

Auditory streaming and vertical localization: Interdependence of "what" and "where" decisions in audition

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

When a sequence of two tones is presented over headphones in an ascending or descending order of pitch, it is heard as correspondingly ascending or descending in space. The illusion of spatial change that accompanies pitch change can be induced onto a pair of noise bursts by presenting them in synchrony with the tones. When cues known to produce stream segregation are introduced, the perceived position of the noises is less influenced by the tones. Stream organization is seen to be implicated in the ability to separately localize concurrent sources of sound. This suggests that "what" and "where" decisions are highly interactive and that neurological evidence that suggests separate pathways for these decisions must be interpreted with caution.

In 1976, Deutsch and Roll reported an illusion obtained by playing a dichotic tonal sequence composed of three 800-Hz tones followed by two 400-Hz tones to one ear, and, in synchrony, three 400-Hz tones followed by two 800-Hz tones to the other ear. Many listeners heard only one sequence of tones that had the sequence of pitches received by the right ear (e.g., 400, 400, 400, 800, 800) but the sequence of positions occupied by the higher frequency, whatever its ear of arrival (e.g., L L L R R). This produced a percept in which, for example, the 400-Hz tone was heard at the left on the first event of the sequence, when in actual fact it had been presented to the right ear. Since the illusion seemed to represent a decoupling of the sources from which the listeners derived positional and pitch information, the authors suggested links between this effect and certain physiological findings that suggest there is a localizing system (the ventral route) and a nearly separate system (the dorsal route) involved in stimulus discrimination (Evans, 1974; Evans & Nelson, 1973a, 1973b). Is this why location and frequency information can be decoupled in the illusion mentioned earlier?

One view of these systems might be that each one makes its own decision about the attribute with which it is concerned, and then decisions of the two systems are combined to represent sources of sound. Deutsch

and Roll (1976) point out that when more than one stimulus is presented at a time, "we have not just one value of each attribute to combine together, but several, and we have to know which values to connect with which." The present paper investigates this problem further by means of another illusion of spatial position. Again, spatial information derived from one source is combined with the "identity" attributes of another source. Using this illusion, we look at some factors that may normally produce the correct conjunction of identity and position.

It has been argued earlier (Bregman, 1978, in press) that in normal environments our auditory systems are usually dealing with complex mixtures of sounds derived from several active sources. There have to be auditory processes that "parse" the mixture so that the correct features are assigned to each source of sound. The feature of "location" would be easy to assign if each incoming signal were a pure tone. Then the output of the filter that picked up a particular frequency in the left ear could be compared (for loudness and time of arrival) with the output of the corresponding filter in the right ear to arrive at a location for that pure tone. Unfortunately, the world not being so obliging, real sources may each have many frequency components and these are likely to be interleaved or partially overlapping in frequency. In deciding upon "what" and "where," we also must decide on "how many."

It appears that such problems may be dealt with by the use of information that is spread out temporally in the incoming signal and that may help to determine the correct decomposition of any particular moment of sound (Bregman, 1978, in press). An appropriate decomposition will group the correct subset of incoming spectral components and assign

We acknowledge the assistance of Mimi Israel in data analysis, and the provision of a synthesizer by Richard Robinson and Lois Henderson. Some of the stimuli were made in the Computer Based Laboratory, Department of Psychology, McGill University. The research was supported in part by grants from the National Research Council of Canada and the FCAC program of the Quebec Ministry of Education. Reprints are available from Albert S. Bregman, Department of Psychology, McGill University, Montreal, Quebec, Canada H3A 1B1.

them to the proper location. The present experiment is a study of one principle that assists in this "parsing." It may be described as follows: The spectral content of any short period of time should be perceived whenever possible as (a) the continuation of the spectral content of the immediately preceding period of time and (b) the addition of some new content. The continued content and the added content (the "remainder") will each be assigned a separate identity. This can be called the "remaindering principle." A special case of it occurs when two simultaneously occurring frequency components have slightly asynchronous onsets. In such a case, two time periods exist, a first one in which only the temporally leading component is present, and a second period in which the mixture of the two components is present. An asynchrony of as little as 10 msec can promote the segregation of the two components into separate perceptual streams. This principle of decomposition has been discussed elsewhere (Bregman, 1978; Bregman & Pinker, 1978). It has been found to be capable of partially eliminating the masking of one tone by a concurrently played tone (Rasch, 1978). In the experiments that follow, we wish to show that the segregation that arises from this principle can influence spatial localization. If it can, we will have discovered a situation in which decisions about "how many" can influence the decision about "where." In the final discussion, we will see that we already know from past research that decisions about "what" and "how many" are highly interactive. Furthermore, we know that our perception of the grouping of spatial positions ("where") can influence our decisions about melodic properties of a pattern ("what"). Therefore, the results expected from the present experiments would contribute to a picture of a system that computes a description involving "how many of what are where" in a highly interactive manner.

We chose to employ a phenomenon of localization in which two-ear comparisons are not involved—namely, the case of up-down localization when the source of sound lies straight ahead of the listener. We did so because it appears that such localizations depend, in part, upon an analysis of the frequency spectrum, and the remaindering principle is one that is believed to analyze spectral information. There is evidence in the literature on vertical localization that the notion of a high tone in music is not simply a metaphor but is associated with real (although inaccurate) spatial localization: Research exists that shows that higher frequencies are localized as higher in space (Pratt, 1930). It has been suggested by recent researchers (Butler, 1969; Roffler, 1968; Searle, Braida, Cuddy, & Davis, 1975) that this perceptual bias is making use of the fact that the outer ear filters an incoming waveform differently depending on its angle of incidence. Regardless of its cause, however,

we have here a phenomenon in which the monaural spectrum is linked to localization. Therefore, the "remaindering" principle should show a definite effect.

The first experiment reported below simply demonstrated that high- and low-pitch tones, presented through headphones, were perceptually localized at higher and lower locations, respectively. The second experiment presented high and low tones over headphones synchronously with a white noise burst presented over a loudspeaker. Normally, the tone induced a mislocalization of the noise, but if the tonal component was caused to be perceptually segregated from the noise, by being captured into a sequence of tones of the same frequency, its effects on the localization of the noise were reduced. In the third experiment, the subject was again presented with a simultaneous noise and tone. This time the segregated condition was created by introducing asynchronies in the amplitude envelopes of the noise and tone. This segregation again reduced the influence of the tone on the judgment of the position of the noise.

EXPERIMENT 1

This experiment was aimed solely at verifying the effects of pitch on spatial localization. It used pairs of tones of the same or different frequencies presented over headphones and asked subjects to judge the change in spatial position of the second tone relative to the first. Subjects were persuaded that the tones were partially coming from a vertical array of speakers in the room.

Method

Procedure. Subjects were tested individually, seated directly facing a pair of dummy speakers affixed on a wall near the corner of a 15 × 20 ft room. The speakers were placed one directly above the other at a separation of 10 ft. An "X" was affixed to the wall at the midpoint between these speakers, so that three spatial locations were indicated, one at the "X" and two at the speakers themselves. The subjects sat 10 ft from this vertical configuration, such that both speakers were equidistant from their heads. Thus, the auditory angle between the two speakers was approximately 53 deg.

Then the headphones were put on, and the subject was told that a signal would be presented partially over the headphones and partially over the higher, the lower, or both speakers. The listeners were warned that the headphones would seem to dominate and that they should try their best to ignore them. They were told that they would be able to hear the speakers despite the headphones because the latter were "acoustically transparent." This latter statement was plausible because of the type of headphones used, Sennheiser Model 424, which have no circumaural cushion and shut out little sound from the environment. The subjects were asked to indicate the source of each tone at the speakers, with a forced choice of either the higher or lower speaker, or the center location "if both speakers were activated at once" (they were told that the sound would appear to come from the middle position when both speakers were active together "due to a common stereo effect"). In reality, all the sounds were presented binaurally through the headphones alone.

On each trial, the subjects heard two tones in succession and

indicated their judgments on a response sheet on which was printed a "vertical" row of three dots. If the two tones had come from different locations, they were to connect two of the three dots on their response sheets by an arrow to indicate the positions of the first and second tones. If the tones both came from the same spatial position, they were to circle the appropriate dot.

The room was then darkened to reduce visual distractions and the experimenter went out. A tape recording with all the conditions on it was run without interruption for about 5 min.

Design of stimuli. Each stimulus was a pair of tones recorded monaurally. The first tone was on for 1.5 sec, then there was a .5-sec silence, and the second tone appeared for 1.5 sec. Each pair was followed by a 7-sec period in which the listener was to mark his or her judgment before the next pair occurred. (These times are only approximate because the tape was prepared manually.)

Each tone was a pitch produced by modulating the frequency of an audio range sine wave with a subaudio (15 Hz) sine wave on an ARP 2600 synthesizer. Each tone therefore wavered across a full-tone range, a semitone above and below the nominal test frequency, and sounded like a rich sound with somewhat ambiguous chroma. This type of sound was used to try to focus the subject's attention on pitch height rather than chroma. (As it turned out, the subjects felt they were hearing musical intervals anyway; therefore, our subsequent studies used tones of steady pitches.)

The five nominal frequencies were at equal logarithmic separations within the octave ranging from 262 to 524 Hz (middle C to the C one octave higher). They were at 262, 311, 370, 440, and 524 Hz. The ascending two-tone sequences always began with the 262-Hz tone and ended with either the same 262-Hz tone or one of the others. Similarly, the descending sequences always began with the highest tone and ended with either itself or one of the lower tones.

Subjects. Twenty-three subjects, ranging in age from 18 to 33, both male and female, were recruited. Of these, two who misunderstood instructions were rejected from the analysis.

Results

The subjects' responses were positions and changes on a 3-unit scale that represented 53 deg of auditory angle. Positive values were assigned to judgments that indicated ascending spatial change, and negative values were assigned to indications of descending spatial change. The resulting scale for judgments of change, therefore, can range from +53 to -53 deg. The mean change judgments, converted to degrees, are plotted in Figure 1 as a function of the size of the pitch interval expressed in semitones. In this figure, descending pitch intervals are shown as negative values.

The analysis of variance in this and later experiments treated the 10 pitch intervals as a single 10-valued variable. The alternative, treating magnitude and direction as separate, would have generated a large Magnitude by Direction interaction that really should be treated as part of the main effect. "Sound moves up spatially if it is increasing in pitch *but* moves downward if decreasing in pitch." The "but" in the previous sentence would be the statistical interaction effect. It does not make sense to treat this as a real interaction, that is, as a restriction on the generality of a main effect.

Figure 1 shows that the pitch interval influences the judgment of spatial change in a strong and regular

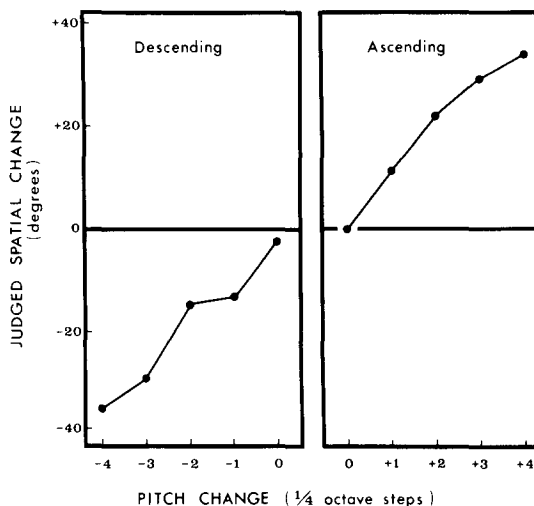


Figure 1. Mean judgments of spatial change of a tone (in degrees) as a function of its frequency change (in one-quarter-octave steps).

manner [$F(9,171) = 38.23$, $p < .001$]. The present study, therefore, was able to show a clear effect of spectral changes upon judged changes in vertical location. The fact that the result can be obtained with sounds presented through headphones enables us to rule out artifacts related to the acoustic properties of the room.

EXPERIMENT 2

This experiment went on to use the effect demonstrated in the first experiment to determine whether the "remaindering" principle that segregates the monaural spectrum could affect spatial localization. A pure tone and a noise burst were played at the same time. The noise was played from speakers at one of two vertical locations, while the tones were always played through earphones. It was expected that the illusory localization of the tones would be transferred to the noise burst when both occurred in synchrony. However, we expected to find that factors known to affect segregation of acoustic components should inhibit the transfer of properties from the tone to the noise burst. In Experiment 2, we segregated the tone from the noise in some conditions by adding a stream of tones, before and after the critical one, to "capture" it and inhibit its fusion with the noise burst (see Figure 2).

The principle of "remaindering" mentioned in the introduction is involved in this design as follows: The burst-plus-tone stimulus contains a component (the tone) that has just previously occurred alone. Accordingly, there should be a bias toward removing this element from the mixture, leaving the noise burst as an independent remainder.

We know already that we can cause the perceptual separation of two concurrent acoustic components (A and B) by capturing A into a sequential stream in

such a way that B is left behind in its own stream (Bregman & Pinker, 1978; Dannenbring & Bregman, 1978). But the "segregation" produced by this method has only been indexed by asking the listener two types of questions: (1) Can A (the component that is pulled out) be heard as part of a sequential pitch stream together with the preceding and following tones? (2) Does B (the component that is left behind) sound purer when segregation has occurred? While these two questions provide evidence for segregation, there is no a priori reason to believe that just because some features derived from two concurrent sounds will segregate, all features must therefore segregate. In particular, if the "where" decision is made by one neural system and the "what" decision by another, factors that produce segregation in the case of "what" decisions need not produce a corresponding segregation of "where" decisions. The judgments used previously to measure streaming have been concerned only with "what" was present, and the "whats" have been defined by their frequency characteristics. It is possible that decisions about location are not based upon the same heuristics that influence segregation in the "what" system; they may rely only on dichotic spectrum/phase relations for left-right information and on the content of the monaural spectrum for up-down information. In this case, we would expect this system to have little ability to reject the influence of a concurrent tone upon localization of a noise burst, even when it is given temporal clues that the two sounds are arising from different sources.

On the other hand, if location decisions do respond to information derived from the sequence of tonal components in exactly the same way that previously investigated judgments of streaming have done, then it is likely that auditory recognition is a global process that utilizes information (but not "decisions") of various types to come up with a single, joint decision of "what" and "where."

Method

Procedure. Subjects were tested individually in an 11 × 6.5 ft acoustically live room. They sat directly facing a wall of the room at a distance of 8 ft. On this wall, two loudspeakers were fixed, one directly above the other at a separation of 9 ft. A thin textile sheet was draped from the ceiling to the floor in front of the speakers to conceal them, and four Xs were printed on the sheet, one in front of the higher speaker, one at the lower one, and two at equal spacings in between. This produced a four-position spatial scale to be used by the subjects in giving their judgments. The scale represented a total of about 59 deg of auditory angle.

Subjects were told that they would, on each trial, hear a double burst of noise at any of four speaker locations marked by Xs, and that they should judge their locations. A response sheet was provided on which, on each trial, they were to circle the appropriate point or points on a diagram of the four Xs and to mark which came first and second. They were warned that simultaneous tones, designed to distract them, would be heard over the headphones and that these should be ignored so that only the noise bursts were localized. The headphones

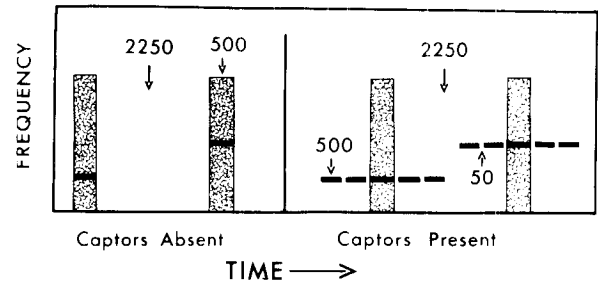


Figure 2. Stimuli used in Experiment 2. The vertical bars represent noise bursts and the dark horizontal bars are tones. The numbers represent the durations in milliseconds of the events or intervals.

were the same as those used in Experiment 1. They consisted of speakers, held from touching the ears by thin, flat foam cushions. The headphones had no circumaural pads and did not entirely eliminate the functioning of the pinna for environmental sounds, which could be heard clearly through the cushions. In the present experiment, in contrast with Experiment 1, the sounds that were to be judged actually did come from the speakers, but, despite the four Xs on the wall, only appeared at either the highest or lowest position.

Stimuli. The stimuli whose positions were to be judged were two short bursts of white noise. These could be presented by either the high (H) or low (L) speaker. A pure tone could be presented over the headphones in synchrony with the noise burst from the loudspeaker. The noise bursts occurred in four positional sequences: H-L, L-H, H-H, and L-L.

On trials in which tones were added to the noise bursts, there were basically two conditions, as shown in Figure 2. Either the tone had captors preceding and following it or it did not. The right-hand section of Figure 2 shows the "captor present" condition. In this condition, each tone-noise combination was preceded and followed by two tones at the same frequency as the one that was synchronous with the noise burst, to capture it into a sequential stream. It was expected that this capturing would remove the tone's contribution to the spectrum of the noise burst for the purposes of localization. Some trials containing two noise bursts without simultaneous tones were also included.

All tones and noise bursts were 500 msec in duration, and the silence between successive tones was 50 msec. The two noise bursts were separated by 2,250 msec. Trials were separated by 7 sec of silence during which subjects were to respond. The speaker signal was adjusted to a comfortable listening level and the loudness of the headphone signal was adjusted in accordance with judgments of three observers to seem equal in loudness to the speaker signal.

The tones used were at 260, 347, 433, and 520 Hz. On the trials with tones, the tone pairs were selected so that there were equal numbers ascending from 260 Hz to each of the higher frequencies and descending from 520 Hz to each of the lower ones. Pairs in which both tones were at 260 Hz or at 520 Hz were also included.

Subjects. Twenty-nine volunteers from the McGill undergraduate and graduate student body served as subjects.

Results

In Figure 3, the mean judgments of spatial change (converted to degrees) for the fused and segregated conditions are shown as a function of the pitch change. There is a significant association between the direction and size of pitch change and the judged spatial change [$F(7,196) = 6.27, p < .001$]. Furthermore, the interaction with the fused/segregated treatment, shown in

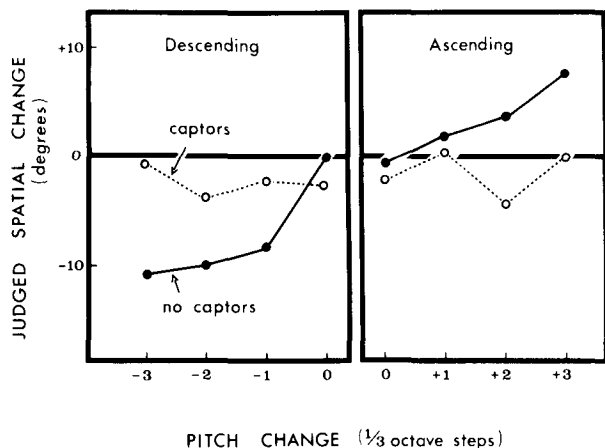


Figure 3. Experiment 2: Mean judgments of spatial change of burst location (converted to degrees) as a function of the pitch change of the accompanying tone (in one-third-octave steps) for conditions with and without captor tones. Results for ascending and descending tonal intervals are shown separately.

Figure 3, is significant [$F(7,196)=7.92$, $p < .001$]. The tonal effect in the fused condition is much stronger than in the segregated condition.

Localization of the unaccompanied pairs of noise bursts was also examined. Judgments were systematically related to the actual change in position of the pair of bursts. The average positional changes indicated by subjects on the four-point response scale for the four types of burst sequence were: -0.04 for high-high, -0.12 for low-low, -0.70 for high-low, and $+0.21$ for low-high. While related to the true spatial changes, the judgments show a rather restricted range of response compared with the range offered by the scale (-3 to $+3$). While the effect of true spatial change on the unaccompanied burst trials taken alone was not tested statistically, a test was made for the effect of true spatial change on all the trials, accompanied by tones or not. The subjects were picking up information about true spatial change [$F(3,84)=15.9$, $p < .001$], although their range of response was restricted. This rather limited localization may have been due to the wearing of headphones or to the acoustic properties of the room. Whatever its cause, the poor localization undoubtedly helped to create a somewhat ambiguous situation in which we were better able to see the effects of the added tonal stimulus. Needless to say, if the acoustic situation had provided strong cues for exact localization, the effects of the tone would have been much smaller and harder to measure. One finds, in general, that one observes stronger effects of any one type of perceptual cue whenever other possible cues are weaker. However, the point of the present experiment was not to estimate the relative size of the effects of the tonal cues as compared with naturally occurring cues, but to determine whether conditions exist in

which "melodic capturing" can influence spatial localization. One set of such conditions was found in the present experiment.

EXPERIMENT 3

This experiment had the same general form as the previous one, but the method of inducing segregation was changed. A pure tone accompanied a noise burst, but rather than causing the tone to segregate from it by the use of captor tones, we caused the noise and the tone to segregate due to a difference in their amplitude envelopes. We based this manipulation on published demonstrations that the onset and offset asynchrony of two tones played at once will influence their segregation (Bregman & Pinker, 1978; Dannenbring & Bregman, 1978; Deutsch, 1979) and will reduce the ability of the tones to mask one another's presence (Rasch, 1978). Imposing an asynchrony in the onset of two tones is similar to making their amplitude envelopes different. In both cases, one frequency or set of frequencies can be heard to increase in amplitude at a time when another frequency or set of frequencies is not. Helmholtz (1877/1954) had also suggested that the independence of loudness changes in frequency components could be a basis upon which the mind could determine which ones went together.

Method

In the present experiment, subjects were again asked to judge the spatial positional change in a sequence of two noise bursts coming from loudspeakers. Again, tones presented over headphones accompanied the bursts. This time, however, in the conditions in which segregation of the tone and burst was desired, the two sounds had different amplitude envelopes.

All procedures with subjects were identical to those in Experiment 2; however, different subjects received the treatments designed to fuse or not fuse the identities of the tone and noise. The same response method was used as in Experiment 1 (joining dots with arrows) except that a 5-point scale was employed to represent the 53 deg of auditory angle between the upper and lower speakers. Also, there were no noise bursts presented without tones. In the present experiment, the five frequencies used in Experiment 1 appeared as steady-state sine tones. There were 80 trials, subdivided into four types of trials in which the two bursts came from a different sequence of speakers: high-high, high-low, low-high, and low-low. They were also subdivided into trials involving 10 possible pitch sequences of the accompanying tones (the 10 sequences of Experiment 1). Two replications of each condition were given to each subject. The presentation apparatus was the same as was used in Experiment 1.

Stimuli. The stimuli were prepared on an ARP 2600 synthesizer and a mini-KORG monophonic synthesizer. In the conditions in which fusion of tone and noise was desired, the sine tone and white noise burst had identical amplitude envelope characteristics, consisting of roughly a 100-msec linear rise time followed immediately by an 800-msec linear decay. In the condition in which segregation was desired, the noise burst had a 200-msec linear rise followed immediately by a 750-msec linear decay; in this condition, the tone envelope rose more rapidly (100 msec) and then immediately began a slow 1,110-msec linear decay. The interburst interval was 1 sec. Successive trials were separated by

7 sec of silence during which subjects recorded their judgments. Because of variations introduced by manual setting of the apparatus and by the manual timing of the interburst intervals, all the values given above were subject to a $\pm 20\%$ nonsystematic variation from trial to trial.

Subjects. Fifty-six undergraduate and graduate students at McGill University served as subjects. All reported normal hearing.

Results

All data were treated in the same manner as the data from Experiment 2. The results are shown in Figure 4. This figure shows that the present experiment replicated the observations in Experiments 1 and 2 that the spatial change of the pair of bursts was strongly influenced in the expected direction by the pitch change in the accompanying tones [$F(9,486) = 23.12, p < .001$].

The theoretically relevant interaction effect came out as expected and replicated the one found in Experiment 2. There was a small, but regular and highly significant, tendency for subjects to be influenced more by the accompanying tones in the fused (same envelope) condition than in the segregated (different envelope) condition [$F(9,486) = 3.80, p < .001$].

Discussion

The results resembled those of Experiment 2. Although there was a substantial temporal overlap of tone and noise burst even in the segregated condition, the auditory system seemed to show some ability to reject the effects of the accompanying tone when the onsets and offsets were not synchronous. Notice that this rejection was not perfect. Even in the segregated condition, there was some tendency for spatial judgments to follow the tonal direction (see Figure 4).

GENERAL DISCUSSION

In Experiments 2 and 3, we instructed subjects to

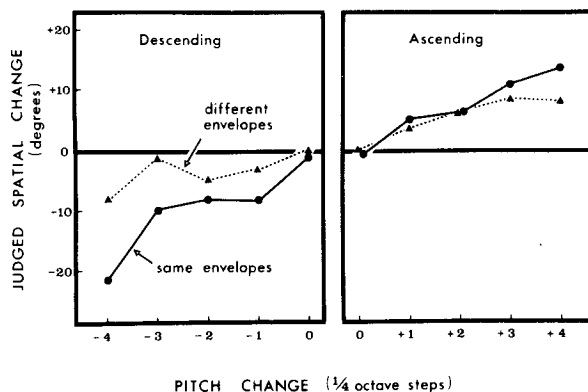


Figure 4. Experiment 3: Mean judgments of spatial change of burst location (converted to degrees) as a function of the pitch change of the accompanying tone (in one-quarter-octave steps) for conditions in which the noise burst and tone had the same or different amplitude envelopes. Results for ascending and descending tonal intervals are shown separately.

try to separately localize the noise burst, but apparently their voluntary attentional processes could not succeed in this task. However, we were able to help voluntary attention substantially by introducing two types of stimulus manipulation that are known to promote segregation of spectral components. The predicted interactions between condition of fusion and the effects of the tone were found quite clearly in the judgment data. We interpret these results as showing that sequential information can affect the parsing of the information for purposes of spatial localization. However, there are alternative interpretations that must be considered.

One such interpretation might appeal to some principle of sequential inhibition, as follows: In both Experiments 2 and 3, the listener was exposed to the pure tone in isolation shortly before the mixture; perhaps this served to attenuate the response of the auditory system to that pure-tone component in the mixture. This would explain why, in the segregated conditions, the tones had less influence upon localization.

We can, however, rule out this explanation, for it would imply that in the segregated condition the tones should be less audible in the mixture. Yet we know from other experiments, related to Experiment 2, that when an isolated pure-tone component precedes a mixture in which it occurs again, it is more *easily* heard as a component of that mixture (Dannenbring & Bregman, 1978; Van Noorden, 1975). Somehow, our perceptual "discounting" of the contribution of that tone to the mixture goes along with our actually hearing that tone *better* in the mixture. The segregation is bidirectional, influencing both the perception of the component that is pulled out of the mixture and the perception of the remainder itself. The latter half of this statement is supported by the fact that when one tone is perceptually segregated from a mixture of two pure tones, the remaining tone sounds more like a separate pure tone. So, a simple explanation based on inhibition of the perception of the "leading" tone in the subsequent mixture is not sufficient; it does not explain why *both* components are heard more clearly.

A second, alternative explanation would argue as follows: The main question underlying the present studies was, "How closely interconnected are the perceptual computations of location and identity?" It is claimed that the results demonstrate an intimate connection; under conditions in which features contributing to the perceptual identity are known to fuse, location information (spurious or valid) is also brought together. Yet, (the argument continues) a correlation does not necessarily imply a causal relationship; the present data do not directly prove the existence of an interactive process that computes a global "what is where" decision; they have simply shown that the "what" and the "where" processes are controlled by the same variables; that is, the fusion

of identity information takes place under the same circumstances that lead to the fusion of location information.

While this argument is logically consistent for Experiment 3, it faces problems with Experiment 2. In that experiment, the captor tones united with the noise-embedded tone because of the similarity of their frequencies. There is research (Bregman & Pinker, 1978) that shows that such capturing depends directly on the closeness in frequency of the capturing and captured tones. It is this same proximity in frequency that governs streaming and the formation of melodies. We have, then, in Experiment 2 a case in which what we can call "melodic capturing" influences the ability of the synchronized tone and noise burst to maintain independent spatial locations. Hence, it would appear that an identity-type process (melodic capturing) is directly influencing localization.

The opposite influence (of location on identity) is demonstrated in a study reported by Judd (1979), who alternated tones rapidly between the two ears. He found that when location information was clear, notes at each ear formed an independent melody. In his experiment, then, location is influencing the calculation of identity. Presumably, with Judd's procedure, the notes forming a familiar melody would only be recognized as such when they appeared in sequence in the same ear.

Deutsch and Roll's illusion (1976) may not reflect the independence of "what" and "where" decisions at all. The illusion probably came into existence due to a strong conflict of two principles of auditory processing. The first is that the auditory system tries to group sequential events that are similar in location. The second is that it tries to group sequential events that are similar in frequency. The result is supposed to be an auditory stream that is consistent with respect to both location and frequency pattern. When the two principles are put into conflict, as when the high and low tones are switched rapidly back and forth between the ears, other principles of localization may dominate. One such principle is that the higher harmonics of a complex tone can be used more reliably than the lower ones to detect the position of the source. Since Deutsch and Roll's 800-Hz high tone can be viewed as a harmonic of their 400-Hz low tone, the auditory system may have computed location from the more reliable higher harmonic. Rather than viewing this illusion as evidence for the separation of "what" and "where" computations, we can view it as evidence for their interconnection: The illusion only occurs when the streaming process cannot succeed in integrating two sources of evidence in the computation of coherent sequential streams. Bregman (1978), Deutsch (1975, 1979), and Deutsch and Roll (1976) have argued that illusions are side effects of a system trying to decide "what is where."

However, the ultimate point is not how often the two mechanisms for "what" and "where" act compatibly and arrive at the right answer about the world, but whether or not they exist as separate mechanisms at all, or perhaps whether each aspect of a perceptual "decision" is controlled by a separate type of information. Surely "what" and "where" are different types of *facts* about an environmental sound, and, unarguably, the auditory system's attempt to obtain a consistent description of one can lead to errors on the other. But these observations need not argue for two separate mechanisms. They could merely point to two different outputs from an integrated mechanism that computes these two aspects of our perceptual descriptions in a highly interactive manner.

It seems, then, that we should interpret the available physiological evidence about "what" and "where" pathways with caution. It may help if we draw a distinction between "perceived location," which is a psychological event, and "information usually used for determining location," which is a physical event. The same distinction can be made for identity. While it may be true that separate systems extract and carry the two types of physical information, nevertheless, in complex animals like ourselves, each type of information may be involved in the building of more than one aspect of the resulting psychological representation of the environment. Location-type information may help to specify perceived identity, as in Judd's (1979) experiment, and identity-type information, as in the melodic capturing of the present Experiment 2, may help to specify perceived location.

REFERENCES

- BREGMAN, A. S. The formation of auditory streams. In J. Requin (Ed.), *Attention and performance VII*. Hillsdale, N.J.: Erlbaum, 1978.
- BREGMAN, A. S. Asking the "what for" question in auditory perception. In M. Kubovy & J. Pomerantz (Eds.), *Perceptual organization*. Hillsdale, N.J.: Erlbaum, in press.
- BREGMAN, A. S., & PINKER, S. Auditory streaming and the building of timbre. *Canadian Journal of Psychology*, 1978, **32**, 19-31.
- BUTLER, R. Monaural and binaural localization of noise bursts vertically in the median sagittal plane. *Journal of Auditory Research*, 1969, **3**, 230-235.
- DANNENBRING, G. L., & BREGMAN, A. S. Streaming vs. fusion of sinusoidal components of complex tones. *Perception & Psychophysics*, 1978, **24**, 369-376.
- DEUTSCH, D. Musical illusions. *Scientific American*, 1975, **233**, 92-104.
- DEUTSCH, D. Binaural integration of melodic patterns. *Perception & Psychophysics*, 1979, **25**, 399-405.
- DEUTSCH, D., & ROLL, P. Separate "what" and "where" decision mechanisms in processing a dichotic tonal sequence. *Journal of Experimental Psychology: Human Perception and Performance*, 1976, **2**, 23-29.
- EVANS, E. F. Neural processes for the detection of acoustic patterns and for sound localization. In F. O. Schmitt & F. T.

- Worden (Eds.), *The neuro-sciences, third study program*. Cambridge, Mass: M.I.T. Press, 1974.
- EVANS, E., & NELSON, P. On the relationship between the dorsal and ventral cochlear nucleus. *Experimental Brain Research*, 1973, **17**, 428-442. (a)
- EVANS, E. F., & NELSON, P. G. The responses of single neurones in the cochlear nucleus of the cat as a function of their location and the anaesthetic state. *Experimental Brain Research*, 1973, **17**, 402-427. (b)
- HELMHOLTZ, H. L. *On the sensations of tone as a physiological basis for the theory of music* (2nd English ed.). New York: Dover, 1954. (Originally published, 1877.)
- JUDD, T. Comments on Deutsch's musical scale illusion. *Perception and Psychophysics*, 1979, **26**, 85-92.
- PRATT, C. C. The spatial character of high and low tones. *Journal of Experimental Psychology*, 1930, **13**, 278-285.
- RASCH, R. The perception of simultaneous notes such as polyphonic music. *Acoustica*, 1978, **40**, 1-72.
- ROFFLER, S. Factors that influence the localization of sound in the vertical plane. *Journal of the Acoustical Society of America*, 1968, **43**, 1255-1259.
- SEARLE, C., BRAIDA, L., CUDDY, D., & DAVIS, M. Binaural pinna disparity: Another auditory localization cue. *Journal of the Acoustical Society of America*, 1975, **57**, 448-455.
- VAN NOORDEN, L. P. A. S. *Temporal coherence in the perception of tone sequences*. Unpublished doctoral dissertation, Eindhoven University of Technology, Eindhoven, The Netherlands, 1975.

(Received for publication October 29, 1979;
revision accepted August 20, 1980.)