analogous to the SS captor of Experiment 1.

These stimuli were designed for the purpose of examining the effects of correspondence in glide orientation on the streaming of successive glides. Captors A to E formed a continuum of slopes deviating gradually from that of the target. As a result, they were expected to show a gradually decreasing ability to capture the target glide, Y. While such a result, if obtained, would be most plausibly interpreted as an orientation effect, it was necessary to consider another interpretation, as follows: While conditions

A to E gradually deviated from Y in orientation, they also gradually departed from Y in their initial frequencies and in their terminal frequencies. There is reason to believe that, in glides, terminal frequencies may be more heavily weighted than are other frequencies. For example, Brady, House, and Stevens (1961) reported that the subjective pitch of a short-duration glided formant transition is identified with the terminal frequency of the transition. Again, Nabelek, Nabelek, and Hirsh (1970, 1973) reported that the pitch of short-duration glided tones could be predicted from a weighted averaging of the frequencies within them, with a greater weight given to later frequencies. The conditions A to E described above would differ increasingly from Y on any unequal weighting of earlier and later frequencies. For comparison purposes, then, two conditions, high steady state (HSS) and low steady state (LSS) were added, one at the highest and one at the lowest frequency in each glided target. The HSS and LSS captors had the same orientation as the C captor, but, while C matched the center frequency of the target, the HSS and LSS captors matched the target's endpoints. It was, therefore, assumed that they would enable us to evaluate the relative importance of the initial and terminal frequencies of glides in causing them to be perceptually organized with other events that precede and follow them.

Method

Stimuli. Ascending and descending glide complexes were produced by modulating a pair of tones up or down in parallel for half an octave. In ascending complexes, the upper component glided exponentially from 724 to 1,024 Hz and the lower component glided from 362 to 512 Hz. In descending complexes, the components glided from 1,024 to 724 Hz and from 512 to 362 Hz, respectively. Figure 4 shows each complex, with the target (which was always the upper component) labeled Y, and the other tone labeled Z.

Conditions A through E were created by "rotating" the captor through five equal increments on log-frequency-by-time coordinates, using as their midfrequency 861 Hz. All tones had the same waveform and envelope properties as in Experiment 1. Successive tones were again separated by 20 msec silences. However, only tones of 230 msec duration were used.

Using the same procedure as in Experiment 1, four stimuli were generated for use in practice trials. Four different randomizations of 42 trials were recorded on audiotape. The 42 trials consisted of three replications of each of the seven captor conditions, used with ascending and descending YZ complexes. Again, each trial consisted of a warning beep, followed by 30 repetitions of the appropriate pure-tone/complex-tone cycle. The output of the playback equipment was measured and found to be flat at 80-dB SPL for a pure tone glided across the entire frequency range employed in the stimuli. Human auditory frequency response is also flat (on average) across the same range at 80 dB (Fletcher & Munson, 1933). The stimuli were therefore presented so that each sinusoidal component had a level of 80 dB. The procedure and apparatus were the same as in Experiment 1.

Listeners. Twenty-four volunteers, ranging in age from 20 to 37 years, were tested. All listeners reported having normal hearing and were drawn from the same population as were those in Experiment 1.
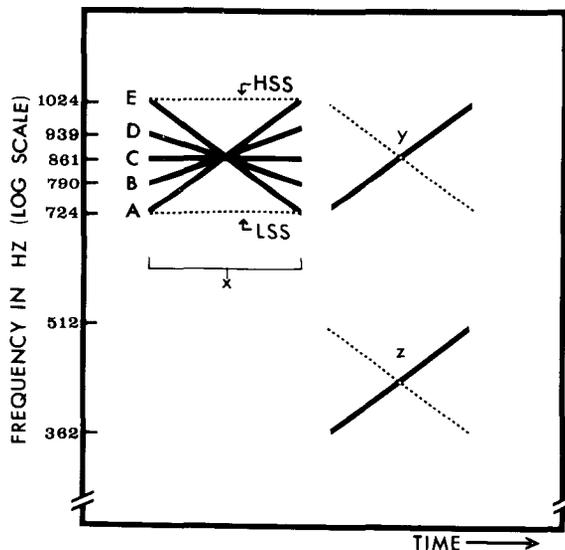

Figure 4. Stimulus patterns used in Experiment 2. The solid lines Y and Z depict the ascending complex, and dotted lines Y and Z depict the descending complex. The captors, labeled A and E, are those for the ascending complex. The same tones in reverse order were the captors referred to as A through E for the descending complex.

Results

The totals of listeners' ratings on the three publications of each condition were subjected to ANOVA. The values of eight omitted responses were estimated with cell means. To compensate, 8 df were subtracted from each error term. No more than two missing values ever occurred in any one of the 14 conditions of the design.

The overall mean ratings on captor conditions A to E are shown in Figure 5. The data suggest that as the correspondence in orientation between captor and target tones was reduced (across conditions A through E), ratings of perceived decomposition became lower. A test for linearity performed on the means of conditions A to E showed the apparent trend to be very reliable [$F(1,130) = 11.43$, $p < .001$]. No significant deviations from linearity were observed. The main effect of the "captor type" factor was highly significant [$F(6,130) = 6.35$, $p < .0001$].

To examine the importance of the initial and terminal frequencies of glides, we can look at Figure 6, which shows the means on the HSS, LSS, and A captor conditions for both ascending and descending targets. The stimulus pattern for the condition associated with each mean is also shown.

Figure 6. Mean ratings (decomposition) on HSS, LSS, and A captor conditions for both ascending and descending glide complexes. The pooled standard deviation for each condition is shown in brackets.

HSS and LSS captors are shown separately because there is reason to believe that their effects might have been different, quite independently of whether or not they matched the initial or terminal frequency of their target. The HSS captor was at the extreme of the frequency range occupied by tones Y and Z taken together, while the LSS captor was well inside this range. There is, therefore, some reason to expect that the LSS captor might have interacted with Z as well as with Y and possibly have exhibited less of a capturing effect than the HSS captor for this reason. Such a pattern is, indeed, visible in Figure 6, the two LSS captors showing less capturing than the two HSS captors. The difference was tested and found to be significant [$F(1,130) = 10.86$, $p < .01$].

Despite this factor, each SS captor captured better when it was at the initial rather than the terminal frequency of its target. This difference appeared in the statistical analysis as a significant interaction between the factor of high vs. low steady state and the factor of ascending vs. descending glide complex [$F(1,30) = 9.37$, $p < .01$]. This latter result argues against the possibility that the later frequencies of a glide are more central in predicting its grouping with other events than are its earlier frequencies. One would expect, by extrapolation of the results of Nabelek, Nabelek, and Hirsh (1970, 1973), that a match to the later frequencies of a glide would be more, rather than less, effective. However, these authors studied pitch estimation rather than grouping, and these two processes may respond to different factors. For example, it seems reasonable that the grouping with an earlier tone should be best predicted by the earlier frequencies in a glide, whereas grouping with a later tone should be predicted best by its later frequencies.
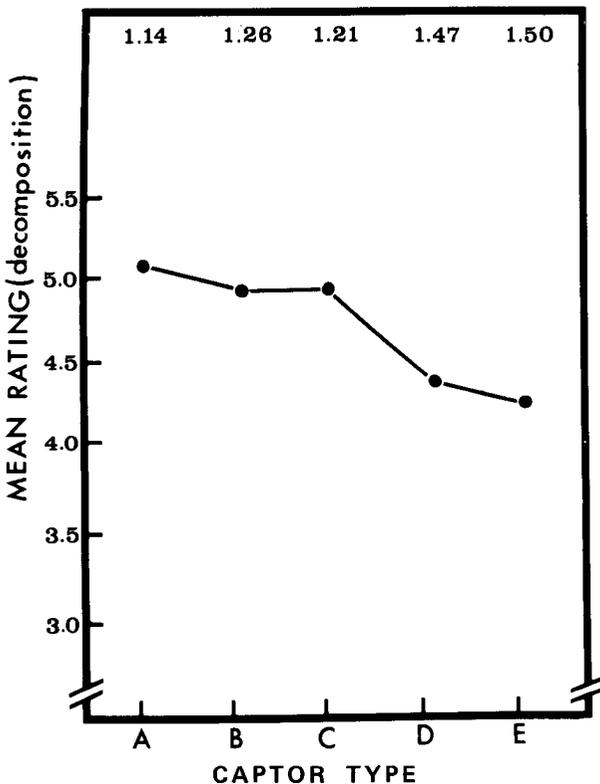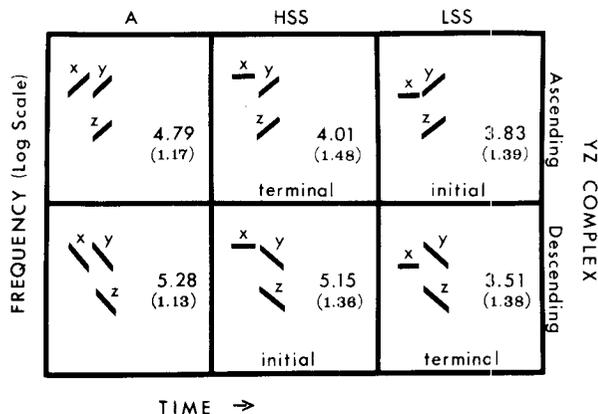

Figure 5. Mean ratings (decomposition) in conditions A through E. The pooled standard deviation for each captor condition is shown at the top of the figure.

In order to see whether or not having the same slope as the target (as in the A conditions) improved the sequential attraction between the captor and its target over and above the attraction that resulted when the captor was matched to the initial or terminal frequency of the target, a statistical analysis was done to compare the performance of the LSS and HSS captors against the A captors. Due to the interaction between the HSS/LSS factor and the ascending/descending complex factor, separate F values were computed for comparisons of each HSS and each LSS captor with the A captor with the corresponding-orientation target (i.e., the A captor shown in each row of Figure 6 was compared with the LSS captor and with the HSS captor shown in the same row). With the exception of the HSS in the descending YZ complex condition, each SS captor showed significantly less capturing than the corresponding A (same slope) captor (at the .05 level or better).

Discussion

The pattern of data in Figures 5 and 6 is consistent with the following set of assumptions:

(1) The strength of sequential grouping between sinusoidal glides is a function of the correspondence in orientation of the glides. As the orientations of successive glides are made more discrepant, perceived decomposition (i.e., sequential grouping) is reduced. While we cannot rule out the possibility that some simple additive function of the frequencies in the glides themselves is responsible for the grouping, simply matching the captor to the initial or terminal frequency of a glide is not sufficient to account for the effect observed in condition A. Glide orientation itself seems like a plausible candidate for the basis of grouping in the glided captor conditions. We are encouraged in this interpretation by reports concerning channels in the auditory system sensitive to rate of frequency modulation (Evans & Whitfield, 1964; Vartanian, 1974). Increasing the rate of FM is analogous to increasing the slope of glides. Therefore both the studies on FM channels and the present studies may be tapping the same glide-coding mechanism.

(2) The auditory system seems to be able to select a target out of a complex more easily when the captor provides a priori evidence as to where the target begins rather than where it ends. This is reasonable on logical grounds.

(3) There is a suggestion in the data that for good capturing to occur the captor must not have any tendency to group with tones other than the target. The HSS captors, which were close to Y but far from Z, were the best of the steady state captors. The fact that proximities among tones compete with one another to achieve the final stream structure of a sequence

of tonal elements has already been documented (Bregman, 1978b).

EXPERIMENT 3

While Experiment 1 attempted to detect a trajectory effect (i.e., a greater tendency for successive glides to group into a stream when they are aligned on a common trajectory), the effect would have been visible only if it had been stronger than the sequential grouping effects due to other factors present in the experiment. Since each of a pair of glides aligned on a common trajectory must necessarily fall in a different frequency range, any such glides must also necessarily benefit less from streaming based on the frequency proximity principle than would glides in the same, or a more similar, frequency range. This fact makes the search for a trajectory effect (which Experiment 1 established to be weaker than the frequency proximity effects) difficult. The present experiment dealt with this problem by causing the trajectory effect to "ride on top of" the frequency proximity effect in a way that should have been visible. The conditions used are shown in Figure 7. Captors A to G were positioned relative to the target Y (the lower tone of the two-glide complex) such that they formed a series in which A was identical to Y, while the other captors became progressively more unlike Y. One of these captors, D(t), lined up with the target on a common trajectory on log-frequency-by-time coordinates. We expected that if we were to plot the ratings of decomposition of the YZ complex for this series of conditions, we
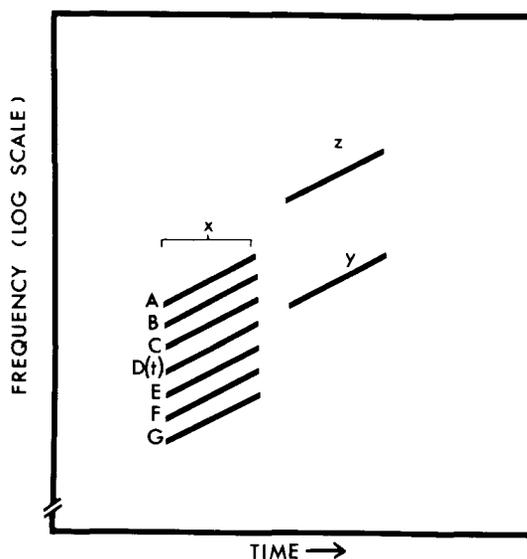

Figure 7. Stimulus patterns used in Experiment 3. Captors are labeled X, the target, Y, and the third tone (designed to fuse with Y), Z.

would see a monotonic curve showing a gradual decrease as the captors moved away from the target in frequency. However, we assumed that there would be a bump in this curve at D(t) representing the additional contribution to effective capturing that was made by the fact that D(t) and Y were aligned on a common trajectory.

Experiment 3 differed from the earlier ones in that it did not employ continuously repeating cycles of the captor and the complex tone, for the following reason: In a continuous cycle, one has to consider not only the relations between the tones within each cycle, but also the relations between the tones of one cycle and those of the next; any within-cycle trajectory relation between X and Y would have to compete with other across-cycle grouping tendencies. To avoid this problem, we introduced silences between cycles. Bregman (1978c) has suggested that streaming mechanisms accumulate evidence over successive events and use the cumulative strength of this evidence to arrive at decisions about grouping. He found that the introduction of a 4-sec silent interval into a sequence cleared most of the bias favoring a particular stream organization which had developed due to previously accumulated information. Therefore, by embedding a 4-sec silence between each repetition of the captor/complex pair in the present experiment, we assumed that grouping tendencies would become evident only within cycle. Repetitions of the pattern were presented not to promote streaming effects, but simply to enable subjects to hear each sequence of a pure tone and a complex tone often enough to make a reliable judgment of the resulting percept.

Due to the fact that stimuli were not presented in continuous cycles, it became difficult (in pretesting) to achieve clear decomposition effects, since such effects benefit from the accumulation of evidence over repeated cycles (Bregman, 1978c). However, it was found that decomposition of the mixture was perceptible when a single alternation of captor and complex glides was presented at a 130-msec tone duration (i.e., the duration that produced the stronger decomposition in Experiment 1). Therefore, only that duration was employed in this experiment.

Method

Stimuli. A complex glide was generated by frequency modulating a pair of synchronous sinusoids upward in parallel. The upper tone (Z) was glided from 1,448 to 1,920 Hz, and the lower tone (Y), which always served as the target tone, was glided 1 octave lower, from 724 to 960 Hz (see Figure 7). To produce a set of captors with identical glide orientations to those of the tones in the complex glide, tones glided from 724 to 960 Hz, from 645 to 855 Hz, from 575 to 762 Hz, from 512 to 679 Hz, from 456 to 604 Hz, from 406 to 539 Hz, and from 362 to 480 Hz were used. The waveform and amplitude envelope characteristics of tones were identical to those employed previously. Each stimulus pattern consisted of a 130-msec pure-tone glide followed 20 msec later by a pair of simultaneous 130-msec pure-tone glides.

Using the same procedure as in the previous experiments, four practice trials were generated. Four randomized trial orders were recorded on audiotape, each consisting of 21 trials (i.e., three replications of each of the seven captor tone conditions). Each trial consisted of two warning tone bursts, followed by 2 sec of silence. Subsequently, eight repetitions of one stimulus pattern were presented, with 4 sec of silence between the repetitions. A single tone burst was used to signal the beginning of a 6-sec response period between trials.

All captor and target tones fell within the frequency region between 362 and 960 Hz, for which the auditory system shows a flat frequency response at 80-dB SPL (Fletcher & Munson, 1933). The response of playback equipment was measured at the headphones and also found to be flat at 80 dB. Thus, the tones were presented at this level and should have had equal loudness (on average for a group of normal subjects). The upper tone Z fell in a region of heightened frequency sensitivity. However, since the tone was not varied across conditions, and no attempt to capture the tone was ever made, it was not necessary to adjust its loudness.

Subjects. Twenty volunteers, ranging in age from 19 to 29 years, were tested. All listeners reported having normal hearing, and were drawn from the same population as those employed in previous experiments.

Procedure. All experimental procedures were similar to those employed in the previous experiments. However, listeners were required to report the percept produced by only one cycle of each pattern, and therefore "fusion" and "decomposition" were redefined. Fused stimuli were described as sounding like a pair of glides, a pure glide followed by a rich glide. Decomposed stimuli were described as sounding like a pattern containing a pair of frequency glides in the lower range, with a third isolated sound (Z) audible in the higher range. Listeners were informed that they would receive eight repetitions of each pattern, and that each presentation of the pattern would be followed by a 4-sec silence. Due to the subtlety of effects in the present experiment, the listeners required considerably more exposure to practice trials than they had in previous experiments, before they could detect decomposition effects at all. The procedure employed to train listeners on the task was as follows: The first practice trial (similar to the presumably most decomposed stimulus of condition A) was presented. Only one listener rated the stimulus as decomposed on the first exposure. For every other listener, the same trial was repeated (approximately four times) until he/she reported hearing three clearly audible tones on each presentation of the pattern. Then the remaining three practice trials were presented. If the listener felt uncertain about the qualitative differences between stimuli, all four practice trials were presented once more. By this point, all listeners reported that they perceived differences between the patterns, except for one who was rejected from the experiment.

Apparatus. All equipment used during the synthesis and presentation of stimuli was identical to that employed in the previous experiments.

Results

The totals of ratings on the three replications in each condition were analyzed in a one-way ANOVA with repeated measures. Figure 8 shows the mean rating for each of the seven captor conditions. The data suggest that there was a monotonic decrement in the ratings, correlated with increasing separation between the average frequencies of captor and target tones (across conditions A through G). Note that there appears to be no upward deviation from the trend in the D(t) condition; we find no evidence for a "trajectory effect" which might have raised the rating on that condition by adding a separate
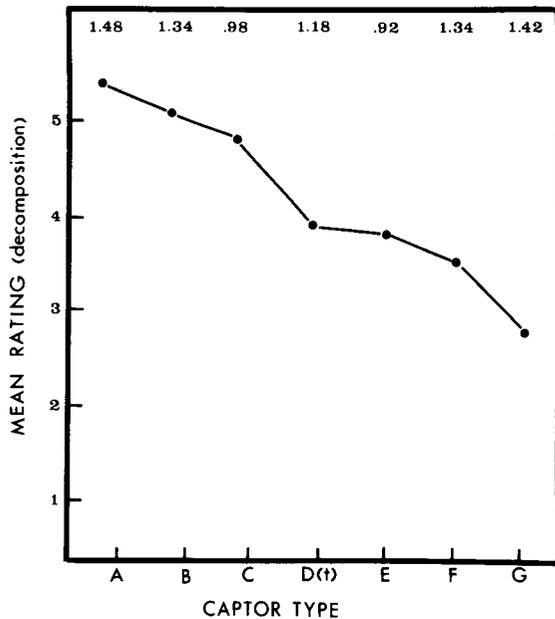
Figure 8. Mean ratings (decomposition) at each level of "captor type." The standard deviation for each condition is shown at the top of the figure.

influence favoring decomposition. The main effect of "captor type" was very reliable [F(6,114) = 11.61, p < .0001]. A trend analysis revealed a significant linear trend [F(1,114) = 66.70, p < .0001] and no significant deviations from linearity.

Discussion

The results of this experiment support the findings of Experiment 1, which suggested that the sequential grouping of successive glides is a direct function of the correspondence in the frequency ranges traversed by the glides. The data also suggest that the auditory system does not extrapolate from trajectories of frequency change when it performs the types of grouping processes that we have looked at in this experiment. It should also be noted that even if the auditory system does not group glides that fall on a common trajectory, the D(t) captor should be stronger than the others because its end point is closer in frequency to the initial frequency of the target glide (i.e., it should benefit from its continuity with the target). Therefore, the results of the present experiment cast doubt on the continuity principle as well as on the common trajectory principle.

GENERAL DISCUSSION

The preceding experiments have shown that a synchronous pair of sinusoids, gliding in parallel on a logarithmic frequency scale will fuse perceptually to produce the experience of a single rich-sounding glide. However, if this pair is preceded by a single gliding (or steady) sinusoid (i.e., the captor), in a frequency range near one of the components of the pair of gliding tones (i.e., the target), this captor tone will act to produce a bias against the perceptual fusion of the simultaneous components and in favor of a sequential organization of the captor tone with the target tone. It seemed that it was the nearness of the average frequencies of the captor and target tones that was important. No effect of the frequency proximity between the end of the captor and the beginning of the target was detected. The second most salient effect was related to the directions of the glides in the captor and target tones. When these corresponded, the "capturing" effect was stronger. However, there was no evidence that lining up the captor tone with its target on a common trajectory promoted the grouping of the captor with the target component.

These results seem to present a contradiction: apparently, the *correspondence* in slope of successive glides affected their grouping, but the *continuity* of their slopes (alignment on a common trajectory) did not. To account for this difference, we can think of a number of alternative explanations. These are presented in increasing order of complexity:

(1) Sequential grouping depends only on the similarity between successive glides in terms of some weighted average of the frequencies in them. According to this explanation, there is no necessity to evoke "glide orientation" to account for the observed effects. For instance, in Experiment 2, captors A to E gradually departed in slope from the orientation of their targets. However, they also departed gradually from the target in any weighted sum of their frequencies that one could compute (except for that particular weighted sum that defined the mean on log frequency coordinates, on which they were all equated). Thus, this explanation accounts for the absence of a trajectory effect by suggesting that there is never any effect of orientation per se on streaming.

(2) The orientation of glides is computed, but this computation is terminated by a silence. Given this assumption, it follows that in our experiments the orientation of X was computed as was the orientation of Y, but that the orientation of the sequence XY was never computed. Therefore, the auditory system could *not* determine that X and Y were aligned on a common trajectory. A minor variation of this explanation is to say that, while the slope of individual events is computed, it is not used to predict the spectral location of later events.

(3) The most complex explanation appeals to a notion of hierarchical encoding. Discontinuities (such as silence) may serve as brackets that define smaller units inside larger ones. In this view, X and

Y would have been treated as smaller units inside the larger unit XY. One might, therefore, speculate that the properties (e.g., slope) of each of the glides, X and Y, may have been computed independently of the properties (e.g., slope) of the glide sequence XY. As a result, descriptions of X and Y may have been assigned to one level of a hierarchical description (i.e., that appropriate for glides), while the XY description may have been assigned to another level (i.e., that appropriate for glide sequences). If perceptual organization processes do not compare the properties of units at different levels in such a hierarchy, then this would account for the absence of a trajectory effect. The fact that the slope of X extended onto that of Y (which would have required a comparison of each with the slope of the XY unit) would simply have been invisible to the system.

Whatever the explanation, the results of the present experiment are in conflict with the conclusions of Bregman and Dannenbring (1973), who demonstrated that the segregation of high from low tones in a rapid sequence was reduced when the ends of each tone consisted of a brief frequency glide "pointing" to the frequency of the next tone. These authors interpreted the greater perceptual coherence of such a sequence as the result of the activity of a predictive tracking mechanism that could use the slope of the terminal glide portion of each tone to predict the frequency region of the next tone. However, one could in retrospect offer an alternative interpretation of that study; for example, it may be that the glide in frequency simply brought the successive tones closer in average frequency. According to our present interpretation, what was observed by Bregman and Dannenbring was an effect of proximity in frequency, not an effect of perceptual predictions based on the estimation of a trajectory. This interpretation is also the one most consistent with the Nabelek, Nabelek, and Hirsh (1970, 1973) finding that the pitch of a glide is a weighted average of the frequencies in it.

A confirmation that the auditory system does not extrapolate glides comes from a study by Dannenbring (1976). The stimulus that was used was a connected series of alternately rising and falling frequency transitions (glides). In one experiment, the peaks of the glides (the points at which the glides reached their maximum frequencies and then began to fall in frequency) were deleted and replaced by white noise. With this stimulus, a perceptual restoration occurred and the subject heard the sinusoidal tone as gliding right through the noise. The question of interest in this experiment concerned the perceived pitch of the perceptually restored peak of the glides. Would the auditory system extrapolate the pitch of the glides upward into the noise? The answer was no. The highest pitch actually heard was somewhat lower than the one generated by the highest frequency actually present in the glide. The missing portion of the glide was perceived as having a pitch determined by an average of nearby frequencies rather than being an extrapolation. Dannenbring's experiment was in some sense a direct test of whether or not the auditory system extrapolates the trajectory of glides because, since there was no dominant pitch in the noise, the auditory system was free to put any pitch there that it wished. Thus, any extrapolation tendency should have been visible.

We can conclude that there is no evidence that the auditory system extrapolates trajectories, although it seems to measure them.

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

BECK, J. Similarity grouping and peripheral discriminability under uncertainty. *American Journal of Psychology*, 1972, 85, 1-20.

BÉKÉSY, G. VON. *Experiments in hearing*. New York: McGraw-Hill, 1960.

BRADY, P. T., HOUSE, A. S., & STEVENS, K. N. Perception of sounds characterized by a rapidly changing resonant frequency. *Journal of the Acoustical Society of America*, 1961, 33, 1357-1362.

BREGMAN, A. S. The formation of auditory streams. In J. Requin (Ed.), *Attention and performance VII*. Hillsdale, N.J: Erlbaum, 1978. (a)

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

BREGMAN, A. S. Auditory streaming is cumulative. *Journal of Experimental Psychology*, 1978, 89, 244-249. (c)

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

BREGMAN, A. S., & DANNENBRING, G. L. The effect of continuity on auditory stream segregation. *Perception & Psychophysics*, 1973, 13, 308-312.

BREGMAN, A. S., & PINKER, S. Auditory streaming and the building of timbre. *Canadian Journal of Psychology*, 1978, 32, 19-31.

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

DANNENBRING, G. L., & BREGMAN, A. S. Streaming vs. fusion of sinusoidal components of complex tones. *Perception & Psychophysics*, 1978, 24, 369-376.

EVANS, E. F., & WHITFIELD, J. C. Responses of cortical neurones to acoustic stimuli varying periodically in frequency. *Journal of Physiology* (London), 1964, 172, 52P-53P.

FLETCHER, H., & MUNSON, W. A. Loudness, its definition, measurement and calculation. *Journal of the Acoustical Society of America*, 1933, 5, 82-108.

GARDNER, R. B., & WILSON, J. P. Evidence for direction-specific channels in the processing of frequency modulation. *Journal of the Acoustical Society of America*, 1979, 66, 704-709.

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

NABELEK, I. V., NABELEK, A. K., & HIRSH, I. J. Pitch of tone bursts of changing frequency. *Journal of the Acoustical Society of America*, 1970, **48**, 536-553.

NABELEK, I. V., NABELEK, A. K., & HIRSH, I. J. Pitch of sound bursts with continuous or discontinuous change of frequency. *Journal of the Acoustical Society of America*, 1973, **53**, 1305-1312.

NORMAN, D. A. Temporal confusions and limited capacity processors. *Acta Psychologica*, 1967, **27**, 293-297.

PARSONS, T. W. Separation of speech from interfering speech by means of harmonic selection. *Journal of the Acoustical Society of America*, 1976, **60**, 911-918.

POLLACK, J. Detection of rate of change of auditory frequency. *Journal of Experimental Psychology*, 1968, **77**, 535-541.

RASCH, R. A. The perception of simultaneous notes, such as in polyphonic music. *Acustica*, 1978, **40**, 21-33.

THOMAS, J. B., & FITZGIBBONS, P. J. Temporal order and perceptual classes. *Journal of the Acoustical Society of America*, 1971, **50**, 86-87. (Abstract)

VAN NOORDEN, L. P. A. S. *Temporal coherence in the perception of tone sequences*. Doctoral thesis, Institute for Perception Research, Eindhoven, The Netherlands, 1975.

VARTANIAN, T. A. On mechanisms of specialized reactions of central auditory neurons to frequency-modulated sounds. *Acustica*, 1974, **31**, 305-310.

WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.