9

GROUPING MECHANISMS IN MUSIC

DIANA DEUTSCH

Department of Psychology
University of California, San Diego
La Jolla, California

I. INTRODUCTION

Music provides us with a complex, rapidly changing acoustic spectrum, often derived from the superposition of sounds from many different sources. Our auditory system has the task of analyzing this spectrum so as to reconstruct the originating sound events. This is analogous to the task performed by our visual system when it interprets the mosaic of light impinging on the retina in terms of visually perceived objects. Such a view of perception as a process of “unconscious inference” was proposed in the last century by Helmholtz (1909–1911/1925), and we shall see that many phenomena of music perception can be viewed in this way.

Two types of issue can be considered here. First, given that our auditory system is presented with a set of first-order elements, we can explore the ways in which these are combined so as to form separate groupings. If all first-order elements were indiscriminately linked together, auditory shape recognition operations could not be performed. There must, therefore, be a set of mechanisms that enable us to form linkages between some elements and that inhibit us from forming linkages between others. Simple mechanisms underlying such linkages are examined in the present chapter. The second issue concerns the ways in which higher order abstractions are derived from combinations of first-order elements so as to give rise to perceptual equivalences and similarities. This issue is explored in Chapter 10, and we shall see that higher-order abstractions are also used as bases for grouping.

In considering the mechanisms whereby we combine musical elements into groupings, we can also follow two lines of inquiry. The first concerns the *dimensions* along which grouping principles operate. When presented with a complex pattern, the auditory system groups elements together according to some rule based on frequency, amplitude, temporal position, spatial location, or some multidimensional attribute such as timbre. As we shall see, any of these attributes can be used as a basis for grouping, but the conditions determining which attribute is used are complex ones.

Second, assuming that organization takes place on the basis of some dimension such as frequency, we can inquire into the *principles* that govern grouping along this dimension. The early Gestalt psychologists proposed that we group elements into configurations on the basis of various simple rules (see, for example, Wertheimer, 1923). One is proximity: closer elements are grouped together in preference to those that are spaced further apart. An example is shown in Figure 1a, where the closer dots are perceptually grouped together in pairs. Another is similarity: in viewing Figure 1b we perceive one set of vertical rows formed by the filled circles and another formed by the unfilled circles. A third, good continuation, states that elements that follow each other in a given direction are perceptually linked together: we group the dots in Figure 1c so as to form the two lines AB and CD. A fourth, common fate, states that elements that change in the same way are perceptually linked together. As a fifth principle, we tend to form groupings so as to perceive configurations that are familiar to us.

It has been shown that such laws operate in the perception of visual arrays, and we shall see that this is true of music also. It seems reasonable to assume—as argued by R. L. Gregory (1970), Sutherland (1973), Hochberg (1974), Deutsch (1975a), Bregman (1978, 1990), and Rock (1986)—that grouping in conformity with such principles enables us to interpret our environment most effectively. In the case of vision, elements that are close together in space are more likely to belong to the same object than are elements that are spaced further apart. The same line of reasoning holds for elements that are similar rather than those that are dissimilar. In the case of hearing, similar sounds are likely to have originated from a common source, and dissimilar sounds from different sources. A sequence that changes smoothly in frequency is likely to have originated from a single source, whereas an abrupt frequency transition may reflect the presence of a new source. Components of a complex spectrum that arise in synchrony are likely to have

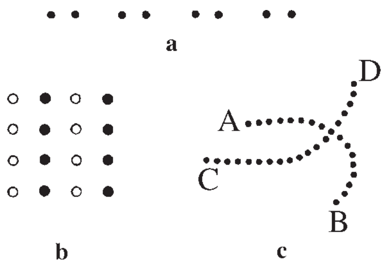


FIGURE 1 Illustrations of the Gestalt principles of proximity, similarity, and good continuation.

emanated from the same source, and the sudden addition of a new component may signal the emergence of a new source.

Another general question to be considered is whether perceptual grouping results from the action of a single decision mechanism or whether multiple decision mechanisms are involved, each with its own grouping criteria. There is convincing physiological evidence that the subsystems underlying the attribution of various characteristics of sound become separate very early in the processing system (Edelman, Gall, & Cowan, 1988). Such evidence would lead us to hypothesize that auditory grouping is not carried out by a single mechanism but rather by a number of mechanisms, which at some stage act independently of each other. As we shall see, the perceptual evidence strongly supports this hypothesis. and further indicates that the different mechanisms often come to inconsistent conclusions. For example, the parameters that govern grouping to determine perceived pitch can differ from those that determine perceived timbre, location, or number of sources (Darwin & Carlyon, 1995; Hukin & Darwin, 1995a). Further evidence comes from various illusions that result from incorrect conjunctions of different attribute values (Deutsch, 1974, 1975a, 1975b, 1980a, 1981, 1983a, 1983b, 1987, 1995). From such findings we shall conclude that perceptual organization in music involves a process in which elements are first grouped together so as to assign values to different attributes separately, and that this is followed by a process of perceptual synthesis in which the different attribute values are combined—either correctly or incorrectly.

II. FUSION AND SEPARATION OF SPECTRAL COMPONENTS

In this section, we consider the relationships between the components of a sound spectrum that lead us to fuse them into a unitary sound image and those that lead us to separate them into multiple sound images. In particular, we shall be exploring two types of relationship. The first is harmonicity. Natural sustained sounds, such as produced by musical instruments and the human voice, are made up of components that stand in harmonic, or near-harmonic, relation (i.e., their frequencies are integer, or near-integer multiples of the fundamental). It is reasonable to expect, therefore, that the auditory system would exploit this feature so as to combine a set of harmonically related components into a single sound image. To take an everyday example, when we listen to two instrument tones playing simultaneously, we perceive two pitches, each derived from one of the two harmonic series that together form the complex.

A second relationship that we shall be exploring is onset synchronicity. When components of a sound complex begin at the same time, it is likely that they have originated from the same source; conversely, when they begin at different times, it is likely that they have originated from different sources. As an associated issue, we shall be exploring temporal correspondences in the fluctuations of components in the steady-state portion of a sound.

The importance of temporal relationships for perceptual fusion and separation was recognized by Helmholtz in his treatise *On the Sensations of Tone* (1859/1954), in which he wrote:

> Now there are many circumstances which assist us first in separating the musical tones arising from different sources, and secondly, in keeping together the partial tones of each separate source. Thus when one musical tone is heard for some time before being joined by the second, and then the second continues after the first has ceased, the separation in sound is facilitated by the succession in time. We have already heard the first musical tone by itself and hence know immediately what we have to deduct from the compound effect for the effect of this first tone. Even when several parts proceed in the same rhythm in polyphonic music, the mode in which the tones of the different instruments and voices commence, the nature of their increase in force, the certainty with which they are held and the manner in which they die off, are generally slightly different for each.… When a compound tone commences to sound, all its partial tones commence with the same comparative strength; when it swells, all of them generally swell uniformly; when it ceases, all cease simultaneously. Hence no opportunity is generally given for hearing them separately and independently. (pp. 59–60).

A. HARMONICITY

Musical instrument tones provide us with many informal examples of perceptual grouping by harmonicity. Stringed and blown instruments produce tones whose partials are harmonic, or close to harmonic, and these give rise to strongly fused pitch impressions. In contrast, bells and gongs, which produce tones whose partials are nonharmonic, give rise to diffuse pitch impressions (Mathews & Pierce, 1980).

Formal experiments using synthesized tones have confirmed this conclusion. De Boer (1976) found that tone complexes whose components stood in simple harmonic relation tended to produce single pitches, whereas nonharmonic complexes tended instead to produce multiple pitches. Bregman and Doehring (1984) reported that placing simultaneous gliding tones in simple harmonic relation enhanced their perceptual fusion. They presented subjects with three simultaneous glides and found that the middle glide was more easily captured into a separate melodic stream when its slope differed from that of the other two. Furthermore, when the slope of the middle glide was the same as the others, it was less easily captured into a separate melodic stream when it stood in harmonic relationship with them.

How far can a single component of a complex tone deviate from harmonicity and still be grouped with the others to determine perceived pitch? Moore, Glasberg, and Peters (1985) had subjects judge the pitches of harmonic complex tones and examined the effects of mistuning one of the components to various extents. When the component was mistuned by less than 3%, it contributed fully to the pitch of the complex. As the degree of mistuning increased beyond 3%, the contribution made by the mistuned component gradually decreased, and at a mistuning of 8%, the component made virtually no contribution to the pitch of the complex.

Darwin and Gardner (1986) obtained analogous effects in the perception of vowel quality. Mistuning a harmonic in the first formant region of a vowel produced shifts in its perceived quality, with increasing shifts as the amount of mistuning increased. For mistunings of around 8%, the direction of the shift was such as would be expected had the component been perceptually removed from the calculation of the formant.

Other investigators have studied the perception of simultaneous complexes that were built on different fundamentals. They varied the relationships between the fundamentals, and examined how well listeners could separate out the complexes perceptually, as a function of these relationships. For example, Rasch (1978) used a basic pattern that consisted of a pair of two-tone chords that were presented in succession. All the tones were composed of a fundamental together with a series of harmonics. The lower tones of each chord were built on the same fundamental, whereas the higher tones differed by a fifth, in either the upward or the downward direction. The subject judged on each trial whether the higher tones formed an ascending or a descending pattern. The threshold amplitude for obtaining reliable judgments was taken as a measure of the degree to which the subject could separate out the tones forming each chord. As shown in Figure 2, as the higher tones were mistuned from simple harmonic relation with the lower ones, detection thresholds fell accordingly, reflecting an enhanced ability to separate out the pitches of the tones comprising the chords.

Huron (1991b) has related such findings on harmonicity and spectral fusion to polyphonic music. One objective of such music is to maintain the perceptual independence of concurrent voices. In an analysis of a sample of polyphonic keyboard

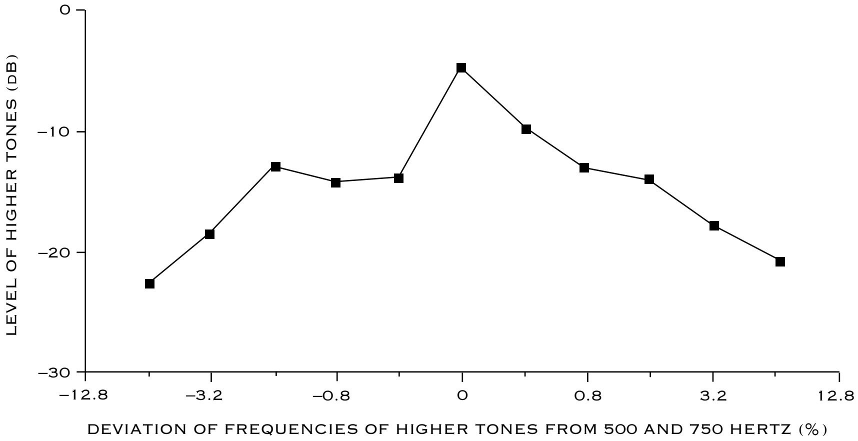


FIGURE 2 Detection thresholds for higher tones in the presence of lower ones. Two chords were presented in sequence. The lower tones of the chords were identical while the higher tones differed by a fifth, in either the upward or the downward direction. Subjects judged whether the higher tones formed a "high-low" or a "low-high" sequence. Detection thresholds fell as the higher tones deviated from simple harmonic relation with the lower ones. (Adapted from Rasch, 1978.)

works by J. S. Bach, Huron showed that harmonic intervals were avoided in proportion to the strength with which they promoted tonal fusion, and he concluded that Bach had used this compositional strategy in order to optimize the salience of the individual voices.

Other composers have focused on the creation of perceptual fusion rather than separation. Particularly in recent times, there has been much experimentation with sounds that were produced by several instruments playing simultaneously, and were configured so that the individual instruments would lose their perceptual identities and together produce a single sound impression. For example, Debussy and Ravel in their orchestral works made extensive use of chords that approached timbres. Later composers such as Schoenberg, Stravinsky, Webern, and Varese often used highly individualized structures, which Varese termed "sound masses" (Erickson, 1975). Here the use of tone combinations that stood in simple harmonic relation proved particularly useful.

To return to the laboratory experiments, findings related to those of Rasch (1978) have also been obtained for speech perception. A number of studies have shown that simultaneous speech patterns could be more easily separated out perceptually when they were built on different fundamentals—in general, the amount of perceptual separation reached its maximum when the fundamentals differed by roughly one to three semitones (Assmann & Summerfield, 1990; Brokx & Nooteboohm, 1982; Scheffers, 1983). Furthermore, formants built on the same fundamental tended to be grouped together so as to produce a single phonetic percept, whereas a formant built on a different fundamental tended to be perceived as distinct from the others (Darwin, 1981; see also Gardner, Gaskill, & Darwin, 1989)

The number of sources perceived by the listener provides a further measure of grouping. Moore, Glasberg, and Peters (1986) reported that when a single component of a harmonic complex was mistuned from the others, it was heard as standing apart from them. In other studies, simultaneous speech sounds were perceived as coming from a larger number of sources when they were built on different fundamentals (Broadbent & Ladefoged, 1957; Cutting, 1976; Darwin, 1981; Gardner et al., 1989).

Interestingly, less mistuning is required to produce the impression of multiple sources than to produce other effects. For example, a slightly mistuned component of a tone complex might be heard as distinct from the others, yet still be grouped with them in determining perceived pitch (Moore et al., 1986) or vowel quality (Darwin, 1981, Gardner et al., 1989). As argued by Darwin and Carlyon (1995), this type of disparity indicates that perceptual grouping involves a number of different mechanisms, which depend on the attribute being evaluated, and these mechanisms do not necessarily use the same criteria.

B. ONSET SYNCHRONICITY

So far we have been considering sounds whose components begin and end at the same time, and we have explored the spectral relationships between them that are conducive to perceptual fusion. In real musical situations, temporal factors also come into play. One such factor is onset synchronicity. The importance of this factor can be shown in a simple demonstration, in which a harmonic series is presented in such a way that its components enter at different times. For example, take a series that is built on a 200-Hz fundamental. We can begin with the 200-Hz component sounding alone, then 1 sec later add the 400-Hz component, then 1 sec later add the 600-Hz component, and so on until all the components are sounding together. As each component enters, its pitch is initially heard as a distinct entity, and then it gradually fades from perception, so that finally the only pitch that is heard corresponds to the fundamental.

Even a transient change in the amplitude of a component can enhance its perceptual salience. This was shown by Kubovy (1976) who generated an eight-tone chord whose components were turned off and on again abruptly, each at a different time. On listening to this chord, subjects perceived a melody that corresponded to the order in which the amplitude drops occurred.

Darwin and Ciocca (1992) have shown that onset asynchrony can influence the contribution made by a mistuned harmonic to the pitch of a complex. They found that a mistuned component made less of a contribution to perceived pitch when it led the others by more than 80 msec, and it made no contribution when it led the others by 300 msec.

Onset asynchrony can also affect the contribution of a component to perceived timbre. Darwin (1984) found that when a single harmonic of a vowel that was close in frequency to the first formant led the others by roughly 30 msec, there resulted an alteration in the way the formant frequency was perceived; this alteration was similar to the one that occurred when the harmonic was removed from the calculation of the formant (see also Darwin & Sutherland, 1984).

Interestingly, Darwin and colleagues have found that the amount of onset asynchrony that was needed to alter the contribution of a component to perceived pitch was greater than was needed to alter its contribution to perceived vowel quality. Hukin and Darwin (1995a) showed that this discrepancy could not be attributed to differences in signal parameters, but rather to the nature of the perceptual task in which the listener was engaged; again arguing, as did Darwin and Carlyon (1995), that such disparities reflect the operation of multiple decision mechanisms in the grouping process.

Onset asynchrony has been found to have higher level effects also. In one experiment, Bregman and Pinker (1978) presented listeners with a two-tone complex in alternation with a third tone, and they studied the effects of onset-offset asynchrony between the simultaneous tones. As the degree of onset asynchrony increased, the timbre of the complex tone was judged to be purer, and it became more probable that one of the tones in the complex would form a melodic stream with the third tone (see also Dannenbring & Bregman, 1978).

Using yet a different paradigm, Deutsch (1979) presented subjects with rapid melodic patterns whose components switched from ear to ear, and with each component accompanied by a drone in the contralateral ear. An onset asynchrony of 15 msec between the melody component and the drone significantly improved identification of the melody, indicating that the melody components were more easily combined together sequentially when they did not occur synchronously with other tones.

When two complex tones are played together, they are perceptually more distinct when their onsets are asynchronous than when they begin to sound at the same time. Rasch (1978) demonstrated this effect using the basic patterns and detection task described earlier. He showed that detection of higher tones in the presence of lower ones was strongly affected by onset asynchrony: Each 10 msec of delay of the lower tones was associated with roughly a 10-dB reduction in detection threshold. At a delay of 30 msec, the threshold for perception of the higher tones was roughly the same as when they were presented alone.

Rasch further observed that the subjective effect of this onset asynchrony was very pronounced. When the onsets of the tones were synchronous, a single fused sound was heard; however, when onset disparities were introduced, the tones sounded very distinct perceptually. This, as Rasch pointed out, is an example of the continuity effect (see Section II,C).

Rasch (1988) later applied the results of this study to live ensemble performances. He made recordings of three different trio ensembles (string, reed, and recorder) and calculated the onset relations between tones when they were nominally simultaneous. He found that asynchrony values ranged from 30 to 50 msec, with a mean asynchrony of 36 msec. Relating these findings to his earlier perceptual ones, Rasch concluded that such onset asynchronies enabled the listener to hear the simultaneous sounds as distinct from each other. According to this line of argument, such asynchronies should not be considered as performance failures, but rather as characteristics that are useful in enabling listeners to hear concurrent voices distinctly.

On this line of reasoning, larger amounts of asynchrony should produce even better and more reliable separation of voices. One might hypothesize, then, that compositional practice would exploit this effect—at least in polyphonic music, where it is intended that the individual voices should be distinctly heard. Evidence for this hypothesis was found by Huron (1993) in an analysis of J. S. Bach's 15 two-part inventions. He found that for 11 of these inventions, values of onset asynchrony were such that there were no other permutations of the rhythms of the voices (with duration, rhythmic order, and meter controlled for) that produced more onset asynchrony than occurred in Bach's actual music. For the remaining four inventions, values of asynchrony were still significantly higher than would be expected by chance. Huron concluded that Bach had deliberately produced such onset asynchronies so as to optimize the perceptual salience of the individual voices.

C. AUDITORY CONTINUITY

Auditory continuity is perhaps the most dramatic effect to result from temporal disparities within tone complexes. Consider the visual analogue shown in the upper portion of Figure 3, which was adapted from Vicario (1982). Line A could, in

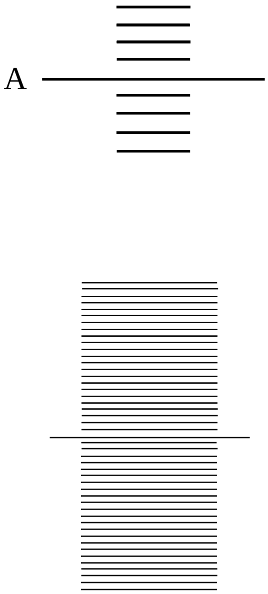


FIGURE 3 Visual analogue of an auditory continuity effect. Line A in the upper illustration could, in principle, be seen as having three components (a line to the left of the rectangle, a line to its right, and a line that forms part of the rectangle itself). However, it is instead seen as a single, continuous line. This effect is weaker in the lower illustration, in which the rectangle is wider, and the lines to its left and right are shorter. (Adapted from Vicario, 1982.)

principle, be viewed in terms of three components: a line to the left of the rectangle, a line to its right, and a line that forms part of the rectangle itself. However, our visual system instead treats all three components as a single line, which is independent of the remaining parts of the rectangle.

Vicario produced a musical equivalent of this demonstration. He generated a chord that consisted of components corresponding to C_4, $D\sharp_4$, $F\sharp_4$, A_4, C_5, $D\sharp_5$, and $F\sharp_5$; with A_4 both preceding and following the other components of the chord. Just as line A in Figure 3 is seen as continuing through the rectangle, so the listener heard a pitch corresponding to A_4 continue right through the chord.

This continuity effect is sensitive to the precise temporal parameters of the various components. To return to Vicario's visual analogue, when the lines forming the rectangle are lengthened and the lines to its left and right are shortened, as in the lower portion of Figure 3, the impression of continuity is reduced. Similarly, when the duration of the lengthened component of the chord is reduced, and the duration of the full chord is lengthened, the impression of auditory continuity is diminished.

In general, demonstrations of auditory continuity have existed for some time (see Warren, 1984, for a review). In an early study, Miller and Licklider (1950) rapidly alternated a tone with a noise burst, and subjects reported that the tone appeared to continue right through the noise. The authors called this the "picket fence effect," because in observing a landscape through a picket fence we see it as continuous rather than as broken up by the pickets. Vicario (1960) independently reported a similar phenomenon, which he called the "acoustic tunnel effect."

A different type of continuity effect was described by Warren, Obusek, and Ackroff (1972). When a broadband noise was repeatedly presented at different intensity levels, listeners heard the fainter noise as persisting without interruption, while the louder noise appeared to come on and off periodically. The authors found that analogous effects occurred with other signals also, such as narrowband noise, and pure and complex tones.

More elaborate continuity effects have also been reported. Dannenbring (1976) generated a pure-tone glide that rose and fell repeatedly. In some conditions, the glide was periodically interrupted by a loud broadband noise; however, it was perceived as though continuous. In contrast, when the glide was periodically broken, leaving only silent intervals during the breaks, listeners heard a disjunct series of rising and falling glides. Visual analogues of these two conditions, and their perceptual consequences, are shown in Figure 4.

Sudden amplitude drops between signals and intervening noise bursts may reduce, or even destroy, continuity effects. For example, Bregman and Dannenbring (1977) presented subjects with a gliding tone such as just described, and found that brief amplitude drops before and after the intervening noise bursts decreased the tendency to perceive the glide as continuous. Similarly, Warren et al. (1972), using noise bursts of alternating loudnesses, found that brief silences between the different bursts reduced the impression of continuity.

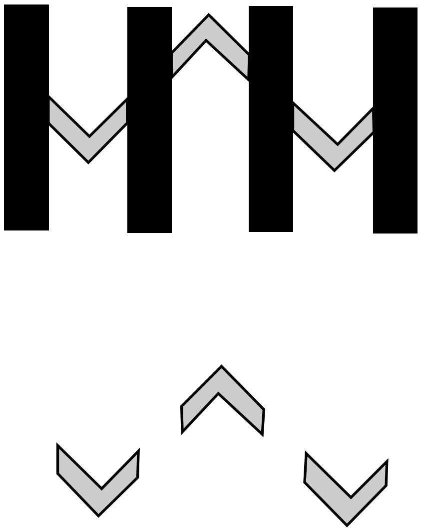


FIGURE 4 Visual illustration of an auditory continuity effect using gliding tones. See text for details. (Adapted from Bregman, 1990, which illustrates an experiment by Dannenbring, 1976.)

Amplitude drops do not, however, necessarily preclude the emergence of continuity effects. For example, tones produced by plucked instruments are characterized by rapid increases followed by decreases in amplitude. In music played by such instruments, when the same tone is rapidly repeated many times, and it is periodically omitted and replaced by a different tone, the listener may perceptually generate the omitted tone. Many examples of this phenomenon occur in 20th century guitar music, such as Tarrega's *Recuerdos de la Alhambra*, shown in Figure 5, and Barrios' *Una Limosna por el Amor de Dios*. Here the strong expectations set up by the rapidly repeating notes cause the listener to "hear" these notes even when they are not being played. Interestingly, at the end of the Barrios piece, the tempo is gradually slowed down, so that the gaps in the repeating presentations become apparent. In this way, the listener is drawn to realize that the gaps had been there, although imperceptibly, throughout the work.

A number of authors, such as Vicario (1973) and Warren (1983), have shown that listeners make use of both prior and subsequent contextual information in determining the strength and nature of continuity effects. In one experiment, Sasaki (1980) generated melodic patterns in which certain tones were omitted and replaced by loud noise bursts. Under some circumstances, listeners "heard" the

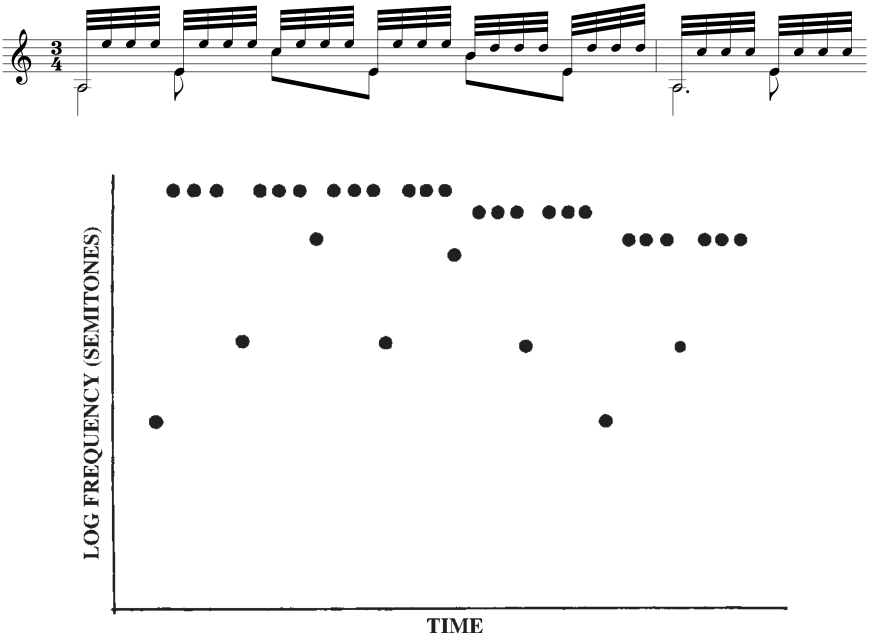


FIGURE 5 The beginning of *Recuerdos de la Alhambra*, by Tarrega. Although the tones are presented one at a time, two parallel lines are perceived, organized in accordance with pitch proximity. (Adapted from Deutsch, 1996.)

missing tone appear through the noise. This percept was most likely to occur when the omitted tone was predictable from the musical context; for example, when it formed part of a well-known melody. In this way, the experiment also provided evidence for grouping in accordance with the principle of familiarity.

In another experiment, Ciocca and Bregman (1987) presented listeners with a gliding tone that was interrupted by a noise burst. When the entering and exiting portions of the glide fell either in the same frequency range, or on a common trajectory, listeners tended to hear the glide as continuing through the noise. Later, Tougas and Bregman (1990) generated two simultaneous glides, one ascending and the other descending, with the two crossing in the middle. Previous studies had shown that global frequency proximity strongly influenced how crossing pitch patterns were perceived (Deutsch, 1975a, 1975b; Tougas & Bregman, 1985; Van Noorden, 1975; see also Section V, this chapter). As expected from these findings, Tougas and Bregman (1990) found that frequency proximity dominated over trajectory in determining the type of perceptual restoration that was obtained: Listeners tended to hear a higher glide that fell and then rose again, together with a lower glide that rose and then fell again, with the two meeting in the middle.

D. FREQUENCY MODULATION

Natural sustained sounds, such as those generated by musical instruments and the singing voice, constantly undergo small frequency fluctuations that preserve the ratios formed by their components (Cardozo & Van Noorden, 1968; Flanagan, 1972; Grey & Moorer, 1977; Lieberman, 1961; MacIntyre, Schumacher, & Woodhouse, 1981, 1982). It has been surmised that the auditory system uses such coherent frequency modulation as a cue for grouping spectral components together; and conversely uses incoherent frequency modulation as a cue for separating them out perceptually (for a discussion, see Bregman, 1990). Indeed, composers such as Chowning (1980) and McNabb (1981) have produced informal demonstrations that coherent frequency modulation, when imposed on synthesized singing voices or musical instrument tones, enhances perceptual fusion.

The issue, however, is theoretically a complex one. It has been argued that because information concerning frequency modulation is severely degraded in reverberant environments, the reliance on incoherent frequency modulation as a cue for perceptual separation could cause us to separate out components when they should in fact be perceptually grouped together. Furthermore, incoherent frequency modulation necessarily causes the frequency relationships between components to depart from harmonicity. Because the perceptual system already uses such departures as cues for perceptual segregation, the usefulness of invoking incoherent frequency modulation as an additional cue is debatable (Summerfield, 1992).

The experimental evidence on this issue is also complex. McAdams (1989) explored the effect of frequency modulation on the perceptual separation of three simultaneous sung vowels, which were built on different fundamentals. He found that when target vowels were frequency modulated, this increased their perceptual prominence. However, the perceived prominence of these vowels was not affected by whether the nontarget vowels were modulated coherently or incoherently with them, or even by whether the nontarget vowels were modulated at all.

In related experiments, Gardner and Darwin (1986) and Gardner et al. (1989) found that incoherent frequency modulation of the components of different vowels did not enhance their perceptual salience. Furthermore, when one component of a vowel was frequency modulated incoherently with the others, this manipulation did not reduce its contribution to the vowel's phonetic categorization.

Other negative findings were obtained by Carlyon (1991, 1992), who found that listeners were insensitive to frequency modulation incoherence when it was independent of departures from harmonicity. When the components of tones stood in nonharmonic relation, listeners were unable to judge whether they were modulated coherently or incoherently with each other (see also Summerfield & Culling, 1992).

Such negative findings raise the question of why frequency modulation can nevertheless enhance a vowel's perceptual salience. A possible explanation was advanced by McAdams (1984), who pointed out that when the harmonics of a vowel are frequency modulated, they also undergo amplitude modulation that traces the vowel's spectral envelope. In this way, the listener is provided with more complete information about the vowel's identity. Such spectral tracing might therefore be responsible for the enhanced perceptual salience of frequency-modulated vowels.

To test this hypothesis, Marin and McAdams (1991) synthesized sung vowels that were frequency modulated in either of two ways. In some conditions, the amplitudes of the components remained constant as their frequencies were modulated, and in other conditions, their amplitudes were varied so as to trace the vowel's spectral envelope.

Subjects were presented with chords consisting of three sung vowels that were built on different fundamentals, and they judged on each trial how prominent each vowel sounded within its chord. Although frequency-modulated vowels were heard as more prominent than unmodulated ones, spectral tracing did not have an effect.

Marin and McAdams' study therefore provided evidence against the spectral tracing hypothesis. As an alternative explanation for the enhanced prominence of frequency-modulated vowels, we may advance the direct hypothesis that neural units involved in the attribution of vowel quality are more strongly activated by frequency-modulated sounds than by unmodulated ones.

E. AMPLITUDE MODULATION

Because many natural sounds consist of spectral components whose amplitudes rise and fall in synchrony with each other, one might conjecture that coherent amplitude modulation would be used by the auditory system as a cue for perceptual fusion. On the other hand, coherent amplitude modulation is by no means universal—the partials of many musical instrument tones do not rise and fall in synchrony with each other. So the use of amplitude modulation incoherence as a cue for perceptual separation could cause the listener to erroneously separate out components when they should be perceptually fused together.

The experimental evidence on this issue is also equivocal. Bregman, Abramson, Doehring, and Darwin (1985) found evidence that coherent amplitude modulation could promote perceptual fusion; however, the modulation rates used here were so high that their findings could instead be interpreted as related to mechanisms involved in pitch perception. At slower rates, convincing evidence that coherent amplitude modulation leads to perceptual fusion has been difficult to obtain (Darwin & Carlyon, 1995).

F. EAR OF INPUT

Because all the components of a sound necessarily originate from a common location, and the components of different sounds originate from different locations, one might expect that the inferred spatial origins of components would strongly influence how they are perceptually grouped together. The issue arises, however, of how the spatial origin of a component should be inferred in the first place. In natural environments, sound waves are subjected to numerous distortions as they travel from their sources to our ears. So if we were to rely on first-order localization cues alone (such as differences in amplitude and phase between the ears), we would risk separating out components when they should instead be combined perceptually.

Given this line of reasoning, we might expect the auditory system not to use first-order localization cues as primary bases for grouping, but instead to use them only when other supporting cues are present. Indeed, we can go further and hypothesize that factors such as harmonicity and onset synchronicity, which indicate that components have originated from a common source, might cause us to hear these components as arising from the same spatial location.

Experimental evidence supporting this view has been obtained from studies in which different components of a complex were presented to each ear. Beerends and Houtsma (1989) had subjects identify the pitches of two complex tones, when their partials were distributed across the ears in various ways. They found that pitch identification was only weakly affected by the way the partials were distributed. Furthermore, Darwin and Ciocca (1992) found that the contribution of a single mistuned harmonic to the pitch of a complex tone was almost as large when this harmonic was delivered to the opposite ear as when it was delivered to the same ear as the other harmonics.

Related effects have been found for the perception of speech sounds. Broadbent and Ladefoged (1957) presented the first two formants of a phrase, with one formant delivered to each ear. Provided that the two formants were built on the same fundamental, subjects were able to identify the speech signal, and they also tended to hear a single voice, so that they were fusing the information from the two ears into a single perceptual image. Later, Hukin and Darwin (1995b) investigated the degree to which a single component contributed to the perceived quality of a vowel when it was presented to the ear opposite the remaining components. They found that this difference in ear of input had only a small effect.

Support has also been obtained for the conjecture that other grouping cues, such as harmonicity and asynchrony of onset, can influence the perceived spatial origin of a component of a complex (Hill and Darwin, 1993). Later we shall see that when two sequences of tones are presented simultaneously, one to each ear, a number of factors influence whether or not ear of input is used as a localization cue, and also influence the perceived spatial origins of the different tones.

III. LARGER SCALE GROUPINGS

So far, we have been focusing on situations in which single tone complexes are presented, and have identified various cues that the listener uses to sort their components into groupings. We now turn to the situation in which sequences of tones are presented instead. Here the auditory system abstracts relationships between successive tones, and uses these relationships as additional grouping cues.

One cue that we use here is pitch proximity: We tend to form sequential linkages between tones that are close in pitch and to separate out those that are further apart. Where rapid sequences of tones are concerned, researchers have frequently drawn an analogy with apparent motion in vision: When two lights that are in spatial proximity are flashed on and off in rapid succession, the observer obtains the illusion that a single light has moved from one location to the other. A second cue is temporal proximity: When pauses are placed between tones within a sequence, we use these as markers for grouping the tones into subsequences. A third cue is similarity of sound quality: When different types of instruments play together, we tend to form linkages between tones of similar timbre. We also invoke other principles in grouping tones together sequentially, such as good continuation and common fate.

IV. GROUPING OF RAPID SEQUENCES OF SINGLE TONES

A. PITCH PROXIMITY AND STREAM FORMATION

When a sequence of tones is presented at a rapid tempo, and the tones are drawn from two different pitch ranges, the listener perceives two melodic lines in parallel, one corresponding to the higher tones and the other to the lower ones. This perceptual phenomenon is frequently exploited by composers in the technique of pseudopolyphony, or compound melodic line. The passage from Tarrega's *Recuerdos de la Alhambra* shown in Figure 5 provides an example. In this Figure, the passage is also represented with pitch and time mapped into the vertical and horizontal dimensions of visual space, and it can be seen that two separate lines emerge in the visual representation, corresponding to the two pitch lines that are perceived by the listener.

This phenomenon of perceptual dissociation has been investigated in a number of studies. Miller and Heise (1950) presented listeners with two alternating tones, at a rate of 10 tones per second. When the pitch difference between these tones was small, listeners heard the sequence as a trill (i.e., as a single string of related tones). However, when the pitch difference was large, listeners instead heard the sequence as two interrupted and unrelated tones. In a further experiment, Heise and Miller (1951) used rapid sequences of tones that were composed of several pitches. When one of the tones in a sequence differed sufficiently in pitch from the others, it was heard in isolation from them.

A related phenomenon was demonstrated by Dowling (1973a). He presented two well-known melodies at a rapid tempo, such that the tones were taken from each melody in alternation. When the melodies were in closely overlapping pitch ranges, their components were perceptually combined into a single stream, with the result that subjects had considerable difficulty in identifying them. However, when the alternating melodies were instead in different pitch ranges, they were readily separated out perceptually, and so were easily identified.

B. TEMPORAL COHERENCE AS A FUNCTION OF PITCH PROXIMITY AND TEMPO

The term *temporal coherence* is used to describe the perceptual impression of a connected series of tones. The conditions giving rise to temporal coherence were studied by Schouten (1962). He found that as the frequency separation between successive tones increased, it was necessary to reduce their presentation rate in order to maintain the impression of a connected series.

Van Noorden (1975) examined this phenomenon in detail. Listeners were presented with sequences consisting of two tones in alternation, and they attempted either to hear temporal coherence or to hear *fission* (i.e., two streams of unrelated tones). Two boundaries were determined by these means. The first was defined as the threshold frequency separation as a function of tempo that was needed for the listener to hear the sequence as connected. The second established these values when the listener was attempting to hear fission. As shown in Figure 6, when listeners were attempting to hear coherence, decreasing the tempo from 50 to 150 msec per tone increased the frequency separation within which coherence could be heard from 4 to 13 semitones. However, when the listeners were instead attempting to hear fission, decreasing the tempo had little effect on performance. Between these two boundaries, there was a large region in which the listener could alter his listening strategy at will, and so hear either fission or coherence. So within this region, attentional set was important in determining how the sequence was perceived.

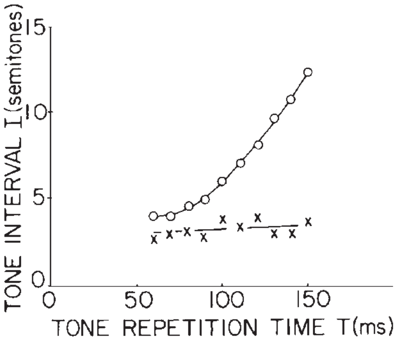


FIGURE 6 Temporal coherence boundary (o), and fission boundary (x) as a function of the frequency relationship between alternating tones and the presentation rate. (Adapted from Van Noorden, 1975).

Bregman and Bernstein (cited in Bregman, 1978) confirmed the interaction between frequency separation and tempo in judgments of temporal coherence. They found that as the frequencies of two alternating tones converged, a higher rate of alternation was required for the sequence to split perceptually into two different streams.

C. GROUPING BY PITCH PROXIMITY BUILDS WITH REPETITION

A number of studies have shown that the splitting of tonal sequences into two streams based on pitch proximity builds with repetition. Van Noorden (1975) compared the temporal coherence boundary for two-tone, three-tone, and long repetitive sequences. With three-tone sequences, the pitch change could be either unidirectional or bidirectional. As shown in Figure 7, for unidirectional three-tone sequences, temporal coherence occurred at rates that were equal to, or even higher than, those for two-tone sequences. However for bidirectional three-tone sequences, the rate of pitch change had to be set much lower than for two-tone sequences in order for coherence to be perceived. For long repetitive sequences, the rate of pitch change had to be set lower still.

In a related experiment, Bregman (1978) presented listeners repeatedly with two high tones together with a single low tone. When this sequence split perceptually into two streams, listeners heard two high tones in alternation, together with a single low tone that was steadily repeated. Bregman varied the number of tones that were packaged between 4-sec periods of silence, and listeners adjusted the speed of the sequence until the point of splitting was determined. As shown in Figure 8, as the number of tones in the package increased, the speed required for perception of separate streams decreased.

To explain this finding, Bregman argued that stream segregation is the product of a mechanism that groups together components of a spectrum so as to recon-

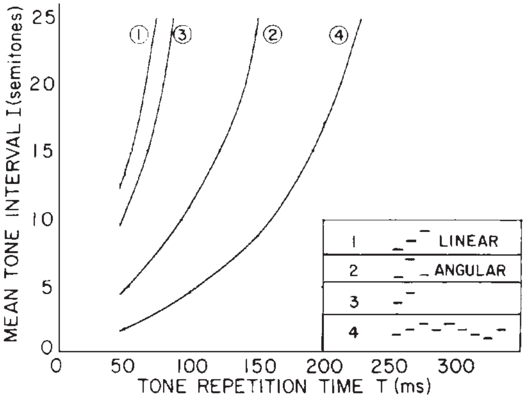


FIGURE 7 Temporal coherence boundary for two-tone (Curve 3), three-tone unidirectional (Curve 1), three-tone bidirectional (Curve 2), and continuous (Curve 4) sequences. (Adapted from Van Noorden, 1975.)

struct the original sounds. Such a mechanism would be expected to accumulate evidence over time, so that the segregation of components into different streams should build up with repetition (see also Bregman, 1990).

Further evidence that stream segregation results from such a parsing mechanism was provided by Bregman and Rudnicky (1975). Listeners judged the orders of two test tones that were embedded in a four-tone pattern that was itself flanked

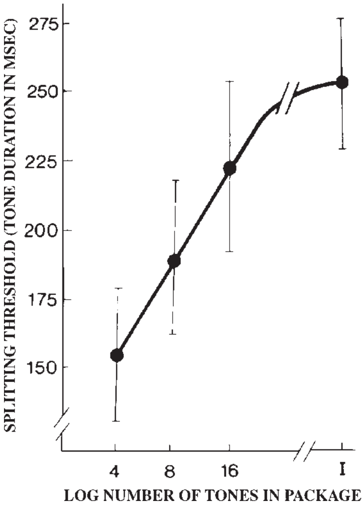


FIGURE 8 Threshold for stream segregation as a function of number of tones per package. Two “high” tones were presented in alternation with a single “low” tone. (Adapted from Bregman, 1978.)

by two "distractor tones." The presence of the distractor tones made the order of the test tones difficult to judge. However, when another stream of tones, called "captor tones," was moved close in frequency to the distractor tones, the distractors then combined with the captors to form a single stream, leaving the test tones in a stream of their own. In consequence, it became easy to judge the order in which the test tones appeared. The authors argued that the listeners were here presented with two simultaneously structured streams, and that the distractor tones could, in principle, belong to either one, but not to both simultaneously.

D. PITCH PROXIMITY AND THE PERCEPTION OF TEMPORAL RELATIONSHIPS

One consequence of the formation of separate perceptual streams is that temporal relationships between elements of the different streams become difficult to process. This has been shown in several ways. Bregman and Campbell (1971) presented a repeating sequence consisting of six tones: three from a high pitch range and three from a low one. When the tones occurred at a rate of 10 per second, it was difficult for listeners to perceive a pattern of high and low tones that was embedded in the sequence.

In a related experiment, Dannenbring and Bregman (1976) alternated two tones at high speeds so that they formed separate perceptual streams, and found that the tones from the two streams appeared to be overlapping in time. Further, Fitzgibbon, Pollatsek, and Thomas (1974) explored the perception of temporal gaps between tones that occurred in rapid sequence. Detection of a 20-msec gap was easy when the gap was placed between tones in the same frequency range, but difficult when it was placed between tones in different ranges (see also Neff, Jesteadt, & Brown, 1982).

Another reflection of such breakdown of temporal processing was found by Van Noorden (1975), who studied the detection of temporal displacement of a tone that alternated continuously with another tone of different frequency. As the rate of presentation of the tones increased, the threshold for detecting temporal displacement also increased. This rise in threshold was substantial when the tones were widely separated in frequency, but only slight when their frequencies were similar.

An effect of frequency disparity on temporal processing has also been found for two-tone sequences. Divenyi and Hirsh (1972) found that discrimination of the size of a temporal gap between tones within a pair deteriorated with increasing frequency separation between the tones. Williams and Perott (1972) also found that the minimum detectable gap between successively presented tones increased with increasing frequency difference between them. However, Van Noorden (1975) showed that the deterioration in temporal processing that he measured was considerably greater for long repetitive sequences than for two-tone sequences, so that it emerged as a consequence of stream formation (Figure 9).

E. GROUPING BY TIMBRE

Tones can also be grouped together on the basis of sound quality, or timbre. This is an instantiation of the principle of similarity: Just as we perceive the array in Figure 1b as four columns, two formed by the filled circles and two by the unfilled ones, so we group together tones that are similar in timbre and separate out those that are dissimilar. As a result, when different instruments play in parallel, we may form groupings based on their timbres even when their pitch ranges overlap heavily. An example is given in Figure 10, which is taken from Beethoven's Spring Sonata for violin and piano. Here the listener perceives two melodic lines that correspond to the tones played by each instrument, rather than linking the tones in accordance with pitch proximity.

A striking consequence of this grouping tendency was demonstrated by Warren, Obusek, Farmer, and Warren (1969). These authors generated a sequence of four unrelated sounds, and they presented it repeatedly without pause. The sounds, each 200 msec in duration, consisted of a high tone, a hiss (noise burst), a low tone, and a buzz (square wave). At this presentation rate, subjects were unable to name the orders in which the sounds occurred; for correct ordering to be achieved, the duration of each sound had to be longer than 500 msec.

Another consequence of grouping by timbre was demonstrated by Wessel (1979). He presented subjects with a repeating pattern consisting of a three-tone ascending pitch line, with successive tones composed of alternating timbres, as defined by their spectral energy distribution. When the timbral difference between successive tones was small, listeners heard the pattern as composed of ascending lines. However, when the timbral difference was large, listeners linked the tones together on the basis of timbre and so heard two, interwoven, descending lines instead.

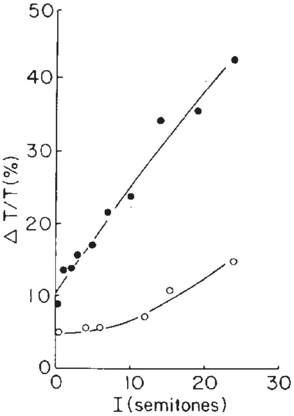


FIGURE 9 o Just noticeable displacement ΔT/T of the second tone of a two-tone sequence as a function of tone interval I. • Just noticeable displacement ΔT/T of one tone in a continuous sequence of alternating tones as a function of tone interval I. (Adapted from Van Noorden, 1975.)

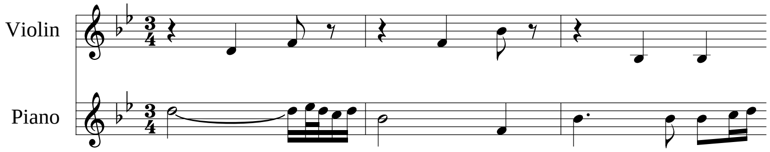


FIGURE 10 Passage from the beginning of the second movement of Beethoven's *Spring Sonata* for violin and piano. The tones played by the two instruments overlap in pitch; however, the listener perceives two melodic lines in parallel, which correspond to those played by each instrument. This reflects perceptual grouping by similarity.

Because timbre is multidimensional in nature, with both static and dynamic characteristics, we can ask which of its aspects are most conducive to perceptual grouping. Concerning static characteristics, Van Noorden (1975) found that tones with the same fundamental but different harmonics segregated into different streams on the basis of their harmonic structure. In a further demonstration, he generated complex tones that were filtered in various ways, and found that listeners segregated them on the basis of their filter characteristics, regardless of whether they were built on the same or different fundamentals.

In a further experiment, Singh (1987) generated sequences of tones that were based on different fundamentals and that also differed in harmonic content. Subjects were able to form streams based on either pitch or timbre. A related result was obtained by Bregman, Liao, and Levitan (1990), who generated tones that differed in both fundamental frequency and peak of spectral envelope. They found that both these factors could be used as bases for grouping.

The literature is equivocal concerning dynamic aspects of timbre. Wessel (1979) found that although streaming was clearly influenced by spectral content, variations in onset transients did not have a similar effect. Similarly, Hartmann and Johnson (1991) reported that although subjects easily segregated tones on the basis of harmonic content, they did not do so by type of temporal envelope, even though the envelopes they used were easily distinguished from each other.

Different conclusions were arrived at by Iverson (1995), who carried out an experiment using a number of different instrument tones. Subjects were able to form melodic streams on the basis of timbre, considering both the tones' static spectral characteristics and also their dynamic ones. The reasons for the differences between these findings remain to be determined.

F. GROUPING BY TEMPORAL PROXIMITY

When a sequence of sound elements is presented with pauses interspersed between them, we readily group the elements into subsequences that are defined by the pauses. In one experiment, Povel and Okkerman (1981) generated sequences consisting of tones of identical frequency, amplitude, and duration that were separated by gaps of two alternating durations. Subjects perceived these sequences as repeating groups of two tones that were segmented in accordance with the temporal gaps.

Other experiments have shown that grouping by temporal proximity can have a pronounced effect on the way that pitch patterns are perceived. Handel (1973) had subjects identify repeating patterns that consisted of dichotomous elements of differing pitch. Identification of the patterns was easy when they were temporally segmented in accordance with their pitch structure (e.g., when an eight-element pattern was segmented into groups of two), but difficult when the patterns were segmented inconsistently with their pitch structure (e.g., when an eight-element pattern was segmented into groups of three). In another experiment, Dowling (1973b) presented patterns that consisted of five-tone sequences that were separated by pauses, and subjects made recognition judgments concerning test sequences that were embedded in these patterns. Performance levels were higher when the test sequences had been presented in a single temporal segment than when a pause had been inserted between its elements.

Deutsch (1980b) performed a study on the recall of hierarchically structured pitch sequences. In some conditions, the sequences were divided into subsequences by the insertion of pauses. Performance levels were high when the pauses were in accordance with pitch structure, but low when they conflicted with pitch structure. Other measures showed that the subjects were grouping the patterns so strongly by temporal segmentation that they were unable to take advantage of pitch structure when this conflicted with temporal structure. This experiment is described in detail in Chapter 10.

G. GROUPING BY GOOD CONTINUATION

A few researchers have found evidence for grouping of tone sequences on the basis of good continuation. Divenyi and Hirsh (1974) studied order identification for three-tone sequences, and found that those with unidirectional frequency changes were easier to order than were those whose frequency changes were bidirectional. Analogous results were obtained by Nickerson and Freeman (1974), Warren and Byrnes (1975), and McNally and Handel (1977) for four-tone sequences. Further, Van Noorden (1975) found that a three-tone sequence was more likely to be judged as coherent when its tones formed unidirectional rather than bidirectional frequency changes.

H. GROUPING BY AMPLITUDE

Under some conditions, amplitude can act as a grouping cue. Dowling (1973a) found that loudness differences increased the perceptual distinctiveness of interleaved melodies. Van Noorden (1975) studied perception of sequences consisting of tones of identical frequency that alternated between two amplitudes. A single coherent stream was heard when the amplitude differences between the tones were less than 5 dB. However, two separate streams were heard when the amplitude differences were larger. In the latter case, attention could be directed to the softer stream as well as to the louder one. With even larger amplitude differences, auditory continuity effects were produced, so that the softer tone was heard as continuing through the louder one.

V. GROUPING OF MULTIPLE TONE SEQUENCES IN SPACE

In ensemble performances, we are presented with multiple streams of tones that arise in parallel from different regions of space. We can then inquire into the principles that govern perceptual grouping under such conditions. Do we form linkages between tones that are similar in pitch, in loudness, or in timbre? Alternatively, do we invoke spatial location as a prominent grouping cue? We shall see that all these factors are involved in grouping, but that they interact in complex ways. So given one type of pattern, grouping may be overwhelmingly determined by pitch proximity. But given a slight alteration in this pattern, grouping by spatial location may occur instead.

Powerful illusions also occur in this situation (Deutsch, 1974, 1975a, 1975b, 1980a, 1981, 1983a, 1983b, 1987, 1995). When we hear a tone, we attribute a pitch, a loudness, a timbre, and we hear the tone as coming from a particular spatial location. Each tone, as it is heard, may then be described as a bundle of attribute values. If our perception is veridical, this bundle reflects the characteristics and locations of the sounds that are being presented. However, when multiple sequences of tones are presented simultaneously, these bundles of attribute values may fragment and recombine in different ways, so that illusory conjunctions result. These illusory conjunctions then reflect the operation of multiple decision mechanisms in the grouping process (Deutsch, 1980a, 1981).

A. THE SCALE ILLUSION AND RELATED PHENOMENA

The *scale illusion*, which was first reported by Deutsch (1974, 1975b), provides an example of an illusory conjunction. The pattern that gives rise to this illusion is shown in the upper portion of Figure 11. This consists of a major scale, with successive tones alternating from ear to ear. The scale is played simultaneously in both ascending and descending form, such that whenever a tone from the ascending scale is in the right ear a tone from the descending scale is in the left ear, and vice versa. The sequence is played repeatedly without pause.

When listening to this pattern through earphones, people frequently experience the illusion shown in the lower portion of Figure 11. A melody corresponding to the higher tones is heard as coming from one earphone (in right-handers, generally the one on the right), with a melody corresponding to the lower tones as coming from the other earphone. When the earphone positions are reversed, the apparent locations of the higher and lower tones often remain fixed. This gives rise to the

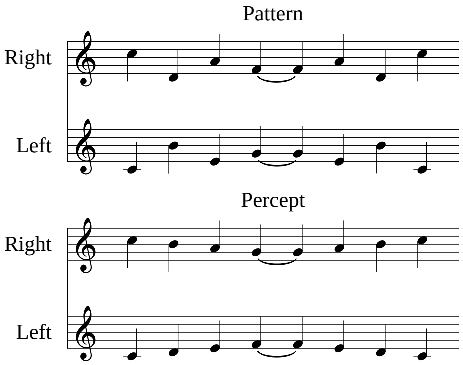


FIGURE 11 The pattern that produces the scale illusion, and the percept most commonly obtained. When this pattern is played through stereo headphones, most listeners hear two melodic lines that move in contrary motion. The higher tones all appear to be coming from one earphone, and the lower tones from the other, regardless of where each tone is coming from.

curious impression that the higher tones have migrated from one earphone to the other, and that the lower tones have migrated in the opposite direction.

Some listeners do not hear all the tones, but instead hear a single melodic line consisting of the higher tones alone, and little or nothing of the lower tones. This, together with other ways the scale illusion is sometimes perceived, is illustrated in Figure 12. The scale illusion and a number of its variants appear in the compact disc by Deutsch (1995).

In listening to the scale illusion, then, grouping by pitch proximity is so powerful that not only are the tones organized melodically in accordance with this principle, but they are frequently reorganized in space to conform with this principle also. Such spatial reorganization is in accordance with other findings showing that, in the absence of further supporting cues, differences in ear of input can have only small effects on how components of a tone complex are grouped together (Beerends & Houtsma, 1989; Darwin & Ciocca, 1992) and that other grouping cues can themselves influence the perceived spatial origins of individual components of a sound complex (Hill & Darwin, 1993). As described earlier, it makes sense that the auditory system would adopt such a listening strategy, because it is conducive to realistic interpretations of our environment. In the present situation, it is probable that a sequence of tones in one pitch range has originated from one source, and that another sequence of tones, in a different pitch range, has originated from a different source. So we exploit pitch proximity as a cue to determine how these tones are grouped together, and even to determine their perceived locations.

Variants of the scale illusion are readily produced. One of these is illustrated in Figure 13. A chromatic scale that ranges over two octaves is presented in both ascending and descending form, with the individual tones switching from ear to

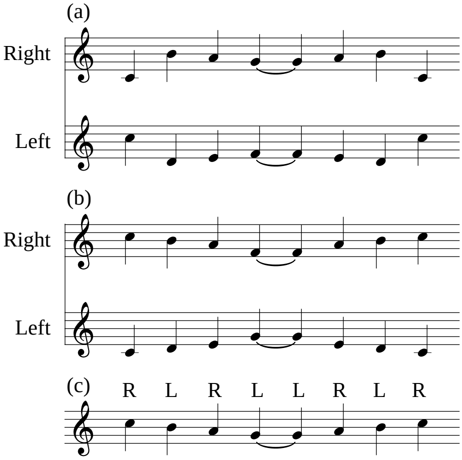


FIGURE 12 Different ways the scale illusion is sometimes perceived. (Adapted from Deutsch, 1995.)

ear in the same way as in the original scale illusion. When the pattern is played in stereo, most listeners hear a higher line that moves down an octave and up again, together with a lower line that moves up an octave and down again, with the two meeting in the middle. Yet when each channel is played separately, it is heard correctly as a series of tones that jump around in pitch. In Figure 13, the smoothing out of the visual representation in the score depicting the percept reflects well the way the sounds are perceptually reorganized.

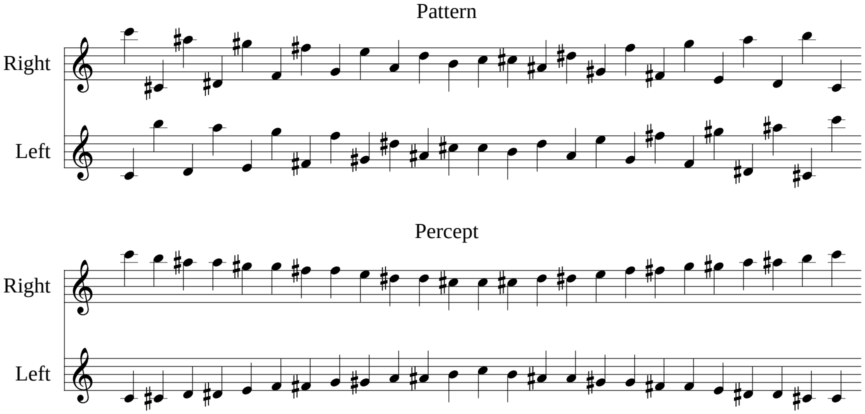


FIGURE 13 The pattern that produces a version of the chromatic illusion, and the way it is most often perceived. (Adapted from Deutsch, 1995.)

Butler (1979a) found evidence that the perceptual reorganization that occurs in the scale illusion also occurs in a broad range of musical situations. He presented the pattern shown in Figure 11 through spatially separated loudspeakers instead of earphones, and asked subjects to notate what they heard. In some conditions, the patterns were composed of piano tones, and differences in timbre were introduced between the sounds coming from the two speakers. Butler found that, despite these variations, virtually all responses reflected channeling by pitch proximity, so that higher and lower melodic lines were perceived, rather than the patterns that were in fact presented. When differences in timbre were introduced between the tones presented through the two speakers, a new tone quality was heard, but it appeared to be coming simultaneously from both speakers.

To determine whether these findings generalize to other configurations, Butler presented listeners with the two-part patterns shown in Figures 14a and 14b. Again, virtually all responses reflected grouping by pitch range. For both of these patterns, a perceptual reorganization occurred, so that a melody corresponding to the higher tones appeared to be coming from one earphone or speaker, with a melody corresponding to the lower tones coming from the other (Figures 14c and 14d).

Such effects even occur on listening to live music in concert halls. There is an interesting passage at the beginning of the final movement of Tchaikovsky's Sixth Symphony (*The Pathetique*). As shown in the upper portion of Figure 15, the notes from the theme are alternated between the first and second violin parts, and the notes from the accompaniment are alternated reciprocally (see Butler, 1979b, for a

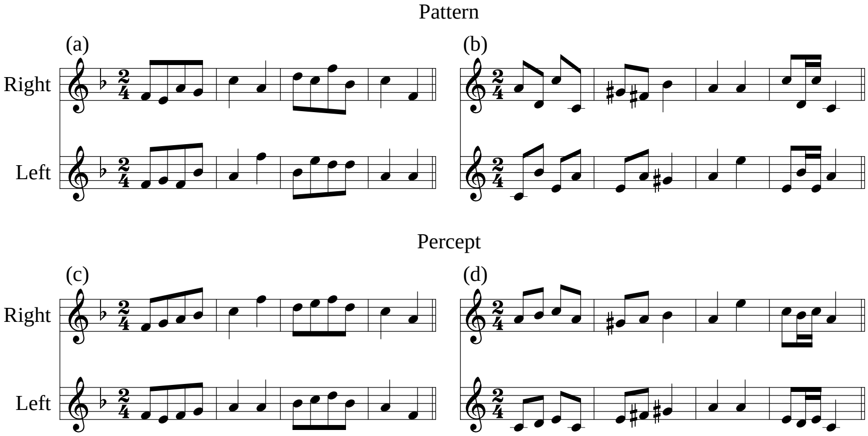


FIGURE 14 The upper portion of the figure shows two-part patterns that were presented to subjects through stereo headphones or loudspeakers. The lower portion shows these patterns as they were most commonly notated. (Adapted from Butler, 1979a.)

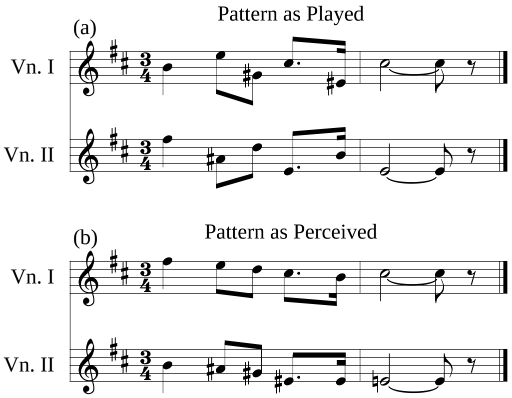


FIGURE 15 Beginning of the final movement of Tchaikovsky's Sixth Symphony *(The Pathetique)*. The upper portion of the figure shows the pattern as it is played, and the lower portion shows how it is generally perceived.

discussion). The passage, however, is not perceived as it is performed; rather, one violin part appears to be playing the theme and the other the accompaniment, as in the lower portion of Figure 15. This is true even with the orchestra arranged in 19th century fashion, so that the first violins are to the left of the audience, with the second violins to their right.

Whether it was Tchaikovsky's intention to produce a spatial illusion here, or whether he expected the audience to hear the theme waft back and forth between the two sides of space, we shall never know. However, there is a legend that the conductor Arthur Nikisch urged Tchaikovsky to rescore this passage so that the first violins would play the entire theme alone and the second violins the accompaniment. Tchaikovsky refused to change his scoring, but Nikisch rescored the passage anyway, and so created a second school of performance of this passage. The reasons for the argument between these two great musicians are unknown, but some conductors still prefer to perform the rescored version rather than Tchaikovsky's original one (Carlson, 1996).[1]

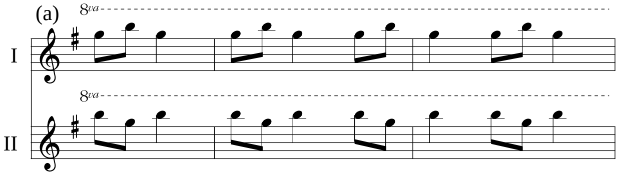
Another example of such spatial reorganization occurs at the end of the second movement of Rachmaninoff's Second Suite for Two Pianos. Here the first and second pianos play the two patterns shown in the upper portion of Figure 16. However, it appears to the listener that one piano is consistently playing the higher tone, and the other piano the lower one, as in the lower portion of this figure (Sloboda, 1985).

To return to the experiment of Deutsch (1975b), it is noteworthy that all the subjects formed perceptual groupings on the basis of overall pitch range. Rather

---

[1] I first came across this legend when it was relayed to me by David Butler, and it was later confirmed by the conductor George Zack, who had heard it from an independent source.

Pattern as Played



Pattern as Perceived

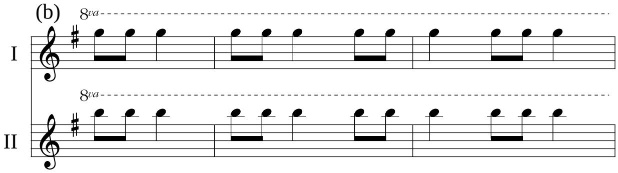


FIGURE 16 A passage from the second movement of Rachmaninoff's *Second Suite for Two Pianos*. The upper portion of the figure shows the pattern as it is played, and the lower portion shows how it is generally perceived.

than following the pattern purely on the basis of local (note-to-note) proximity, they either heard all the tones as two nonoverlapping pitch streams, or they heard the higher tones and little or nothing of the lower ones. No subject reported hearing a full ascending or descending scale as part of the pattern.

A related finding was obtained by Van Noorden (1975), who presented an ascending sequence of tones in rapid alternation with a descending one. Subjects heard this pattern as higher and lower melodic lines that moved in contrary motion. Similar findings were obtained by Tougas and Bregman (1985, 1990), who found an analogous perceptual organization with simultaneous ascending and descending glides.

The perceptual tendency to form melodic streams based on overall pitch range is reflected in the avoidance of part crossing in polyphonic music. Huron (1991a) documented this effect in an analysis of the polyphonic works of J. S. Bach. Interestingly, although when writing in two parts Bach avoided part crossing, he avoided it more assiduously when writing in three or more parts. Huron concluded that Bach was attempting to minimize the perceptual confusion that might otherwise have occurred as the density of sound images increased.

Do differences in timbre affect perception of the scale illusion? As described earlier, Butler (1979a) found that moderate differences in timbre did not alter the basic effect. However, Smith, Hausfield, Power, and Gorta (1982) used tones with substantial differences in timbre (one stream was generated by a synthesized piano and another by a synthesized saxophone) and found that timbre was then used as a basis for grouping. In a further experiment, A. L. Gregory (1994) generated a number of different instrument tones, and used these in various combinations to construct ascending and descending scales. When the tones were of identical timbre, listeners perceived nonoverlapping pitch streams, as described in Deutsch (1975b). However, when substantial differences in timbre were introduced, listeners tended to use these as cues for streaming.

Composers frequently exploit timbre as a carrier of melodic motion (Erickson, 1975) and place different instrument tones in the same pitch range, recognizing that listeners form groupings on the basis of instrument type when the timbre differences are sufficiently large. This is illustrated in the passage from Beethoven's *Spring Sonata* shown in Figure 10.

So far, we have been considering situations in which the tones coming from two sources are simultaneous, and this leads us to inquire what happens when temporal disparities are introduced. As we saw earlier, one would expect the listener to interpret such temporal disparities as indicators that the sounds were originating from different sources, and so to separate them out perceptually. As a result, we would expect streams to be formed on the basis of spatial location rather than pitch proximity.

As a test of this hypothesis, Deutsch (1979) presented subjects via earphones with the melodies shown in Figure 17. One of these patterns was repeatedly presented on each trial, and the subjects identified which of the two they had heard.

The four conditions in the experiment, together with their associated error rates, are shown in Figure 18. In Condition A, the melody was delivered to both ears simultaneously, and the error rate was here very low. In Condition B, the tones within each melody were switched haphazardly between ears, and the error rate here was considerably higher. On listening to patterns configured in this fashion, one feels compelled to attend to the tones delivered to the left ear or to the right, and it is very difficult to integrate them into a single coherent stream. Condition C was exactly as Condition B, except that the melody was accompanied by a drone. Whenever a tone from the melody was delivered to the right ear, the drone was delivered to the left ear, and vice versa. So both ears again received input simultaneously, even though the melody was switching from ear to ear, just as in Condition B. It can be seen that identification of the melody in the presence of the contralateral drone was again at a high level. In Condition D, the drone again

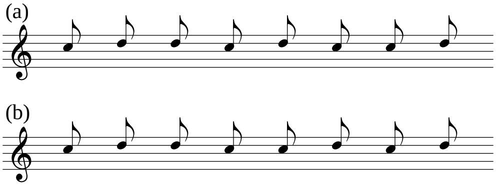


FIGURE 17 Basic pattern used to study the effects on melody identification of rapid switching between ears. All tones were 30 msec in duration, and tones within a sequence were separated by 100-msec pauses. (From Deutsch, 1979.)

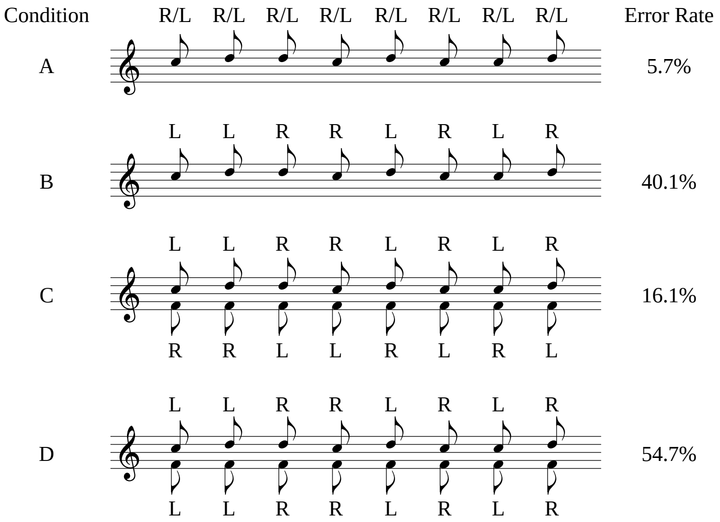


FIGURE 18 Examples of the way the tones were distributed to the two ears in the different conditions of the experiment of Deutsch (1979). Also shown are the error rates in these different conditions. See text for details.

accompanied the melody, but it was now delivered to the same ear as the melody component. So in this condition, input was again to one ear at a time, and as a result, performance again dropped substantially.

We can conclude that when tones emanate from different spatial locations, temporal relationships between them are important determinants of how they are perceptually grouped together. When the tones arrive at the two ears simultaneously (as in the scale illusion, and in Conditions A and C of the drone experiment) they are organized sequentially on the basis of pitch proximity. However, when the tones at the two ears are clearly separated in time, grouping by spatial location is so powerful as to virtually abolish the listener's ability to integrate them into a single melodic stream.

A similar conclusion was reached by Judd (1979), who generated two repeating patterns consisting of tones that were presented to the left and right ears in alternation. Comparing the two patterns, the orders of the tones were identical when each channel was played separately, but different when the channels were played together. Subjects listened to pairs of these patterns, and they judged on each trial whether the members of the pair were the same or different. On half the trials, the tones within each channel were separated by silent gaps, and on the other half, the gaps were filled with noise. Judd found that identification performance was better in the presence of the noise, and interpreted this finding as due to the noise degrading the localization information, which in turn discouraged grouping by spatial location.

To return to the study of Deutsch (1979), a second experiment was performed to explore intermediate cases, in which the tones arriving at the two ears were not strictly simultaneous but instead overlapped in time. Specifically, in some conditions of this experiment, the components of the melody and the drone were offset from each other by 15 msec. These intermediate conditions produced intermediate results: Identification of the melody in the presence of the contralateral drone when the two were asynchronous was more difficult than when the melody and drone were strictly synchronous, but easier than when the melody components switched between the ears without an accompanying drone.

It is interesting that Berlioz (1948) came to a similar conclusion from a composer's perspective. In his *Treatise on Instrumentation* he wrote:

> I want to mention the importance of the different points of origin of the tonal masses. Certain groups of an orchestra are selected by the composer to question and answer each other; but this design becomes clear and effective only if the groups which are to carry on the dialogue are placed at a sufficient distance from each other. The composer must therefore indicate on his score their exact disposition. For instance, the drums, bass drums, cymbals, and kettledrums may remain together if they are employed, as usual, to strike certain rhythms simultaneously. But if they execute an interlocutory rhythm, one fragment of which is given to the bass drums and cymbals, the other to kettledrums and drums, the effect would be greatly improved and intensified by placing the two groups of percussion instruments at the opposite ends of the orchestra, that is, at a considerable distance from each other.

Findings from the scale illusion and its variants, together with the drone experiment, indicate that the perception of musical passages can indeed be influenced profoundly by the spatial arrangements of instruments. When a pattern of tones is played at a rapid tempo, and the tones comprising the pattern are distributed between different instruments, the listener may be unable to integrate them into a single coherent stream. Such integration is more readily accomplished when the tones played by the different instruments overlap in time. However there is a tradeoff: as the amount of temporal overlap increases, our ability to identify the spatial origins of the different instrument tones decreases, and when the tones are simultaneous, spatial illusions should occur.

We now return to the question of how perception of simultaneous patterns of tones may be influenced by whether the higher tones are to the listener's right and the lower tones to the left, or the other way round. We saw earlier that in the scale illusion, right-handers tend to hear higher tones on their right and lower tones on their left, regardless of where the tones are indeed coming from. This means that tone combinations of the "high-right/low-left" type tend to be correctly localized, whereas combinations of the "high-left/ low-right" type tend to be localized less correctly.

Deutsch (1985) examined this effect in detail. Musically trained subjects were presented with simultaneous sequences of tones, one to each ear, and they transcribed them in musical notation. A sequence such as used in the study is shown in Figure 19. Each ear received a haphazard ordering of the first six tones of the F

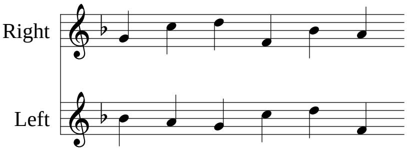


FIGURE 19 Example of a passage used to determine accuracy of pitch perception for chords of the "high-right/low-left" type, and of the "high-left/low-right" type. (Reprinted with permission from Deutsch, 1985. © 1985 by The Regents of the University of California.)

major scale, so that for some chords the tone fed to the right ear was higher and the tone to the left was lower (high-right/low-left chords), and for others this spatial disposition was reversed (high-left/low right chords). Subjects were asked to notate the tones that were presented to one ear, and to ignore those presented to the other.

When the subjects were attending to the right ear, they notated more higher tones than lower ones correctly. Furthermore, more higher tones than lower ones intruded from the left ear into their notations. In contrast, when the subjects were attending to the left ear, they correctly notated virtually the same number of higher and of lower tones, with a marginal advantage to the lower ones. Furthermore, more lower tones than higher ones intruded from the right ear into their notations. In other words, just as in the scale illusion, tones comprising high-right/low-left chords were correctly localized more often than were tones comprising high-left/low-right chords.

This finding raises the question of whether there might also be a perceptual advantage to high-right/low-left chords when localization accuracy is not at issue. In a further experiment, subjects listened to patterns that were designed in the same way as before. However, instead of focusing attention on one ear and ignoring the other, they were asked to notate the entire pattern, ignoring ear of input. It was found that more tones were correctly notated when they came from high-right/low-left chords than from high-left/low-right chords, showing that pitches forming chords with a high-right/low-left spatial disposition were more accurately perceived.

To the extent that effects of this sort occur in live musical situations, the following line of reasoning may be advanced. In general, contemporary seating arrangements for orchestras are such that, from the performers' point of view, instruments with higher registers are to the right and those with lower registers to the left. As an example, Figure 20 shows a seating plan for the Chicago Symphony, viewed from the back of the stage. Considering the strings, the first violins are to the right of the second violins, which are to the right of the violas, which are to the right of the cellos, which are in turn to the right of the basses. Consider also the brasses: The trumpets are to the right of the trombones, which are to the right of the tuba. Furthermore, the flutes are to the right of the oboes, with the clarinets to the right of the bassoons. It is interesting that the same principle tends to hold for other musical ensembles also. We may speculate that this type of spatial disposition has evolved by trial and error because it is conducive to optimal performance.

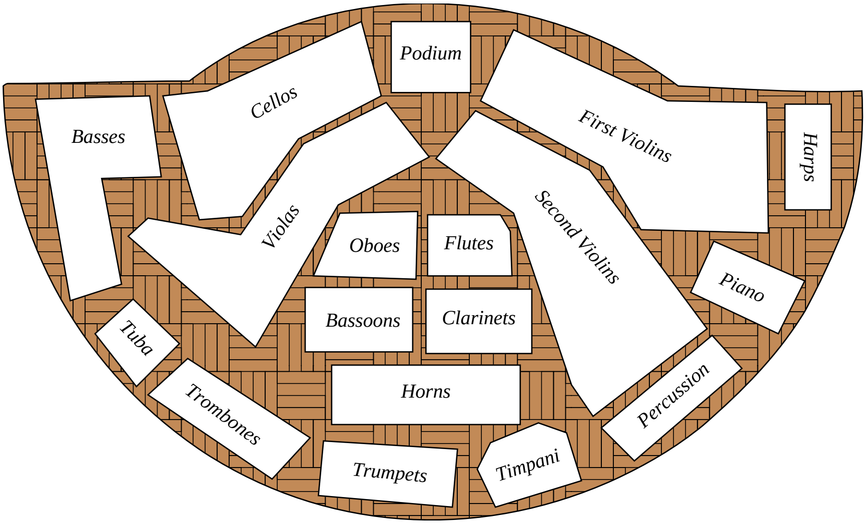


FIGURE 20 Seating plan for the Chicago Symphony, as viewed from the orchestra. (Adapted from Machlis, 1977.)

However, this presents us with a paradox. Because the audience sits facing the orchestra, this disposition is mirror-image reversed from their point of view: Instruments with higher registers tend to be to the audience's left, and those with lower registers, to their right. So for the audience, this spatial arrangement should cause perceptual difficulties. In particular, instruments with low registers, which are to the audience's right, should be less well perceived and localized. As described in Deutsch (1987), it is unclear how this problem can be resolved so as to produce an optimal arrangement for both the performers and the audience.

A further example of the spatial reorganization that we have been discussing was developed by Deutsch (1995), and is called the *glissando illusion*. The pattern that produces this illusion consists of a single oboe tone, which is played together with a sine wave that glides up and down in pitch. The two sounds are switched from ear to ear in such a way that when the oboe tone is in the left ear the glissando is in the right ear, and vice versa. This illusion appears on the compact disc by Deutsch (1995).

Many people hear the oboe tone as jumping back and forth from ear to ear, while the glissando appears to be joined together quite seamlessly. People localize the glissando in a variety of ways. For example, it is sometimes heard as coming from a fixed location, and sometimes as traveling from left to right as its pitch moves from low to high and then back from right to left as its pitch moves from high to low. The apparent spatial location of the glissando does not jump around as its components shift from ear to ear—the smoothness of its pitch transition is taken by the auditory system as a cue to assign it either a constant or a gradually changing location.

B. THE OCTAVE ILLUSION

In the experiments on simultaneous sequences so far described, grouping by pitch proximity was the rule when both ears received input simultaneously; grouping by spatial location occurred only when temporal disparities were introduced between the tones presented to the two ears. The *octave illusion*, which was first reported by Deutsch (1974), provides an interesting exception, because here following by spatial location occurs even when the tones presented to the two ears are strictly simultaneous. We shall see that this principle is adopted under special conditions of frequency relationship between tones that are presented in sequence at the two ears.

The pattern that gives rise to the octave illusion is shown in the upper portion of Figure 21. As can be seen, two tones that are spaced an octave apart are repeatedly presented in alternation. The identical sequence is played to both ears simultaneously but out of step with each other, so that when the right ear receives the high tone the left ear receives the low tone, and vice versa. The octave illusion appears on the compact disc by Deutsch (1995).

The octave illusion can take a number of different forms (Deutsch, 1974; 1975a, 1980a, 1981, 1983a, 1983b, 1987, 1995). Many people hear a single tone that switches from ear to ear, while its pitch simultaneously shifts back and forth between high and low. So it appears that one ear is receiving the pattern “high tone - silence - high tone - silence” (in right-handers, this is generally the right ear) while the other is receiving the pattern “silence - low tone - silence - low tone” (in right-handers, this is generally the left ear). This percept is illustrated in the lower portion of Figure 21. When the earphone positions are reversed, the apparent locations of the high and low tones often remain fixed: The tone that had appeared in the right ear continues to appear in the right ear, and the tone that had appeared in the left ear continues to appear in the left ear.

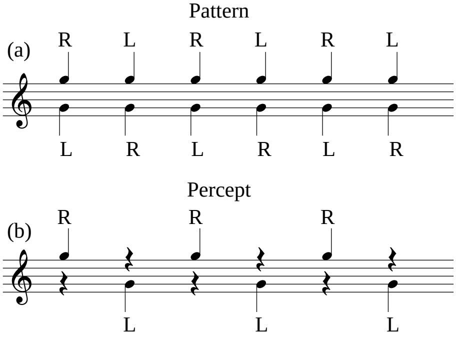


FIGURE 21 The pattern that produces the octave illusion, and the percept most commonly obtained. When this pattern is played through stereo headphones, most listeners hear an intermittent high tone in the right ear, which alternates with an intermittent low tone in the left ear.

Deutsch (1975a) hypothesized that the octave illusion results from the combined operation of two different decision mechanisms; one determines what pitch we hear, and the other determines where the tone appears to be coming from. The model is depicted in Figure 22. To provide the perceived pitches (i.e., the melodic line), the frequencies arriving at one ear are followed, and those arriving at the other ear are suppressed. However, each tone is localized at the ear that receives the higher frequency, regardless of whether a pitch corresponding to the higher or the lower frequency is in fact perceived.

We can take the case of a listener who follows the pitches delivered to his right ear. When the high tone is presented to the right and the low tone to the left, this listener hears a high tone, because this is presented to the right ear. The listener also localizes the tone in the right ear, because this ear is receiving the higher frequency. However, when the low tone is presented to the right ear and the high tone to the left, this listener now hears a low tone, because this is presented to the right ear, but localizes the tone in the left ear instead, because this ear is receiving the higher frequency. So the entire pattern is heard as a high tone to the right that alternates with a low tone to the left.

It can be seen that, on this model, reversing the positions of the earphones would not alter the basic percept. However, for the case of a listener who follows the pitches presented to the left ear instead, holding the localization rule constant, the identical pattern would be heard as a high tone to the left alternating with a low tone to the right. Later experiments have provided further evidence for this model (Deutsch, 1978, 1980a, 1981, 1987, 1988; Deutsch & Roll, 1976).

We can note that the octave illusion is a striking example of an illusory conjunction; a case in which the perceptual system incorrectly binds different attribute values together. Such incorrect binding raises the thorny issue of how we

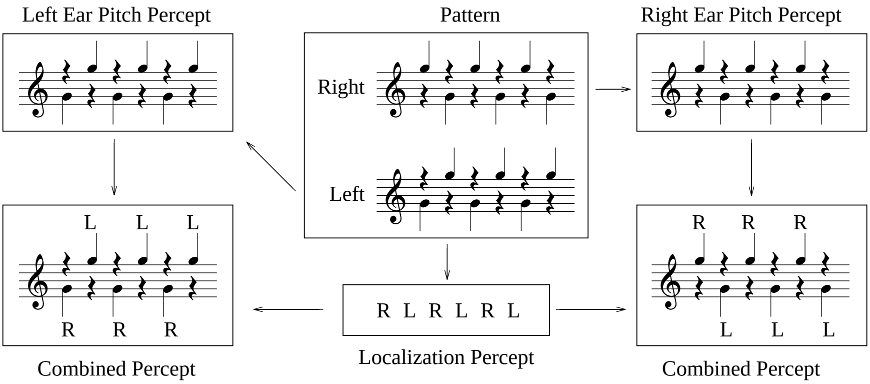


FIGURE 22 Model showing how the outputs of two decision mechanisms, one determining perceived pitch, and the other determining perceived location, combine to produce the octave illusion. See text for details.

generally manage to arrive at correct binding solutions—a problem that has been addressed particularly for the case of vision, and has recently begun to receive attention for hearing also. Deutsch (1980a, 1981) provided an early parallel processing model of how correct binding can occur for the case of values of two attributes (pitch and location) and that also explains the incorrect binding that occurs in the octave illusion. This model has been expanded to account for the correct binding of values of three or more attributes, such as pitch, location, and loudness (Deutsch, 1998).

Setting these issues aside, and considering only what pitches are perceived in the octave illusion, we note that grouping is here based on spatial location: The pitches that are heard correspond to the tones presented either to the listener's right ear or to his left. Further experiments have shown that the factor responsible for perceiving pitch on the basis of spatial location is that the two ears receive the same frequencies in succession, rather than different frequencies (Deutsch, 1978, 1980a, 1981, 1988).

Why should such a perceptual strategy have evolved? We may argue that it enables us to follow new, ongoing information with a minimum of interference from echoes or reverberation. In everyday listening, when the same frequency emanates successively from two different regions of space, the second occurrence may well be due to an echo. So it is a useful perceptual strategy to suppress the second occurrence of the sound from conscious perception. A similar argument has been advanced for the precedence effect: In listening to music, a single sound image may be obtained when the waveforms arriving from two different spatial locations are separated by time intervals of less than around 70 msec (see also Haas, 1951; Wallach, Newman, & Rosenzweig, 1949; Zureck, 1987).

C. MELODY PERCEPTION FROM PHASE-SHIFTED TONES

Another type of configuration that produces grouping by spatial location was described by Kubovy and colleagues. Kubovy, Cutting, and McGuire (1974) presented a set of simultaneous and continuous sine wave tones to both ears. They then phase-shifted one of the tones in one ear relative to its counterpart in the opposite ear. When these tones were phase-shifted in sequence, as shown in Figure 23, a melody was heard that corresponded to the phase-shifted tones; however, the melody was undetectable when the signal was played to either ear alone. Subjectively, the dichotically presented melody was heard as occurring inside the head but displaced to one side of the midline, while a background hum was heard as localized to the opposite side. So it appeared as though a source in one spatial position was producing the melody, while a source in a different spatial position was producing the background hum.

Kubovy (1981) pointed out that there are two potential interpretations of this effect. First, the segregation of the melody from the noise could have been based on concurrent difference cues; that is, the target tone may have been segregated because at that time its interaural disparity—or apparent spatial location—dif-

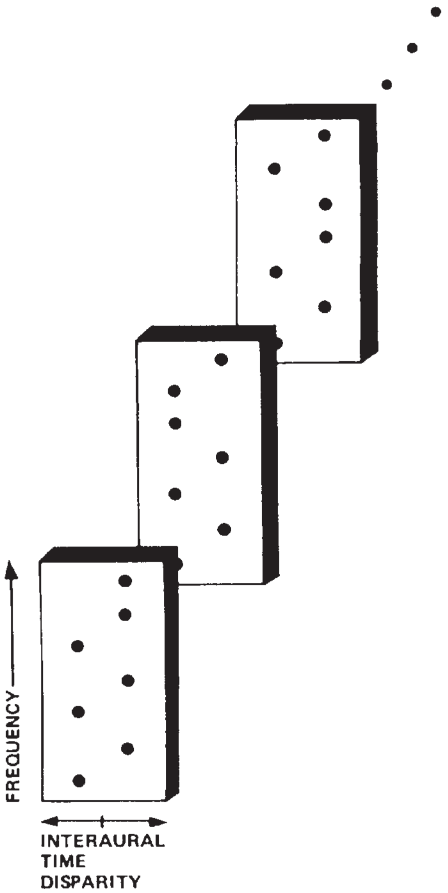


FIGURE 23 Pattern such as used by Kubovy, Cutting, and McGuire (1974) to show grouping of phase-shifted tones. See text for details. (From Kubovy & Pomerantz, 1981.)

fered from that of the background tones. Alternatively, the effect could have been based on successive difference cues; that is, the target tone may have been segregated because it had shifted its apparent position in space.

Kubovy devised two further configurations to determine which of these factors was responsible for the effect. In the first, the target tones shifted while the locations of the background tones remained constant, so producing a successive difference cue. In the second, the target tones themselves did not shift, but the background tones did, so producing a concurrent difference cue. Kubovy found that although both types of cue were effective in producing segregation, the successive difference cue was considerably more effective than the concurrent difference cue (see also Kubovy & Howard, 1976).

It is interesting that in Kubovy's paradigm, the listener creates melodic patterns from a fixed set of frequencies that are continuously present and simply shift their positions in space. A similar principle appears to be operating here as in the octave illusion, where melodic groupings are formed by the changing spatial positions of tones of identical frequency. It may then be hypothesized that Kubovy's effect also reflects the operation of a perceptual mechanism that has evolved to suppress echoes and reverberation.

D. HANDEDNESS CORRELATES

Handedness correlates have been obtained for both the octave and the scale illusions. Deutsch (1974) found that, in perceiving the octave illusion, right-handers tended strongly to hear the high tone on the right and the low tone on the left, and to maintain this percept when the earphone positions were reversed. However, there was considerable variation among left-handers in terms of where the high and low tones appeared to be localized, and what type of illusion was obtained. From further studies, it was concluded that these findings reflected a tendency to perceive the pitches that were presented to the dominant side of space rather than the nondominant side (Deutsch, 1975a, 1980a, 1981; 1983a, 1983b; Deutsch & Roll, 1976).

In a further study, Deutsch (1983a) divided the subject population into three groups—right-handed, mixed handed, and left-handed. For all three groups the tendency to perceive the high tone on the right and the low tone on the left was stronger among subjects with only right-handed parents and siblings, than among those who had a left- or mixed-handed parent or sibling. This pattern of results is in accordance with the neurological literature relating patterns of cerebral dominance to handedness and familial handedness background (Ettlinger, Jackson, & Zangwill, 1956; Luria, 1969; Subirana, 1958).

The handedness correlates obtained for the scale illusion may be viewed as reflecting more activity in the dominant hemisphere on the part of neural units underlying the higher tones, and more activity in the nondominant hemisphere on the part of units underlying the lower tones. Justification for this view comes in part from neurological studies showing that patients who experience palinacousis tend to perceive the illusory sound as located on the side of auditory space contralateral to the lesion (Jacobs, Feldman, Diamond, & Bender, 1973). Further, patients who obtain auditory sensations upon stimulation of the temporal lobe generally refer these sensations to the contralateral side of space (Penfield & Perot, 1963). This also explains the perceptual anisotropy found by Deutsch (1985) on listening to dichotic tonal sequences: right-handers perceived dichotic chords more accurately when the high tone was presented to the right ear and the low tone to the left, rather than the reverse.

VI. EQUAL-INTERVAL TONE COMPLEXES

Perceptual grouping principles emerge strongly with the use of tone complexes whose components are separated by equal intervals. Octave-related complexes have been explored most extensively (see also Chapter 10, this volume); however, tones whose components are related by other intervals have also been explored, as have chords produced by combinations of two or more octave-related complexes.

In a seminal experiment, Shepard (1964) generated a series of tones, each of which was composed of 10 sinusoidal components that were related by octaves. The amplitudes of the components were scaled by a fixed, bell-shaped spectral envelope, so that those in the middle of the musical range were highest, and those at the extremes were lowest. Shepard then varied the pitch classes of the tones by shifting all the components up or down in log frequency.

Subjects listened to ordered pairs of such tones and judged whether they formed ascending or descending patterns. When the second tone was removed one or two steps clockwise from the first along the pitch class circle (see Figure 21 in Chapter 10, this volume), listeners heard an ascending pattern; when the second tone was removed one or two steps counterclockwise, listeners heard a descending pattern instead. When the tones within a pair were separated by larger distances along the pitch class circle, the tendency for judgments to be determined by proximity gradually lessened, and when the tones were separated by exactly a half-octave, ascending and descending judgments occurred equally often.

Based on these findings, Shepard produced a compelling demonstration. A series of tones was played that repeatedly traversed the pitch class circle in clockwise steps, so that it appeared to ascend endlessly in pitch: C♯ sounded higher than C, D as higher than C♯, D♯ as higher than D, … , A♯ as higher than A, B as higher than A♯, C as higher than B, and so on without end. Counterclockwise motion produced the impression of an endlessly descending series of tones.

Risset (1969, 1971) produced a number of striking variants of Shepard's demonstration. In one variant, a single gliding tone was made to traverse the pitch class circle in clockwise direction, so that it appeared to move endlessly upward in pitch. When the tone was made to glide in counterclockwise direction, it appeared to move endlessly downward. In another variant, a tone was made to glide clockwise around the pitch class circle, while the spectral envelope was made to glide downward in log frequency; in consequence the tone appeared both to ascend and to descend at the same time (see also Charbonneau & Risset, 1973).

Effects approaching pitch circularity have been generated by composers for hundreds of years, and can be found in works by Gibbons, Bach, Scarlatti, Haydn, and Beethoven, among others. In the 20th century, effective pitch circularities have been produced by composers such as Stockhausen, Krenek, Berg, Bartok, Ligeti, and in particular Risset, using both natural instruments and computer-generated sounds. Braus (1995) provides an extensive discussion of such works.

Returning to the experimental evidence, the work of Shepard and Risset showed that when other cues to height attribution are weak, listeners will invoke proximity in making judgments of relative height for successively presented tones. We can then ask whether the perceptual system might invoke proximity in making judgments of relative height for simultaneously presented tones also.

In an experiment to examine this issue, Deutsch (1991) presented subjects with patterns such as shown in Figure 24 . Each pattern consisted of two simultaneous pairs of tones. In one pair, the second tone was a semitone clockwise from the first; in the other, it was a semitone counterclockwise. As expected from the earlier work, subjects organized these patterns sequentially in accordance with pitch proximity, so that they heard two melodic lines, one of which ascended by a semitone while the other descended by a semitone. So, for example, the pattern shown in Figure 24 was heard as the descending line D-C♯, together with the ascending line A♯-B. However, the descending line could in principle be heard as higher and the ascending line as lower (Percept A), or the ascending line could be heard as higher and the descending line as lower (Percept B). Figure 25 shows these two alternative perceptual organizations in musical notation.

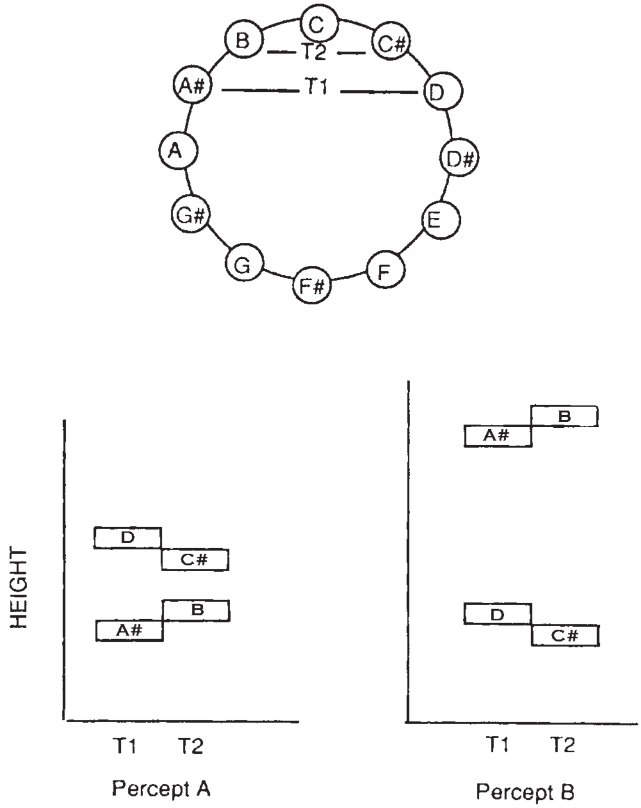


FIGURE 24 Example of pattern used to study the effect of pitch proximity in grouping simultaneous tones, together with two alternative perceptual organizations. Here the descending pair (D-C♯) is played together with the ascending pair (A♯-B). In principle, the descending line could be heard as higher and the ascending line as lower, reflecting perceptual organization in accordance with pitch proximity (Percept A). Alternatively, the ascending line could be heard as higher and the descending line as lower, reflecting perceptual organization that runs counter to pitch proximity (Percept B). (From Deutsch, 1988.)

In the experiment, one such pattern was presented on each trial, and subjects judged whether the line that was higher in pitch ascended or descended. From these judgments, it was inferred which tones were heard as higher and which as lower. All subjects showed a strong tendency to organize the patterns so that they formed compact harmonic groupings. For example, the pattern in Figure 24 tended to be heard as Percept A rather than Percept B. It appears, therefore, that the perceptual system tends to organize tone patterns in accordance with proximity along the simultaneous, or harmonic, dimension as well as along the sequential one.

In all the experiments described so far, the patterns were such that proximity along the pitch class circle co-occurred with proximity based on the spectral properties of the tones. The question then arises as to which of these two factors was responsible for the proximity effects that were obtained. This question was addressed by Pollack (1978) with respect to Shepard's original experiment. He presented subjects with complex tones whose components were related by octaves or octave multiples, and found that as the spectral overlap between successively presented tones increased, the tendency to follow by proximity increased also. Pollack concluded that proximity along the spectral dimension was responsible for Shepard's results. A similar conclusion was reached by Burns (1981), who found that the tendency to follow pairs of such tones in accordance with spectral proximity was no greater when the tones were composed of octave-related components than when their components were related by other intervals.

Spectral proximity effects have also been used to produce other striking illusions. Risset (1986) described an illusion produced by a complex tone whose components were spaced at intervals that were slightly larger than an octave. He played this tone first at one speed and then at twice the speed, so that each component of the first tone had a corresponding component of the second tone with a slightly lower frequency. Listeners heard the second tone as lower than the first, indicating that they were invoking proximity between successive spectral components in making their judgments (see also Risset, 1969, 1971, 1978). A similar finding was independently reported by Schroeder (1986), who pointed out that this effect is analogous to certain phenomena in fractal geometry.

In addition to proximity, another grouping principle has been shown to operate here. Teranishi (1982) generated a set of major triads that were composed of octave-related complexes and were generated under a trapezoidal spectral envelope. When a subset of these triads was played in the succession as shown in Figure 26,

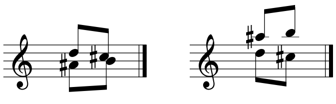


FIGURE 25 Representation in musical notation of the alternative perceptual organizations shown in Figure 24.

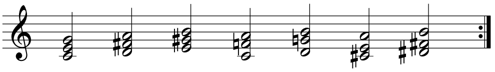


FIGURE 26 Representation of the pattern used to obtain an endlessly ascending scale from a sequence of chords. The tones were octave-related complexes, generated under a trapezoidal spectral envelope. A global pitch movement was here perceived, reflecting perceptual organization by common fate. (Reprinted with permission from Nakajima et al., 1988; data from Teranishi, 1982. ©1988 by The Regents of the University of California.)

listeners obtained the impression of an endlessly ascending scale. However, as can be seen by perusal of Figure 26, the most proximal relationships between components of successive tones were not uniformly in the ascending direction. For example, taking the first two chords, the descending line G-F♯ follows proximity more closely than the ascending line G-A. It appears, therefore, that the listeners were basing their relative pitch judgments on an impression of global pitch movement, or "common fate."

A follow-up study by Nakajima, Tsumura, Matsuura, Minami, and Teranishi (1988) also examined perception of successions of major triads that were produced by octave-related complexes. Paired comparison judgments involving such triads showed that whereas some subjects showed a pitch circularity of an octave, others showed a pitch circularity of roughly 1/3 octave. The authors concluded that the subjects were basing their judgments on the perception of global pitch movement, even when the precise melodic intervals between successive components were not preserved (see also Nakajima, Minami, Tsumura, Kunisaki, Ohnishi, & Teranishi, 1991).

In a related study, Allik, Dzhafarov, Houtsma, Ross, and Versfeld (1989) generated random chord sequences that were composed of octave-related complexes. When such chords were juxtaposed in time in such a way that a sufficient number of successive components were related by proximity in the same direction, a global pitch movement in this direction was heard.

In general, composers have frequently made use of a perceptual effect of common fate, by creating sequences of chords whose components moved in the same direction and by similar degrees while the precise intervals between successive tones were varied. An example is given in Figure 27, which shows a passage from Debussy's prelude *Le vent dans la plaine.* Here, the grouping of successive pitches by proximity alone should cause the listener to hear a number of repeating pitches, together with the falling-rising sequence D♭-C-D♭-C; however, these percepts are discarded in favor of an impression of a descending series of chords.

VII. CONCLUSION: RELATIONSHIPS TO MUSIC THEORY AND PRACTICE

In this chapter, we have explored a number of findings that elucidate the way our auditory system groups the components of music into perceptual configura-

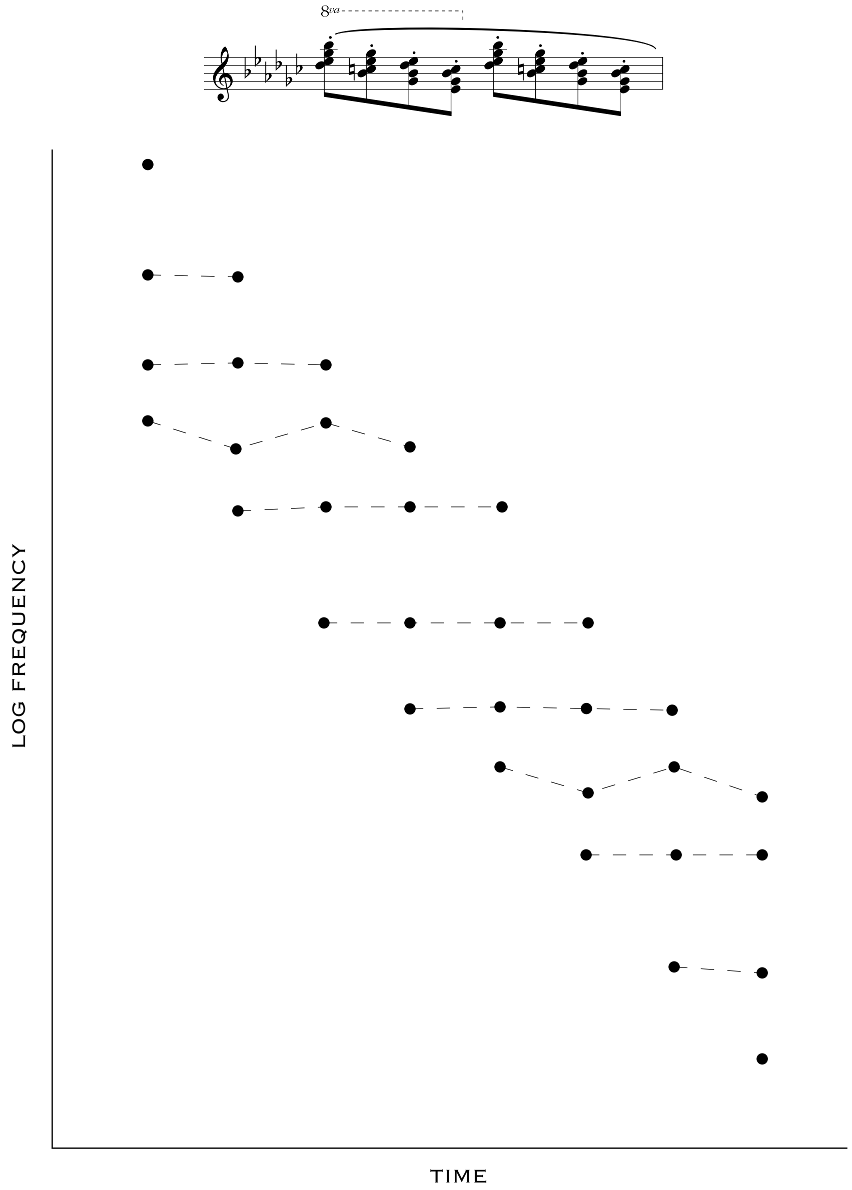


FIGURE 27 A passage from Debussy's prelude *Le vent dans la plaine.* The listener perceives this passage globally, as a downward pitch movement, in accordance with the principle of common fate.

tions. Beyond their interest to psychologists, these findings have implications for music theory and practice (Deutsch, 1984; Huron, 1991a, 1991b, 1993; Wright & Bregman, 1987).

In treatises on music theory, we encounter a number of rules that instruct the student in the art of composition. Among these are the “law of stepwise progression,” which states that melodic progression should be by steps (i.e., a half step or a whole step) rather than by skips (i.e., more than a whole step) because stepwise progression is considered to be in some way “stronger” or “more binding.” Another law prohibits the crossing of voices in counterpoint. What is left unspecified is why these precepts should be obeyed: It is assumed that the reader will either follow them uncritically or recognize their validity by introspection. The findings that we have been reviewing provide such laws with rational bases by demonstrating the perceptual effects that occur when they are violated. This in turn enables musicians to make more informed compositional decisions.

As a related point, with the advent of computer music, the composer is no longer bound by the constraints of natural instruments, but is instead faced with an infinity of compositional possibilities. As a result, the understanding of certain basic perceptual phenomena has become of critical importance, such as the factors that lead us to fuse together the components of a spectrum so as to obtain a unitary sound image, and the factors that lead us to separate out components so as to obtain multiple sound images. Such knowledge is a necessary first step in the creation of new musical timbres. For similar reasons, we need to understand the principles by which we form simultaneous and successive linkages between different sounds, so that listeners will perceive musical patterns as intended by the composer.

Finally, the illusions we have been exploring show that listeners do not necessarily perceive music in accordance with the written score, or as might be imagined from reading a score. Musical rules that have evolved through centuries of practical experience provide some ways of protecting the composer from generating music that could be seriously misperceived. However, with our new compositional freedom there has emerged a particularly strong need to understand how music as it is notated and performed maps on to music as it is heard. The findings reviewed here have brought us closer to realizing this goal, although much more remains to be learned.

REFERENCES

Allik, J., Dzhafarov, E. N., Houtsma, A. J. M., Ross, J., & Versfeld, N. J. (1989). Pitch motion with random chord sequences. *Perception & Psychophysics , 46,* 513–527.

Assmann, P. F., & Summerfield, A. Q. (1990). Modelling the perception of concurrent vowels: Vowels with different fundamental frequencies. *Journal of the Acoustical Society of America, 88,* 680–697.

Beerends, J. G., & Houtsma, A. J. M. (1989). Pitch identification of simultaneous dichotic two-tone complexes. *Journal of the Acoustical Society of America, 85,* 813–819.

Berlioz, H. (1948). *Treatise on instrumentation* (R. Strauss, Ed.). New York: Kalmus.

Braus, I. (1995). Retracing one’s steps: An overview of pitch circularity and Shepard tones in European music, 1550–1990. *Music Perception, 12,* 323–351.

Bregman, A. S. (1978). The formation of auditory streams. In J. Requin (Ed.), *Attention and performance* (Vol. VII, pp. 63–76). Hillsdale, NJ: Erlbaum.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound.* Cambridge, MA: MIT Press.

Bregman, A. S., Abramson, J., Doehring, P., & Darwin, C. J. (1985). Spectral integration based on common amplitude modulation. *Perception & Psychophysics, 37,* 483–493.

Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology, 89,* 244–249.

Bregman, A. S., & Dannenbring, G. L. (1977). Auditory continuity and amplitude edges. *Canadian Journal of Psychology, 31,* 151–159.

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

Bregman, A. S., Liao, C., & Levitan, R. (1990). Auditory grouping based on fundamental frequency and formant peak frequency. *Canadian Journal of Psychology, 44,* 400–413.

Bregman, A. S., & Pinker, S. (1978). Auditory streaming and the building of timbre. *Canadian Journal of Psychology, 32,* 20–31.

Bregman, A. S., & Rudnicky. A. (1975). 1. Auditory segregation: Stream or streams? *Journal of Experimental Psychology: Human Perception and Performance, 1,* 263–267.

Broadbent, D. E., & Ladefoged, P. (1957). On the fusion of sounds reaching different sense organs. *Journal of the Acoustical Society of America, 29,* 708–710.

Brokx, J. P. L., & Nootebohm, S. G. (1982). Intonation and the perceptual separation of simultaneous voices. *Journal of Phonetics, 10,* 23–36.

Burns, E. (1981). Circularity in relative pitch judgments for inharmonic complex tones: The Shepard demonstration revisited, again. *Perception & Psychophysics, 30,* 467–472.

Butler, D. (1979a). A further study of melodic channeling. *Perception & Psychophysics, 25,* 264–268.

Butler, D. (1979b). *Melodic channeling in a musical environment.* Paper presented at the Research Symposium on the Psychology and Acoustics of Music, Kansas.

Cardozo, B. L., & Van Noorden, L. P. A. S. (1968). Imperfect periodicity in the bowed string. *IPO Program Reports, 3,* 13–15.

Carlson, S. (1996). Dissecting the brain with sound. *Scientific American, 275*(6), 112–115.

Carlyon, R. P. (1991). Discriminating between coherent and incoherent frequency modulation of complex tones. *Journal of the Acoustical Society of America, 89,* 329–340.

Carlyon, R. P. (1992). The psychophysics of concurrent sound segregation. *Philosophical Transactions of the Royal Society of London, Series B,* 336, 347–355.

Charbonneau, G., & Risset, J. C. (1973). Circularité de jugements de hauteur sonore. *Comptes Rendus de l'Academie des Sciences, Serie B, 277,* 623.

Chowning, J. M. (1980). Computer synthesis of the singing voice. In *Sound generation in winds, strings, computers* (pp. 4–13). Stockholm: Royal Swedish Academy of Music Stockholm, Publ. No. 29.

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

Cutting, J. E. (1976). Auditory and linguistic processes in speech perception: Inferences from six fusions in dichotic listening. *Psychological Review, 83,* 114–140.

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

Dannenbring, G. L., & Bregman, A. S. (1976). Stream segregation and the illusion of overlap. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 544–555.

Dannenbring, G. L., & Bregman, A. S. (1978). Streaming vs. fusion of sinusoidal components of complex tones. *Perception & Psychophysics, 24,* 369–376.

Darwin, C. J. (1981). Perceptual grouping of speech components differing in fundamental frequency and onset-time. *Quarterly Journal of Experimental Psychology, 33A,* 185–207.

Darwin, C. J. (1984). Perceiving vowels in the presence of another sound: Constraints on formant perception. *Journal of the Acoustical Society of America, 76,* 1636–1647.

Darwin, C. J., & Carlyon, R. P. (1995). Auditory grouping. In B. C. J. Moore (Ed.), *Hearing* (pp. 387–424). San Diego: Academic Press.

Darwin, C. J., & Ciocca, V. (1992). Grouping in pitch perception: Effects of onset asynchrony and ear of presentation of a mistuned component. *Journal of the Acoustical Society of America , 91,* 3381–3390.

Darwin, C. J., & Gardner, R. B. (1986). Mistuning a harmonic of a vowel: Grouping and phase effects on vowel quality. *Journal of the Acoustical Society of America, 79,* 838–845.

Darwin, C. J., & Sutherland, N. S. (1984). Grouping frequency components of vowels: When is a harmonic not a harmonic? *Quarterly Journal of Experimental Psychology, 36A,* 193–208.

De Boer, E. (1976). On the 'residue' and auditory pitch perception. In W. D. Keidel & W. D. Neff (Eds.), *Handbook of sensory physiology* (Vol. 5, Part 3, pp. 479–583). New York: Springer-Verlag.

Deutsch, D. (1974). An auditory illusion. *Nature, 251,* 307–309.

Deutsch, D. (1975a). Musical illusions. *Scientific American, 233*(4), 92–l04.

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

Deutsch, D. (1978). Lateralization by frequency for repeating sequences of dichotic 400-Hz and 800-Hz tones. *Journal of the Acoustical Society of America, 63,* 184–186.

Deutsch, D. (1979). Binaural integration of melodic patterns. *Perception & Psychophysics, 25,* 399–405.

Deutsch, D. (1980a). The octave illusion and the what-where connection. In R. S. Nickerson & R. W. Pew (Eds.), *Attention and performance* (Vol. VIII, pp. 575–594). Hillsdale, NJ: Erlbaum.

Deutsch, D. (1980b). The processing of structured and unstructured tonal sequences. *Perception & Psychophysics, 28,* 381–389.

Deutsch, D. (1981). The octave illusion and auditory perceptual integration. In J. V. Tobias & E. D. Schubert (Eds.), *Hearing research and theory* (Vol. I, pp. 99–142). New York: Academic Press.

Deutsch, D. (1983a). The octave illusion in relation to handedness and familial handedness background. *Neuropsychologia, 21,* 289–293.

Deutsch, D. (1983b). Auditory illusions, handedness, and the spatial environment. *Journal of the Audio Engineering Society, 31,* 607–620.

Deutsch, D. (1984). Psychology and music. In M. H. Bornstein (Ed.), *Psychology and its allied disciplines* (pp. 155–193). Hillsdale, NJ: Erlbaum.

Deutsch, D. (1985). Dichotic listening to melodic patterns, and its relationship to hemispheric specialization of function. *Music Perception, 3,* 1–28.

Deutsch, D. (1987). Illusions for stereo headphones. *Audio Magazine, 71*(3), 36–48.

Deutsch, D. (1988). Lateralization and sequential relationships in the octave illusion. *Journal of the Acoustical Society of America, 83,* 365–368.

Deutsch, D. (1991). Pitch proximity in the grouping of simultaneous tones. *Music Perception, 9,* 185–198.

Deutsch, D. (1995). *Musical illusions and paradoxes* [CD]. Philomel Records, P. O. Box 12189, La Jolla, CA 92039-2189.

Deutsch, D. (1996). The perception of auditory patterns. In W. Prinz & B. Bridgeman (Eds.), *Handbook of perception and action* (Vol. 1, pp. 253–296). San Diego, CA: Academic Press.

Deutsch, D. (1998). *A parallel processing solution to the binding problem, with particular reference to hearing.* Manuscript in preparation.

Deutsch, D., & Roll, P. L. (1976). Separate 'what' and 'where' decision mechanisms in processing a dichotic tonal sequence. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 23–29.

Divenyi, P. L., & Hirsh, I. J. (1972). Discrimination of the silent gap in two-tone sequences of different frequencies. *Journal of the Acoustical Society of America, 52,* 166S.

Divenyi, P. L., & Hirsh, I. J. (1974). Identification of temporal order in three-tone sequences. *Journal of the Acoustical Society of America, 56,* 144–151.

Dowling, W. J. (1973a). The perception of interleaved melodies. *Cognitive Psychology, 5,* 322–337.

Dowling, W. J. (1973b). Rhythmic groups and subjective chunks in memory for melodies. *Perception & Psychophysics, 4,* 37–40.

Edelman, G. M., Gall, W. E., & Cowan, W. M. (1988). *Auditory function: Neurobiological bases of hearing.* New York: Wiley.

Erickson, R. (1975). *Sound structure in music.* Berkeley: University of California Press.

Ettlinger, G., Jackson, C. V., & Zangwill, O. L. (1956). Cerebral dominance in sinistrals. *Brain, 79,* 569–588.

Fitzgibbon, P. J., Pollatsek, A., & Thomas, I. B. (1974). Detection of temporal gaps within and between perceptual tonal groups. *Perception & Psychophysics, 16,* 522–528.

Flanagan, J. L. (1972). *Speech, analysis, synthesis, and perception.* Berlin: Springer.

Gardner, R. B., & Darwin, C. J. (1986). Grouping of vowel harmonics by frequency modulation: Absence of effects on phonemic categorization. *Perception & Psychophysics, 40,* 183–187.

Gardner, R. B., Gaskill, S. A., & Darwin, C. J. (1989). Perceptual grouping of formants with static and dynamic differences in fundamental frequency. *Journal of the Acoustical Society of America, 85,* 1329–1337.

Gregory, A. L. (1994). Timbre and auditory streaming. *Music Perception, 12,* 161–174.

Gregory, R. L. (1970). *The intelligent eye.* New York: McGraw-Hill.

Grey, J. M., & Moorer, J. A. (1977). Perceptual evaluation of synthesized musical instrument tones. *Journal of the Acoustical Society of America, 62,* 454–462.

Haas, H. (1951). Über den einfluss eines Einfachechos auf die Hörsamkeit von Sprache. *Acustica, 1,* 49–52.

Handel, S. (1973). Temporal segmentation of repeating auditory patterns. *Journal of Experimental Psychology, 101,* 46–54.

Hartmann, W. M., & Johnson, D. (1991). Stream segregation and peripheral channeling. *Music Perception, 9,* 155–184.

Heise, G. A., & Miller, G. A. (1951). An experimental study of auditory patterns. *American Journal of Psychology, 64,* 68–77.

Helmholtz, H. von (1925). *Helmholtz's physiological optics* (Translated from the 3rd German ed., 1909–1911 by J. P. C. Southall, Ed.). Rochester, NY: Optical Society of America.

Helmholtz, H. von (1954). *On the sensations of tone as a physiological basis for the theory of music* (2nd English ed.). New York: Dover. (Original work published 1859)

Hill, N. J., & Darwin, C. J. (1993). Effects of onset asynchrony and of mistuning on the lateralization of a pure tone embedded in a harmonic complex. *Journal of the Acoustical Society of America, 93,* 2307–2308.

Hochberg, J. (1974). Organization and the Gestalt Tradition. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception* (Vol. I, pp. 180–211). New York: Academic Press.

Hukin, R. W. & Darwin, C. J. (1995a). Comparison of the effect of onset asynchrony on auditory grouping in pitch matching and vowel identification. *Perception & Psychophysics, 57,* 191–196.

Hukin, R. W. & Darwin, C. J. (1995b). Effects of contralateral presentation and of interaural time differences in segregating a harmonic from a vowel. *Journal of the Acoustical Society of America, 98,* 1380–1386.

Huron, D. (1991a). The avoidance of part-crossing in polyphonic music: Perceptual evidence and musical practice. *Music Perception, 9,* 93–104.

Huron, D. (1991b). Tonal consonance versus tonal fusion in polyphonic sonorities. *Music Perception, 9,* 135–154.

Huron, D. (1993). Note-onset asynchrony in J. S. Bach's two-part inventions. *Music Perception, 10,* 435–444.

Iverson, P. (1995). Auditory stream segregation by musical timbre: Effects of static and dynamic acoustic attributes. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 751–763.

Jacobs, L., Feldman, M., Diamond, S. P., & Bender, M. B. (1973). Palinacousis: Persistent or recurring auditory sensations. *Cortex, 9,* 275–287.

Judd, T. (1979). Comments on Deutsch's musical scale illusion. *Perception & Psychophysics, 26,* 85–92.

Kubovy, M. (1976). *The sound of silence: A new pitch segregation phenomenon.* Paper presented at the meeting of the Psychonomic Society, St. Louis.

Kubovy, M. (1981). Concurrent pitch segregation and the theory of indispensable attributes. In M. Kubovy & J. Pomerantz (Eds.), *Perceptual organization* (pp. 55–98). Hillsdale, NJ: Erlbaum.

Kubovy, M., Cutting, J. E., & McGuire, R. M. (1974). Hearing with the third ear: Dichotic perception of a melody without monaural familiarity cues. *Science, 186,* 272–274.

Kubovy, M., & Howard, F. P. (1976). Persistence of a pitch-segregating echoic memory. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 531–537.

Kubovy, M., & Pomerantz, J. (Eds.). (1981). *Perceptual organization.* Hillsdale, NJ: Erlbaum.

Lieberman, P. (1961). Perturbations in vocal pitch. *Journal of the Acoustical Society of America, 33,* 597–603.

Luria, A. B. (1969). *Traumatic aphasia.* Mouton: The Hague.

Machlis, J. (1977). *The enjoyment of music* (4th ed.). New York: Norton.

MacIntyre, M. E., Schumacher, R. T., & Woodhouse, J. (1981). Aperiodicity in bowed string motion. *Acustica, 49,* 13–32.

MacIntyre, M. E., Schumacher, R. T., & Woodhouse, J. (1982). Aperiodicity in bowed string motion: On the differential-slipping mechanism. *Acustica, 50,* 294–295.

Marin, C. M. H., & McAdams, S. (1991). Segregation of concurrent sounds: II. Effects of spectral envelope tracing, frequency modulation coherence, and frequency modulation width. *Journal of the Acoustical Society of America, 89,* 341–351.

Mathews, M. V., & Pierce, J. R. (1980). Harmony and nonharmonic partials. *Journal of the Acoustical Society of America, 68,* 1252–1257.

McAdams, S. (1984). The auditory image: A metaphor for musical and psychological research on auditory organization. In W. R. Crozier & A. J. Chapman (Eds.), *Cognitive processes in the perception of art* (pp. 298–324). Amsterdam: North-Holland.

McAdams, S. (1989). Segregation of concurrent sounds: I. Effects of frequency modulation coherence. *Journal of the Acoustical Society of America, 86,* 2148–2159.

McNabb, M. M. (1981). Dreamsong: the composition. *Computer Music Journal, 5,* 36–53.

McNally, K. A., & Handel, S. (1977). Effect of element composition on streaming and the ordering of repeating sequences. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 451–460.

Miller, G. A., & Heise, G. A. (1950). The trill threshold. *Journal of the Acoustical Society of America, 22,* 637–638.

Miller, G. A., & Licklider, J. C. R. (1950). The intelligibility of interrupted speech. *Journal of the Acoustical Society of America, 22,* 167–173.

Moore, B. C. J., Glasberg, B. R., & Peters, R. W. (1985). Relative dominance of individual partials in determining the pitch of complex tones. *Journal of the Acoustical Society of America, 77,* 1853–1860.

Moore, B. C. J., Glasberg, B. R., & Peters, R. W. (1986). Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *Journal of the Acoustical Society of America, 80,* 479–483.

Nakajima , Y., Minami, H., Tsumura, T., Kunisaki, H., Ohnishi, S., & Teranishi, R. (1991). Dynamic pitch perception for complex tones of periodic spectral patterns. *Music Perception, 8,* 291–314.

Nakajima, Y., Tsumura, T., Matsuura, S., Minami, H., & Teranishi, R. (1988). Dynamic pitch perception for complex tones derived from major triads. *Music Perception, 6,* 1–20.

Neff, D. L., Jesteadt, W., & Brown, E. L. (1982). The relation between gap discrimination and auditory stream segregation. *Perception & Psychophysics, 31,* 493–501.

Nickerson, R. S., & Freeman, B. (1974). Discrimination of the order of the components of repeating tent sequences: Effects of frequency separation and extensive practice. *Perception & Psychophys–ics, 16,* 471–477.

Penfield, W., & Perot, P. (1963). The brain's record of auditory and visual experience. *Brain, 86,* 595–696.

Pollack, I. (1978). Decoupling of auditory pitch and stimulus frequency: The Shepard demonstration revisited. *Journal of the Acoustical Society of America, 63,* 202–206.

Povel, D. J., & Okkerman, H. (1981). Accents in equitone sequences. *Perception & Psychophysics, 30,* 565–572.

Rasch, R. A. (1978). The perception of simultaneous notes such as in polyphonic music. *Acustica, 40,* 22–33.

Rasch, R. A. (1988). Timing and synchronization in ensemble performance. In J. A. Sloboda (Ed.), *Generative processes in music: The psychology of performance, improvization, and composition* (pp. 71–90). Oxford: Oxford University Press.

Risset, J. C. (1969). Pitch control and pitch paradoxes demonstrated with computer-synthesized sounds. *Journal of the Acoustical Society of America, 46,* 88(A).

Risset, J. C. (1971). Paradoxes de hauteur: Le concept de hauteur sonore n'est pas le meme pour tout le monde. *Proceedings of the Seventh International Congress on Acoustics, Budapest, S10,* 613–616.

Risset, J. C. (1978). *Paradoxes de hauteur* (with sound examples), IRCAM Rep. 10, Paris.

Risset, J. C. (1986). Pitch and rhythm paradoxes: Comments on "Auditory paradox based on fractal waveform" [*J. Acoust. Soc. Am., 79,* 186–189 (1986)]. *Journal of the Acoustical Society of America, 80,* 961–962.

Rock, I. (1986). The description and analysis of object and event perception. In K. Boff, L. Kaufman, & J. Thomas (Eds.), *Handbook of perception and human performance* (pp. 1–77). New York: Wiley.

Sasaki, T. (1980). Sound restoration and temporal localization of noise in speech and music sounds. *Tohuku Psychologica Folia, 39,* 79–88.

Scheffers, M. T. M. (1983). *Sifting vowels: Auditory pitch analysis and sound segregation.* Unpublished doctoral thesis, Groningen University, The Netherlands.

Schouten, J. F. (1962). On the perception of sound and speech: Subjective time analysis. *Fourth International Congress on Acoustics, Copenhagen Congress Report II,* 201–203.

Schroeder, M. R. (1986). Auditory paradox based on fractal waveform. *Journal of the Acoustical Society of America, 79,* 186–189.

Shepard, R. N. (1964). Circularity in judgments of relative pitch. *Journal of the Acoustical Society of America, 36,* 2345–2353.

Singh, P. (1987). Perceptual organization of complex tone sequences: A tradeoff between pitch and timbre? *Journal of the Acoustical Society of America, 82,* 886–899.

Sloboda, J. A. (1985). *The musical mind.* New York: Clarendon (Oxford University Press).

Smith, J., Hausfield, S. Power, R. P. & Gorta, A. (1982). Ambiguous musical figures and auditory streaming. *Perception & Psychophysics, 32,* 454–464.

Subirana, A. (1958). The prognosis in aphasia in relation to cerebral dominance and handedness. *Brain, 81,* 415–425.

Summerfield, Q. (1992). Roles of harmonicity and coherent frequency modulation in auditory grouping. In M. E. H. Schouten (Ed.), *The auditory processing of speech: From sounds to words* (pp. 157–165). Berlin: Mouton.

Summerfield, Q., & Culling, J. F. (1992). Auditory segregation of competing voices: Absence of effects of FM and AM coherence. *Philosophical Transactions of the Royal Society of London, Series B, 336,* 357–366.

Sutherland, N. S. (1973). Object recognition. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception* (Vol. III, pp. 157–186). New York: Academic Press.

Teranishi, R. (1982). Endlessly ascending/descending chords performable on a piano. *Reports of the Acoustical Society of Japan, H,* 82–68.

Tougas, Y., & Bregman, A. S. (1985). Crossing of auditory streams. *Journal of Experimental Psychology: Human Perception and Performance, 11,* 788–798.

Tougas, Y., & Bregman, A. S. (1990). Auditory streaming and the continuity illusion. *Perception & Psychophysics, 47,* 121–126.

Van Noorden, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences.* Unpublished doctoral dissertation, Technische Hogeschoel Eindhoven, The Netherlands.

Vicario, G. (1960). L'effetto tunnel acustico. *Revista di Psicologia, 54,* 41–52.

Vicario, G. (1973). *Tempo Psicologia ed Eventi.* Florence: C.-E Giunti-G. Barbera.

Vicario, G. (1982). Some observations in the auditory field. In J. Beck (Ed.), *Organization and representation in perception* (pp. 269–283). Hillsdale, NJ: Erlbaum.

Wallach, H., Newman, E. B., & Rosenzweig, M. R. (1949). The precedence effect in sound localization. *American Journal of Psychology, 62,* 315–336.

Warren, R. M. (1983). Auditory illusions and their relation to mechanisms normally enhancing accuracy of perception. *Journal of the Audio Engineering Society, 31,* 623–629.

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

Warren, R. M., & Byrnes, D. L. (1975). Temporal discrimination of recycled tonal sequences: Pattern matching and naming of order by untrained listeners. *Journal of the Acoustical Society of America, 18,* 273–280.

Warren, R. M., Obusek, C. J., & Ackroff, J. M. (1972). Auditory induction: Perceptual synthesis of absent sounds. *Science, 176,* 1149–1151.

Warren, R. M., Obusek, C. J., Farmer, R. M., & Warren, R. P. (1969). Auditory sequence: Confusions of patterns other than speech or music. *Science, 164,* 586–587.

Wertheimer, M. (1923). Untersuchung zur Lehre von der Gestalt II. *Psychologische Forschung, 4,* 30l–350.

Wessel, D. L. (1979). Timbre space as a musical control structure. *Computer Music Journal, 3,* 45–52.

Williams, K. N., & Perrott, D. R. (1972). Temporal resolution of tonal pulses. *Journal of the Acoustical Society of America, 51,* 644–647.

Wright, J. K., & Bregman, A. S. (1987). Auditory stream segregation and the control of dissonance in polyphonic music. *Contemporary Music Review, 2,* 63–92.

Zureck, P. M. (1987). The precedence effect. In W. A. Yost & G. Gourevitch (Eds.), *Directional hearing* (pp. 85–106). New York: Springer-Verlag.