Hearing Musical Streams

Stephen McAdams
Hearing and Speech Sciences
Stanford University School of Medicine
Stanford, California 94305

Albert Bregman
Department of Psychology
McGill University
Montreal, Quebec, Canada

Introduction

The perceptual effects of a sound are dependent upon the musical context in which that sound is imbedded. That is, a given sound's perceived pitch, timbre, and loudness are influenced by the sounds that precede it, coincide with it, and even follow it in time. Thus, this context influences the way a listener will associate the sound with various melodic, rhythmic, dynamic, harmonic, and timbral structures within the musical sequence. It thus behooves the composer and interpreter to understand the various perceptual organizing principles that affect the derivation of musical context from sequences of acoustic events. We include the interpreter here because several musical dimensions, such as timbre, attack and decay transients, and tempo, are often not specified exactly by the composer and are controlled by the performer.

In this article we shall discuss principles that describe how various musical dimensions affect the perceived continuity of music. Leon van Noorden has stated that "in sequences where the tones follow one another in quick succession, effects are observed which indicate that the tones are not processed individually by the perception system. On the one hand we find various types of mutual interaction between successive tones, such as forward and backward masking, loudness interactions and duration interactions. On the other hand, a kind of connection is found between the successive perceived tones." [28, p.1]. As for simultaneous sonic events, Bregman [5, 9] has suggested that different sounds are extracted according to various perceptual and cognitive organizational mechanisms from the superimposed acoustic vibrations. While some researchers would like to attribute these phenomena to mechanisms in the peripheral or early central nervous system (which are surely involved to some extent, see [24, 33]), prominent members of this more psychophysically oriented "school" are beginning to think the organization is too complex to be so easily explained. However, rather than delve into theoretical explanations of these phenomena, it will suffice for the present purpose to describe them in general and to briefly quantify some salient parameters that have intrigued us with compositional possibilities. This article thus presents, in a tutorial fashion, a review of research (including our own) which has direct implications for musicians, especially for composers working with computer music. In all of our research and in most other research cited, computers (predominantly PDP-11 minicomputers) were used to synthesize the sounds presented. They were also used for presentation of sound stimuli, collection of responses from the listeners, and analysis of the data. For thorough summaries and theoretical treatments of this area of research, see [2, 5, 9.].

What Is An Auditory Stream?

Auditory stream formation theory is concerned with how the auditory system determines whether a sequence of acoustic events results from one, or more than one, "source." A physical "source" may be considered as some sequence of acoustic events emanating from one location. A "stream" is a psychological organization that mentally represents such a sequence and displays a certain internal consistency, or continuity, that allows that sequence to be interpreted as a "whole." By way of example, two possible perceptual organizations of a repeating six-tone sequence are illustrated in Figure 1. Time is represented on the horizontal axis and frequency is represented on the vertical axis. The dotted lines connecting the tones in the figure indicate the stream percepts. In the first configuration, six tones are heard one after the other in a continuously repeating cycle (Taped Illustration 1a)[1]; it is easy to follow the entire melodic pattern. In the second percept, though, one might hear two separate three-tone patterns which appear to have little relationship to each other (Taped Illustration 1b). It is difficult in this case to follow the original six-tone pattern. Note that in the first example one stream is heard, and in the second, two are heard.

---

(1) A tape containing sound examples is available from Mr. McAdams. Descriptions of the taped illustrations are found in Appendix 1.

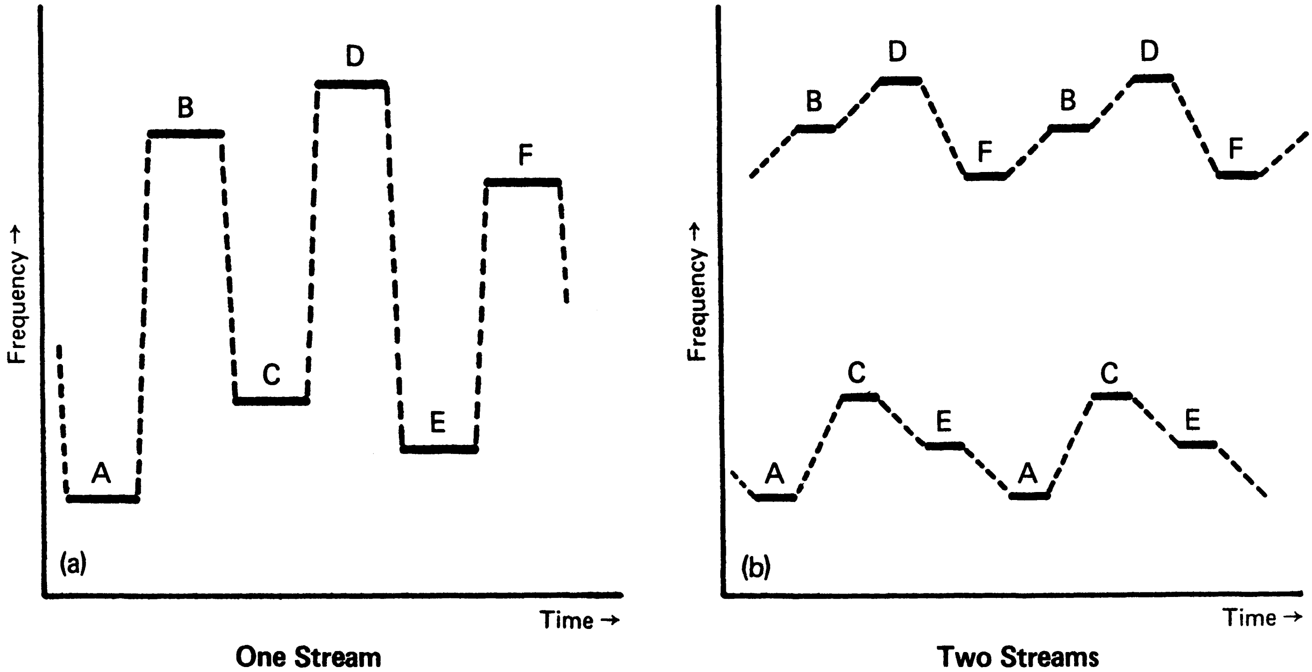


Figure 1. A repeating 6-tone sequence composed of interspersed high and low tones can result in different percepts. In Figure 1a, with high and low tones alternating at a tempo of 5 tones/sec., one perceptual stream is heard. In Figure 1b, at a tempo of 10 tones/sec., the high tones perceptually segregate from the low tones to form two streams (*cf.* Taped Illustration 1).

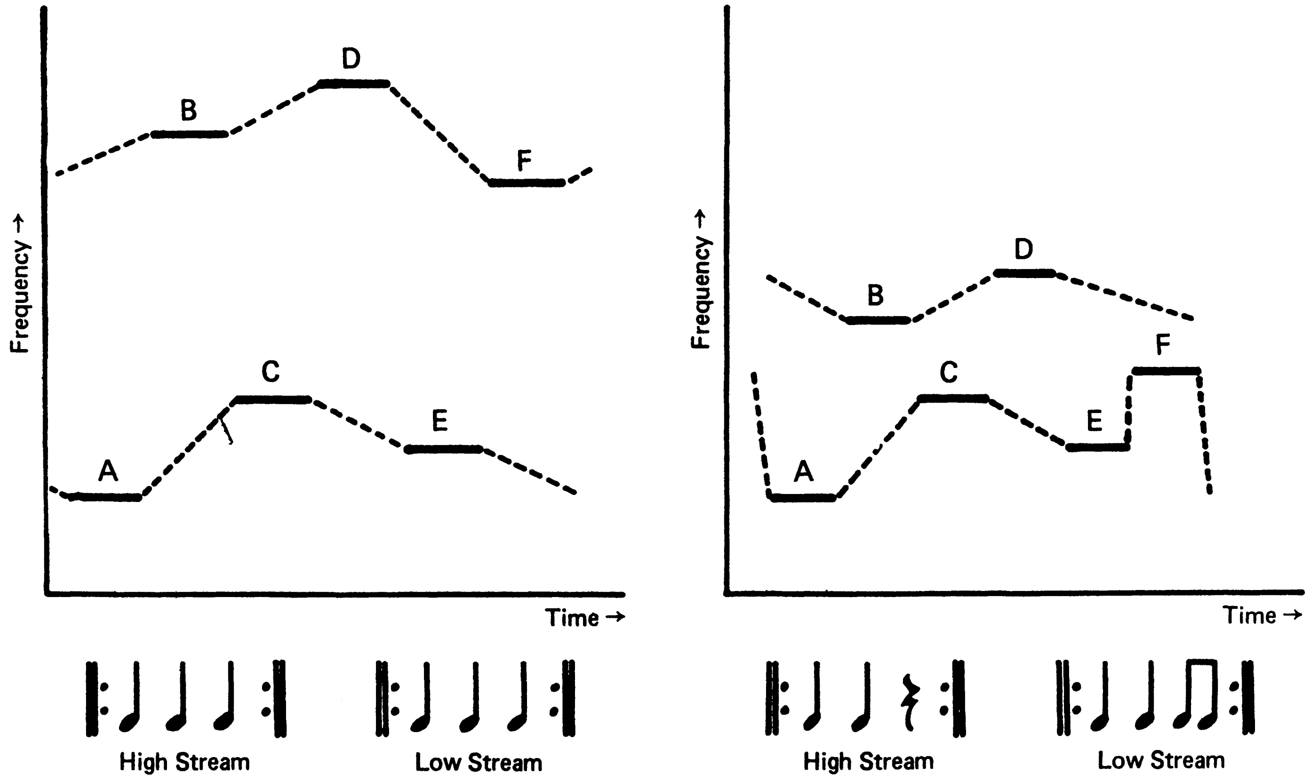


Figure 2. Due to the competition among stream organizations, tone F may be perceived as belonging to either the higher stream or the lower stream but not to both. The organization of the streams changes, among other things, the perceived rhythmic structure, as indicated under each diagram (*cf.* Taped Illustration 2).

In the rest of this section we will discuss some of the properties exhibited by a stream. For example, it is possible to focus one's attention on a stream and follow it through time [6, 28]. In the first taped example one can follow the six-tone pattern without any trouble. Thus a stream must exhibit a certain coherence over time. However, in the second example, it is difficult to follow the six-tone pattern, but it is easy to follow either three-tone pattern. Notice that one can pay attention to either the higher or lower stream, switching between them at will, but that it is not possible to attend to both simultaneously (Taped Illustration 1b). Indeed, each three-tone pattern in Figure 1b constitutes a separate stream and maintains its own temporal coherence. While a listener is paying attention to one coherent stream, other acoustic information is perceptually relegated to the background. If one group of sounds is distinct enough, the foreground-background relation may be almost involuntary and it may require a great deal of attentional effort to focus on streams initially relegated to the background.

The information-processing nature of the stream segregation process is suggested by the observation that the segregation of a sequence into smaller streams takes time to occur [4, 13]. In Figure 1b, notice that one can hear a six-tone pattern for the first few cycles before it segregates into two separate streams (Taped Illustration 1b). It thus appears that the perceptual system assumes things are coming from one source until it acquires enough information to suggest an alternate interpretation.

When temporal coherence is lost in a sequence, it becomes more difficult to order the events of that sequence in time [6, 11, 14, 28]. In Figure 1a, it would be easy to tell the order of tones A, B, and C. But as this larger stream breaks down into the smaller streams of Figure 1b, *i.e.*, as temporal coherence is lost, it becomes more difficult to judge the order of these tones. In such a case one might notice that tone A comes before tone C but it would be hard to tell whether tone B came before A, between A and C, or after C. This statement should be qualified by noting that in the tone sequences mentioned all of the tones are of equal loudness and timbre. The tone sequences are furthermore continually recycling and are faded in so that the listener cannot label tone A as being the first tone in the sequence. Information in a musical context, such as the first beat being emphasized as a downbeat, may give the listener an anchor point against which to relate the temporal positions of other tones. Thus one can judge the order of events in time within a given perceptual stream but not necessarily across streams. Accompanying this is the observation that different streams can appear to overlap in time even when they do not. Again in Figure 1b, notice that one *can* hear two three-tone patterns apparently going on at the same time even though the tones are alternating (Taped Illustration 1b).

If an event is potentially a member of more than one "competing" stream, one may perceive it as belonging to one stream or another but not to both simultaneously [3, 10, 11]. This is not to say that a musician cannot hear several simultaneous lines. The point is that it is impossible to use several *parsing schemes* at the same time. Figure 2 illustrates the effect on our second example of moving the higher triplet into closer frequency proximity to the lower triplet. We can find an intermediate position where the lowest tone, tone F, can group with either the higher or lower stream (Taped Illustration 2). Notice that this regrouping results in a rhythmic transformation as shown under each configuration. Some non-musical examples of this phenomenon are the face-vase illusion and the reversible Necker cubes; Escher and Vasarely have produced art works based on such principles of perceptual organization.

Finally, this brings up the relationship between "source" and "stream." A stream is perceived as emanating from a single source. So, in the first example, the pattern was fairly continuous and was easily recognized as coming from one source; but in Figure 1b, the large frequency distance between the two three-note groups introduced a sort of discontinuity that caused the perceptual system to interpret the sequence as resulting from two sources. Since at any given moment the composite pressure variations stimulating the ear result from several sources, the auditory system needs a battery of heuristics to parse, or segregate, the information into separate streams. It thus needs to build a description of the acoustic environment from separate descriptions of the various streams and the relationships between them [2]. Factors which the perceptual system uses to build descriptions of streams, and subsequently sources, are frequency, rate of occurrence of events (or tempo), intensity, timbre, and attack/decay transients. In the rest of the paper these will be discussed in some detail. Of course it is obvious that sounds are assigned perceptually to different sources when the physical sources are at different spatial positions. In this case, intensity, spectral, and temporal cues are all utilized to parse the sound into separate sources. However, we will primarily confine our discussion to the *illusion* of many sources which occurs due to the organizations within a single emanation of sound.

Frequency and Tempo

Consider that a repetitive cycle of tones spread over a certain frequency range may be temporally coherent, or integrated, at a particular tempo. It is possible to gradually increase the tempo until certain tones group together into separate streams on the basis of frequency, as discussed above. The faster the tempo, the greater the degree of breakdown or decomposition into narrower streams until ultimately every given frequency might be beating along in its own stream. This last possibility is dependent upon a number of other factors which will be discussed later.

Figure 3 illustrates the possible stages of perceptual decomposition for a recycling six-tone pattern as one gradually increases the tempo (Taped Illustration 3). Note that streams *per se* are not tracked beyond a certain point, but a texture or timbre is perceived since the ability to temporally resolve the individual tones degenerates altogether. Of course, one could hold the tempo constant and gradually expand the frequency relationships to achieve a similar streaming effect [3, 6, 14, 17], but the musical consequences would be vastly different as one can hear in Taped Illustration 4. Dowling [17] used simple melodies to illustrate this frequency-based streaming principle. He interleaved two melodies in the same frequency range thereby making it very difficult, without prior knowledge of the melodies, to separate them perceptually. But as they were pulled apart in frequency, *i.e.* when all the tones of one of the melodies were transposed upward, each melody became apparent (Taped Illustration 5).

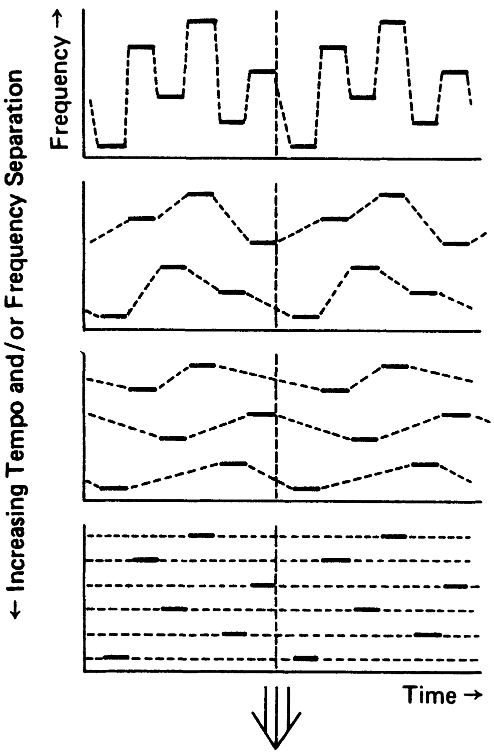


Figure 3. This figure illustrates the decomposition of an acoustic sequence into smaller and smaller perceptual streams as the frequency separation between the tones or the tempo of the sequence increases. In the latter case, a point is ultimately reached where one can no longer perceive individual tonal events; a texture or timbre is heard instead (*cf.* Taped Illustrations 3 and 4).

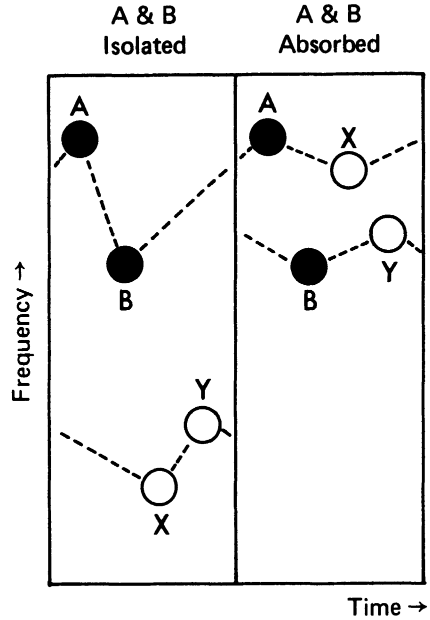


Figure 4. This figure illustrates the effect of frequency context as opposed to frequency separation on stream formation. In the first part, tones A and B form one perceptual stream which is unaffected by the simultaneous stream composed of tones X and Y. When tones X and Y are moved into the proximity of tones A and B, which remain unchanged, an alternate perceptual interpretation results whereby tones A and B are assigned to separate streams on the basis of a new frequency context (*cf.* Taped Illustration 6).

The frequency range within which the perceptual system groups tones on the basis of frequency proximity is not constant; the grouping can vary with the particular pattern of frequencies presented. For example (see Figure 4), two tones, A and B, are arranged with given frequency and temporal separations such that they will always stream together when no other tones are present. We can create a similar organization with tones X and Y in another frequency range. These two groups will each form a stream as can be heard in Taped Illustration 6a. By bringing the pairs into the same frequency range, new streams can be formed, such as A–X and B–Y, on the basis of an alternative proximity organization [3] (Taped Illustration 6b). Thus, the particular relationships between frequencies in a tonal pattern, and not just the frequency separation between adjacent tones, plays a vital role in the formation of streams.

It appears from our examples that there is an essentially inverse and strictly interdependent relationship between tempo and frequency relationships among individual tones [26, 28]: the faster the tones follow one another, the smaller the frequency separation at which they segregate into separate perceptual streams. Conversely, the greater the frequency separation, the slower the tempo at which segregation occurs. This illustrates yet another aspect of music which, with the aid of the computer, can come under the composer's control.

Suppose we make a graph, as in Figure 5, which relates frequency separation of two alternating tones on one axis to the rate of alternation of the tones on the other axis. Note that the horizontal axis indicates increasing tone repetition time, which corresponds to decreasing tempo. One can draw boundaries on this graph indicating the frequency-tempo regions in which the tones cohere as a single stream and those in which they segregate into two simultaneous streams of different frequencies. There are two such boundaries. Leon Van Noorden [28] has termed these the *fission boundary* and *temporal coherence boundary* and noted that they are slightly different for each person. In Figure 5 the upper curve is the temporal coherence boundary. Above this boundary it is impossible to integrate the two alternating tones into one stream. Below this boundary lies the temporal coherence region, where it is possible to integrate events into a single perceptual stream. The lower boundary is the fission boundary. Below this it is impossible to hear more than one stream. Note that the regions of fission and coherence also overlap, creating an ambiguous region where either percept may be heard.

As the tempo decreases temporal coherence can be maintained at greater frequency separations. However, the frequency separation required for fission remains fairly constant with decreasing tempo. The temporal coherence and fission boundaries above and below a given tone are symmetrical with respect to pitch. This indicates that the phenomena of temporal coherence and fission may occur depending on the absolute musical interval between the tones but irrespective of the direction of pitch change. While there are substantial quantitative differences in these boundaries

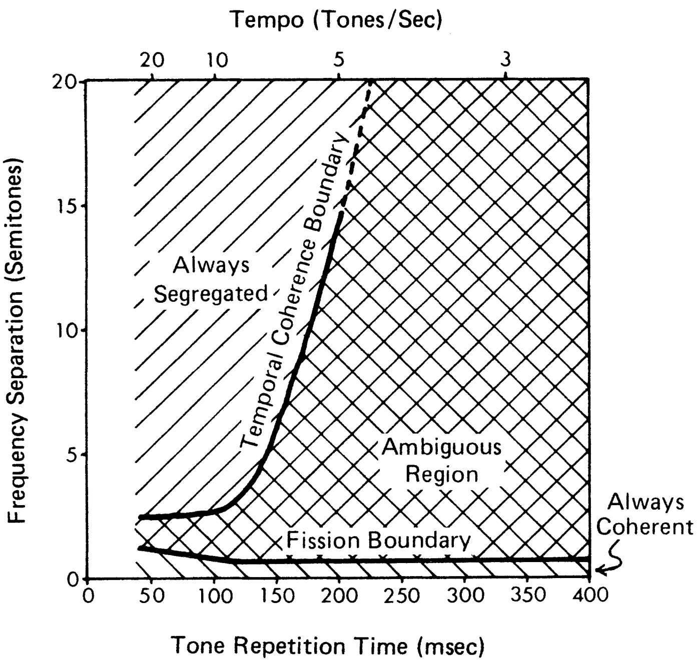


Figure 5. Boundaries of temporal coherence (upper curve) and fission (lower curve) define three perceptual regions in the stream relationship between two alternating tones each lasting 40 msec. This relationship is a function of both the tempo of alternation and the frequency separation between the tones (after [28]).

between listeners, the qualitative trends tend to be similar [28]. Note that the difference between boundaries is increasingly substantial for tempi below about 10 tones/sec. (tone repetition time = 100 msec.). This means that at very slow tempi of, say, five tones/sec., a separation of more than a minor 10th is necessary to induce streaming. Below this it becomes virtually impossible to induce streaming. In the limiting case, if the temporal distance between tones is too great, they do not seem to be connected at all but sound as isolated events.

A musically relevant aspect of these boundaries should be mentioned. The region between the two boundaries may be considered to be an ambiguous region since either a segregated or an integrated percept may be heard. The primary determining factor in this region is attention. In other words, it is possible to shift one's attention back and forth between the two percepts when the sequence falls in this region. For example, one may focus either on a whole stream percept or on smaller individual streams. It appears that the closer the sequence lies to one of the boundaries, the easier it is to focus on the percept which is predominant beyond that boundary. Conversely, it is very difficult in this situation to shift one's attention back to the other percept. Once the physical values go beyond either of these boundaries, attaining the complementary percept may be considered to be impossible. The role of attention will be discussed in more detail later.

Frequency Trajectories

Another frequency-based effect involves frequency trajectories. These are important on two levels. The first involves trajectories between tones (see Figure 6). Bregman and Dannenbring [7] have found that tones that are connected by glissandi are much less likely to segregate under given conditions of tempo and frequency separation than those which make abrupt frequency transitions. Intermediate situations beget intermediate results. Yet even when using a sine wave for frequency modulation of a sine tone, it is possible to discern a sort of streaming effect of the higher and lower peaks of the modulation at certain modulation frequencies (Taped Illustration 7). At lower modulating frequencies one can track the modulation, but higher modulating frequencies result in the effect of a texture or timbre. This continuum is found for modulation involving discrete changes in frequency as well.

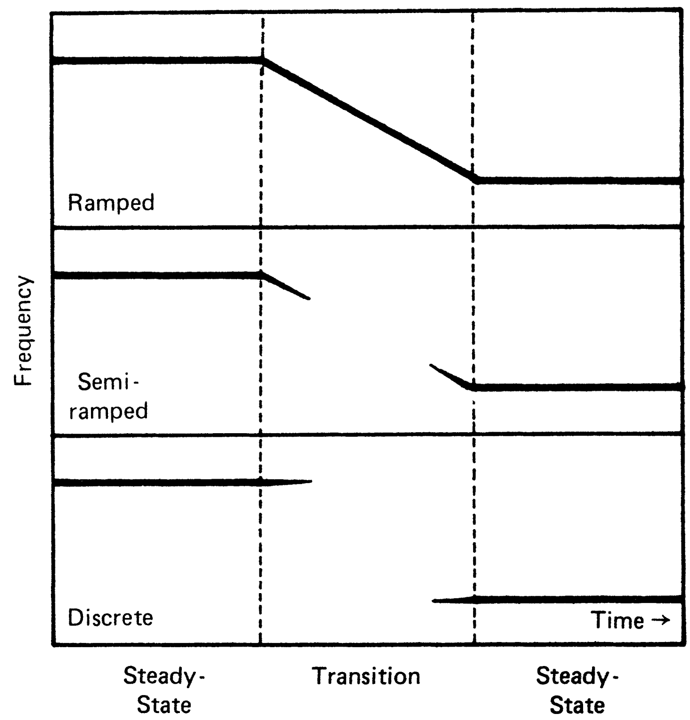


Figure 6. These are the 3 types of frequency transitions between tones used by Bregman and Dannenbring (1973). In the top section the tones are completely connected by a frequency glissando. In the middle section an interrupted glissando is directed towards the succeeding tone. No frequency glide occurs in the bottom section; the first tone ends on one frequency and the next tone begins on another (after [7]). The unconnected tones segregate more readily than the others.

The second type of trajectory might be called a melodic trajectory. The basic rule goes: large jumps and sudden changes in direction of a melody produce discontinuity in that melody. In terms of stream formation, one or two tones in a melody that are removed from the melodic continuity of the rest could be perceived as coming from a different source and would not be integrated into the melody as a whole, perhaps leaving a rhythmic gap in the phrase depending on the nature of the main sequence. It has been reported, however, that the excluded tones are sometimes noticed in the background, with their absence having little effect on the main melody line [23]. Implied polyphony in a solo compound melody line, as in the suites for solo instruments by Bach, is a compositional use of this principle.

Schouten [30] reported that if an ascending and descending major scale fragment played with sine tones is continually repeated, temporal coherence is maintained up to a tempo of about 20 tones/sec. However, if these same tones are arranged at random, the maximum tempo at which coherence still occurs is reduced to about 5-10 tones/sec. It might be inferred that the reduction of predictability reduced the pitch boundary within which the auditory system could successfully integrate the incoming information. But van Noorden [28] found that previous knowledge of the order of tones had no effect on the coherence boundary. It might be that the small frequency jumps in Schouten's demonstration are effective in holding the stream together.

Heise and Miller [23] investigated four melodic contours: a V-shaped contour, an inverted V, and rising and falling scale patterns; the V-shaped patterns which change direction can be thought of as being less "predictable" than the scale patterns which move in only one direction. Each pattern was eleven tones long and the frequency of the middle tone was variable. They found that the degree to which the variable tone could be separated in frequency from the rest of the pattern before it segregated into its own stream was a function both of the shape and of the steepness of the pattern. The steepness was varied by keeping the tempo constant and varying the interval between successive tones. As the rate of frequency change for the entire pattern increases over time, so does the amount of frequency separation of the middle tone required to produce segregation. Less separation of the middle tone is required to produce segregation with the V-shaped patterns than with the scale patterns, possibly indicating that the perceptual system can follow "predictable" patterns more quickly than "unpredictable" ones. (However, certain anomalies in their data, which will not be discussed here, might suggest other interpretations.)

Van Noorden found similar results for patterns with as few as three tones [28]. He investigated the relative temporal coherence of so-called linear and angular three-tone sequences. For linear sequences, *i.e.*, those with two tone intervals in the same direction, the temporal coherence boundary occurs at faster tempi than were found with angular sequences in which the melodic pattern changed direction at the middle tone. In this latter case the first and third tones are more contiguous in frequency than are the first and second, or second and third tones, which facilitates a loss of temporal coherence similar to that found by Schouten. It should be noted that while these melodic trajectory effects do seem to play a role in musical stream formation beyond the simple frequency separation effect, they are certainly confounded by other contextual organizations.

Both of these trajectory examples illustrate a principle of perceptual organization that uses pattern continuity, in some form or another, as a criterion for "source" distinction. Figure 7 illustrates, however, that frequency proximity may sometimes compete with trajectory organization; Deutsch found that simultaneous ascending and descending scale patterns presented to opposite ears segregate into upright and inverted V-shaped melodic contours [16]. Each contour is heard as if being presented to one ear. The same result was found by Halpern [22] for simultaneous ascending and descending sine tone glissandi, as can be heard in Taped Illustration 8; in this case both glissandi were presented to both ears. In these two examples a stream boundary is established at the pitch where the two lines cross.

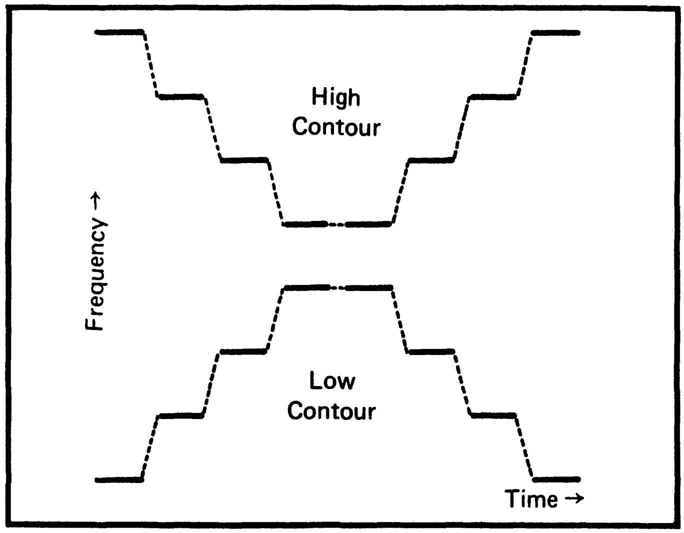


Figure 7. This stimulus, used by Deutsch (1975), is composed of ascending and descending scales presented simultaneously to opposite ears. Two V-shaped patterns (high and low contours outlined with dashes) are perceived rather than the complete ascending and descending scale patterns.

Loudness and Continuity Effects

Fission can be obtained by alternating sequences of tones similar in frequency and timbre but differing in intensity. Figure 8 illustrates the range of percepts achieved as the amplitude level of tone A is varied relative to that of tone B in the alternating sequence ABAB . . . . The reference level for tone B used by van Noorden in these experiments [28] was 35 db SPL; the frequency of both tones was 1 KHz and each tone lasted 40 msec. If the level of tone A is below the auditory threshold (approximately 6 db SPL at 1 KHz), only the B stream is heard at half tempo, as might be expected (see Figure 8a). When tone A is loud enough to be heard and is at least 5 db below tone B, two separate streams of different loudness can be perceived, each at half tempo, with A being the softer stream (see Figure 8b). When A is within 5 db of B, a "pulsing" stream is heard and neither the A nor the B stream can be heard independently, *i.e.*, temporal coherence is inevitable in this range (see Figure 8c). As the level of the A tone is increased above that of the B tone, three different percepts may result, depending on the alternation tempo of A and B. If this tempo is less than about 13 tones/sec., fission is the next percept heard. This time B is the softer stream (see Figure 8d). Thus a certain degree of loudness difference allows us to focus on either stream using only this information. If the tempo is greater than about 12.5-13 tones/sec., the percept encountered is the "roll" effect discovered by van Noorden. It sounds as if stream A, the louder stream, were pulsing at half tempo as in the fission percept, but the B stream sounds as if it were pulsing at full

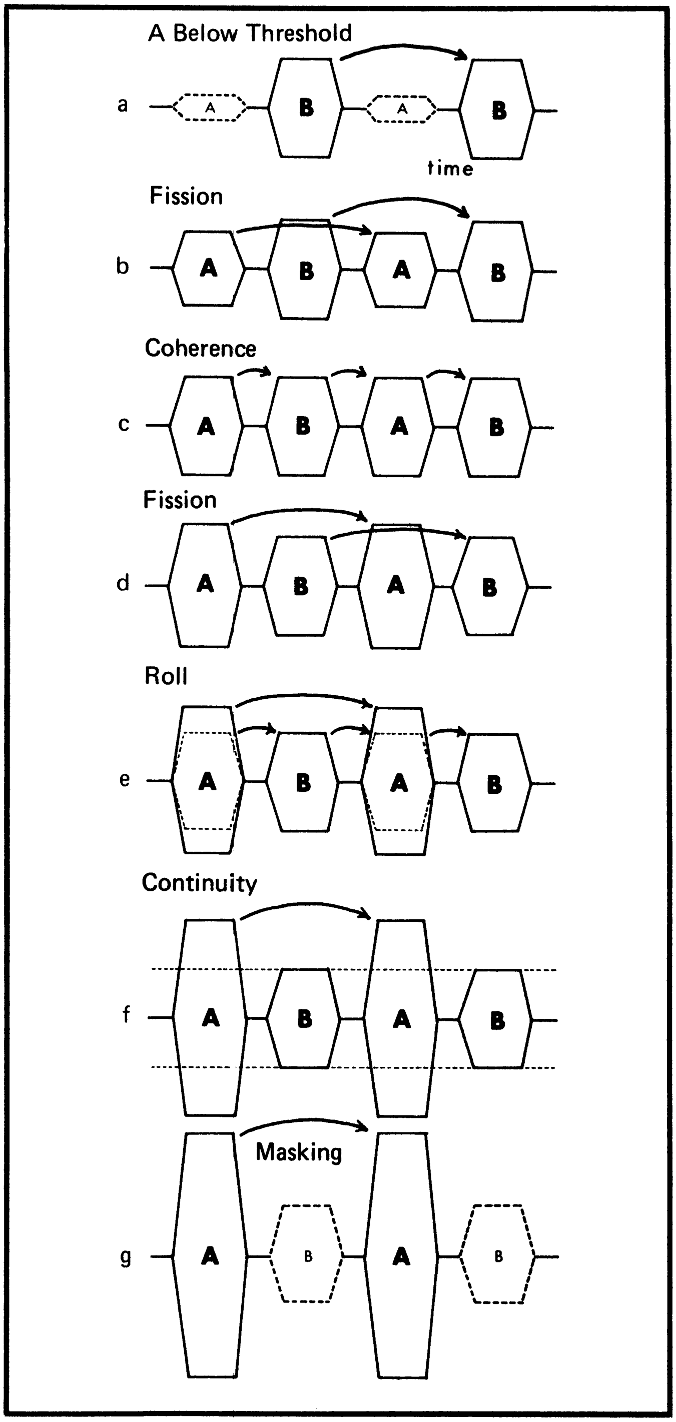


Figure 8. This figure illustrates the range of possible percepts found by van Noorden (1975) for two alternating sine tones differing in each case only in their intensities. Tone B was kept at 35 db SPL throughout; both tones were 40 msec. long and had a frequency of 1 KHz. a) The level of tone A is below the auditory threshold. b) Tone A is at least 5 db below tone B. c) Tone A is within 5 db of tone B. d) Tone A is louder than tone B with an alternation tempo less than about 13 tones/sec. e) Tone A is louder than tone B with more than 13 tones/sec. f) Tone A is about 18-30 db louder than tone B and the tempo is still above 13 tones /sec. g) Tone A is more than 30 db louder than tone B. The arrows indicate the percepts reported; a more complete description is given in the text.

tempo (see Figure 8e). Thus the A stream may be heard independently, but not the B stream. In other words, it is as if the A tones consisted of two parts: one that combines with the B stream to give a full tempo roll, and another, at half tempo, which can be perceived separately. At a tempo of about 13 tones/sec. another effect emerges when the level of A is about 18-30 db above that of B. This is the continuity effect shown in Figure 8f, so named because tone B is not heard as pulsing but rather as a fairly soft, continuous tone under the louder, pulsing A stream; this is an example of a class of effects that will be discussed below. Finally, if the level of the A stream is incremented still further, this stream completely masks the B stream (Figure 8g). Again, this set of loudness-based phenomena exhibits an ambiguous region between the coherence and fission boundaries where one might pay attention to either the A or B streams individually, or to the AB stream as a whole. There are thus three perceptual regions for alternating tones at the same frequency where the tempo is above about 12.5 tone/sec.: the *roll region*, the *continuity region*, and the *temporal coherence region*.

Van Noorden made quantitative measurements of the fission boundary (see Figure 9). For tempi of about 2.5 to 10 tones/sec., the fission boundary is more or less horizonal, *i.e.*, the intensity difference (ΔL) necessary for a segregated percept does not change with tempo over this range, but lies about 2 to 4 db on either side of the reference tone level. For tempi less than 2.5 tones/sec., the minimum level difference for fission increases with decreasing tempo (or longer inter-tone intervals of silence) and is symmetric about $\Delta L = 0$ (no difference in level). For tempi greater than 10 tones/sec. the level difference at which fission occurs increases with increasing tempo but the situation is not sym-

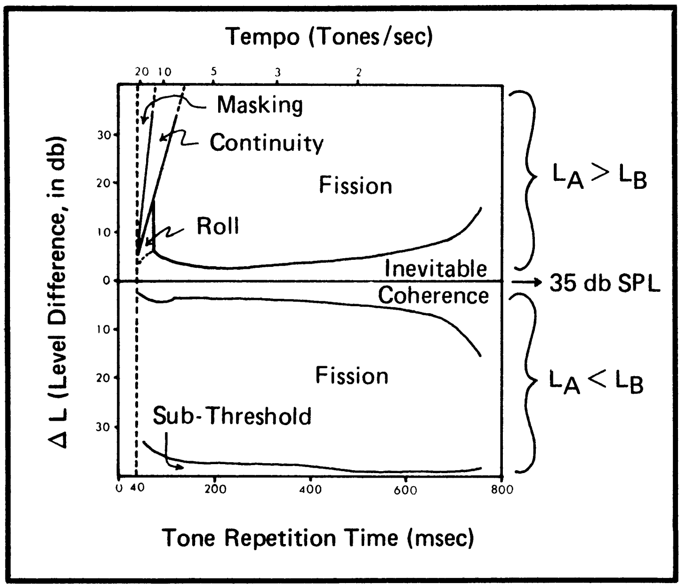


Figure 9. The percepts shown in Figure 8 fall into different regions of tempo-intensity combinations. The vertical dotted line corresponds to the duration of each tone. The level differences between tones A and B are relative to the level of tone B at 35 db SPL. The horizontal axis is shown as tone repetition time on the bottom and as its reciprocal, tempo, on the top. It should be pointed out that Figure 5 and Figure 9 cannot be directly compared, because the alternating tones in Figure 5 were played at different frequencies.

metrical about $\Delta L = 0$; for tempi greater than this, there is the complex interaction of level difference and tempo which results in the roll, continuity, and masking regions.

A combination of frequency separation and level difference was used to obtain the roll and pulsation thresholds shown for two different tempi in Figure 10. A great deal of psychoacoustic research has been devoted to the boundary between the roll and continuity regions. The "pulsation threshold," as this boundary has been termed [24, 28, 33], requires a greater loudness difference at slower tempi than does the roll threshold, as can be seen in Figure 10.

The continuity effect may operate according to principles very similar to those which govern *occlusion* in vision. The interpretation that one object occludes another is given by information derived from the intersecting edges of the objects. In Figure 11, the fragments have no apparent structure since information regarding the shape of the occluding structure is not present. However, in Figure 12, with the appropriate edge information, the fragments can be grouped into meaningful chunks. The same process may be postulated to occur in audition. That is, the "edges" of a louder sound and a contiguous softer sound might be used to induce the perception that the less prominent sound is partially occluded, or masked, by the louder. In such a case, the illusion of continuity of the softer sound "behind" the louder sound may result. This has been found to be true for sine tones masking sine tones, noise masking sine tones, and noise masking noise [8, 14, 24, 28, 32, 33, 34].

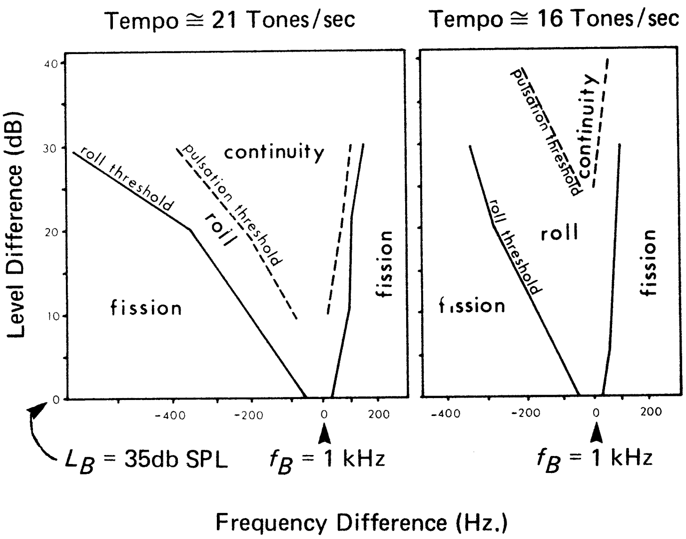


Figure 10. The boundaries between the fission, roll, and continuity regions already discussed in Figures 8 and 9 are shown here as a function of frequency and level differences. There is an increase in slope for roll and pulsation thresholds and an upward shift in the latter at slower tempi. The frequency and level differences are given relative to the frequency and level of tone B, *i.e.* 1000 Hz. and 35 db SPL, respectively. 0 KHz. frequency separation corresponds to the vertical dashed line in Figure 9. (Note that Figure 9 cannot be mapped directly onto Figure 10; in the former illustration, listeners were asked to try to hear fission, but in Figure 10, the test subjects were listening for roll. The different focus on the part of the listener results in a shift in the boundaries shown.)

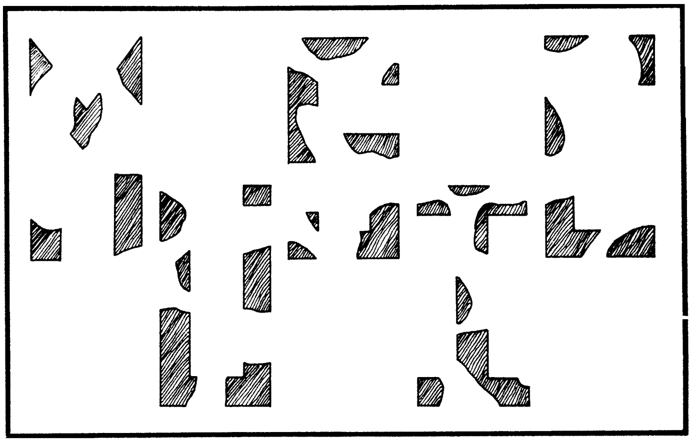


Figure 11. These fragments have little apparent structure and meaning since information regarding the shape of the occluding structure is not present.

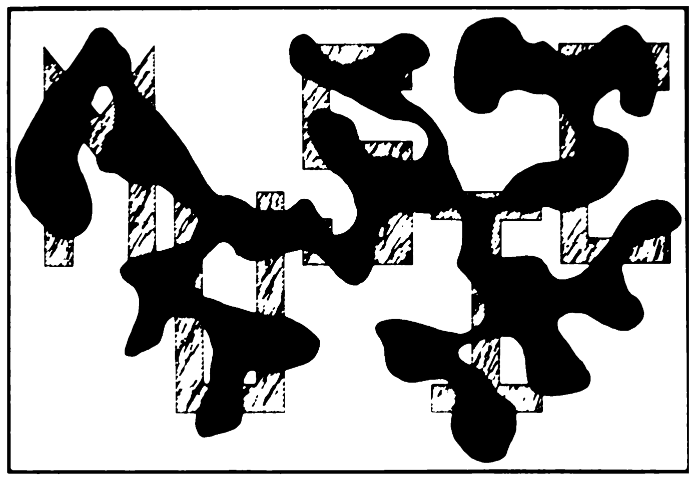


Figure 12. With appropriate information about the edges of the occluding structure, the fragments of Figure 11 can be grouped into meaningful chunks.

That the nature of the temporal "edge" between the two sounds is the important factor was illustrated by Bregman and Dannenbring [8]. As Figure 13 illustrates, they varied the transitional amplitude of a tone both leading into the attack and coming out of the decay of a noise masker. In reality, the tone stopped the instant the noise began and then began again the instant the noise stopped. The greatest continuity was found when there was no change in the amplitude of the tone. When the level of the tone was either increased or decreased before the attack of the noise, the illusion of continuity was substantially diminished. The amount of change in judged continuity was greater for the initially decreasing amplitude ramp than for the initially increasing ramp. The change in the amplitude of the tone may suggest a possible termination of the tone sometime during the noise burst. The length of the ramp was also found to have an effect on perceived continuity. The least continuity was found when the amplitude ramps lasted 50 msec., and for longer ramps there was less of a break in perceived continuity. This may suggest that a very long ramp would be interpreted as no ramp since the rate of change of intensity would approach zero.

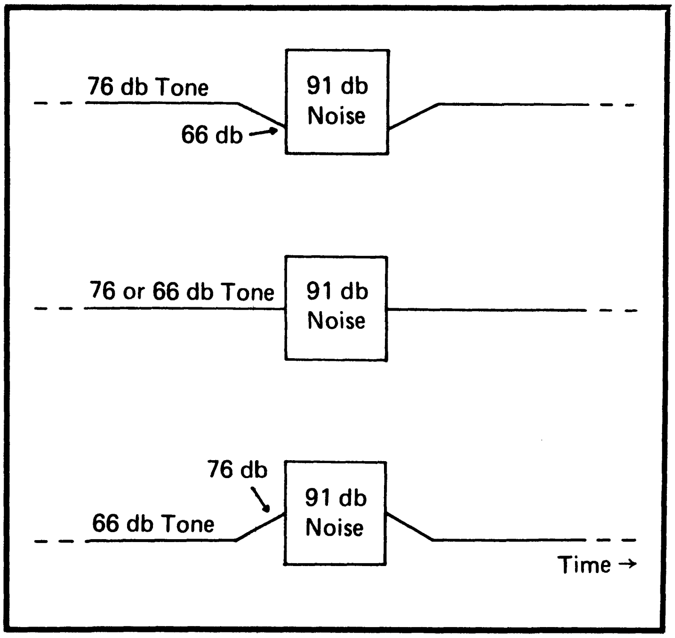


Figure 13. The stimuli used by Bregman and Dannenbring (1977) consisted of a pure tone interrupted by noise. The level of the tone leading into and coming out of the noise burst was varied as shown. The greatest degree of continuity was found in the center case where there was no change in level.

It has also been reported that alternation speed has almost no effect on the perceived continuity [14].

An important finding in relation to this phenomenon was reported by Warren, *et al.* [34]. They noted that there must be some degree of spectral overlap in the two signals to obtain the continuity effect. With sine tones, for example, the effect is greatest when the tones are identical in frequency, and falls off monotonically as the frequency separation on either side of the masker increases. The effect was even found to occur through a burst of noise as long as 50 seconds [34].

Though some people think the pulsation threshold reflects only a neural excitation pattern that decays gradually, with the perception of continuity resulting from overlap of excitation at smaller time intervals [24, 33], the following demonstration illustrates that higher-level context-dependent predictive mechanisms may be involved. These seem to parallel the predictive nature of the stream formation mechanisms mentioned previously. In Taped Illustration 9, one hears speech that has been gated; it is difficult to understand the spoken message. Then the gaps in the speech are filled with white noise. This effect was first verified experimentally by Miller and Licklider in 1950 [27]. The same sort of phenomenon with gated music and then gated-plus-noise-filled music is presented in Taped Illustration 10. We have chosen a familiar composition to dramatize the effect. It may be very difficult to recognize it in the gated condition, but should become easily recognizable when the gaps are filled with noise. In both of these demonstrations the amount of information removed and then replaced, or induced by the perceptual system, is far too large and far too complex to be entirely attributed to interaction between localized excitation patterns.

Timbre

Another musically useful dimension that can play a role in streaming effects is timbre. Timbre tends to be the psychoacoustician's multidimensional waste-basket category for everything that cannot be labeled pitch or loudness, including short-term spectral changes such as onset transients, long-term spectra, those dynamic qualities which a musician would term "texture," and so on. The important thing to notice about stream formation is that any group of simultaneous spectral components which move roughly parallel to each other in the pitch (logarithmic frequency) domain are most likely to result in a *fused* percept and thus to be interpreted as emanating from one source [22, 25]. So even though the individual spectral envelopes of two instruments or sources change over time, they each follow patterns which are sufficiently consistent internally to be fused and yet different enough, in most cases, to distinguish them one from the other. There are, of course, many situations in which especially sensitive instrumentalists can blend their tones and become as one voice, but these situations usually involve very slow, smooth changes where onset transients cannot give the listener the information needed to separate the sources. As an example, how often is a performance of Ravel's *Bolero* precise enough so that the instruments playing the various "harmonic" parts are not heard individually, but fuse into one voice? The fact that one typically hears them separately points to the extreme sensitivity of the auditory system to even minute variations in temporal and spectral elements.

Let us examine one of the many aspects of timbre listed above in more detail, using a simplistic steady-state timbre derived from summed sinusoids, as was done to test the effect of timbre on stream segregation [25]. If the frequencies of a sequence of sinusoidal tones, as shown in Figure 14, are within close proximity (such that they would form one stream if none of them were enriched), it is possible to cause a subset of these tones to segregate into its own stream by inducing a timbre difference. In this figure, the vertical line indicates that timbre is created by, or at least co-exists with, the fusion of components into a single percept (Taped Illustration 11). We will discuss the nature of perceptual fusion later.

A musical use of timbre-based streaming might work as follows: Figure 15 shows an isotimbral melodic phrase containing three sub-melodies that are slowly to emerge as contrapuntal melodies on their own. Each sub-melody is marked with a different symbol under the notes that are to belong to it. Now we might cause the "X" melody to slowly change in timbre to a point where its timbre differs sufficiently from the rest of the stream to induce separation. At this point there would be two contrapuntal lines distinguished by their respective timbres as discussed by Wessel [35]. Alternatively, several simultaneous timbral modifications might cause an equal number of streams to emerge as separate contrapuntal melodies, as is illustrated in the second part of Figure 15 by the "X" and "O" streams. Note that as these two imbedded melodies emerge, the original melody is changed since several of its elements have left and joined other groups. (It is important to note that certain types of inharmonic enrichment of a tone might change the pitch as well as the timbre, which would be another case altogether). An understanding of stream formation based on timbre modification is important if one intends to use multitimbral melodies compositionally, as

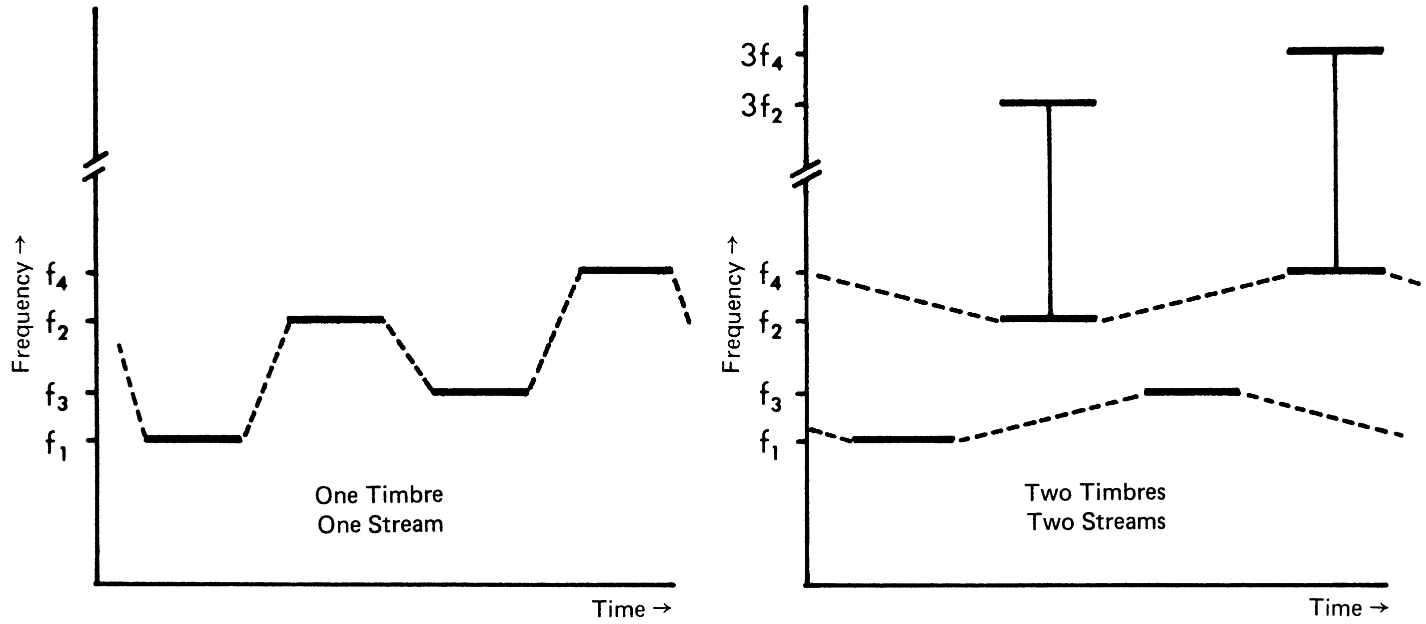


Figure 14. Timbre differences between groups of tones have been shown by McAdams (1977) to influence segregation of those groups. A repeating 4-tone sequence can be decomposed into two streams when two of the tones are enriched with their respective third harmonics. The dotted lines indicate the stream percepts and the vertical solid lines represent the fusion and timbre of the complex tone (*cf.* Taped Illustration 11).

Varèse and Schoenberg, among others, have done. A problem potentially arises when large changes in both timbre and pitch occur simultaneously. It becomes very difficult for the listener to follow the "melody" line if the tempo is above the listener's fission boundary for that set of events. One's attention may tend to drift from one timbre set to another and relegate the rest to unattended streams rather than following the sequence the composer intended.

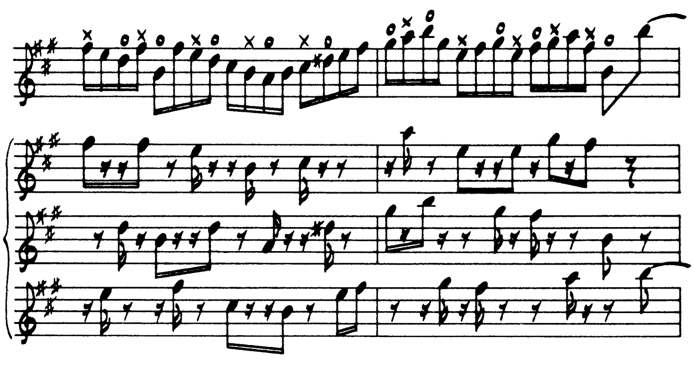
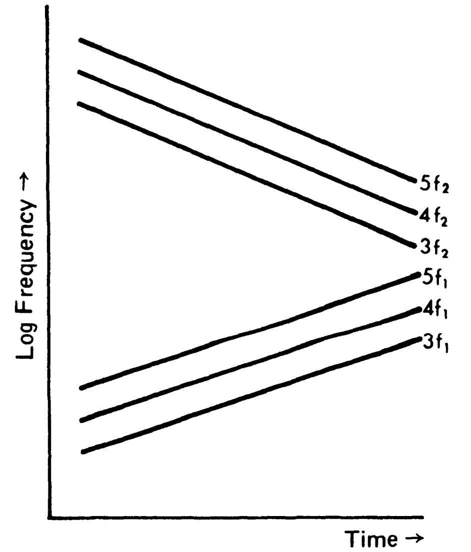
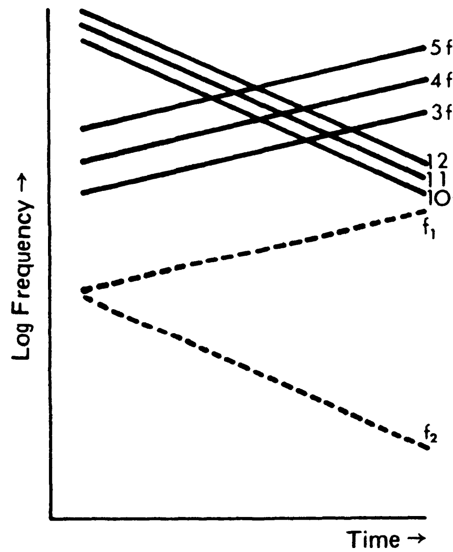


Figure 15. This figure illustrates the possible effect of changing the timbre of different sub-groups of a melody line. The notes marked with X's and O's in the original stream are shown on separate staves below to indicate the rhythmic transformation that takes place after segregation occurs. (Taken from the A Major Prelude, *Well-Tempered Clavier*, Book I, by J.S. Bach).

Timbral Trajectories

This leads into a discussion of the effects of timbral trajectories on stream segregation. As was mentioned before, if two glissandi of identical spectral structure cross in frequency, it is reportedly easier for listeners to hear high and low V-shaped contours than to hear the ascending and descending paths actually taken by each glissando. However, if the glissandi do not cross, or if they differ in spectral structure, it becomes possible to segregate them and hear the rising and falling lines separately. In Taped Illustration 12 and in Figure 16, harmonics 3, 4, and 5 of two different "missing" fundamentals are modulated exponentially in opposite directions. It is easy to hear them separately since neither the spectral components nor the "missing" fundamentals cross. In the next example (see Figure 17) two different spectral structures are used in the same frequency range. The descending glissando is composed of harmonics 10, 11, and 12 and the ascending glissando is composed of harmonics 3, 4, and 5 (Taped Illustration 13). One might argue that it is not a timbre difference that allows one to distinguish these patterns but the fact that the pitches do not cross, as is illustrated in Figure 17 by the dotted lines which show the "missing" fundamentals. Halpern [22] tested for this by using the same harmonic composition (see Figure 18) as in the last example and crossing the pitches of the glissandi, which would correspond to the "missing" fundamentals, rather than crossing the frequency components. The rising and falling glissandi were still perceived, which suggests that the segregating effect was based on spectral content rather than on pitch (Taped Illustration 14). For comparison, listen again to the pure tone crossing in Taped Illustration 8. Some pre-tests done for both Halpern's





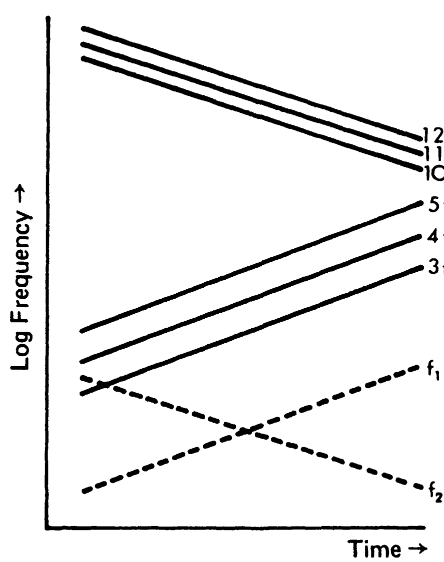
[22] and McAdams' [25] studies bore out the intuition that it is necessary to use constant ratio relationships among components to attain fusion and constant timbre. Further research on the fusion of harmonics for the building of timbre is currently in progress in Bregman's laboratory at McGill University.

Timbre Movement Versus Pitch Movement

One pitch phenomenon has a scientific history which is almost coextensive with that of pitch research in general. This phenomenon has been termed the "missing" fundamental for reasons that will become obvious. It is possible to construct a tone from several sinusoidal components whose frequencies are integer multiples of some fundamental frequency which may or may not be actually present in the final signal's spectrum. For example, frequencies of 1000, 1250, 1500, and 1750 Hz would represent the 4th, 5th, 6th and 7th harmonics, respectively, of a fundamental at 250 Hz. Whether or not a spectral component is present at the fundamental frequency, the perceived pitch almost always corresponds to the fundamental. Of interest in our discussion of timbre and streaming is the fact that tonal events with various spectral structures can elicit the same "missing" fundamental as long as the components are integer multiples of that fundamental and are within a certain so-called "existence region" [29] of harmonics that can actually contribute to the pitch. Thus different tonal events can elicit the same pitch but with different timbres.

Some of the first author's personal experience in working with the pitch of complex tones has led him to the conclusion that it is possible for changes in pitch and timbre to contribute competing information to the overall perceived movement in tone sequences. Figure 19 illustrates that is is possible for timbre to get "sharper" or "brighter" [see for example 1, 18, 20], *i.e.* the spectral components move to a higher frequency range, while the pitch moves in the opposite direction. In Figure 19, while both complexes will elicit pitches corresponding to the "missing" fundamental, *i.e.* F1 and F2, the

Figure 16 (above, left). Harmonics 3, 4, and 5 of two different "missing" fundamentals are modulated exponentially in frequency in opposite directions. Neither the frequencies nor the fundamentals of the two glissandi cross each other. Two separate glissandi are heard (*cf.* Taped Illustration 12).
Figure 17 (above, right). An ascending glissando composed of harmonics 3, 4, and 5 of one "missing" fundamental and a descending glissando composed of harmonics 10, 11, and 12 of another "missing" fundamental are crossed in frequency, but the pitches of the "missing" fundamentals do not cross. Two different glissandi are heard (*cf.* Taped Illustration 13).
Figure 18 (below). Here the same harmonic structure as in Figure 17 is used but now the pitches of the "missing" fundamentals cross each other and the partials do not. Two separate streams are still heard (*cf.* Taped Illustration 14).



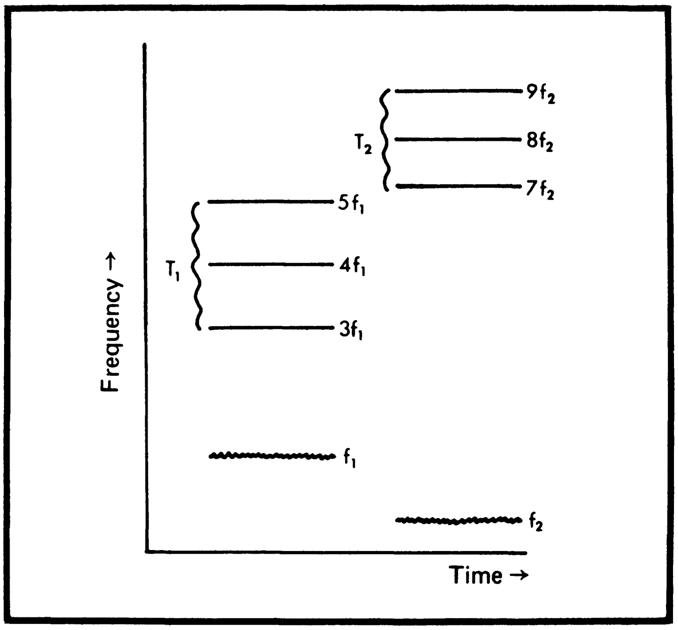


Figure 19. It is possible to construct two tones whose respective spectral "centers of gravity" (see text) and "missing" fundamental pitches move in opposite directions when these tones are presented in succession.

"center of gravity" of the spectral envelope, T1, of the first tone is lower than that of the second tone, T2. Spectral "center of gravity" may be roughly defined as the weighted mean of the spectral components in an acoustic event averaged over the duration of the event, their weights being derived from their relative amplitudes.

Van Noorden investigated the possibility of streaming based on pitch contiguity, as opposed to streaming based on frequency contiguity, *i.e.* continuity due to the continued presence of spectral components. Figure 20 shows the stimuli he used. The first consisted of two different three-component complex tones with harmonics 3, 4, and 5 for the first tone and harmonics 8, 9, and 10 for the second tone. Both tones have the same fundamental frequency. Here an attempt was made to get "missing" fundamentals to stream together. In the second stimulus a pure tone of frequency f alternated with a complex containing harmonics 3 through 10 of the same fundamental frequency; this was an attempt to stream a pure tone with a "missing" fundamental of the same frequency. He found that it was not possible with either stimulus to obtain a coherent stream percept. However, it was necessary that these stimuli have vastly different spectral structures: if one were to partially overlap the spectral structures of the respective tones, it would be impossible to separate the percept of streaming based on frequency contiguity from that based on pitch contiguity. For example, if the first tone consisted of harmonics $3f$, $4f$, and $5f$, and the second tone consisted of harmonics $5f$, $6f$, and $7f$, how is one to know if

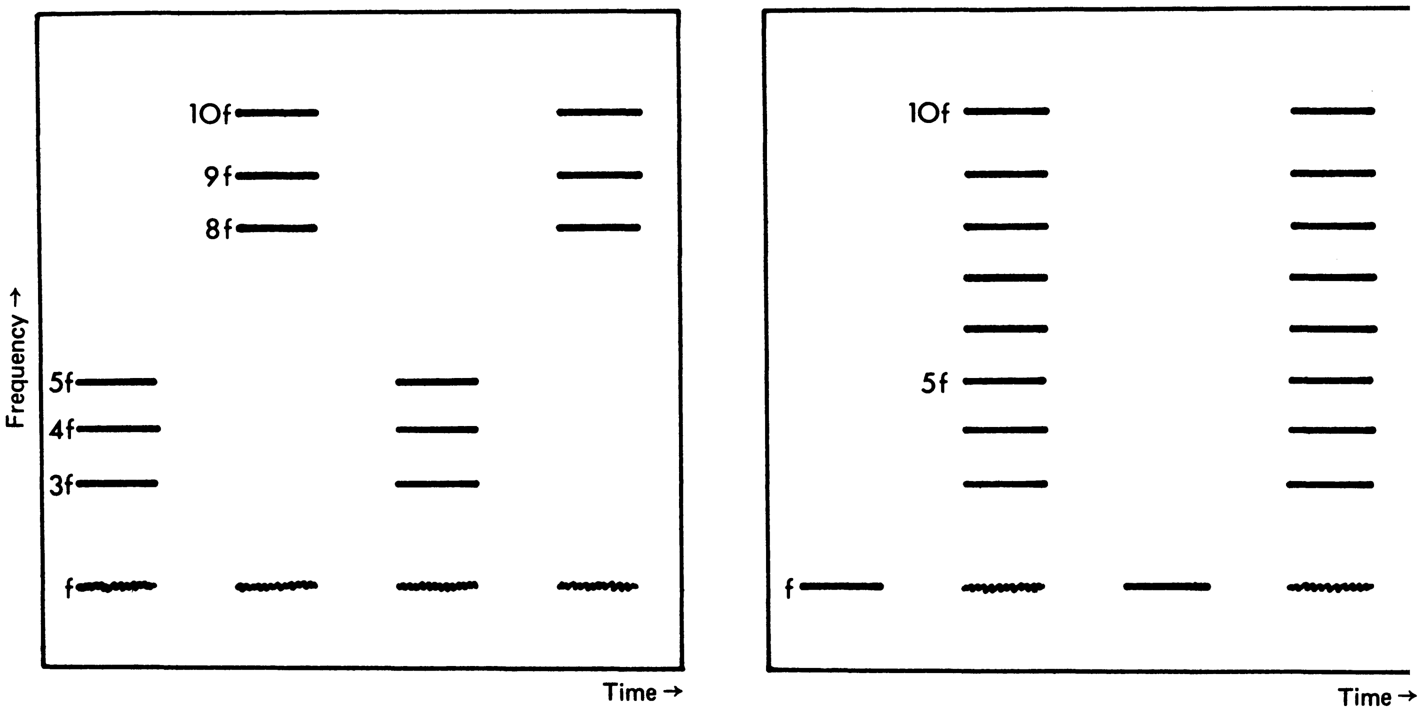


Figure 20. Van Noorden (1975) tried alternating two tone complexes of differing harmonic structure but identical "missing" fundamental and also tried alternating a sine tone with an 8-tone complex whose "missing" fundamental was at the same pitch as the pure tone. He was not able to obtain streaming of the fundamentals in the first example or of the sine tone and the "missing" fundamental in the second example.

the streaming effects reported are due to a repetition of $5f$ or to the perception of a repeating "missing" fundamental, f. A primary result of this arrangement of harmonics was that van Noorden used substantially different timbres in the two tones, and this condition may have contributed substantially to the prevention of stream formation. Equally likely is the possibility that stream formation mechanisms do not use pitch information but use frequency information. The following findings may lend support to the former notion.

Our work on timbre effects in stream formation [25] found an anomalous result that may have been due to pitch effects interacting with timbre effects. In Figure 21, two of the test cases used in that study are shown. In the first, the higher tone pair of the repeating four-tone sequence is enriched with the third harmonic and in the second case the lower pair is similarly enriched. The first case was found to segregate significantly more than the second. It is McAdams' contention that this is the result of an interaction between perceived pitch "height" and timbral "brightness." This "brightness" seems to be related primarily to spectral "center of gravity." In our example with the enriched high pair, the pitch "height" is already greater than that of the low pair and the addition of a higher component increases the "brightness" of its timbre. The perceptual integration of these two factors may facilitate the segregation of the tone pairs by resulting in an overall increase in perceived "height" of the tones in the high pair. However, in the second condition where the low pair was enriched, the increase in "brightness," interacting with perceived pitch "height," apparently served to increase the overall perceived "height" of the lower pair. This would bring it into closer perceptual proximity to the higher pair and thus attenuate the effect of the difference in timbre between the two pairs. But this condition still evoked more segregation than two isotimbral control cases, where all of the tones were composed of either one or two components. Of interest here, again, is the fact that pitch and timbre movement may work against each other.

While these timbre changes have proven to be a consistent irritant to the experimenter pursuing the elusive "missing" fundamental, the implied interrelationships between timbral "brightness" and pitch "height" may prove interesting to investigate in terms of stream formation and to apply (both experimentally and compositionally) to melodic sequence construction. Grey [18] has delineated a three-dimensional timbre space roughly defined by 1) spectral distribution, 2) spectral fluctuation and synchronicity of higher harmonic transients, and 3) low-amplitude, high-frequency attack transients. How does the "distance" between events in this timbre space affect the temporal coherence and fission boundaries? Do the effects obtained vary differently along these different dimensions associated with timbre? Research by Wessel [35] and Grey and Gordon [19, 20] has provided information which would allow the composer to deal with timbre as a compositional parameter. This information also allows a certain degree of prediction of the perceptual relationships between sounds resulting from such timbral variations.

Fusion, Timbre, and Frequency Streaming

The idea of competing perceptual organizations is an exciting one for music composition and theory. Bregman and Pinker [10] investigated the competition between sequential

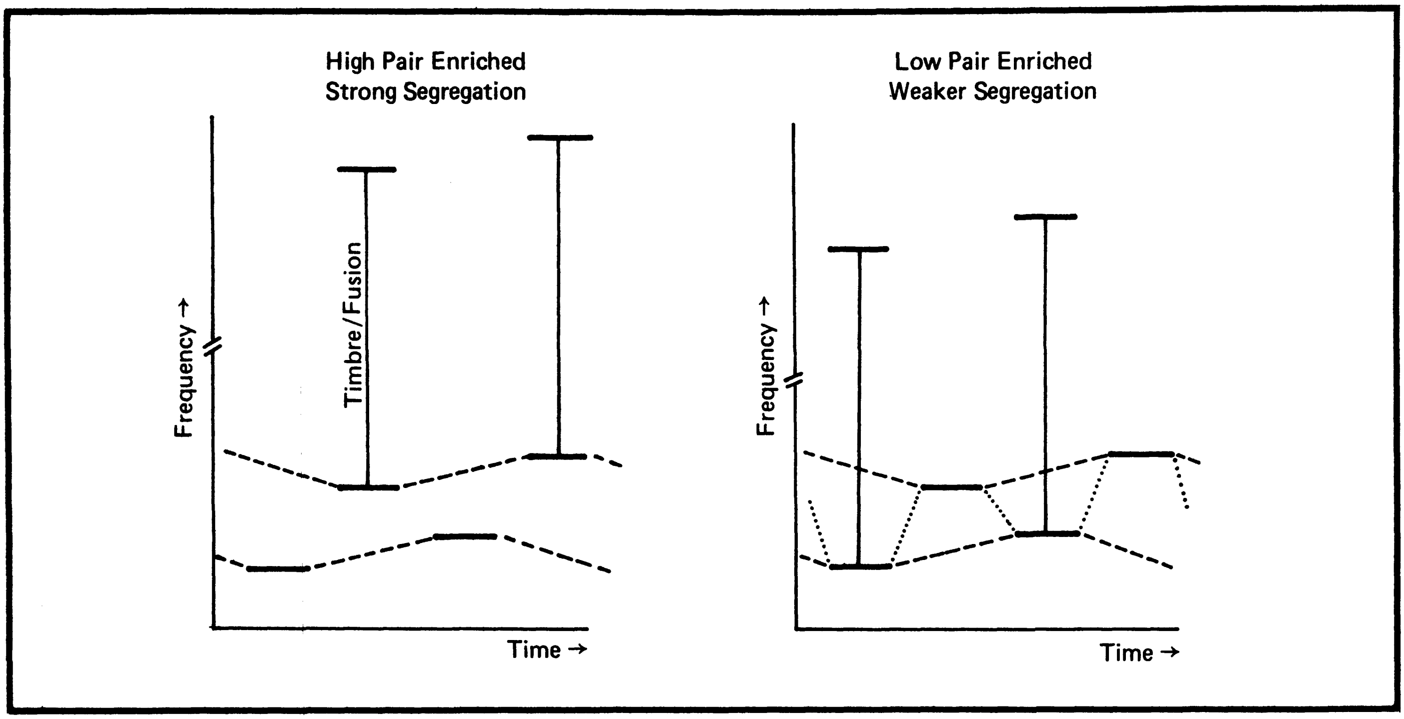


Figure 21. One result of McAdams' study (1977) suggested that there is an interaction between pitch "height" and timbral "sharpness" (see text). In a repeating 4-tone sequence, one of the pairs of tones was selectively enriched by adding the third harmonic. A greater degree of segregation of the high and low streams was found for the former case. The dashed lines indicate the two-stream percepts and the dotted lines indicate a potential one-stream percept. The vertical solid lines represent the fusion and timbre of the 2-tone complex.

(or frequency) organizations and simultaneous (or timbral) organizations in the formation of auditory streams. The stimulus used by Bregman and Pinker was a sine tone alternating with a two-tone complex. In Figure 22, tones A and B would represent the sequential organization and tones B and C would represent the simultaneous organization. The harmonicity and synchronicity of tones B and C in the complex were varied as was the frequency separation between the sine tone A and the upper component B of the complex. The rationale for varying these two parameters was as follows. Tones with frequency relationships derived from simple ratios, *i.e.* those that exhibit "consonance," should tend to fuse more readily than combinations considered to be dissonant. While the evidence for this was very weak in the Bregman and Pinker study, work currently in progress in Bregman's laboratory strongly suggests that this is indeed the case. The new evidence further suggests that the effect of harmonicity itself is relatively weak and may be overridden by stronger factors such as frequency contiguity and synchronicity of attack and decay. Tones with synchronous and identically shaped attack and decay ramps are more likely to fuse than those with asynchronous or dissimilar attacks and decays [12, 15]. This may be a major cue in being able to parse out the different instruments playing together in an orchestra since they all have substantially different attack characteristics. In addition, there is a very low probability of several people precisely synchronizing their attacks.

In light of this work, one might make the following predictions for the perception of the stimuli used by Bregman and Pinker: 1) Sequential streaming is favored by the frequency proximity of tones A and B (as we have illustrated in the earlier examples). 2) The simultaneous (or timbral) fusion of tones B and C is favored by the synchrony of their attacks. 3) These two effects "compete" for tone B's membership in their respective perceptual organizations. 4) Finally, when tone B is "captured" by tone A, it is removed from the timbral structure and tone C sounds less rich. Thus, it is reasoned that if the two simultaneous sine tones B and C are perceived as belonging to separate streams, they should be heard as sine tones. But if they are heard as one stream, they should sound like one rich tone.

It would be appropriate to introduce the notion of "belongingness" at this point, since we talk of tones belonging to streams, and of frequency components and the timbre resulting from their interaction belonging to a perceived tonal event. "Belongingness" (a term used in the perceptual literature of Gestalt psychology) may be considered as a principle of sensory organization which serves to reconstruct physical "units" into perceptual events by grouping sensory attributes

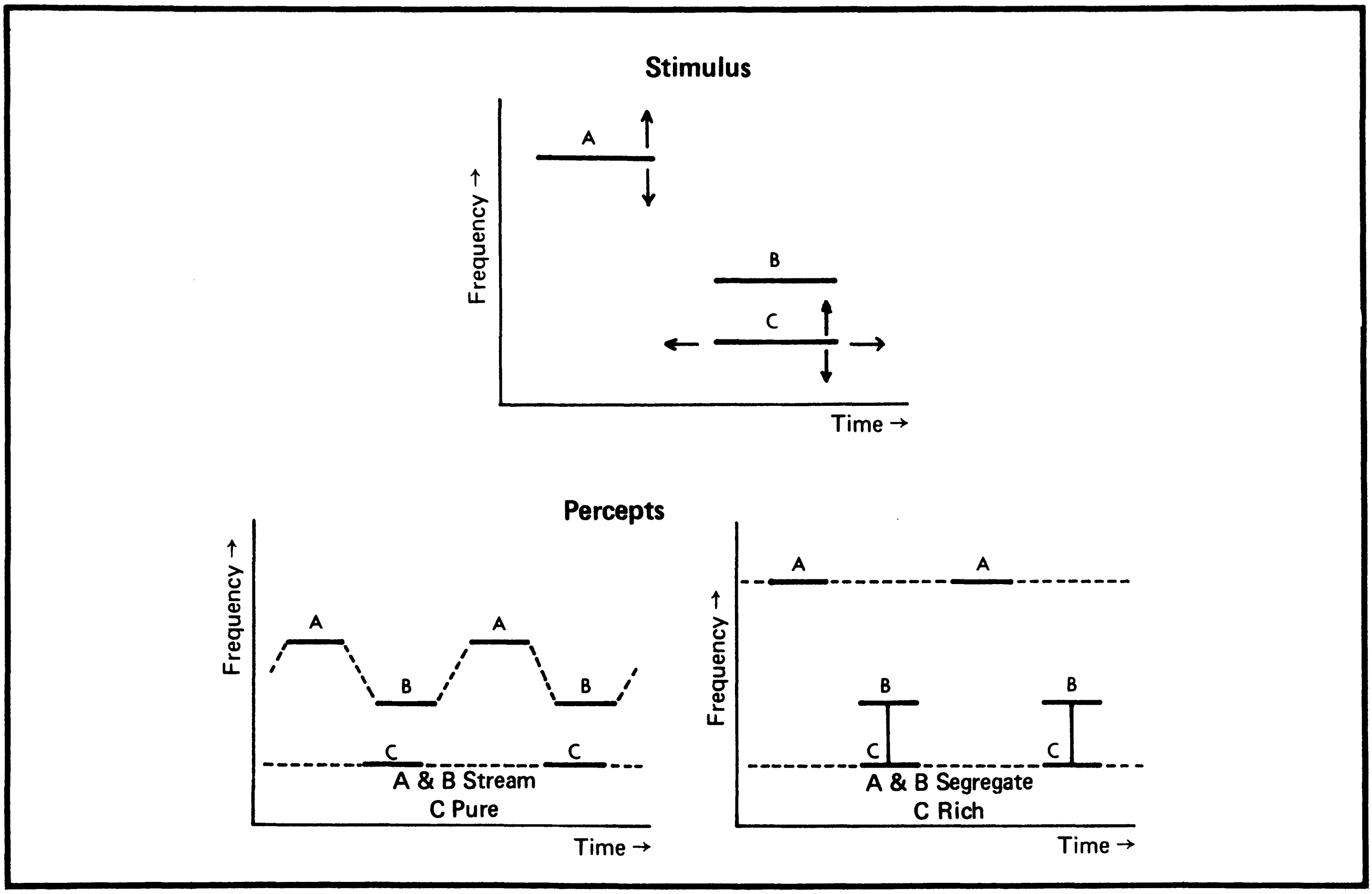


Figure 22. The competition between sequential and simultaneous organizations in the formation of auditory streams is shown here. Tone B can belong either to the sequential organization with tone A or to the simultaneous organization with tone C but not to both at the same time. Bregman and Pinker (1978) varied the frequency separations between tones A and B and between tones B and C and also varied the relative synchrony of onset of tones B and C. The dotted lines in the figure indicate the stream percepts and the vertical solid lines represent the fusion of tones B and C (*cf.* Taped Illustration 15).

of those events into unified percepts. As Bregman [2] points out, "belongingness is a necessary outcome of any process which decomposes mixtures, since any sensory effect must be assigned to some particular source." In this case, when the simultaneous tones B and C are segregated, the timbre resulting from their interaction still exists and can be heard if one listens for it, but it is not perceptually assigned to either of the tones B or C, and thus does not affect the perception of them. The nature of a stream is such that its qualities are due to the perceptual features assigned to it.

In Taped Illustration 15 one can hear a case in which A is close to B in frequency and C is asynchronous with B. Then a case is heard where A is further away from B in frequency and C is synchronous with B. Tones B and C have the same frequencies in both cases. Listen for both the A-B stream and the richness of tone C. The listeners in Bregman and Pinker's study reported perceiving C as being richer when B and C were synchronous, and this judged richness dropped off with an increase in asynchrony, *i.e.* as C either preceded or followed B by 29 or 58 msec. As the frequency separation between A and B was increased, C was reportedly perceived as being increasingly rich.

The Role of Context in Determining Timbre

These findings indicate that the perceived complexity of a moment of sound is context-dependent (see [19] as another example of the trend toward viewing timbre as depending on context). Context may be supplied by a number of alternative organizations that compete for membership of elements not yet assigned. *Timbre is a perceived property of a stream organization rather than the direct result of a particular waveform, and is thus context-dependent.* In other words, two frequency components whose synchronous and harmonic relationships would cause them to fuse under isolated conditions may be perceived as separate sine tones if another organization presents stronger evidence that they belong to separate sequential streams.

A very compelling demonstration of the decomposition of a timbre organization by alternate frequency streaming organizations is illustrated in Figure 23. When tone A is presented by itself it elicits a timbre, denoted T_A. If this tone is preceded by tone B eliciting timbre T_B, and succeeded by tone C eliciting timbre T_C, one notices that timbre T_A completely disappears and is replaced by timbres T_B and T_C. Here the highest and lowest components of tone A are streamed with those of tone B and subsequently assume a timbre identical to that of tone B. Also, the inner components of tone A stream with those of tone C and assume a like timbre (Taped Illustration 16).

An important question concerning the assignment of timbre and pitch to a tonal event arises. Both timbre and pitch have been found to be context-dependent [10, 21, respectively]. But each may be determined by different ongoing contextual-organizations; as such they may be considered to be associated perceptual dimensions of a sound but may not be inextricably bound to one another. How relevant to music theory, then, are studies that deal with the perceived pitch and timbre of tones in isolation? This question is not meant to insinuate that sensory and psychophysical experimentation are useless. Far from it. If we think in terms of investigating the experience of music at different levels of processing (Figure 24), we see how important all of these areas of research are in building the whole picture. The raw physical input is modified by the limits of the sense organ whose output is still further modified by cognitive processes. But by studying steady-state, or at least relatively simple signals, we can find the limits and interactions of the sense organs and peripheral processes, such as temporal and spectral resolution, lateral inhibition, and masking, which limits affect the final percept. Beyond these, the central perceptual processes such as pitch extraction, timbre buildup, and coherence and fission modify the initial neural result of stimulation of the sensory system. Further interactions in higher brain processes such as attentional processes, memory and comparison of pitch, timbre and loudness, context extrac-

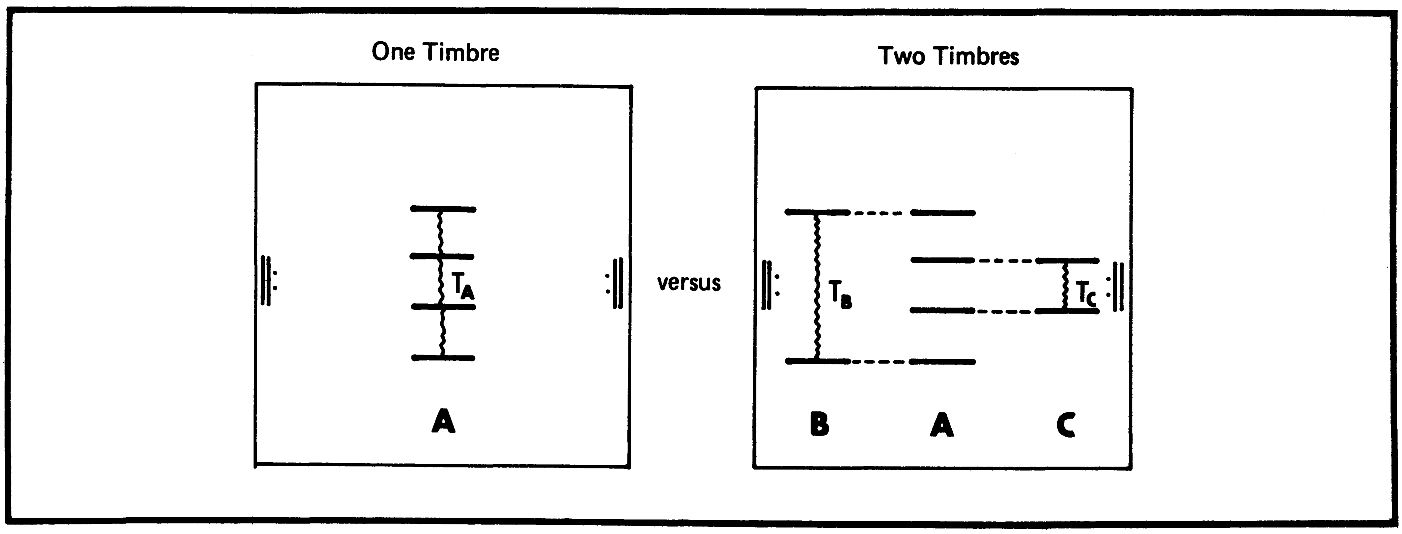


Figure 23. The first part of this figure shows a repeating tone (A) consisting of 4 harmonics. This tone would elicit a certain timbre percept, T_A. In the second part this tone is preceded by tone B, consisting of the top and bottom harmonics, and is succeeded by tone C consisting of the two inner harmonics. Tone B elicits timbre T_B and tone C elicits timbre T_C. However, due to the streaming of tone A's components with those of tones B and C, T_A totally disappears and is replaced by T_B and T_C (*cf.* Taped Illustration 16).

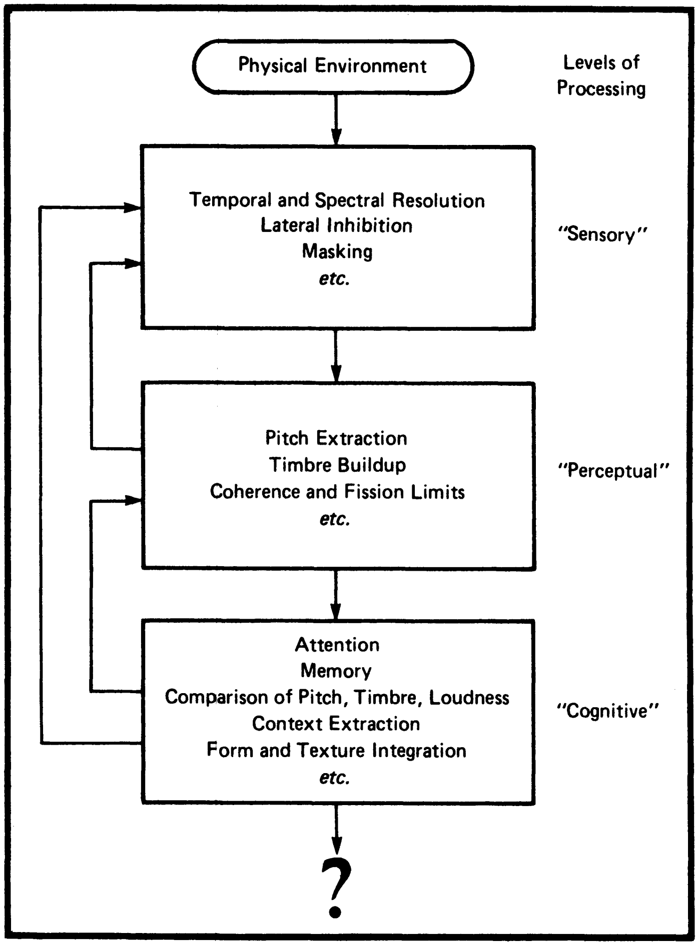


Figure 24. This block diagram suggests a possible arrangement of the processing of acoustic information at different interconnected levels of the auditory system.

tion, and form and texture integration, sculpt the transduced information into meaningful percepts. It is felt that all of these levels of processing feed into each other in a sort of heterarchical (as opposed to a hierarchical) system. The point being made is that in the framework of music where all of these complex interactions are of great importance, the context that is created may be the essential determinant of the musical result of a given sound. One sound is potentially perceivable in a great number of ways, depending on its context.

Melody

This leads us to believe that the fundamental perceptual element in music may be the "melody" rather than the isolated tone. Or in the terminology of auditory perception, the fundamental structure is the auditory stream. This is not, of course, a new notion; but an empirical approach may allow us to clarify and delimit the concept to the extent that we may predict the perceptual results. That, in the mind of the first author, is the primary concern of the composer. Let us, then, examine melody and its relation to attention.

For our purposes, we can think of melody as a connected and ordered succession of tones [28]. It follows then that temporal coherence is necessary for a sequence of tones to be perceived as a whole. On the other hand, a sequence of tones may segregate into two or more separate streams which are individually coherent. In this case, we would perceive several simultaneous melodies rather than one.

If such an operational definition of melody is tenable, we must then question how it is that some of the extensions of elements other than pitch for "melodic" material are valid perceptually. For example, when a composer uses timbre as thematic material, do we still perceive the sequence as maintaining a temporal continuity? Sometimes coherence is maintained by sensitive performers and sometimes it can be very difficult to perceive. Of course, we have not included other elements which affect the perception of melody, such as underlying harmonic structure, but we are only attempting to convey the notion that temporal coherence should be considered essential to melody formation. Conversely, one may use the principles of fission to develop rules for creating polyphony and counterpoint in sequences of acoustic events.

Attention and Musical Structure

It has become apparent to the first author that musical structure, as it is perceived in real-time, is inextricably bound to attentional processes. A bit of introspection will reveal that there are at least two kinds of attentional processes, which we might call active or willful attention, and passive or automatic attention. One might willfully direct one's attention to some object or sequence of events, such as listening to particular events within a piece of music. Or, some unusual event might attract one's attention unexpectedly, such as the honking horn of an oncoming car you had not noticed as you stepped into the street absorbed deep in thought. That horn demands your attention and in all probability gets it in a hurry. In particular, both kinds of attention may participate in the process of listening to auditory streams. For instance, in Figure 5 when a sequence of tones lies above the temporal coherence boundary, no amount of active attention can extract the percept of one coherent stream. Here perception is limited by passive attentional processes [28]. This has important consequences for composers who intend to use fast melodic sequences, since it suggests that there are tempi at which the listener may not be able to follow as a melody the sequence you have constructed, regardless of the attentional will power invoked. An example of these effects may be found in the sequences presented at different rates in Charles Dodge's *Earth's Magnetic Fields.* Without pretending to know the composer's intent, it can be amusing to listen to the same sequence decompose and re-integrate itself during various tempo changes. In a multi-streamed sequence one can relax attentional effort with the result that attention might randomly alternate among the available streams. Or one might selectively focus attention on any one of them individually and even play them against one another.

Van Noorden reported that the temporal coherence boundary (the boundary below which all tones may belong to one stream) is not affected by previous knowledge of the sequence, and considered it to be a function of a passive attentional mechanism, given that the listener "wants to hear coherence." However, what happens between the boundaries of temporal coherence and fission depends upon a number of factors, such as context, and seems to be under the influence of attention. Dowling [17], for example, reported that if listeners knew beforehand which melodies were being interleaved, they could, with a bit of practice, extract the appropriate melody even at very small separations of the ranges traversed by each melody. It is currently assumed that this ability would degenerate at faster tempi. In addition, van Noorden found an effect at very fast tempi where it was virtually impossible to hear sequences within a range of two or three semitones as other than temporally coherent. At very fast tempi (about 12.5 tones/sec.) the tones of such narrow patterns are not heard as separate members of a sequence, but actually merge into a continuously rippling texture, as one can hear in Taped Illustration 17. There are thus attentional limits in the ability of the auditory system to track a sequence of events. When events occur too quickly in succession, the system uses the various organizational rules discussed in this article to reorganize the events into smaller groups. It may then track events within a particular group if the listener is paying attention to it, but this narrowing of focus necessarily causes a loss of information. One result is the inability to make fine temporal order judgments between streams. These organizational mechanisms reflect the tendency of the auditory system to simplify things in the face of excessive complexity. In the example where the fast sequence of tones merges into a continuous "ripple," the auditory system is unable to successfully integrate all of the incoming information into a temporal structure and simplifies the situation by interpreting it as texture (see also [31]). Thus the auditory system, beyond certain tempi, may interpret the sequence as a single event and assign to it the texture or timbre created by its spectral and temporal characteristics.

An understanding of (or intuition about) these organizational processes can lead to new dimensions of control over musical structure. For example, one might construct contrapuntal sequences that play across various stream boundaries and through different borderline regions between temporal coherence and fission. (It may be that composers such as Bach were already using perceptual ambiguity consciously in their work.) Any or all of the relevant musical parameters might be used to accomplish this. Then, an appropriate use of events that vie for or demand the listener's attention can be used by the composer to "sculpt" the attentional processes of the listener. Since some events seem more striking to some persons than to others, this attentional sculpture in time would lead different listeners through different paths of auditory experience. Further, perception is bound to vary from time to time within a single person, so the experience would be different with each listening. A composition of sufficient, controlled complexity might thus be perceptually infinite for a given listener.

Conclusion

An attempt has been made here to point out that composers and music theorists should thoroughly examine the relationship between the "musical" principles they use and espouse, and the principles of sensory, perceptual, and cognitive organization that operate in the human auditory system. Many of the principles discussed in this article extend to higher-level perceptual analysis of musical context and structure and may well represent a scientific counterpart to some extant music theoretical principles. In other cases, though, this group of phenomena suggests perceptual organizations which have little relation to methods currently used to construct or analyze musical structure. To ignore the evidence from the real life system in developing a theory of music or a musical composition is to take the chance of relegating one's work to the realm of what might be termed "paper music."

Acknowledgments

This paper is based on a workshop entitled "The Perceptual Factoring of Acoustic Sequences into Musical Streams" delivered at the 1978 International Computer Music Conference, Evanston, Illinois, and published in the *Proceedings* of the Conference. The authors would like to thank Dr. Leon van Noorden for his many helpful suggestions and valuable criticisms of the manuscript; Figures 5, 8, 9, 10 and 20 were redrawn, with permission, from Dr. van Noorden's thesis. The present version of this article was prepared while Mr. McAdams was a Graduate Fellow of the National Science Foundation. Dr. Bregman's research has been supported by grants from the National Research Council of Canada, the Quebec Ministry of Education, and the McGill University Faculty of Graduate Studies and Research. Much of the research used the facilities of the Computer-Based Laboratory of the McGill University Department of Psychology. The editors of *Computer Music Journal* would like to thank Mr. McAdams for preparing the illustrations in this article.

Appendix 1

Description of Taped Illustrations

1. A repeating sequence of three high tones (1600, 2000, 2500 Hz.) is interspersed with three low tones (350, 430, 550 Hz.) used by Bregman and Campbell (1971). a) At a tempo of five tones/sec. the sequence is perceived as one stream of alternating high and low tones (*cf.* Figure 1a). b) At a tempo of ten tones/sec. the sequence segregates perceptually into one stream of high tones and one stream of low tones (*cf.* Figure 1b).
2. Another repeating six-tone sequence is played at a tempo of ten tones/sec., but the higher triplet is closer in frequency to the lower. Tone F may be perceived as belonging to either the high stream or the low stream depending on the listener's focus. Note that tone F cannot belong to both streams at once (*cf.* Figure 2).
3. A repeating six-tone sequence starts at a slow tempo. As the tempo is gradually increased, the sequence is progressively decomposed into smaller perceptual streams until it is no longer possible to follow the tonal events which merge into the percept of timbre or texture (*cf.* Figure 3).
4. Using the same initial sequence as in Taped Illustration 3, the frequency separation between temporally adjacent tones is gradually increased. The same sort of decomposition into smaller streams occurs. The limits in this example are determined not by temporal resolution but by the audible frequency range (*cf.* Figure 3).
5. The tones of two familiar melodies are interleaved in the same frequency range. As the frequency range of one melody is shifted away from that of the other, the melodies become recognizable.
6. a) In the first part of this example (see Figure 4), tones A and B, alternating in one frequency range, form a separate stream from tones X and Y, alternating in another frequency range. b) In the second part, tones X and Y are moved into the proximity of the frequencies of tones A and B with a sub-

sequent perceptual change. Now tones A and X stream, as do tones B and Y. Note also that the rhythms of the respective streams change between the two examples.

7. This example demonstrates the streaming potential of a sine wave used for frequency modulation of a sine tone. The progression heard is from a very slow, trackable modulation, through a point where the high and low peaks of the modulated sound segregated and then to the point where a timbre is perceived. The example then returns to the initial modulation frequency.

8. Two sine waves are modulated exponentially in frequency. One ascends in frequency, the other descends in frequency and they cross at one point. Most listeners report hearing high and low V-shaped contours as opposed to hearing the full glissandi.

9. First, gated speech is heard. Note the difficulty in discerning the message. However, when the gaps are filled with white noise, the spoken message sounds continuous and a pulsing noise is heard in the background.

10. The same demonstration as in Taped Illustration 9 is used on musical material. It is difficult to induce the information lost in the gaps until those gaps are filled with white noise.

11. The first part of this example is a repeating four-tone pattern, all of the tones of which have the same timbre. This represents a pattern usually perceived as being temporally coherent. However, when the third harmonic is added to the second and fourth tones of the pattern, thus changing their timbre, two streams of differing timbre may be heard. The frequencies of the fundamentals are not changed (*cf.* Figure 14).

12. Two glissandi with identical spectral structures moving in opposite directions are heard separately since none of their spectral components cross in frequency (*cf.* Figure 16).

13. Two glissandi with differing spectral structures moving in opposite directions are made to cross each other in frequency. Their differing timbres still allow the listener to segregate them perceptually (*cf.* Figure 17).

14. The same timbres used in Taped Illustration 13 are arranged such that their pitches cross but none of the frequency components cross. They are still heard as separate glissandi (*cf.* Figure 18).

15. In the first part of this example (see Figure 22) tones A and B are close to each other in frequency while tones B and C are asynchronous (tone C leads tone B by 29 msec.). Tones A and B form a sequential stream and tone C is perceived as being fairly pure. In the second part, tone A's frequency is separated from that of tone B and tones B and C are synchronous in attack and decay. Tone A forms a stream by itself and tones B and C fuse into a single, richer percept.

16. First, a repeating four-component tone is heard. Note the timbre of this tone (see Figure 23). Then this tone A (at the same repetition rate) is preceded by a tone B consisting of the top and bottom components of tone A and is followed by a tone C consisting of the two inner components of tone A. Two streams are formed and the components of A are split into separate streams with tones B and C. Accompanying this stream organization is a decomposition of tone A's timbre into the timbres of tones B and C.

17. In this example a sequence of tones very close to each other in frequency is slowly accelerated in tempo until the percept of one continuous, rippling tone is achieved. Then the tempo is decelerated until the individual tones can be heard once again.

References

1. Bismarck, G. von. (1974). "Sharpness as an attribute of the timbre of steady sounds." *Acustica,* Vol. 30, pp. 159-172.
2. Bregman, A. S. (1978a). "Asking the 'what for' question in auditory perception," in Kubovy, M. and J. Pomerantz, eds., *Perceptual Organization.* Hillsdale, N. J.: Lawrence Earlbaum Associates.
3. Bregman, A. S. (1978b). "Auditory streaming: Competition among alternative organizations." *Perc. and Psychophys.*, Vol. 23, pp. 391-398.
4. Bregman, A. S. (1978c). "Auditory streaming is cumulative." *J. Exp. Psych./Human Perc. Perf.,* Vol. 4, pp. 380-387.
5. Bregman, A. S. (1978d). "The formation of auditory streams." in J. Requin, ed., *Attention and Performance VII.* Hillsdale, N. J.: Lawrence Earlbaum Associates.
6. Bregman, A. S. and J. Campbell (1971). "Primary auditory stream segregation and the perception of order in rapid sequences of tones." *J. Exp. Psych*, Vol. 89, pp. 244-249.
7. Bregman, A. S. and G. L. Dannenbring (1973). "The effect of continuity on auditory stream segregation." *Perc. and Psychophys.*, Vol. 13, pp. 308-312.
8. Bregman, A. S. and G. L. Dannenbring (1977). "Auditory continuity and amplitude edges." *Can. J. Psych./Rev. Can. Psych.*, Vol. 31, pp. 151-159.
9. Bregman, A. S. and G. L. Dannenbring (1978). "Auditory streaming." *Sci. Am.*, in press.
10. Bregman, A. S. and S. Pinker (1978). "Auditory streaming and the building of timbre." *Can. J. Psych.*, Vol. 32, pp. 19-31.
11. Bregman, A. S. and S. I. Rudnicky (1975). "Auditory segregation: stream or streams?" *J. Exp. Psych./Human Perc. Perf.*, Vol. 1, pp. 263-267.
12. Bregman, A. S. and H. Steiger (1978). "Auditory streaming and vertical localization: interdependence of 'what' and 'where' decisions in audition." Unpub. manuscript, McGill Univ. Dept. of Psych., Montreal, Quebec, Canada.
13. Dannenbring, G. L. and A. S. Bregman (1976). "Effect of silence between tones on auditory stream segregation." *J. Acous. Soc. Am.*, Vol. 59, pp. 987-989.
14. Dannenbring, G.L. and A.S. Bregman (1976)). "Stream segregation and the illusion of overlap." *J. Exp. Psych./Human Perc. Perf.*, Vol. 2, pp. 544-555.
15. Dannenbring, G. L. and A. S. Bregman (1978). "Streaming vs. fusion of sinusoidal components of complex tones." *Perc. and Psychophys.*, Vol. 24, pp. 369-376.
16. Deutsch, D. (1975) "Two-channel listening to musical scales." *J. Acous. Soc. Am.*, Vol. 57, pp. 1156-1160.
17. Dowling, W. J. (1973). "The perception of interleaved melodies." *Cog. Psych.*, Vol. 5, pp. 322-337.
18. Grey, J. M. (1977) "Multidimensional perceptual scaling of musical timbres." *J. Acous. Soc. Am.*, Vol. 61, pp. 1270-1277.
19. Grey, J. M. (1978), "Timbre discrimination in musical pattern." *J. Acous. Soc. Am*, Vol. 64, pp. 467-472.
20. Grey, J. M. and J. W. Gordon (1978). "Perceptual effects of spectral modifications on musical timbres." *J. Acous. Soc. Am.*, Vol. 63, pp. 1493-1500.
21. Guilford, J. M. and H. M. Nelson (1936). "Changes in the pitch of tones when melodies are repeated." *J. Exp. Psych.*, Vol. 19, pp. 193-202.

(continued on p. 60)

(Buxton article continued from p. 25)

Mathews, M. (1969) *The Technology of Computer Music*, Cambridge: MIT Press.

Murray, D., Beauchamp, J., and Loitz, G. (1978) "Using the PLATO/TI980A Music Synthesis System: The PLACOMP Language," *Proceedings of the 1978 International Computer Music Conference*, Roads (Comp.), Evanston: Northwestern University Press, Vol. 1: 151-166.

New England Digital Corporation (1978). *Synclavier Instruction Manual*, Norwich, Vermont: New England Digital Corp., PO Box 305.

Newman, W. and Sproull, R. (1979) *Principles of Interactive Computer Graphics (Second Edition)*, New York: McGraw-Hill.

Patel, S. (1979) "Score Editor Design: The Foundations of Scriva," Toronto: unpublished manuscript, SSSP/CSRG, University of Toronto.

Pulfer, J. (1972) *MUSICOMP Manual*, Ottawa: N.R.C. Radio and Electrical Engineering Division.

Reeves, W., Buxton, W., Pike, R., and Baecker, R. (1978) "Ludwig: an Example of Interactive Computer Graphics in a Score Editor," *Proceedings of the 1978 International Computer Music Conference*, Roads (Comp.), Evanston: Northwestern University Press, Vol. 2: 392-409.

Ritchie, D. M. and Thompson K. (1974) "The UNIX Time-Sharing System," *Communications of the ACM*, 17.7: 365-375.

Smith, L. (1978) *MUS10 User Manual*, Stanford: work in progress, CCRMA, Dept. of Music, Stanford University.

Smith, L. (1972) "SCORE—A Musician's Approach to Computer Music," *Journal of the Audio Engineering Society*, 20.1: 7-14.

Tanner, P. (1972) "MUSICOMP, An Experimental Aid for the Composition and Production of Music," *ERB-869*, Ottawa: N.R.C. Radio and Electrical Engineering Division.

Tucker, W. H., Bates, R. H. T., Frykberg, S. D., Howrath, R. J., Kennedy, W. K., Lamb, M. R. and Vaughan, R. G. (1977) "An Interactive Aid for Musicians," *International Journal of Man-Machine Studies*, 9: 635-651.

Vercoe, B. (1975) "Man-Computer Interaction in Creative Applications," Cambridge: unpublished manuscript, Studio for Experimental Music, M.I.T.

(McAdams and Bregman article continued from p. 43)

22. Halpern, L. D. (1977). "The effect of harmonic ratio relationships on auditory stream segregation." Unpub. undergrad. research report, McGill University Dept. of Psych., Montreal, Quebec, Canada.
23. Heise, G. A. and G. A. Miller (1951). "An experimental study of auditory patterns." *Am. J. Psych.*, Vol. 64, pp. 68-77.
24. Houtgast, T. (1972). "Psychophysical evidence for lateral inhibition in hearing." *J. Acous. Soc. Am.*, Vol. 51, pp. 1885-1894.
25. McAdams S. E. (1977). "The effect of tone quality on auditory stream segregation." Unpub. honours thesis, McGill University Dept. of Psych., Montreal, Quebec, Canada.
26. Miller, G.A. and G.A. Heise (1950). "The thrill threshold." *J. Acous. Soc. Am.*, Vol. 22, pp. 637-638.
27. Miller, G. A. and J. C. R. Licklider (1950). "Intelligibility of interrupted speech." *J. Acous. Soc. Am.*, Vol. 22, pp. 167-173.
28. van Noorden, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences.* Ph.D. Diss., Tech. Hogeschool, Eindhoven, The Netherlands, Pub. by The Institute of Perception Research, Eindhoven, The Netherlands.
29. Ritsma, R. J. (1962). "Existence region of the tonal residue. Part I." *J. Acous. Soc. Am.*, Vol. 34, pp. 1224-1229.
30. Schouten, J. F. (1962). "On the perception of sound and speech: Subjective time analysis." Fourth Int'l. Cong. Acous., Copenhagen. *Reports* II, pp. 201-203.
31. Steck, L. and P. Machotka (1975). "Preference for musical complexity: Effects of context." *J. Exp. Psych./Human Perc. Perf.*, Vol. 104, pp. 170-174.
32. Thurlow, W. R. (1957). "An auditory figure-ground effect." *Am. J. Psych.*, Vol. 70, pp. 653-654.
33. Verschuure, J. (1978). *Auditory excitation patterns:*

(continued on p. 63)

U.S. POSTAL SERVICE

STATEMENT OF OWNERSHIP, MANAGEMENT AND CIRCULATION

(Required by 39 U.S.C. 3685)

1. TITLE OF PUBLICATION: COMPUTER MUSIC JOURNAL — A. PUBLICATION NO.: 3 5 6 6 9 0 — 2. DATE OF FILING: 28 Sep 1979

3. FREQUENCY OF ISSUE: quarterly — A. NO. OF ISSUES PUBLISHED ANNUALLY: four — B. ANNUAL SUBSCRIPTION PRICE: $20.00

4. LOCATION OF KNOWN OFFICE OF PUBLICATION *(Street, City, County, State and ZIP Code) (Not printers)*: 1263 El Camino Real, Menlo Park, CA 94025 (San Mateo County)

5. LOCATION OF THE HEADQUARTERS OR GENERAL BUSINESS OFFICES OF THE PUBLISHERS *(Not printers)*: same

6. NAMES AND COMPLETE ADDRESSES OF PUBLISHER, EDITOR, AND MANAGING EDITOR

PUBLISHER *(Name and Address)*: Willard J. Holden, 1263 El Camino Real, Menlo Park, CA 94025

EDITOR *(Name and Address)*: Curtis Roads (same)

MANAGING EDITOR *(Name and Address)*: Curtis Roads (same)

7. OWNER *(If owned by a corporation, its name and address must be stated and also immediately thereunder the names and addresses of stockholders owning or holding 1 percent or more of total amount of stock. If not owned by a corporation, the names and addresses of the individual owners must be given. If owned by a partnership or other unincorporated firm, its name and address, as well as that of each individual must be given.)*

NAME	ADDRESS
People's Computer Co.	1263 El Camino Real, Menlo Park, CA 94025

8. KNOWN BONDHOLDERS, MORTGAGEES, AND OTHER SECURITY HOLDERS OWNING OR HOLDING 1 PERCENT OR MORE OF TOTAL AMOUNT OF BONDS, MORTGAGES OR OTHER SECURITIES *(If there are none, so state)*

NAME	ADDRESS
none	

9. FOR COMPLETION BY NONPROFIT ORGANIZATIONS AUTHORIZED TO MAIL AT SPECIAL RATES *(Section 132.122, PSM)*
The purpose, function, and nonprofit status of this organization and the exempt status for Federal income tax purposes *(Check one)*

[XX] HAVE NOT CHANGED DURING PRECEDING 12 MONTHS — [ ] HAVE CHANGED DURING PRECEDING 12 MONTHS — *(If changed, publisher must submit explanation of change with this statement.)*

10. EXTENT AND NATURE OF CIRCULATION	AVERAGE NO. COPIES EACH ISSUE DURING PRECEDING 12 MONTHS	ACTUAL NO. COPIES OF SINGLE ISSUE PUBLISHED NEAREST TO FILING DATE
A. TOTAL NO. COPIES PRINTED *(Net Press Run)*	4,840	3,300
B. PAID CIRCULATION 1. SALES THROUGH DEALERS AND CARRIERS, STREET VENDORS AND COUNTER SALES	853	625
2. MAIL SUBSCRIPTIONS	1,984	1,800
C. TOTAL PAID CIRCULATION *(Sum of 10B1 and 10B2)*	2,837	2,425
D. FREE DISTRIBUTION BY MAIL, CARRIER OR OTHER MEANS SAMPLES, COMPLIMENTARY, AND OTHER FREE COPIES	797	100
E. TOTAL DISTRIBUTION *(Sum of C and D)*	3,634	2,525
F. COPIES NOT DISTRIBUTED 1. OFFICE USE, LEFT OVER, UNACCOUNTED, SPOILED AFTER PRINTING	1,051	775
2. RETURNS FROM NEWS AGENTS	155	Ø
G. TOTAL *(Sum of E, F1 and 2—should equal net press run shown in A)*	4,840	3,300

11. I certify that the statements made by me above are correct and complete. — SIGNATURE AND TITLE OF EDITOR, PUBLISHER, BUSINESS MANAGER, OR OWNER: Willard J. Holden

12. FOR COMPLETION BY PUBLISHERS MAILING AT THE REGULAR RATES *(Section 132.121, Postal Service Manual)*

39 U. S. C. 3626 provides in pertinent part: "No person who would have been entitled to mail matter under former section 4359 of this title shall mail such matter at the rates provided under this subsection unless he files annually with the Postal Service a written request for permission to mail matter at such rates."

In accordance with the provisions of this statute, I hereby request permission to mail the publication named in Item 1 at the phased postage rates presently authorized by 39 U. S. C. 3626.

SIGNATURE AND TITLE OF EDITOR, PUBLISHER, BUSINESS MANAGER, OR OWNER: Willard J. Holden PUBLISHER

(McAdams and Bregman · continued from p. 60)

The significance of the pulsation threshold method for the measurement of auditory nonlinearity. Ph.D. Diss., Erasmus University, Rotterdam, The Netherlands.

34. Warren, R. M., C. J. Obusek, and J. M. Ackroff (1972). "Auditory induction: Perceptual synthesis of absent sounds." *Science*, Vol. 176, pp. 1149-1151.
35. Wessel, David L. (1979). "Timbre Space as a Musical Control Structure." *Computer Music Journal*, Vol. 3, No. 2, pp. 45-52.