6 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—a task often referred to as auditory scene analysis (Bregman, 1990). 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 light.

Several issues are considered here. First, given that our auditory system is presented with a set of low-level elements, we can explore the ways in which these are combined so as to form separate groupings. If all low-level 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 low-level elements and inhibit us from forming linkages between others.

In examining such linkages, we can 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 according to some rule based on frequency, amplitude, timing, 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; however, the conditions that determine which attribute is followed are complex ones.

Second, assuming that organization takes place on the basis of some dimension such as pitch, 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 (Wertheimer, 1923). One is proximity: closer elements are grouped together in preference to those that are 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,
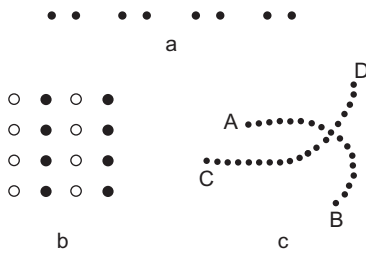
Figure 1 Illustrations of the Gestalt principles of proximity (a), similarity (b), and good continuation (c).

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 is reasonable to assume that grouping in conformity with such principles enables us to interpret our environment most effectively. In the case of vision, elements that are close 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 emanated from the same source, and the sudden addition of a new component may signal the emergence of a new source.

As a related question, we can ask whether the perceptual grouping of elements such as frequency and spatial location results from the action of a single, overarching decision mechanism or from multiple decision mechanisms, each with its own grouping criteria. As will be described, the evidence shows that grouping decisions are not made by a single, internally coherent, system, but rather by a number of different subsystems, which at some stage act independently of each other, and can arrive at inconsistent conclusions. For example, the sound elements that are assigned to different sources so as to determine perceived pitch can differ from those that are assigned to determine perceived timbre, loudness, and location. From such findings, we must conclude that perceptual organization in music involves a process in which elements are first grouped together in various ways 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. Resulting from this two-stage process, the different attribute values are sometimes combined incorrectly, so that illusory conjunctions result (cf. Deutsch, 1975a, 1975b, 1981, 2004; Deutsch, Hamaoui, & Henthorn, 2007; Deutsch & Roll, 1976). Auditory scene analysis cannot, therefore, be regarded as the product of a single, internally coherent system, but rather as the product of multiple systems whose

outputs are sometimes inconsistent with each other (see also Hukin & Darwin, 1995a; Darwin & Carlyon, 1995).

As a further issue, the grouping of sound elements in music involves not only the creation of low-level features such as tones, but also the conjunction of these features at higher levels so as to form intervals, chords, durational relationships, and rhythmic patterns, as well as phrases and phrase groups (see also Chapter 7). As we shall find, auditory grouping is the function of a highly elaborate and multifaceted system, whose complexities are becoming increasingly apparent.

II.   Fusion and Separation of Spectral Components

In this section, we consider the relationships between the components of a musical 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 explore two types of relationship. The first is harmonicity. Natural sustained sounds, such as are produced by musical instruments and the human voice, are made up of components that stand in harmonic, or near-harmonic, relation; that is, their frequencies are integer, or near-integer multiples of the fundamental frequency. 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.

A second relationship that we explore is onset synchronicity. When components of a sound complex begin at the same time, they are likely to have originated from the same source; conversely, when they begin at different times, they are likely to have originated from different sources. As an associated issue, we explore 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.   *Effects of Harmonicity*

Musical instrument tones provide us with many examples of perceptual grouping by harmonicity. String and wind 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 produce tones whose partials are nonharmonic, and these give rise to diffuse pitch impressions. The effect of harmonicity has been explored in numerous experiments using synthesized tones (Carlyon, 2004; Carlyon & Gockel, 2007; Darwin, 2005a; Darwin & Carlyon, 1995).

How far can a single component of a complex tone deviate from harmonicity and still be grouped with the other components to determine perceived pitch? Moore, Glasberg, and Peters (1985) had subjects judge the pitches of harmonic complex tones, and they examined the effects of mistuning one of the harmonics to various extents. When the harmonic 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 this component gradually decreased, and at a mistuning of 8%, the component made virtually no contribution to the pitch of the complex.

The effect of a mistuned harmonic can, however, be made to vary by changing its relationship to the remainder of the complex (Darwin, 2005a). In one experiment, subjects were presented with a harmonic complex tone that contained one mistuned harmonic. When this harmonic was synchronous with the others, the perceived pitch of the complex was slightly shifted. However, when the mistuned harmonic entered sufficiently before the others, it no longer contributed to the pitch of the complex (see also Darwin & Ciocca, 1992; Ciocca & Darwin, 1999). Furthermore, when the complex was preceded by a sequence of four tones at the same frequency as the mistuned harmonic, the pitch shift again disappeared, indicating that the mistuned harmonic had formed a separate stream with the preceding tones. Also, when all the harmonics in the complex (including the mistuned one) were given a common vibrato, larger amounts of mistuning were needed to remove the contribution of the mistuned harmonic to the pitch of the complex, indicating that the common vibrato had caused the harmonics to be bound together more effectively (Darwin, Ciocca, & Sandell, 1994).

Huron (1991b, 2001) has related 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 works by J. S. Bach, Huron showed that harmonic intervals were avoided in proportion to the strength with which they promoted tonal fusion. He concluded that Bach had used this strategy in order to optimize the salience of the individual voices in these compositions.

Composers have also focused on the creation of perceptual fusion of simultaneous tones so as to give rise to unique timbres. For example, at the opening of Schubert's Unfinished Symphony the oboe and clarinet play in unison, with the result (assuming the performers play in strict synchrony) that listeners hear a fused sound with a unique timbre that appears to be emanating from a single instrument. More recently, composers have frequently experimented with sounds produced by

several instruments playing simultaneously, such that the individual instruments lost their perceptual identities and together produced a single sound impression. For example, Debussy and Ravel made extensive use of chords that approached timbres. Later composers such as Schoenberg, Stravinsky, Webern, and Varese often employed highly individualized structures, which Varese termed "sound masses" (Erickson, 1975), and here tone combinations that stood in simple harmonic relation were particularly useful.

To return to laboratory experiments, a number of studies have found that simultaneous speech patterns could be more easily separated out perceptually when they were built on different fundamentals—in general, the amount of useful perceptual separation reached its maximum when the fundamentals differed by roughly one to three semitones (Assmann & Summerfeld, 1990; 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 (Broadbent & Ladefoged, 1957). The relationship of these findings to musical tones was explored by Sandell and Darwin (1996), who generated simultaneous tone pairs taken from five different orchestral instruments (flute, B♭ clarinet, cor anglais, French horn, and viola). The authors found that subjects were better able to separate and identify the tones when their pitches differed by a semitone.

The number of sources perceived by the listener provides a further measure of grouping. For example, Moore, Glasberg, and Peters (1986) found that when a single component of a harmonic complex was slightly mistuned from the others, it was heard as standing apart from them. Interestingly, less mistuning is required to produce the impression of multiple sources than to produce other effects. For example, a slightly mistuned harmonic in a sound 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, Gaskill, & Darwin, 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 employ the same criteria. This issue is discussed further in Section VI, where it is shown that, in listening to simultaneous sequences of tones, separate, and sometimes inconsistent, decision mechanisms are employed to determine the perceived pitch, location, loudness, and timbre of each tone, so that illusory conjunctions result.

B.   *Effects of Onset Synchronicity*

So far we have been considering sounds whose components begin and end at the same time, and we have been exploring the spectral relationships 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 so that its components enter at different times. For example, we can take a series that

is built on a 200-Hz fundamental. Suppose we begin with the 200-Hz component sounding alone, then 1 s later add the 400-Hz component, then 1 s 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 forming a distinct entity, and then it gradually fades from perception, so that finally only a pitch that corresponds to the fundamental is perceived.

Even a transient change in the amplitude of a component can enhance its perceptual salience. If a particular harmonic of a complex tone is alternately omitted and restored, this can cause it to stand out as a pure tone, separately audible from the remainder of the complex, and it can even be heard for a short time after being turned back on (Hartmann & Goupell, 2006; Houtsma, Rossing, & Wagenaars, 1987).

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 harmonic made less of a contribution to perceived pitch when it led the others by more than 80 ms, and it made no contribution when it led the others by 300 ms. Later, Ciocca and Darwin (1999) observed that a mistuned harmonic made a larger contribution to the pitch of a target sound when it occurred following the onset of the target than when it preceded its onset.

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

Interestingly, Darwin and colleagues found that the amount of onset asynchrony that was needed to alter the contribution of a harmonic 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 subject was engaged; again arguing, as did Darwin and Carlyon (1995), that such disparities reflect the operation of multiple decision mechanisms in the grouping process that can act independently of each other.

Onset asynchrony has been shown to have higher level effects also. In one experiment, Bregman and Pinker (1978) presented subjects 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 the probability increased that one of the tones in the complex would form a melodic stream with the third tone (see also Deutsch, 1979, discussed in Section VI,A).

So far, we have been considering the effects of onset asynchrony on the grouping of components of single complex tones; however, asynchronies also influence the grouping of entire tone complexes. Specifically, when two complex tones are presented together, they are perceptually more distinct when their onsets are asynchronous. Rasch (1978) presented subjects with simultaneous pairs of complex

tones, and found that detection of the higher tone of a pair was strongly affected by onset asynchrony: Each 10 ms of delay of the lower tone was associated with roughly a 10-dB reduction in detection threshold, and at a delay of 30 ms, the threshold for perception of the higher tone was roughly the same as when it was presented alone. Further, when the onsets of the higher and lower tones were synchronous, a single fused sound was heard; yet when onset disparities were introduced, the tones sounded very distinct perceptually.

From these findings, Rasch (1988) later applied these findings to live ensemble performances. He made recordings of three different trio ensembles (string, reed, and recorder) and calculated the onset relationships between tones that were nominally simultaneous. He obtained asynchrony values that ranged from 30 to 50 ms, with a mean asynchrony of 36 ms. Relating these findings to those he had obtained earlier on perception, Rasch concluded that such onset asynchronies enabled the listener to hear the simultaneous tones as distinct from each other. According to this line of reasoning, such asynchronies should not be considered as performance failures, but rather as characteristics that are useful in enabling listeners to hear concurrent voices distinctly.

From these findings, one would expect large amounts of asynchrony to be conducive to the separation of voices in an ensemble. One might therefore hypothesize 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 obtained by Huron (1993, 2001) in an analysis of Bach's 15 two-part inventions. He found (controlling for duration, rhythmic order, and meter) that for 11 of the inventions, no other permutations of the rhythms of the voices would have produced more onset asynchrony than occurred in Bach's actual music. For the remaining 4 inventions, values of asynchrony were still significantly higher than would be expected from chance. Huron concluded that Bach had deliberately produced such onset asynchronies so as to optimize the perceptual salience of the individual voices in these compositions.

C. Auditory Continuity

Auditory continuity is a dramatic effect that can result from temporal disparities within tone complexes. This effect is important to the interpretation of our natural environment, where sound signals are frequently masked by other sounds. In order to maintain a stable representation of the auditory world, our perceptual system needs to restore the masked portions of each sound, by estimating their characteristics based on other sounds that occur before, during, and after the masking sound. The continuity effect is widespread, and has been shown to occur in nonhuman species such as cats (Sugita, 1997), monkeys (Petkov, O'Connor, & Sutter, 2003), and birds (Braaten & Leary, 1999; Seeba & Klump, 2009), as well as in human listeners (Houtgast, 1972; Miller & Licklider, 1950; Vicario, 1960; Warren, Obusek, & Ackroff, 1972).

Consider the visual analogue shown in the upper portion of Figure 2. Line A could, in principle, be viewed in terms of three components: a line to the left

Figure 2 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).
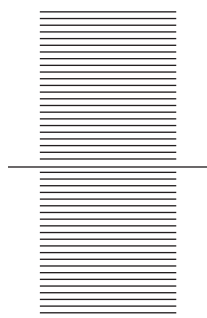
of the rectangle, a line to its right, and a line that forms part of the rectangle itself. However, our visual system instead treats all three components as a single line, which is independent of the remaining parts of the rectangle. Vicario (1982) produced a musical equivalent of this demonstration. 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. Just as line A in Figure 2 is seen as continuing through the rectangle, so a pitch corresponding to A_4 is heard as continuing right through the chord.

The 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 2, 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 increased, the impression of continuity is diminished.

An interesting demonstration of auditory continuity was provided by Dannenbring (1976), who 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 instead heard a disjunct series of rising and falling glides. Visual analogues of these two conditions, and their perceptual consequences, are shown in Figure 3.

Sudden amplitude drops between signals and intervening noise bursts may reduce, or even destroy, continuity effects (Bregman & Dannenbring 1977; Warren et al., 1972); however, this does not necessarily occur. For example, tones produced by plucked instruments are characterized by rapid increases followed by decreases in
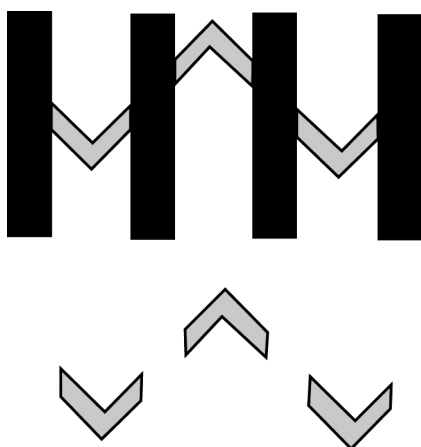
Figure 3 Visual illustration of an auditory continuity effect using gliding tones. Adapted from Bregman (1990), which illustrates an experiment by Dannenbring (1976).

amplitude. In music played by such instruments, when the same tone is rapidly repeated many times, and is periodically omitted and replaced by a different tone, the listener may perceptually generate the missing tone. Many examples of this phenomenon occur in 19th- and 20th-century guitar music, such as Tarrega's *Recuerdos de la Alhambra* (Figure 4), and Barrios's *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 in fact been there, although imperceptibly, throughout the work.

Remijn, Nakajima, and Tanaka (2007) explored auditory continuity across a silent interval from a different perspective. They presented subjects with a pattern consisting of two crossing frequency glides of unequal duration that shared a silent gap of 40 ms or less at the crossing point. The gap was perceived to occur only in the shorter glide, while the longer glide was perceived as continuous.

Both long- and short-term memory can influence the strength and nature of the auditory continuity effect (Vicario, 1973; Warren, 1983). In one experiment, Sasaki (1980) generated melodic patterns in which certain tones were omitted and replaced by loud noise bursts. Under some circumstances, subjects "heard" the 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.

A short-term context effect was demonstrated by Ciocca and Bregman (1987), who presented subjects 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, subjects 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
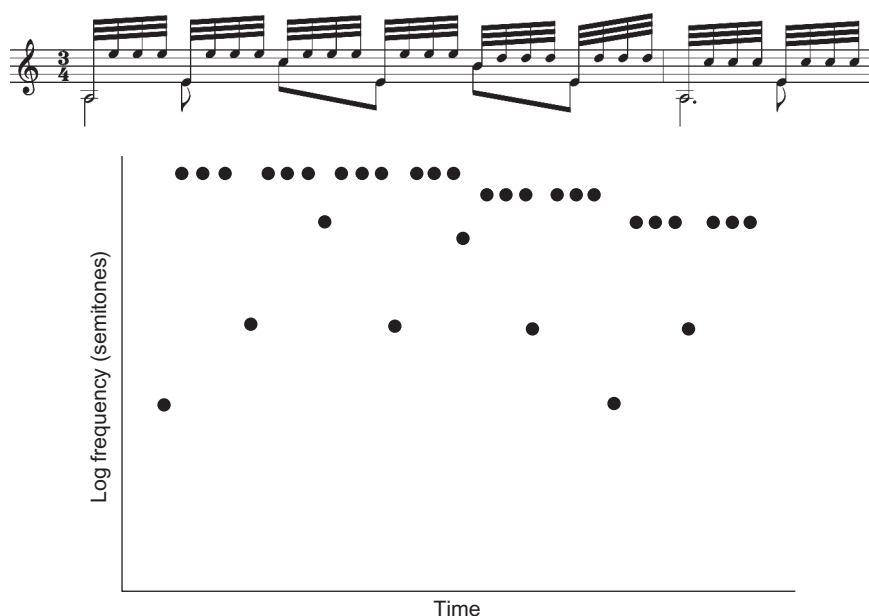
Figure 4 The beginning of *Recuerdos de la Alhambra*, by Tarrega. Although the tones are presented one at a time, two parallel lines are perceived, organized in accordance with pitch proximity.
Adapted from Deutsch (1996).

influenced how crossing pitch patterns were perceived (Deutsch, 1975a, 1975b; Tougas & Bregman, 1985; Van Noorden, 1975; see also Section VI). As expected from these findings, Tougas and Bregman (1990) observed that frequency proximity dominated over trajectory in determining the type of perceptual restoration that was obtained: Subjects 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.

Continuity effects can be influenced by more complex factors. In one experiment, Darwin (2005b) had subjects make judgments on complex tones that alternated with band-pass noise. He found that a quiet complex tone was heard as continuous when all its harmonics fell within the frequency range of the noise band. This impression of continuity was substantially reduced when harmonics were added that were outside the range of the noise; however, it was largely restored when the additional components produced a different fundamental frequency. Darwin concluded that continuity judgments are made on entire simultaneously grouped objects, rather than being determined by local frequency criteria (see also McDermott & Oxenham, 2008).

In other experiments, Riecke, Mendelsohn, Schreiner, and Formisano (2009) demonstrated that continuity illusions can be influenced by preceding sound patterns. Specifically, they found that whether or not the same perceptually ambiguous

glide was heard as continuous could be modulated by the loudness and perceived continuity of preceding glides. These context effects were related less to the spectra of the preceding sounds than to how they had been interpreted by the listener.

The brain mechanisms underlying the continuity illusion have also been explored. Petkov, O'Connor, and Sutter (2007) studied responses in the auditory cortex of macaque monkeys to tones that were interrupted by a loud noise. They found that some neurons responded to discontinuous tones that were interspersed with noise as though the tones were continuous (see also Petkov & Sutter, 2011).

D.   Effects of Vibrato

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. It has been surmised that the auditory system uses such coherent frequency modulation (FM) as a cue for grouping spectral components together, and conversely uses incoherent FM as a cue for separating them out perceptually (Bregman, 1990). Indeed, composers such as Chowning (1980) and McNabb (1981) have produced informal demonstrations that coherent vibrato enhances perceptual fusion when imposed on synthesized singing voices or musical instrument tones. Later Darwin, Ciocca, and Sandell (1994) found that a mistuned harmonic was more likely to contribute to the pitch of a complex tone when it was given a common vibrato with the other harmonics.

The issue with respect to incoherent FM, however, is theoretically complex: Because information concerning FM is severely degraded in reverberant environments, the reliance on incoherent FM as a cue for perceptual separation could cause the listener to separate out components that should instead be grouped together. Furthermore, incoherent FM necessarily causes the frequency relationships between components of a tone to depart from harmonicity. Because the perceptual system already uses such departures as cues for perceptual segregation (as discussed earlier), the utility of incoherent FM as an additional cue is debatable.

The experimental evidence on this issue is also complex. McAdams (1989) explored the effect of vibrato on the perceptual separation of three simultaneous sung vowels which were built on different fundamentals. He found that when target vowels were given a vibrato, this increased their perceptual salience. However, the perceived salience of target vowels was not affected by whether the nontarget vowels were given a vibrato. Other negative findings were obtained by Carlyon (1991, 1992), who found that subjects were insensitive to incoherent vibrato 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.

Such negative findings raise the question of why vibrato can nevertheless enhance a vowel's perceptual salience. McAdams (1984) pointed out that when the harmonics of a vowel are given a vibrato, they also undergo amplitude modulation (AM) that traces the vowel's spectral envelope. In this way, the listener is provided with more

complete information concerning the vowel's identity, and such spectral tracing might therefore be responsible for the enhanced perceptual salience of vowels with vibrato. However, Marin and McAdams (1991) found that although vowels with vibrato were heard more saliently, spectral tracing was not a factor here. As an alternative explanation for the enhanced prominence of frequency-modulated vowels, we may conjecture that neural units involved in the attribution of vowel quality are more strongly activated by frequency-modulated sounds than by unmodulated ones.

E.   Effects of Amplitude Modulation

Because many natural sounds consist of spectral components whose amplitudes rise and fall in synchrony, one might conjecture that coherent AM would be used by the auditory system as a cue for perceptual fusion, and incoherent AM would be used as a cue for perceptual separation. On the other hand, the partials of many musical instrument tones do not rise and fall in synchrony. So the use of AM incoherence as a cue for perceptual separation could cause the listener to erroneously separate out components that should be perceptually fused together.

   The experimental evidence on this issue is also equivocal. Bregman, Abramson, Doehring, and Darwin (1985) concluded that coherent AM could promote perceptual fusion; however, the modulation rates they used were high, and at slower rates, convincing evidence that coherent AM leads to perceptual fusion has been difficult to obtain (Darwin & Carlyon, 1995).

F.   Effects of Spatial Separation

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 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 subject 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 at the two ears), we would risk separating out components that 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 (see also Section VI).

   Experimental evidence supporting this view has been obtained from studies in which different components of a sound 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 ears in various ways. They found that pitch identification was only weakly affected by the way the partials were

distributed, showing that the perceptual system was treating them as coming from the same source. A related effect was found by Darwin and Ciocca (1992), who observed 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 listeners with the first two formants of a phrase, with one formant delivered to each ear. When the two formants were built on the same fundamental, the listeners were able to identify the speech signal, and they also tended to hear a single voice, and so 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, and found that this difference in ear of input had only a small effect on perceived vowel quality.

Support has also been obtained for the conjecture that other grouping cues, such as harmonicity and onset asynchrony, can influence the perceived spatial origin of a component of a sound complex. For example, Hill and Darwin (1993) found that mistuning a harmonic in a sound complex caused its apparent location to be determined separately from the remainder of the complex. We shall see in Section VI 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 that these factors also influence the perceived spatial origins of the individual tones in a complex.

G.   Effects of Statistical Regularities

Listeners are able to form groupings from repeating spectrotemporal structures that are embedded in changing acoustic input, even in the absence of strong cues such as harmonicity, onset synchronicity, and so on. As a result, we can identify complex sounds that are present in a mixture, provided that they occur repeatedly. McDermott, Wrobleski, and Oxenham (2011) synthesized novel sounds that shared some of the structure of natural sounds, but lacked strong grouping cues. In general, subjects were unable to identify such a sound when it was presented in a single mixture. However, when a series of sound mixtures was presented, each containing the same target sound, subjects were able to identify the target reliably—a significant benefit was even obtained from only two presentations of the target sound mixed in with others.

III.   Larger-Scale Groupings

So far, we have been focusing on situations in which single tone complexes are presented, and we have identified various cues that are used by the listener to sort their

components into groupings. We now turn to the situation in which a sequence of tones is presented instead. Here the auditory system abstracts relationships between tones in a sequence, 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. 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, we obtain 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 in a sequence, we use these as cues for grouping the tones into sub-sequences. A third cue is similarity of sound quality: When different types of instrument play together, we tend to form linkages between tones that are similar in timbre. We also invoke other principles, such as good continuation and common fate, in making grouping decisions. In addition, high-level factors are involved, such as memory, attention, and the processing of information at high levels of abstraction.

We first explore the separation of a sequence of single tones into different groupings. Two issues are considered. The first, which has been the subject of considerable research, concerns the ways in which we group a sequence of tones into separate and parallel streams in a polyphonic texture. The second issue, which is of considerable importance but has been the subject of less research, is how we divide a musical stream into coherent segments that are separated by temporal boundaries (see also Chapter 7). Finally, we explore the grouping of simultaneous sequences of tones.

IV.   Auditory Streaming and Implied Polyphony

A.   Streaming by Pitch Proximity

In general, 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. The separation of rapid sequences into different streams is widespread, and occurs in human infants (Demany, 1982; McAdams & Bertoncini, 1997; Winkler et al., 2003) and in nonhuman species, such as monkeys (Izumi, 2002), birds (MacDougall-Shackleton, Hulse, Gentner, & White 1998), and fish (Fay, 1998).

Auditory streaming is frequently exploited in music using the technique of pseudo-polyphony, or "compound melodic line." Baroque composers such as Bach and Telemann frequently employed this technique, particularly in works for solo recorder or string. In more recent times, the technique has been employed to strik-ing effect by composers of classical and romantic guitar music, in which the notes produced by plucked string occur at very rapid tempo. The passage from Tarrega's *Recuerdos de la Alhambra* shown in Figure 4 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. It can be seen that two separate lines emerge in the visual representation, and these correspond to the two melodic 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 subjects with two pure tones at different frequencies (A and B) such that they alternated at a rate of 10 per second, forming an ABAB pattern (Figure 5). When the pitch difference between the tones was small (about one or two semitones), subjects heard the sequence as a trill. However, when this pitch difference was large, subjects instead heard the sequence as two interrupted and unrelated tones. In a further experiment, Heise and Miller
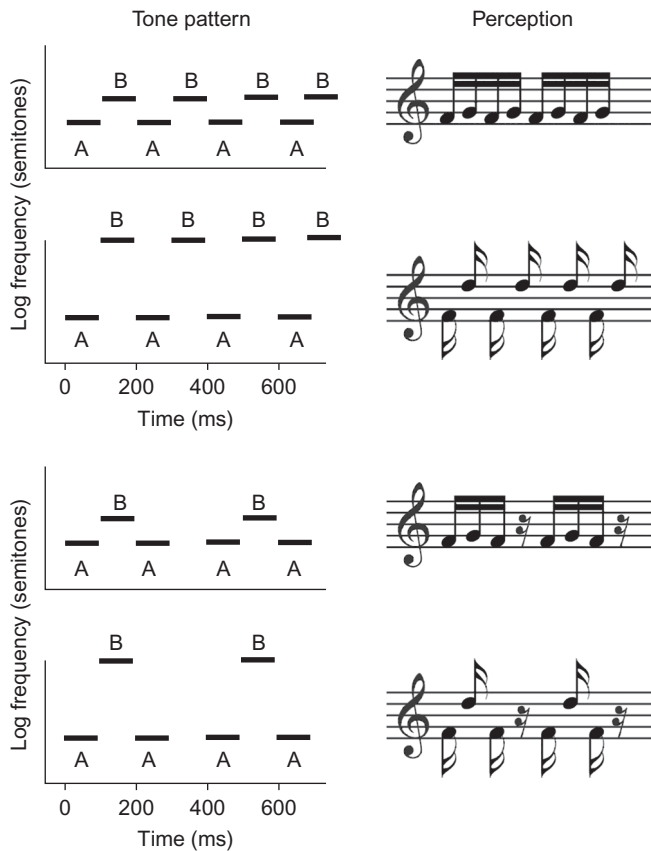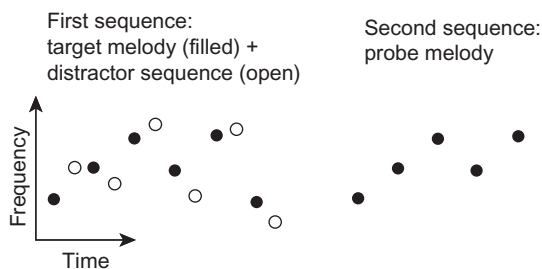

Figure 5 Patterns most frequently employed to study the perceptual segregation of a rapid sequence of tones. The patterns either alternate between the two pitches A and B (forming an ABAB pattern) or they consist of repeating triplets (forming an ABA_ triplet pattern). When the pitch distance between tones A and B is small, a single stream of related tones is perceived. When the pitch distance is large, with continued listening two unrelated pitch streams are perceived, one consisting of the low tones and the other of the high tones.

(1951) employed rapid sequences of tones that were composed of several different pitches. When one of the tones in a sequence differed sufficiently in pitch from the others, it was heard as in isolation from them.

Related phenomena have been demonstrated by Dowling and colleagues (Dowling, 1973b; Dowling, Lung, & Herrbold, 1987). In one paradigm, two well-known melodies were presented at a rapid tempo, such that the tones from each melody occurred in alternation. The melodies were difficult to identify when their pitch ranges overlapped closely, but easy to identify when their pitch ranges differed. In another paradigm, Dowling presented an unfamiliar target melody, followed by a probe melody that was interleaved with a distractor sequence. Subjects judged whether the two melodies were the same or different, and their judgments improved with increasing pitch separation between the probe melody and the distractor tones.

Considering Dowling's second paradigm, Bey and McAdams (2003) studied the effects of interleaving the distractor tones with either the target or the probe melody. In one condition, the target melody was first presented mixed with distractor tones, and this was followed by a probe melody that was presented alone (Figure 6). In a second condition, the target melody was first presented alone, and this was followed by the probe melody mixed with distractors. In the third condition, the target was again first presented alone, but this time in transposed form. In a control condition, both the target and the probe melodies were presented alone. As expected from previous findings, in all conditions, performance improved with increasing frequency separation between the target and the distractors. However, performance was enhanced overall when the target was first presented alone, and

(a) Experimental condition

First sequence:
target melody (filled) +
distractor sequence (open)

Second sequence:
probe melody

Frequency

Time

(b) Control condition

First sequence:
target melody
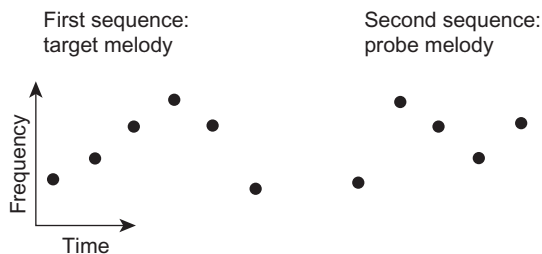
Second sequence:
probe melody

Frequency

Time

Figure 6 Examples of patterns employed to study perception of tone sequences that were interleaved with distractor tones. In the experimental condition depicted here, the target melody was presented interleaved with distractors, and in the control condition, the target and probe melodies were both presented alone. In both conditions, the target melody was followed by a probe melody, and subjects judged whether or not the probe differed from the target. Adapted from Bey and McAdams (2003).

the probe was interleaved with distractors. In this latter condition, the subjects were better able to encode the sequence in memory before judging the mixed sequence. Interestingly, this performance enhancement was not as large when the target was presented in transposed form, indicating that absolute pitch level was also involved in the process.

Van Noorden (1975) carried out a detailed study on the influence of pitch proximity and tempo in the building of perceptual streams. Subjects were presented with sequences consisting of two tones in alternation, and they attempted to hear either a single coherent stream or two separate streams. Two boundaries were determined by these means. The first was the threshold frequency separation as a function of tempo that was needed for the listener to hear a single stream. The second was the threshold frequency separation as a function of tempo when the listener was attempting to hear two streams. As shown in Figure 7, when the subjects were attempting to hear a single stream, decreasing the tempo from 50 to 150 ms per tone increased the range of frequency separation within which one stream could be heard from 4 to 13 semitones. However, when the subjects were instead attempting to hear two streams, 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 one or two streams. So within this region, attention played a role in determining how the sequence was perceived.

The preceding experiments employed either pure tones or harmonic complex tones in which pitch (fundamental frequency) and spectrum were co-varied. We can then ask whether differences in pitch with spectrum held constant, or differences in spectrum with pitch held constant, can alone give rise to streaming. Singh (1987) and Bregman, Liao, and Levitan (1990) explored the streaming of sequences in which tones differed either in spectral envelope, or fundamental frequency, or both; they found that both these factors contributed to stream segregation. Considering spectral region alone, Van Noorden (1975) found that listeners perceptually segregated sequences consisting of a pure tone alternating with a complex tone with the same fundamental frequency, or of two complex tones with
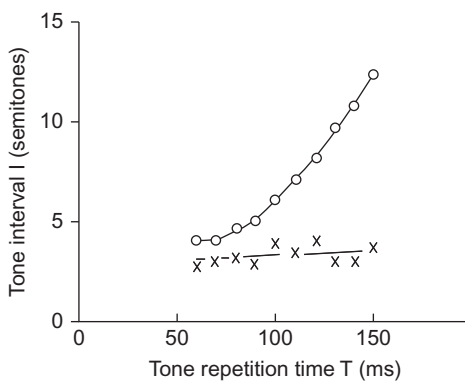

Figure 7 Temporal coherence boundary (o), and fission boundary (x) as a function of the frequency relationship between the alternating tones and presentation rate. Adapted from Van Noorden (1975).

the same fundamental frequency but different harmonics, showing that spectral region here played a role. Considering fundamental frequency alone, Vliegen and Oxenham (1999) employed sequences of tones consisting only of high harmonics, with spectral envelope held constant. Subjects segregated these sequences on the basis of fundamental frequency, and the amount of segregation was comparable to that found for pure tones. However, Vliegen, Moore, and Oxenham (1999) observed that spectral differences also contributed to stream segregation (see also Carlyon & Gockel, 2007; Grimault, Micheyl, Carlyon, Arthaud, & Collett, 2000). The effect of spectrum is an example of streaming by timbre, which is explored in the next section.

B. Streaming by Timbre

The grouping of sounds on the basis of sound quality, or timbre, is an example of the Gestalt 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 8, 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 streaming tendency was demonstrated by Warren, Obusek, Farmer, and Warren (1969). These authors generated a sequence of four unrelated sounds that were presented repeatedly without pause. The sounds, each 200 ms 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 needed to be longer than 500 ms.

Another consequence of streaming 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 that were defined by their spectral energy distribution. When the timbral difference

Figure 8 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.

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. In a related experiment, Cusack and Roberts (2000) employed an interleaved melodies task, and found that target sounds were more easily separated from distractors when they differed in timbre.

Since timbre is multidimensional in nature (see McAdams, Chapter 2), we can ask which of its aspects are most conducive to streaming. Iverson (1995) presented subjects with sequences of orchestral tones that were equated for pitch and loudness, and they rated how strongly the tones were perceptually segregated from each other. Multidimensional scaling analyses of their judgments indicated that both static and dynamic timbral attributes were involved. In a further experiment, Iverson presented subjects with interleaved melodies, and had them attempt to recognize the melodies on the basis of a target timbre. Again, judgments were influenced by both static and dynamic timbral attributes. Attack time was found to be influential in two ways: tones with rapid attacks segregated from each other more strongly, as did tones with contrasting attack times.

Later, Bey and McAdams (2003) had subjects listen to a target melody that was interleaved with a distractor sequence, followed by a probe melody that they compared with the target. Synthesized instrument sounds were employed, and these had earlier been organized perceptually in terms of distance along a multidimensional timbral space. The tones had been found to vary along three dimensions: The first dimension related to spectral centroid, the second to attack quality, and the third to variations or irregularities in the spectral envelope. Melody identification improved with increasing distance between the target and distractor sequence in this multidimensional space.

In a further study, Cusack and Roberts (2004) explored the effect on streaming of dynamic variations in the frequency spectrum. They generated periodic tones that differed in fundamental frequency, with the total amount of spectral flux held constant. Sequential patterns of tones were more likely to segregate perceptually when they contained different patterns of spectral variation, particularly variation in frequency centroid over time. The issue of timbre as a grouping cue is revisited in Section VI.

C.   Building and Resetting of Auditory Streams

A number of studies have shown that the splitting of tone sequences into separate streams based on pitch builds with repetition. Van Noorden (1975) compared the signal parameters that were required for listeners to hear a single connected series of tones, using sequences of different types. Compared with two-tone sequences, bidirectional three-tone sequences needed to be presented at a considerably slower tempo in order for a connected series to be perceived. For long repetitive sequences, the tempo needed to be slower still (Figure 9).

Other studies have confirmed that the percept of two streams rather than one builds over time. Anstis and Saida (1985) found that stream segregation built
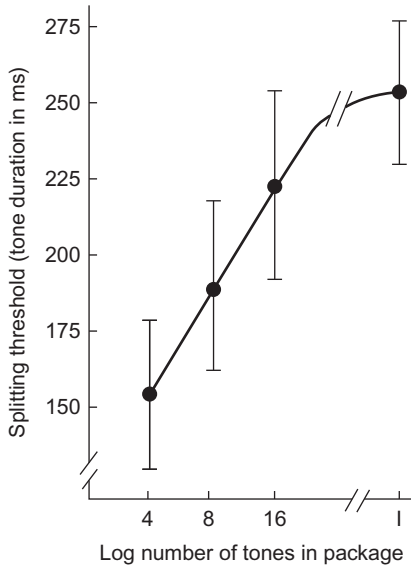
Figure 9 Temporal coherence boundary for two-tone (Curve 3), three-tone unidirectional (Curve 1), three-tone bidirectional (Curve 2), and continuous (Curve 4) sequences. Adapted from Van Noorden (1975).

steeply during the first 10 s of sequence presentation, continued to build at a decreased rate thereafter, and even appeared incomplete when the sequence had continued for 60 s. Roberts, Glasberg, and Moore (2002) also found that stream segregation built rapidly during the first few seconds, and this was followed by a more gradual buildup that appeared incomplete even after 25−30 s.

The streaming process can be reset by various manipulations. Bregman (1978) presented listeners with a repeating sequence that consisted of two high tones together with a single low tone. When the sequence split perceptually into two streams, listeners perceived two high tones in alternation, together with a single low tone that was steadily repeated. The number of tones that were packaged between 4-s periods of silence was varied, and listeners adjusted the speed of the sequence until the point of splitting was determined. As shown in Figure 10, as the number of tones in the package increased, the tempo required for perception of separate streams decreased. Beauvois and Meddis (1997) explored this issue further by having the subjects listen to an induction sequence of repeating tones (AAAA . . .) that were designed to produce the tendency to hear an A stream (see Figure 5). They then presented a silent interval, followed by a short ABAB test sequence. The tendency to hear the sequence as segregated into A and B streams decreased with increasing duration of the silent interval (see also Snyder, Carter, Lee, Hannon, & Alain, 2008).

The preceding findings lead to the further conjecture that the streaming mechanism can be "reset," not only by the interpolation of silent intervals, but also by other changes in the stimulus pattern. As a test of this conjecture, Anstis and Saida (1985) presented listeners with sequences of ABAB tones in one ear, so that stream segregation developed. They then switched the sequence to the other ear, and this produced a reduction in streaming. The authors concluded that the streaming mechanism was "reset" by the change in signal location—a conclusion that was later supported by Rogers and Bregman (1993, 1998) who also produced "resetting" by an abrupt increase in loudness (see also Roberts, Glasberg, & Moore, 2008).

Figure 10 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).

D.   *Streaming and Perception of Temporal Relationships*

One consequence of the formation of separate perceptual streams is that temporal relationships between elements of 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.

Using a different paradigm, Van Noorden (1975) studied the detection of temporal displacement of a tone that alternated continuously with another tone of different frequency. Using a basic pattern that consisted of two tones at frequencies A and B, such that they formed repeating ABA_ triplets (Figure 5), he produced different values of temporal displacement between the A and B tones. As the tempo of the sequence increased, the threshold for detecting temporal displacement between tones A and B also increased. This rise in threshold was substantial when the tones were widely separated in frequency, but only slight when their frequencies were similar. Such deterioration in temporal processing was considerably larger for long repetitive sequences than for two-tone sequences, indicating that it was associated with the process of stream formation (Figure 11). Interestingly, impaired sensitivity to temporal relationships between alternating tones has also been found when harmonic complex tones formed segregated streams based on differences in either spectrum or fundamental frequency (Vliegen et al., 1999).
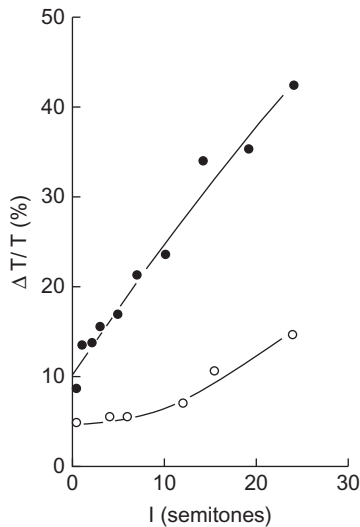
Figure 11 ○ Just noticeable displacement ΔT/T of the second tone of a two-tone sequence as a function of tone interval I. • Just noticeable displacement ΔT/T of one tone in a continuous sequence of alternating tones as a function of tone interval I.
Adapted from Van Noorden (1975).

Van Noorden (1975) showed that the loss of temporal information resulting from stream segregation can have profound effects on the way a sequence is perceived. In an intriguing sound demonstration, he presented listeners with a continuous ABA_ triplet pattern, and gradually altered the pitch relationship between the tones. When tones A and B were close in pitch, a clear "galloping rhythm" was heard, reflecting the temporal relationships between successive tones. However, as the pitch difference between the tones increased, the "galloping rhythm" disappeared, and two unrelated temporal patterns were heard instead, one formed of the A tones and the other of the B tones. Later, Roberts et al. (2008) found that when the streaming process was reset by a change in signal parameters, so that a single stream was again perceived, judgments of the relative timing of temporally adjacent tones of different frequency improved.

E. Streaming by Amplitude and Amplitude Modulation

Amplitude can act as a cue to streaming under some conditions. For example, Dowling (1973b) found that loudness differences increased the perceptual distinctiveness of interleaved melodies. Van Noorden (1975) studied the perception of sequences consisting of tones of identical frequency that alternated between two different amplitudes. A sequence was heard as a single coherent stream when the amplitude difference between the tones was smaller than 5 dB, but as two separate streams when this amplitude difference was larger. With very large amplitude differences, auditory continuity effects were produced, so that the softer tone was heard as continuing through the louder one.

Grimault, Bacon, and Micheyl (2002) carried out a study to determine whether rate of amplitude modulation (AM) might serve as a basis for streaming in the absence of other cues. When subjects were presented with temporal sequences that

consisted of bursts of broadband noise at alternating AM rates, they perceived a single perceptual stream when the difference in AM rate was less than 0.75 octave, but two separate streams when this difference was greater than an octave.

F.   Effects of Short-Term Memory and Context on Streaming

The strength of stream segregation depends not only on the characteristics of the test sequence, but also on the short-term context in which it occurs. Rogers and Bregman (1993) presented listeners with ABA_ triplet patterns that were preceded by "induction sequences" of different types. They found that the streaming of a test sequence was enhanced when it was preceded by an induction sequence whose properties were similar to those of the test sequence, showing that short-term memory played a role here.

As a further context effect, Snyder et al. (2008) found that the range of pitch relationships that had been presented on previous trials affected the stream segregation boundary. Subjects were presented with sequences consisting of ABA_ triplets that were defined by tones of different frequency. It was found that the larger the frequency difference between tones on the preceding trial (and even on trials before the preceding trial), the less the streaming on the current trial.

Whether or not auditory streaming occurs can even be influenced by cues from a different modality. Rahne, Bockmann, Specht, and Sussman (2007) presented subjects with perceptually ambiguous tone sequences that could be perceived either as one or as two streams. Concomitant visual sequences that were designed to promote either an integrated or a segregated percept influenced the perceptual organization of the tone sequences in the direction of the visual cue.

G.   Streaming and Attention

As described earlier, Van Noorden (1975) observed that when sequences were presented that consisted of two tones in alternation, there was a region of ambiguity within which the listener could direct his attention at will, and so hear either one or two streams. More recently, a number of studies have explored the influence of attention on stream segregation in detail.

An experiment by Carlyon, Cusack, Foxton, and Robertson (2001) consisted of several conditions. In one condition, a sequence of ABA_ triplets was presented to the left ear, while no sounds were presented to the right ear. Subjects reported continuously how many streams they heard, and so necessarily attended to the presented sequence. In another condition, the same sequence was presented to the left ear. However, during the first 10 s, the subjects made judgments on a series of noise bursts that were simultaneously presented to the right ear; they then switched attention and made streaming judgments on the left ear sequence. In a further condition, the subjects received the same stimuli as in the two-task condition, but were asked to ignore the noise bursts and concentrate only on the left ear sequence. The buildup of streaming was considerably attenuated in the condition where the subjects attended to the noise bursts compared with the other

conditions, indicating that attention played an important role in the process (see also Carlyon, Plack, Fantini, & Cusack, 2003).

It is possible, however, that in this experiment switching attention from the noise bursts to the tone sequence may have caused a resetting of the streaming process similar to that which occurs when other features of the sequence are abruptly changed (Anstis & Saida, 1985). Evidence for this view was provided by Cusack, Deeks, Aikman, and Carlyon (2004), who showed that interpolating silent gaps in the sequence to be judged had the same effect as attention switching in reducing the buildup of streaming.

Studies involving physiological measures have also indicated that attention can modulate the streaming process. Snyder, Alain, and Picton (2006) presented subjects with repeating ABA_ triplets and recorded their event-related potentials (ERPs) while they either attended to the sequence or watched a silent movie during its presentation. Stream segregation developed when the subjects were attending to the sequence, and this correlated with ERP enhancements; however, the ERP effect was reduced when the subjects ignored the sequence. Elhalali, Xiang, Shamma, and Simon (2009) arrived at a similar conclusion from an experiment in which subjects were required to attend either to a repeating target tone that was surrounded by random "maskers" or to the background maskers themselves. Recordings using magnetoencephalography (MEG) showed that attention strongly enhanced the neural representation of the attended target in the direction of boosting foreground perception.

Other studies using measures such as the mismatch negativity component of the ERP—an index of preattentive acoustic processing—have indicated that stream formation can occur even when subjects are performing a task that draws attention away from the presented sounds (cf. Sussman, Horvath, Winkler, & Orr, 2007; Sussman, Ritter, & Vaughan, 1999). It appears, therefore, that streaming can develop preattentively, though it can also be influenced by attention focusing. The preattentive streaming of sounds on the basis of complex attributes such as timbre, and the involvement of memory in the streaming process, are in accordance with the model of attention advanced by Deutsch and Deutsch (1963), which proposes that attention selection is determined by the analysis of information at very high levels of processing. The issue of attention selection in grouping is revisited in Section VI.

H. Brain Mechanisms Underlying Streaming

During the past decade, there has been a flowering of interest in brain mechanisms underlying auditory streaming. These have involved recordings from neural units in animals, as well as brain scanning in human subjects.

As described earlier, when a rapid sequence of tones is continuously presented, and it forms either an ABA_ triplet pattern or an ABAB pattern, the tendency to hear the sequence as two streams builds with repetition, and builds more strongly as the frequency separation between the tones increases and the tempo and duration of the sequence increase. To investigate the basis of these effects, Fishman, Reser,

Arezzo, and Steinschneider (2001) recorded activity from neural units in the primary auditory cortex of awake macaque monkeys to sequences consisting of ABAB tones that differed in frequency. They adjusted the frequency of the A tone so as to elicit the strongest response at the recording site, and they then varied the frequency of the B tone. At slow tempi, the unit showed responses to both the A and B tones. However, at faster tempi, the unit's responses to the B tones weakened as the frequency separation between the A and B tones increased. At large frequency separations and fast tempi, the unit's responses were predominantly to the A tones. By extrapolation, the same process can be assumed to have occurred in units that responded most strongly to the B tones (see also Fishman, Arezzo, & Steinschneider, 2004).

Based on these findings, Fishman and colleagues proposed that streaming results from a number of response characteristics of the auditory system—frequency selectivity of individual units, forward suppression across units, and adaptation. Arising from these response characteristics, the A and B tones activate more distinct neuronal populations with increasing frequency separation between the tones, and also with increasing tempo and duration of the sequence, so giving rise to stream segregation. Although this model was originally applied to pure tones of differing frequency, it can also be applied to the streaming of complex tones based on pitch, spectral envelope, spatial location, timbre, and so on—indeed wherever different populations of units can be hypothesized to subserve perception of different attribute values (Shamma & Micheyl, 2010).

From another perspective, a number of researchers have investigated the relationship between neural responses to signals that produce streaming and percepts of these signals by human subjects. For example, Micheyl, Tian, Carlyon, and Rauschecker (2005) studied neural responses in the primary auditory cortex of awake rhesus monkeys to tone sequences that would be expected to produce streaming. These responses corresponded well to perceptual changes reported by human subjects when presented with similar stimuli.

An even more direct measure of the neural correlates of streaming involves having human subjects make psychophysical streaming judgments while their patterns of brain activity are simultaneously monitored. Cusack (2005) employed sequences of ABA_ triplets whose parameters were in the region in which percepts fluctuated between a single coherent stream and two segregated streams (Van Noorden, 1975). Subjects made judgments of one or two streams while their brain activity was monitored using functional magnetic resonance imaging (fMRI). More activity was found in the posterior intraparietal sulcus during the two-stream than the one-stream percept, even with the stimulus parameters held constant.

In a similar vein, Gutschalk, Micheyl, Melcher, Rupp, Scherg, and Oxenham (2005) examined the neural bases of streaming in human subjects, using both behavioral measures and concomitant MEG. Employing sequences of ABA_ triplets, they showed that manipulating the tempo and the frequency difference between the alternating tones resulted in changes in the auditory evoked field; these changes corresponded closely to the degree of streaming reported by the subjects. The authors also created sequences consisting of ABA_ triplets in the region

of ambiguity that produced a bistable percept of either one or two streams. They found that even though the stimulus parameters were held constant, patterns of activation covaried with the subjects' percepts. From the patterns of activation they observed, the authors concluded that streaming most likely arose from nonprimary auditory cortex.

Later, Gutschalk, Oxenham, Micheyl, Wilson, and Melcher (2007) presented human subjects with ABBB sequences consisting of harmonic complex tones with varying fundamental frequencies but identical spectral envelopes. As the pitch separation between the alternating tones increased, the subjects were more likely to hear two streams. Along with the development of streaming at the perceptual level, cortical activity as measured by fMRI and MEG increased, both in primary auditory cortex and in surrounding nonprimary areas, with patterns strongly resembling those found for pure tones (see also Wilson, Melcher, Micheyl, Gutschalk, & Oxenham, 2007).

Other studies have produced evidence that streaming occurs in regions below the level of the cortex. Using fMRI, Kondo and Kashino (2009) demonstrated the involvement of the medial geniculate body of the thalamus in streaming by human subjects. An even more striking finding was obtained by Pressnitzer, Sayles, Micheyl, and Winter (2008) from single units in the cochlear nucleus of anaesthetized guinea pigs. (The cochlear nucleus receives input from the cochlear nerve, and so is the first way station along the auditory pathway.) The responses from this region were found to be similar to those from the cortex, and displayed all the functional properties that were needed for streaming to occur. Furthermore, perceptual responses obtained from human subjects correlated well with the neurometric responses obtained from the guinea pig cochlear nucleus. As a possible explanation for the neural substrates of streaming at this low level, the responses from the cochlear nucleus could be modulated by centrifugal projections from higher-level structures, including the cortex (Suga & Ma, 2003; Winer, 2006).

We now ask whether streaming is mediated by activity in the left hemisphere, or the right, or both. A number of studies exploring the neural correlates of streaming based on pitch have found either no difference in activation between the left and right hemispheres, or activation primarily in the right hemisphere (Cusack, 2005; Gutschalk et al., 2005, 2007; Snyder et al., 2006; Wilson et al., 2007). In contrast, Deike, Gaschler-Markefski, Brechmann, and Scheich (2004) and Deike, Scheich, and Brechmann (2010) found activation primarily in the left hemisphere when subjects were asked to segregate A from B tones continuously in sequences where the tones differed in timbre or in pitch. As suggested by these authors, differences in task requirements may have been responsible for the different patterns of hemispheric activity that were obtained in the various studies.

V.   Grouping and Phrase Structure

In the foregoing sections, we have considered ways in which the listener groups sequences of tones into separate parallel streams in a polyphonic texture. We now

turn to the aspect of grouping in music whereby the listener divides sequences of tones into coherent subsequences that are separated by temporal boundaries. It is generally agreed that sequences in Western tonal music are represented by the listener as tonal-temporal hierarchies—notes combine to form motives, which in turn combine to form phrases, which in turn combine to form phrase groups, and so on until the level of the entire piece is reached (Deutsch & Feroe, 1981; Meyer, 1956, 1973; Lerdahl & Jackendoff, 1983; Narmour, 1990, 1999; Salzer, 1962; Schenker, 1956, 1973; Temperley, 2001; Thomson, 1999). The division of the musical surface into hierarchically organized chunks confers enormous processing advantages, as discussed in detail in Chapter 7. Here we discuss the cues whereby such groupings are formed.

Tenney and Polansky (1980), in a computational study of grouping in music, stressed the role of temporal proximity, as well as changes in values of other attributes such as pitch and dynamics. Later, Lerdahl and Jackendoff (1983) proposed that grouping boundaries are placed at longer intervals between note onsets (after rests, at the ends of slurs, at long intervals between attacks, or after long notes), and also at changes in values of attributes such as pitch range and dynamics. In an experimental investigation of this issue, Deliège (1987) asked subjects to listen to excerpts of Western classical music, and to mark boundaries between groupings. The perceived boundaries were found to correspond largely to Lerdahl and Jackendoff's proposed grouping cues, with the strongest effects occurring after long notes, followed by changes in dynamics and timbre (see also Clarke & Krumhansl, 1990; Frankland & Cohen, 2004).

In general, grouping by temporal proximity has emerged as the most powerful cue for the perception of phrase boundaries. Other work has shown that this cue can, in turn, have a pronounced effect on the perception of pitch patterns. Handel (1973) had subjects identify repeating patterns that consisted of dichotomous elements of differing pitch. Identification performance was high when the patterns were temporally segmented in accordance with pitch structure, but low when temporal segmentation and pitch structure were placed in conflict. Further, Dowling (1973a) 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 sequence was presented in a single temporal segment than when a pause was inserted between its elements.

Using more elaborate configurations, Deutsch (1980) presented subjects with sequences of tones, which they recalled in musical notation. The sequences were either hierarchically structured according to the rules of Deutsch and Feroe (1981) (see Chapter 7 for details), or they were composed of the same set of tones but arranged in haphazard fashion. When the tones were presented at equal temporal intervals, performance levels for the structured sequences were very high, whereas they were low for the unstructured sequences. This finding shows that listeners exploit musical knowledge acquired through long-term exposure—in this case knowledge of the pitch alphabets used in Western tonal music—in order to group notes into phrases.

Returning to the issue of temporal segmentation, Deutsch (1980) found that performance levels were further enhanced when structured sequences were temporally segmented in accordance with their pitch structure. However, when the sequences were segmented in conflict with their pitch structure, the subjects instead formed perceptual groupings based on temporal segmentation. In consequence, they were unable to take advantage of the pitch structure inherent in such sequences, so that performance levels were again low. (See Chapter 7, Figure 14, for the types of temporal segmentation that were employed.)

What happens when grouping by pitch proximity and temporal segmentation are set in opposition to each other? Hamaoui and Deutsch (2010) performed a series of experiments to assess the relative strengths of these two cues. Sequences of 12 tones were constructed in which pitch proximity suggested one type of grouping (e.g., four groups of three tones each) and temporal segmentation suggested an opposing type of grouping (in this case, three groups of four tones each). In the default condition, tones were 200 ms in duration and were separated by 100-ms pauses. The tones within a subsequence moved in semitone steps, and the pitch distances employed to suggest grouping by pitch proximity were 2, 5, and 11 semitones. For example, in the sequence shown in the upper part of Figure 12, boundaries between subsequences were marked by distances of 2 semitones, and in the sequences shown in the lower part of the figure, these boundaries were marked by distances of 11 semitones. To suggest grouping by temporal segmentation, the pauses following every third or fourth tone in the sequence were increased by durations varying from 15 to 60 ms. As expected, the larger the pitch distance between groups of tones, the greater the tendency to form groupings based on pitch proximity. However, the temporal cue to grouping was found to be surprisingly powerful, frequently overriding cues provided by large pitch distances when the durations of the pauses were increased by amounts as small as 30 ms. As illustration, the data produced by one subject are shown in Figure 13.

Figure 12 Examples of sequences used to study grouping by temporal segmentation, when this was placed in opposition to grouping by pitch proximity. Here all sequences ascended in semitone steps. (a) Boundaries between subsequences marked by distances of 2 semitones; (b) boundaries between subsequences marked by distances of 11 semitones.
From Hamaoui and Deutsch (2010).

In further experiments, Hamaoui and Deutsch (2010) presented subjects with sequences of tones that were hierarchically structured in accordance with the rules of Deutsch and Feroe (1981), together with control sequences that were unstructured but otherwise matched in terms of pitch relationships. The subjects formed groupings based on hierarchical pitch structure, and these groupings were considerably more resistant to the temporal cue than were the matched unstructured sequences; further, groupings that were based both on pitch proximity and hierarchical structure were even more resistant to the temporal cue.

The influence of pitch proximity on the formation of coherent patterns was also shown in an experiment by Deutsch (1978). Subjects were asked to judge whether two tones were the same or different in pitch when these were separated by a sequence of intervening tones. Performance levels improved as the average pitch distance between the tones in the intervening sequence was reduced (see Chapter 7, Figure 23). This indicated that when the tones in the intervening sequence were proximal in pitch, they formed a network of pitch relationships to which the test tones were anchored.

Statistical evidence that pitch proximity is involved in phrasing was provided by Huron (2006) in a study of musical intervals in roughly 4,600 folk songs. The average interval size within phrases was found to be 2.0 semitones, whereas that between the end of one phrase and the beginning of the next was significantly larger—at 2.9 semitones.
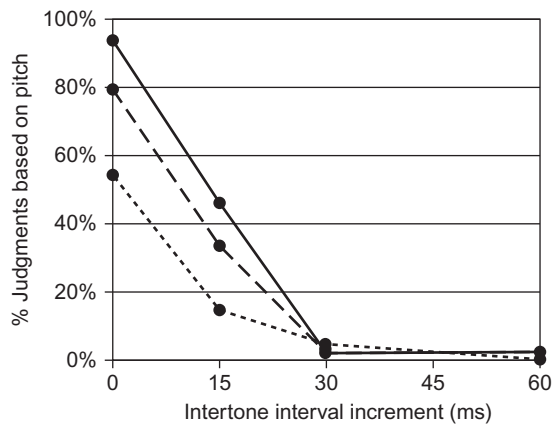
Figure 13 Groupings based on pitch proximity when this cue was placed in conflict with small increments in the pauses between tones. Sequences such as those shown in Figure 12 were presented. In the default condition, tones were 200 ms in duration and separated by 100-ms pauses. Data from one subject are displayed, showing that very small increments in the pauses between tones served as grouping cues, even overriding grouping based on large pitch distances. Solid line indicates grouping by 11 semitone distances; dashed line by 5 semitone distances, and dotted line by 2 semitone distances.
From Hamaoui and Deutsch (2010).

A study by Tan, Aiello, and Bever (1981) provided further evidence that knowledge of musical structure contributes to the grouping of tone sequences into phrases. These authors presented melodies consisting of two phrases that were determined by their implied harmonic structure, with no other cues to phrase boundaries. Each melody was then followed by a two-tone probe. It was found that musically trained subjects were better able to identify the probe when it had occurred within a phrase than when it crossed a phrase boundary.

Interestingly, grouping of pitch patterns can also occur based on short-term learning of statistical probabilities between adjacent tones, even in the absence of long-term structural cues. This was shown by Saffran, Johnson, Aslin, and Newport (1999) who created "words" consisting of three-tone patterns, with the tones within words chosen at random from within an octave range. The words were then presented in random orderings, so that there were high transitional probabilities between tones within words, and low transitional probabilities between tones across words. Listeners rapidly learned to group and recognize the words that were formed in this way.

Although melodic phrase structure frequently coincides with metrical structure, this does not necessarily occur (Lerdahl & Jackendoff, 1983; Temperley, 2001). As illustration, at the beginning of Chopin's Waltz in D♭ major (Op. 64, No. 1) the melody is composed of the repeating four-note pattern (G-A♭-C-B♭). This pattern is at variance with the $\frac{3}{4}$ metrical structure, so that instead of perceiving a repeating four-note pattern, listeners perceive the two alternating six-note patterns (G-A♭-C-B♭-G-A♭) and (C-B♭-G-A♭-C-B♭), as illustrated in Figure 14. So here grouping by metrical structure overrides grouping by repetition of the pitch pattern.

The question also arises as to whether grouping boundaries should be considered to apply to the entire texture of a passage or to different melodic lines separately. Although a rule specifying consistent divisions simplifies many analyses (Lerdahl & Jackendoff, 1983), exceptions can easily be found. Figure 15 shows measures 3−8 of Bach's Italian Concerto. The lower and middle voices alternate between a two-note and a four-note phrase. Yet the phrasing of the higher voice cuts across the boundaries defined by the other voices, so that different groupings are perceived in parallel.

VI.   Grouping of Simultaneous Tone Sequences

When listening to ensemble performances, we are generally presented with multiple sequences of tones that arise in parallel from different regions of space. We can then inquire into the principles that govern the perceptual grouping of such configurations. Do we form parallel linkages between tones that are similar in pitch, in loudness, or in timbre? Do we invoke spatial location as a prominent grouping cue? We shall see that many factors are involved in such grouping, and that they interact in complex ways.

Figure 14 Measures 3−6 of Chopin's Walze in D♭ major (Op. 64, No. 1). Here grouping by metrical structure overrides grouping by repetition of a pitch pattern. The melody is composed of the repeating four-note pattern [G-A♭-C-B♭]. This pattern is at variance with the $\frac{3}{4}$ metrical structure, so that instead of perceiving a repeating four-note pattern, listeners perceive two alternating six-note pitch patterns [G-A♭-C-B♭-G-A♭] and [C-B♭-G-A♭-C-B♭].

Figure 15 Perceptual grouping in measures 4−7 of the second movement of Bach's Italian Concerto. While the lower and middle voices alternate between a two-note and four-note phrase, the phrasing of the higher voice cuts across the boundaries defined by the other voices, so that different groupings are perceived in parallel.

The problem faced by the auditory system in parsing simultaneous streams of sound that emanate from different regions of space is far more difficult than that faced by the visual system in parsing a visual array. The visual system is presented with a spatial layout of elements at the periphery (with the exception of depth perception). In contrast, space in the auditory system is not mapped directly onto the receptor surface, so the listener is required to infer the spatial origins of sounds by indirect means. Inferred sound location must therefore provide a much less reliable cue for the analysis of auditory scenes than is provided by spatial location for the analysis of visual scenes. In addition, sounds are by their nature fleeting, so that scrutinizing each sound element in series is not feasible for auditory scene analysis.

It is not surprising, therefore, that listening to complex music can be quite prone to error, and that powerful illusory conjunctions can occur. When we hear a tone, we attribute a pitch, a loudness, a timbre, and we hear the tone as emanating from a particular spatial location. Each tone, as it is perceived, may then be described as a bundle of attribute values. If our perception is veridical, this bundle reflects the characteristics and location of the emitted sound. However, when multiple sequences of tones are presented simultaneously from different regions of space, these bundles of attribute values may fragment and recombine incorrectly, so that

illusory conjunctions result. These illusory conjunctions can sometimes be due to random error, but they can also reflect the operation of multiple decision mechanisms in the grouping process.

A.   *The Scale Illusion and Related Phenomena*

The scale illusion, which was first devised by Deutsch (1975a, 1975b) results from illusory conjunctions of pitch and location. The pattern that gives rise to this illusion is shown in the upper portion of Figure 16. It 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 16. A melody corresponding to the higher tones is heard as coming from one earphone (in right-handers, this is generally the earphone on the right), while a melody corresponding to the lower tones is heard as coming from the opposite earphone. When the earphone positions are reversed, the apparent locations of the higher and lower tones often remain fixed. This gives rise to the 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. (A minority of listeners instead hear a single melodic line that consists of the higher tones alone, and little or nothing of the lower tones; other listeners obtain yet different illusions, as described in Deutsch, 1975b).

In experiencing 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 also frequently reorganized in space in accordance with their melodic reorganization. Such spatial reorganization is in agreement with other

Figure 16 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.

findings showing that, in the absence of further supporting cues, differences in ear of input may 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 components of a sound complex (Hill & Darwin, 1993).

Although in producing the scale illusion the auditory system arrives at conclusions that are wildly wrong, the illusion is based on a listening strategy that is generally conducive to realistic interpretations of our natural environment. It is unlikely that a source in one location is producing a set of tones that leap around in pitch, while another source in a different location is producing a different set of tones that also leap around in pitch. It is far more 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 should be grouped together, and we infer their perceived locations on this basis (Deutsch, 1975a, 1987).

Variants of the scale illusion are readily produced. One of these, called the chromatic illusion, is illustrated in Figure 17. A chromatic scale that ranges over two octaves is presented in both ascending and descending form, with the individual tones switching from ear to ear in the same way as in the 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, the pattern is heard correctly as a series of tones that leap around in pitch. In Figure 17, the smoothing out of the visual representation of the percept reflects well the way the sounds are perceptually reorganized.

Figure 17 The pattern that produces a version of the chromatic illusion, and the way it is most often perceived.
Adapted from Deutsch (1995).

The cambiata illusion, which was first devised by Deutsch (2003), is based on the same principle. Figure 18 shows the pattern that is presented to the listener via earphones, together with the illusion that is most often obtained. The tones that are presented via each earphone leap around in pitch. However, the percept that is most frequently obtained consists of a higher melody that is composed of three tones that are close in pitch, together with a lower melody that is also composed of three tones that are close in pitch. And again, the higher tones often appear to be emanating from one earphone and the lower tones from the other. Further, as with the scale illusion, there are substantial individual differences in how the cambiata illusion is perceived, with right-handers tending to hear the higher tones as coming from the right earphone and the lower tones from the left one.

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 scale illusion pattern 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 grouping 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 from the two speakers, a new tone quality was heard, but it appeared to be coming simultaneously from both speakers.

Figure 18 The pattern that produces a version of the cambiata illusion, and the way it is most often perceived.
Adapted from Deutsch (2003).

To determine whether these findings generalize to other configurations, Butler presented listeners with the melodic patterns shown in Figure 19. Again, virtually all responses reflected grouping by pitch range. For both these patterns, a perceptual reorganization occurred, so that a melody corresponding to the higher tones appeared to be coming from one earphone or loudspeaker, and a melody corresponding to the lower tones appeared to be coming from the other one.

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 Figure 20, the notes from the theme alternate between the first and second violin parts, and the notes from the accompaniment alternate reciprocally (see Butler, 1979b, for a 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. This is true even with the orchestra arranged in 19th century fashion, so that the first violins are to the left of the audience and 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 and the second violins the accompaniment. Tchaikovsky refused to change his scoring; however, 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).

Figure 19 Patterns used to study grouping of simultaneous sequences in the experiment of Butler (1979a).
Adapted from Butler (1979a).

Figure 20 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.

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 different patterns, each composed of the same two tones. However, it appears to the listener that one piano is consistently playing the higher tone, and the other piano the lower one (Sloboda, 1985).

To return to the experiment of Deutsch (1975b), it is noteworthy that all subjects formed perceptual groupings based on overall pitch range. Rather 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. This shows that the subjects were invoking global pitch range as well as local pitch proximity in making grouping judgments.

A related finding was obtained by Van Noorden (1975), who presented an ascending sequence of tones in rapid alternation with a descending one, and subjects heard higher and lower melodic lines that moved in contrary motion. Tougas and Bregman (1985, 1990) observed an analogous perceptual organization of simultaneous ascending and descending glides. However, the perceptual reorganization of tones in space was not explored in these two studies.

The perceptual tendency to form melodic streams based on overall pitch range is in line with the rule prohibiting voice crossing in counterpoint, and is reflected in the tendency by composers to avoid part crossing in polyphonic music—an effect documented by Huron (1991a) in an analysis of the polyphonic works of J. S. Bach. Interestingly, although Bach avoided part crossing when writing in two parts,

he avoided it even 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, Hausfeld, Power, and Gorta (1982) used tones with substantial timbral differences (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, Gregory (1994) generated a number of different instrument tones and used these in various combinations to construct ascending and descending scales. When there was little or no difference in timbre between the scales, listeners perceived higher and lower pitch streams, as described in Deutsch (1975b). However, when substantial differences in timbre were introduced, listeners tended to use these differences as cues for streaming.

We can here point out that 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—as in the Beethoven passage shown in Figure 8. However, the difference in timbre needs to be salient for this device to be effective. A study by Saupe, Koelsch, and Rubsamen (2010) illustrates the difficulty experienced by listeners in judging simultaneous tone sequences on the basis of timbre, when the timbral differences are small and conflict with grouping by pitch proximity. These authors created brief compositions consisting of three melodic parts, each with a different computer-generated timbre (violin, saxophone, and clarinet). The subjects were asked to focus attention on the violin part and to detect falling jumps in this part, ignoring those in the saxophone and clarinet parts. When the three parts emanated from the same spatial location, the error rates in detecting the pitch jumps were extremely high.

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 listeners to interpret such disparities as indicating 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 here on the basis of spatial location rather than pitch proximity.

As a test of this hypothesis, Deutsch (1979) had subjects identify melodic patterns in which tones were distributed across ears in various ways. There were four conditions in the experiment, and these are illustrated in Figure 21. In Condition A, the melody was delivered to both ears simultaneously, and the performance level here was very high. In Condition B, the tones forming each melody were switched haphazardly between ears, and this difference in ear of input caused performance levels to drop considerably. 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 in this condition, both ears again received input simultaneously, even though
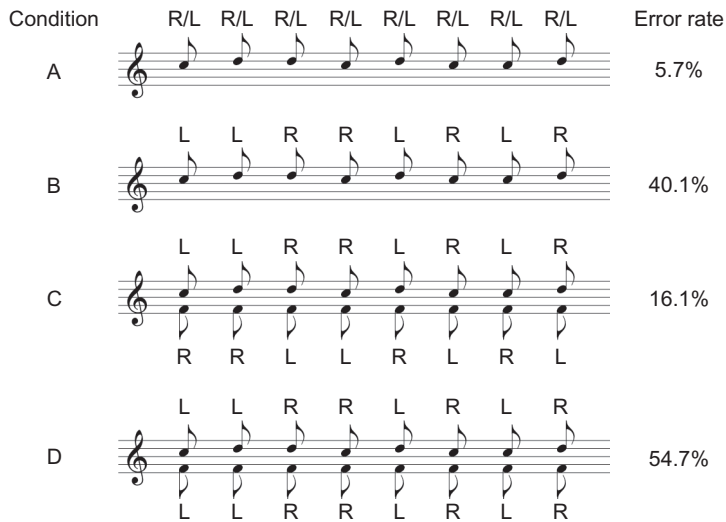
Figure 21 Examples of the ways in which tones were distributed between the two ears in the experiment on binaural integration of melodic patterns. Also shown are the error rates in the different conditions.
Adapted from Deutsch (1979).

the melody was still switching from ear to ear. The presence of the contralateral drone caused identification performance to return to a high level. In Condition D, the drone again accompanied the melody, but it was now delivered to the same ear as the melody component, so that input was again to one ear at a time. In this condition, 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 tones arrive at both ears simultaneously, 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. Subjects listened to pairs of these patterns and judged on each trial whether the members of the pair were the same or different. On half the trials, the tones presented to each ear were separated by silent gaps, and on the other half, the gaps were filled with noise. Judd found that identification performance was enhanced in the presence of the noise, and concluded that the noise degraded the localization information, and so 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 synchronous but instead overlapped in time. Specifically, in some conditions the components of the melody and the drone were offset from each other by 15 ms. These intermediate conditions produced intermediate results: Identification of the melody in the presence of the contralateral drone when the two were asynchronous was poorer than when the melody and drone were strictly synchronous, but better than when the tones from the melody switched between the ears without an accompanying drone.

It is interesting that Berlioz (1948) came to a similar conclusion from the composer's perspective. In *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 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, listeners may be unable to integrate them into a single coherent stream. Such integration is more readily accomplished when the tones played by different instruments overlap in time. However there is a trade-off: As the amount of temporal overlap increases, our ability to identify the spatial origins of the tones decreases, and when the tones are presented simultaneously, spatial illusions are likely to 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 reverse. As described earlier, when listening to 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 coming from. This means that simultaneous 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 the tones in musical notation. Each ear received a haphazard ordering of the first six tones of a major scale, so that for some chords the tone at the right ear was higher and the tone at the left ear was lower ("high-right/low-left" chords),

and for other chords 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 that were presented to the other ear.

When the subjects were attending to the right ear, they notated more higher than lower tones tones correctly. Furthermore, more higher than lower tones 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 lower tones, with a marginal advantage to the lower tones. Further, more lower than higher tones 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 those comprising "high-left/low-right" chords.

In a further experiment, subjects were asked to notate the entire pattern, disregarding ear of input. It was found that they notated more tones correctly when these formed "high-right/low-left" chords than when they formed "high-left/low-right" chords. So we can conclude that there is an advantage to "high-right/low-left" dispositions, both in terms of where the tones appear to be coming from and also how well their pitches are 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 22 shows a seating plan for the Chicago Symphony, viewed from the back of the stage. Considering the strings, the first violins are
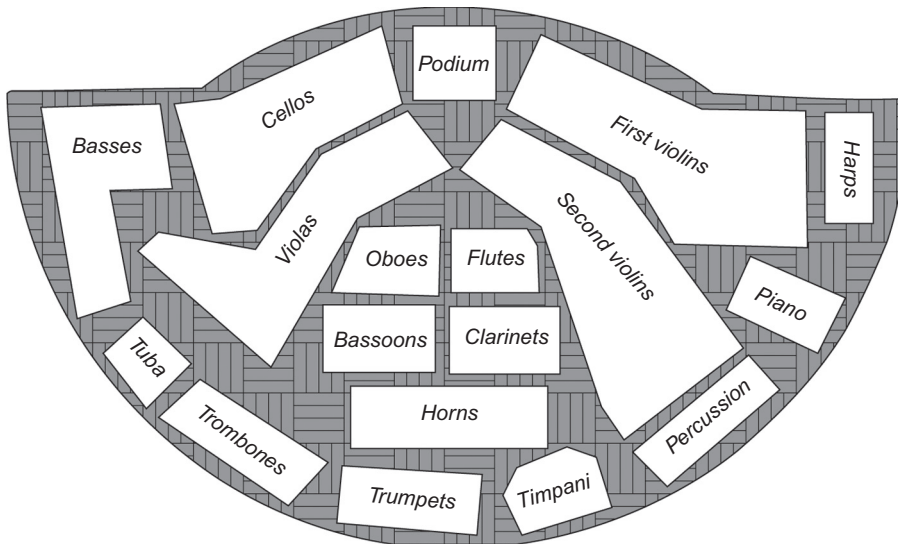

Figure 22 Seating plan for the Chicago Symphony, as viewed from the back of the stage. Adapted from Machlis (1977).

to the right of the second violins, which are to the right of the violas, which are to the right of the cellos, which in turn are 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, and 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.

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 that 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 seating arrangement for both the performers and the audience.

A further illusion in which tones are perceptually reorganized in space was devised by Deutsch (1995), and is called the glissando illusion. The pattern that gives rise to this illusion consists of a synthesized oboe tone of constant pitch, played together with a pure tone whose pitch glides up and down. The listener is seated in front of two stereophonically separated loudspeakers, with one to his left and the other to his right. The signals are repeatedly alternated between the loud-speakers such that when a segment of the oboe tone emanates from one speaker a segment of the glissando emanates from the other one.

On listening to this pattern, the oboe tone is heard correctly as switching between loudspeakers; however, the segments of the glissando appear to be joined together seamlessly, so that it appears to be emanating from a source that moves slowly around in space in accordance with its pitch motion. In a large-scale study, Deutsch et al. (2007) found that listeners localized the glissando in a variety of ways. Most right-handers heard the glissando move between left and right, and also between low and high in space, as its pitch moved between low and high; however nonright-handers were less likely to obtain this percept.

Whereas in the scale illusion, most listeners perceive patterns of tones that appear to be coming from fixed spatial locations, the glissando is always perceived as coming from a source that moves slowly through space. In addition, many sub-jects obtain a percept that results from a synthesis of illusory motion both between left and right and also between low and high. In consequence, the glissando is sometimes heard as tracing an elliptical path between a position low and to the left when its pitch is lowest, and high and to the right when its pitch is highest, as illustrated in Figure 23.

We now turn to hypothesized neurological substrates for these illusions. In all cases, there is a strong tendency for right-handers to hear the higher tones as on the right side of space, and the lower tones as on the left side, whereas left-handers and mixed-handers do not show such a strong tendency. Based on findings relating patterns of cerebral dominance to handedness (Isaacs, Barr, Nelson, & Devinsky,
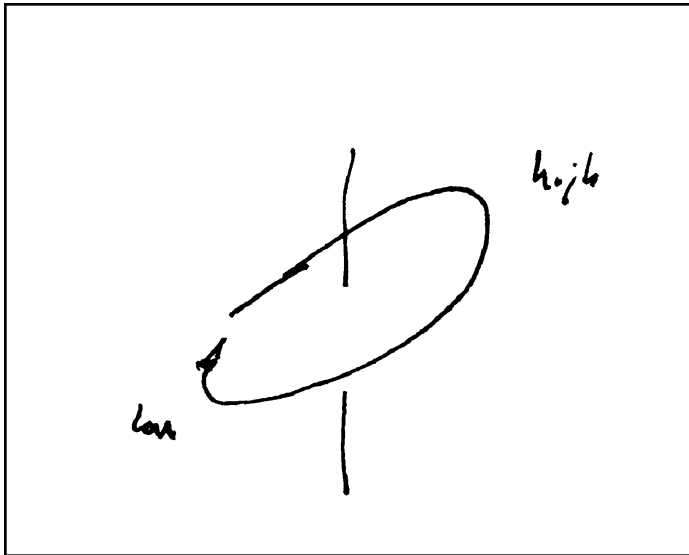
Figure 23 Original drawing by a subject to illustrate his perception of the glissando illusion. The glissando was perceived as tracing an elliptical path through space, from low and to the left when its pitch was lowest, and high and to the right when its pitch was highest. From Deutsch et al. (2007).

2006; Knecht et al., 2000; Luria, 1969; Milner, Branch, & Rasmussen, 1966; Pujol, Deus, Losilla, & Capdevila, 1999), we can conjecture that there is a tendency to perceive higher tones as on the dominant side of space and lower tones as on the nondominant side. This conjecture is supported by other findings indicating that sounds tend to be perceived as on the side of space contralateral to the hemisphere that is more strongly activated (Hari, 1990; Jacobs, Feldman, Diamond, & Bender, 1973; Penfield & Perot, 1963). So we can conclude that these illusory patterns give rise to greater activity in the dominant hemisphere in response to higher tones and to greater activity in the nondominant hemisphere in response to lower tones.

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 that were presented to the two ears. The octave illusion, which was discovered by Deutsch (1974), provides an interesting exception, because here following by spatial location occurs even when the tones delivered to the two ears are strictly simultaneous.
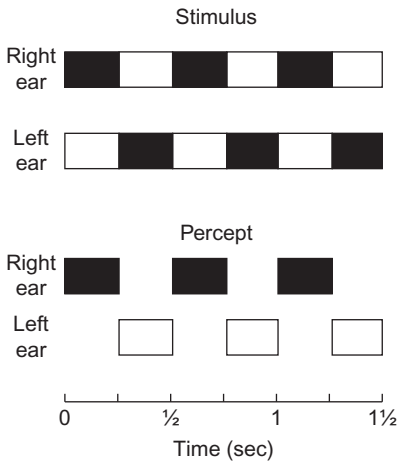
Figure 24 Pattern that produces the octave illusion, together with the percept most commonly obtained. Filled boxes indicate tones at 800 Hz and unfilled boxes tones at 400 Hz. When this pattern is played through stereo headphones, most right-handed listeners perceive an intermittent high tone in the right ear that alternates with an intermittent low tone in the left ear.
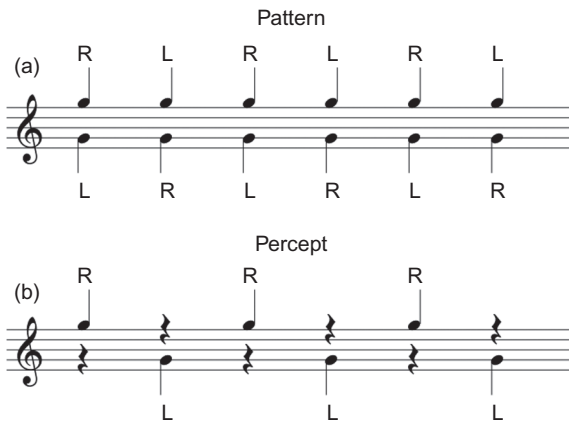Adapted from Deutsch (1974).

Figure 25 Pattern that produces the octave illusion together with the percept most commonly obtained, shown in musical notation.

The pattern that gives rise to the octave illusion is shown in the upper portions of Figures 24 and 25. Two pure tones spaced an octave apart are repeatedly presented in alternation without pause. The identical sequence is presented to both ears simultaneously; however, the tones are 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.

There are strong differences between listeners in how the octave illusion in perceived (Deutsch, 1974, 1975a, 1981, 1983a, 1983b, 1987, 1995, 2004). Most right-handers 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 the right ear is receiving the pattern "high tone−silence−high tone−silence" while the left ear is receiving the pattern "silence−low tone−silence−low tone." This percept is illustrated in the lower portions of Figures 24 and 25. When the
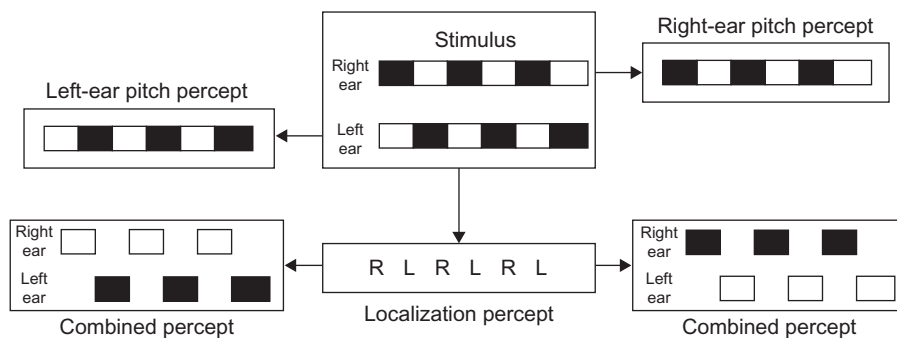
Figure 26 Model showing how the output of two decision mechanisms, one determining perceived pitch and the other determining perceived location, can combine to produce the octave illusion. Filled boxes indicate tones at 800 Hz, and unfilled boxes indicate tones at 400 Hz.
Adapted from Deutsch (1981).

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. This produces the bizarre impression that switching headphone positions has caused the high tone to migrate from one earphone to the other, and the low tone to migrate in the opposite direction.

Deutsch (1975a) hypothesized that the octave illusion results from the combined operation of two separate 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 26. To provide the perceived pitches, 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 perceived.

We can take a listener who perceives 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 it is presented to his right ear. The listener also localizes the tone in his 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 it is presented to his right ear, but he localizes the tone in the left ear instead, because this ear is receiving the higher frequency. The resultant illusory conjunction of pitch and location results in the percept of 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 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 psychophysical experiments have provided further evidence for this model (cf. Deutsch, 1981; Deutsch & Roll, 1976).

Since this model was proposed, substantial evidence for separate *what* and *where* pathways in the auditory system has been obtained, both in nonhuman primates (cf. Kaas & Hackett, 2000; Rauschecker, 1998; Recanzone, Guard, Phan, & Su, 2000; Tian, Reser, Durham, Kustov, & Rauschecker, 2001) and in human subjects (cf. Ahveninen et al., 2006; Altmann, Bledowski, Wibral, & Kaiser, 2007; Arnott, Binns, Grady, & Alain, 2004; Recanzone & Sutter, 2008). These findings provide clues to the neuroanatomical substrate of the octave illusion in terms of separate *what* and *where* decision mechanisms.

Recently, Lamminmaki, Hari, and colleagues have provided evidence concerning the neuronatomical underpinnings of the octave illusion, placing the locus of both the *what* and *where* components in the auditory cortex. Lamminmaki and Hari (2000) focused on the *where* component. Using MEG, they recorded responses to 400-Hz and 800-Hz tones that were presented in different combinations at the two ears. The N100m response (100-ms response from the auditory cortex) at each hemisphere was found to be stronger to tone pairs in which the 800-Hz tone was presented contralaterally and the 400-Hz tone ipsilaterally than when the tone pairs were presented in the opposite configuration. Given that monaural sounds evoke a stronger N100m response in the hemisphere contralateral to the presented sound, and that listeners localize single sounds to the ear contralateral to the hemisphere in which more activation occurs (Hari, 1990), this finding agrees with the proposed lateralization component of the illusion; that is, lateralization of the perceived tone to the ear that receives the higher frequency.

Lamminmaki, Mandel, Parkkonen, & Hari (in press) focused on the *what* component of the illusion—that is, the pattern of pitches that were perceived. Again using MEG, they recorded steady-state responses to all monaural and binaural combinations of 400-Hz and 800-Hz tones, presented as continuous sounds. The subjects were selected for obtaining a typical octave illusion—that is, a high tone in the right ear alternating with a low tone in the left ear. During dichotic presentation of frequencies corresponding to those in the octave illusion, the ipsilateral responses to the left ear tones were weaker, and those to right ear tones were stronger than when both ears received the same tone. Importantly, for the most paradoxical component of the illusion—that is, presentation of the high tone to the left ear and the low tone to the right ear, producing the illusory percept of a low tone in the left ear—responses to the left ear tones were also weaker in the contralateral hemisphere. Taken together, these two sets of findings point to a neuroanatomical instantiation of the octave illusion in terms of separate *what* and *where* decision mechanisms.

In other work on the neural underpinnings of the octave illusion, Ross, Tervaniemi, and Näätänen (1996) questioned whether the illusion was present at the level of the auditory cortex, or whether it was created higher in the processing stream. These authors presented subjects with the sequence producing the illusion, and intermittently inserted illusion-mimicking sequences of single tones that were presented monaurally. The oddball sequences elicited the mismatch negativity

component of the ERP, which is thought to be generated in the auditory cortex, and to reflect perceived changes in sound properties. The authors concluded that the illusion is generated beyond the auditory cortex. However, the sounds as they are perceived in the illusion differ in subtle ways from those mimicking the illusion, and the mismatch negativity could well have picked up on these differences. For example, Sonnadara and Trainor (2005) found that when subjects who heard the illusion as a series of high tones in the right ear alternating with low tones in the left ear were presented with a pattern mimicking the illusion, the mimicking pattern appeared to be louder than the illusion-generating one. This finding is in line with that of Lamminmaki et al., who also showed that steady state responses in the auditory cortex to binaurally presented tones were suppressed compared with responses to tones that were presented monaurally.

We can here note that the octave illusion has pronounced handedness correlates. Deutsch (1974) found that 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 rather than the nondominant ear (Deutsch, 1975a, 1981; 1983a, 1983b, 2004; Deutsch & Roll, 1976).

In a further study, Deutsch (1983b) divided the subject population into three groups—right-handed, mixed-handed, and left-handed. The right-handers tended strongly to hear the high tone on the right and the low tone on the left. This tendency was less strong among mixed-handers and even less strong among left-handers. Furthermore, for all three handedness 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 with a left- or mixed-handed parent or sibling. This pattern of results is in accordance with the literature relating patterns of cerebral dominance to handedness and familial handedness background (Ettlinger, Jackson, & Zangwill, 1956; Isaacs et al., 2006; Luria, 1969; Subirana, 1958), and indicates that in experiencing the octave illusion, listeners generally perceive the high tone on the dominant side of space and the low tone on the nondominant side.

A recent finding has produced evidence that perception of the octave illusion may even serve as a reliable indicator of the direction of cerebral dominance in a given individual. Ferrier, Huiskamp, Alpherts, Henthorn, and Deutsch (in preparation) presented the octave illusion to 17 patients who were scheduled to undergo the Wada test to assess their pattern of cerebral dominance, in preparation for brain surgery for the relief of temporal or frontal lobe epilepsy. It was found that all patients heard the high tone on the side contralateral to the hemisphere that was later determined by the Wada test to be dominant for speech. Based on this finding, we conjecture that the octave illusion could be used as a simple, noninvasive, and reliable test for the assessment of direction of cerebral dominance—a possibility that has considerable clinical potential.

Other work has explored the signal parameters that are necessary to produce the octave illusion. McClurkin and Hall (1981) replaced the 400-Hz pure tone with a high-frequency complex tone with a 200-Hz fundamental. The illusion was found to persist, with the subjects again tending to hear the high-pitched tone in the right ear and the low-pitched tone in the left ear. This finding indicated that pitch rather than frequency region was responsible for the illusory percept most often obtained. Later, Brancucci, Padulo, and Tommasi (2009) observed that the illusion was not confined to tones that were related by exactly an octave, but also occurred with tones that were spaced at intervals that deviated from an octave by one or two semitones; however, the illusion deteriorated as the size of the intervals decreased further. This finding is in accordance with an earlier demonstration by Deutsch (1983a) in which the intervals formed by the two alternating tones were made to vary. In this demonstration, the illusion became degraded as the size of the interval decreased, so that at the minor third an entirely different perceptual impression was produced. Concerning tone duration, while Deutsch (1974, 1983b) used 250-ms tones, Zwicker (1984) found that the illusion sharpened with the use of 200-ms tones; however, Brancucci et al. (2009) reported that the effect was stronger with 500-ms tones. These temporal discrepancies most likely resulted from differences in other signal parameters used for generating the illusion.

In a further study, Brannstrom and Nilsson (2011) replaced the 400-Hz and 800-Hz pure tones with narrow-band noises with overlapping spectra, and had subjects make judgments on the pitch and localization components of the illusion separately. Most subjects perceived an illusion in terms of a dominant ear for pitch and lateralization by frequency, as in the two-channel model shown in Figure 26. They also found that the salience of the illusion increased with an increase in the high-frequency content of the noise signal.

An interesting study by Brancucci, Lugli, Santucci, and Tommasi (2011) showed that once the octave illusion is induced, its effect can persist strongly. The subjects were presented first with a priming sequence consisting of the octave illusion pattern, and then repeatedly with a test sequence consisting of one of the alternating dichotic chords (either 400 Hz right/800 Hz left, or 800 Hz right/400 Hz left) for up to 6 s. For all the test sequences, the repeating dichotic chords continued to be heard as in the illusion.

C.   Illusory Conjunctions and Attention

In the stereo illusions we have been describing—the scale illusion and its variants, the glissando illusion, and the octave illusion—illusory conjunctions of pitch and location occur preattentively in most listeners. We can take a listener who clearly hears the octave illusion as a high tone to the right alternating with a low tone to the left (Figures 24 and 25). This listener can focus attention on either the high tone or the low one, or on either the right ear or the left one. When the listener is focusing attention on the low tone that is perceived as to the left, in reality the low tone is being presented to the right, and the high tone—which is suppressed from perception—is being presented to the left. An illusory conjunction of the low pitch

with the left location therefore occurs despite focused attention on either the low tone or the left ear, so this illusion is not destroyed by attention focusing.

A similar argument applies to the scale illusion: The listener who obtains a strong illusion such as shown in Figure 16 can focus attention on either the higher or the lower tones, and this does not cause the illusion to break down. Similarly, for a listener who obtains a strong glissando illusion, focusing attention on the glissando does not cause it to be heard correctly as leaping between the left and right loudspeakers. Further, in the study of Deutsch (1985), the illusory conjunctions of pitch and location occurred despite explicit instructions to attend to and notate the tones presented to one ear and ignore the other. These findings run counter to the suggestion, proposed for the case of vision, that the conjunction of features requires attentional control and that illusory conjunctions occur when stimuli are outside the focus of attention (Treisman & Gelade, 1980), because here the illusions occur even though the listener is focusing attention on the tone to be identified. Rather, such illusory conjunctions must reflect the outcome of separate decision mechanisms whose outputs combine preattentively so as lead to erroneous percepts. Evidence for preattentive conjunction of different attribute values has also been found by others in studies using mismatch negativity (Gomes, Bernstein, Ritter, Vaughan, & Miller, 1997; Sussman, Gomes, Manette, Nousak, Ritter, & Vaughan, 1998; Takegata, Huotilainen, Rinne, Näätänen, & Winkler, 2001; Winkler, Czigler, Sussman, Horváth, & Balazs, 2005).

A study by Deouell, Deutsch, Scabini, Soroker, and Knight (2008) on two patients with unilateral neglect provides further evidence that illusory conjunctions occur preattentively, and continue to occur when attention is focused on the illusory tones. Unilateral neglect generally occurs with damage to the nondominant hemisphere, and is often accompanied by auditory extinction—the failure to perceive sounds that are presented on one side of space when other sounds are simultaneously presented on the opposite side. The patients were presented with the scale illusion through headphones, and they reported hearing a single stream of tones that smoothly descended and then ascended—as in the scale illusion. However, they also reported hearing all the tones in one ear and silence in the other ear. Since the "extinguished" tones were being perceived, they must have been erroneously grouped preattentively to one side of space.

Other authors have also reported a high prevalence of illusory conjunctions for musical tones. Hall, Pastore, Acker, and Huang (2000) presented subjects with arrays of simultaneous and spatially distributed tones. The subjects were asked to search for specific cued conjunctions of values of pitch and instrument timbre. For example, the target sound could be that of a violin at a fundamental frequency of 509 Hz, followed by an array of simultaneously presented sounds (such as a violin at 262 Hz and a trombone at 509 Hz) that were differentially localized. The subjects judged in separate tasks whether a particular designated feature of timbre or pitch appeared in the array, and whether a combination of two such features appeared. Although the listeners were well able to identify either the pitch or the timbre alone, they made frequent errors in reporting the presence or absence of target conjunctions, with estimates of illusory conjunction rates ranging from 23% to 40%.

Other research has shown that illusory conjunctions of different attribute values can occur with serial presentation also. Thompson, Hall, and Pressing (2001) presented subjects with a target sequence that was followed by a probe tone. When the probe tone matched one target tone in pitch and a different target tone in duration, on over half the trials the subjects responded that the probe tone matched the same target tone in both pitch and duration.

D. Melody Perception from Phase-Shifted Tones

Another configuration that produces grouping of simultaneous pitch patterns by spatial location was described by Kubovy and colleagues. Kubovy, Cutting, and McGuire (1974) presented a set of simultaneous and continuous pure 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, 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 another 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 its interaural disparity—or apparent spatial location—differed 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. In further experiments, Kubovy found that both concurrent and successive difference cues were involved in the effect.

VII. Grouping of Equal-Interval Tone Complexes

A. Grouping by Pitch Proximity

Perceptual grouping principles emerge strongly in tone complexes whose components are separated by equal intervals. Octave-related complexes have been explored most extensively (see also Chapter 7). 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.

Shepard (1964) generated a series of tones, each of which was composed of 10 components that were separated by octaves. The amplitudes of the components were scaled by a fixed, bell-shaped spectral envelope, such 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 their components up or down in log frequency.
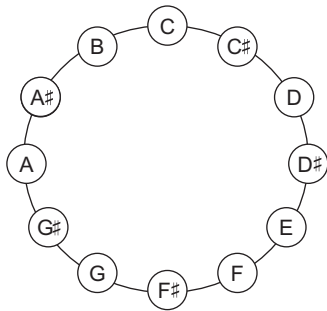
Figure 27 The pitch class circle.

Subjects listened to successive 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 (Figure 27), listeners heard an ascending pattern; when the second tone was removed one or two steps counterclockwise, listeners heard a descending pattern instead. As 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 gave rise to 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, Tenny, and in particular Risset, using both natural instruments and computer-generated sounds. Braus (1995) provides an extensive discussion of such works. Circular pitches have even been put to effective use in movies. Richard King, the sound designer for the Batman movie *The Dark Knight*, employed an ever-ascending glide for the sound of Batman's vehicle, the Batpod. In an article for the *Los Angeles Times*, King wrote "When played on a keyboard, it gives the illusion of greater and greater speed; the pod appears unstoppable."

Percept A   Percept B

Figure 28 Representation of alternative perceptual organizations in the experiment on grouping of simultaneous pairs of Shepard tones. Subjects grouped the pattern in accordance with harmonic proximity (Percept A) in preference to Percept B.
From Deutsch (1988).

Returning to the experimental evidence, the work of Shepard and Risset showed that when other cues to height attribution are weak, listeners invoke proximity in making judgments of relative height for successively presented tones. We can then ask whether the auditory 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 that consisted of two simultaneous pairs of Shepard 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. However, as shown in Figure 28, the descending line could in principle be heard as higher and the ascending line as lower, so forming a harmonic grouping in accordance with proximity (Percept A), or the ascending line could be heard as higher and the descending line as lower, so forming a harmonic grouping that ran counter to proximity (Percept B).

It was found that all subjects showed a strong tendency to organize the patterns so that they were grouped in accordance with proximity along the harmonic dimension. For example, the pattern in Figure 28 tended to be heard as Percept A rather than Percept B.

In all the experiments so far described, the patterns employed 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 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 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 reported by Schroeder (1986), who pointed out that this effect is analogous to certain phenomena in fractal geometry.

In order to achieve pitch circularity, must the choice of materials be confined to highly artificial tones, or to several instrument tones playing simultaneously? If circular scales could be created from sequences of single tones, each of which comprised a full harmonic series, then the theoretical and practical implications of pitch circularity would be broadened.

Benade (1976) pointed out that a good flautist, while playing a sustained note, can vary the relative amplitudes of the odd and even numbered harmonics so as to produce a remarkable effect. Suppose he starts out playing note A at F0 = 440 Hz; the listener hears this note as well defined both in pitch class and in octave. Suppose, then, that the performer changes his manner of blowing so that the amplitudes of the odd-numbered harmonics are gradually reduced relative to the even-numbered ones. At some point the listener realizes that he is now hearing the note A an octave higher—that is, corresponding to F0 = 800 Hz—yet this octave transition had occurred without traversing the semitone scale. We can then conjecture that a tone consisting of a full harmonic series might be made to vary continuously between two octaves without traversing the helical path shown in Figure 29. If this were so, then pitch should be represented as a cylinder rather than as a helix—as indicated by the dashed line between D♯' and D♯'' in Figure 29. Indeed, Patterson, Milroy, and Allerhand (1993) and Warren, Uppenkamp,
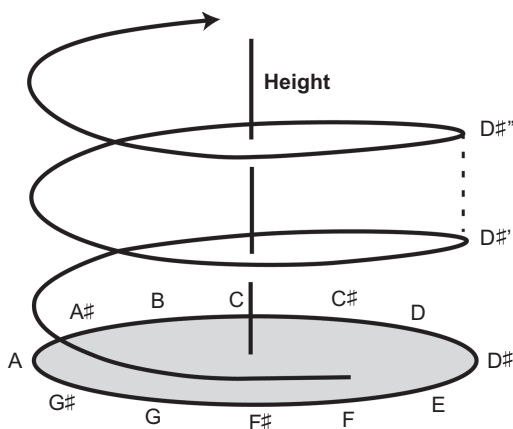

Figure 29 The helical model of pitch. Musical pitch is shown as varying along both a linear dimension of height and also a circular dimension of pitch class. The helix completes one full turn per octave, with the result that tones standing in octave relation are in close spatial proximity. The dashed line from D♯' to D♯'' indicates that the pitch of a tone can also be made to vary within the octave along the height dimension without traversing the helix, pointing to a cylindrical rather than helical representation.

Patterson, and Griffiths (2003) found that attenuating the odd harmonics of a complex tone relative to the even ones resulted in a perceived increase in the pitch height of the tone.

Based on these findings, I reasoned that it should be possible to create pitch circularity from a bank of harmonic complex tones by appropriate manipulations of their odd and even harmonics. One begins with a bank of 12 harmonic complex tones, whose F0s range in semitone steps over an octave. For the tone with the highest F0, the odd and even harmonics are identical in amplitude. Then for the tone a semitone lower, the amplitudes of the odd harmonics are reduced relative to the even ones, so raising the perceived height of this tone. Then for the tone another semitone lower, the amplitudes of the odd harmonics are further reduced relative to the even ones, so raising the perceived height of this tone to a greater extent. One continues this way down the octave in semitone steps, until for the tone with the lowest F0, the odd harmonics no longer contribute to the tone's perceived height. The tone with the lowest F0 is therefore heard as displaced up an octave, and pitch circularity is achieved.

After some trial and error, I settled on the following parameters. Complex tones consisting of the first six harmonics were employed, and the amplitudes of the odd harmonics were reduced by 3.5 dB for each semitone step down the scale. When this bank of tones was presented with F0s in ascending semitone steps, listeners heard the sequence as eternally ascending. When the bank was played in descending semitone steps, the sequence was heard as eternally descending instead. Furthermore, when single gliding tones were used instead of steady-state tones, impressions of eternally ascending and descending glides were obtained.

In a formal experiment, Deutsch, Dooley, and Henthorn (2008) employed such a bank of 12 tones, and created sequential pairings between each tone and each of the other tones. Listeners were then asked to judge for each tone pair whether it ascended or descended in pitch. When the tones within a pair were separated by a short distance along the pitch class circle, judgments were based almost entirely on proximity. This tendency decreased with increasing distance along the circle, but remained high even at a distance of 5 semitones—almost half way around the circle. When the data were subjected to multidimensional scaling, strongly circular configurations were obtained.

The intriguing possibility then arises that this algorithm could be employed to transform banks of natural instrument tones so that they would also exhibit pitch circularity. William Brent, then a graduate student at the University of California, San Diego, achieved considerable success using bassoon samples, and also some success with oboe, flute, and violin samples, and he has shown that the effect is not destroyed by vibrato. The possibility of creating circular banks of tones derived from natural instruments expands the scope of musical materials available to composers and performers. At the theoretical level, these demonstrations of pitch circularity indicate that pitch should be represented as a solid cylinder rather than as a helix (see also Deutsch, 2010.)

Figure 30 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 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.

B.   Grouping by Common Fate

Returning to sequences composed of octave-related complexes, another perceptual grouping principle has been shown to operate. Teranishi (1982) created a set of major triads that were composed of octave-related complexes generated under a trapezoidal spectral envelope. When a subset of these triads was played in succession as shown in Figure 30, listeners obtained the impression of an endlessly ascending scale. However, as can be seen by perusal of Figure 30, 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. However, listeners followed global direction in perceiving this chord succession instead, so that they were basing their relative pitch judgments on an impression of global pitch movement, or "common fate."

In a follow-up study, 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 displayed a pitch circularity of an octave, others displayed a pitch circularity of roughly ⅓ octave. The authors concluded that the subjects were basing their judgments on the perception of global pitch movement (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 so 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 31, 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);
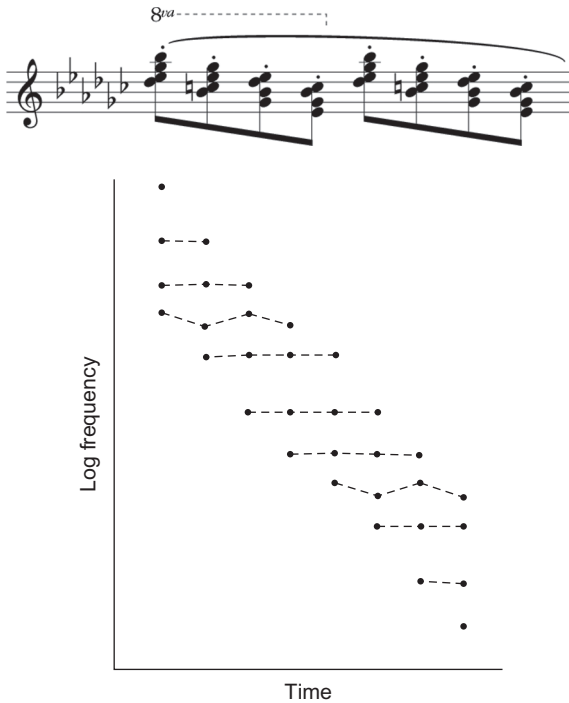
Figure 31 A passage from Debussy's prelude *Le Vent dans la Plaine.* The listener perceives this passage globally as a downward pitch movement, in accordance with the principle of common fate.

however, these percepts are discarded in favor of an impression of a descending series of chords.

VIII.   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 configurations. Beyond their interest to psychologists, these findings have implications for music theory and practice.

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, it has become critically important to understand certain basic perceptual phenomena, 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 neces-sarily 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 free-dom, there has emerged a particularly strong need to understand how music as it is notated and performed maps onto music as it is perceived. The findings reviewed here have brought us closer to realizing this goal, although much more remains to be learned.

Acknowledgments

References

Ahveninen, J., Jääskeläinen, I. P., Raij, T., Bonmassar, G., Devore, S., & Hamalainen, M., et al. (2006). Task-modulated "what'' and ''where'' pathways in human auditory cortex. *Proceedings of the National Academy of Sciences*, *103*, 14608−14613.

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.

Altmann, C. F., Bledowski, C., Wibral, M., & Kaiser, J. (2007). Processing of location and pattern changes of natural sounds in the human auditory cortex. *NeuroImage*, *35*, 1192−1200.

Anstis, S. M., & Saida, S. (1985). Adaptation to auditory streaming of frequency-modulated tones. *Journal of Experimental Psychology: Human Perception and Performance*, *11*, 257−271.

Arnott, S. R., Binns, M. A., Grady, C. L., & Alain, C. (2004). Assessing the auditory dual-pathway model in humans. *NeuroImage*, *22*, 401−408.

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

Beauvois, M. W., & Meddis, R. (1997). Time decay of auditory stream biasing. *Perception & Psychophysics*, *59*, 81−86.

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.

Benade, A. H. (1976). *Fundamentals of musical acoustics*. Oxford, UK: Oxford University Press.

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

Bey, C., & McAdams, S. (2003). Postrecognition of interleaved melodies as an indirect measure of auditory stream formation. *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 267−279.

Braaten, R. F., & Leary, J. C. (1999). Temporal induction of missing birdsong segments in European starlings. *Psychological Science*, *10*, 162−166.

Brancucci, A., Lugli, V., Santucci, A., & Tommasi, L. (2011). Ear and pitch segregation in Deutsch's octave illusion persist following switch from stimulus alternation to repetition. *Journal of the Acoustical Society of America*, *130*, 2179−2185.

Brancucci, A., Padulo, C., & Tommasi, L. (2009). "Octave illusion" or "Deutsch's illusion"? *Psychological Research*, *73*, 303−307.

Brannstrom, K. J., & Nilsson, P. (2011). Octave illusion elicited by overlapping narrowband noises. *Journal of the Acoustical Society of America*, *129*, 3213−3220.

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., 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.

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.

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.

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

Carlyon, R. P. (2004). How the brain separates sounds. *Trends in Cognitive Sciences*, *8*, 465−471.

Carlyon, R. P., Cusack, R., Foxton, J. M., & Robertson, I. H. (2001). Effects of attention and unilateral neglect on auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 115−127.

Carlyon, R. P., & Gockel, H. (2007). Effects of harmonicity and regularity on the perception of sound sources. In W. A. Yost, A. N. Popper, & R. R. Fay (Eds.), *Auditory perception of sound sources* (pp. 191−213). New York, NY: Springer.

Carlyon, R. P., Plack, C. J., Fantini, D. A., & Cusack, R. (2003). Cross-modal and non-sensory influences on auditory streaming. *Perception*, *32*, 1393−1402.

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 J. Sundberg (Ed.), *Sound generation in winds, strings, and computers* (pp. 4−13). Stockholm, Sweden: Royal Swedish Academy of Music.

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

Ciocca, V., & Darwin, C. J. (1999). The integration of nonsimultaneous frequency components into a single virtual pitch. *Journal of the Acoustical Society of America*, *105*, 2421−2430.

Clarke, E. F., & Krumhansl, C. L. (1990). Perceiving musical time. *Music Perception*, *7*, 213−251.

Cusack, R. (2005). The intraparietal sulcus and perceptual organization. *Journal of Cognitive Neuroscience*, *17*, 641−651.

Cusack, R., Deeks, J., Aikman, G., & Carlyon, R. P. (2004). Effects of location, frequency region, and time course of selective attention on auditory stream analysis. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 643−656.

Cusack, R., & Roberts, B. (2000). Effects of differences in timbre on sequential grouping. *Perception & Psychophysics*, *62*, 1112−1120.

Cusack, R., & Roberts, B. (2004). Effects of differences in the pattern of amplitude envelopes across harmonics on auditory stream segregation. *Hearing Research*, *193*, 95−104.

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.

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. (2005a). Pitch and auditory grouping. In C. J. Plack, A. J. Oxenham, R. R. Fay, & A. N. Popper (Eds.), *Springer handbook of auditory research: Pitch neural coding and perception* (pp. 278−305). New York: Springer.

Darwin, C. J. (2005b). Simultaneous grouping and auditory continuity. *Perception & Psychophysics*, *67*, 1384−1390.

Darwin, C. J., & Carlyon, R. P. (1995). Auditory grouping. In B. C. J. Moore (Ed.), *Hearing* (pp. 387−424). San Diego, CA: 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., Ciocca, V., & Sandell, G. R. (1994). Effects of frequency and amplitude modulation on the pitch of a complex tone with a mistuned harmonic. *Journal of the Acoustical Society of America*, *95*, 2631−2636.

Deike, S., Gaschler-Markefski, B., Brechmann, A., & Scheich, H. (2004). Auditory stream segregation relying on timbre involves left auditory cortex. *Neuroreport*, *15*, 1511−1514.

Deike, S., Scheich, H., & Brechmann, A. (2010). Active stream segregation specifically involves the left human auditory cortex. *Hearing Research*, *265*, 30−37.

Deliège, I. (1987). Grouping conditions in listening to music: an approach to Lerdahl & Jackendoff's grouping preference rules. *Music Perception*, *4*, 325−360.

Demany, L. (1982). Auditory stream segregation in infancy. *Infant Behavior & Development*, *5*, 261−276.

Deouell, L. Y., Deutsch, D., Scabini, D., Soroker, N., & Knight, R. T. (2008). No disillusions in auditory extinction: perceiving a melody comprised of unperceived notes. *Frontiers of Human Neuroscience*, *1*, 1−6.

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

Deutsch, D. (1975a). Musical illusions. *Scientific American*, *233*, 92−104.

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

Deutsch, D. (1978). Delayed pitch comparisons and the principle of proximity. *Perception & Psychophysics*, *23*, 227−230.

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

Deutsch, D. (1980). 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, NY: Academic Press.

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

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

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*, 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]. La Jolla, CA: Philomel Records.

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. (2003). *Phantom words, and other curiosities* [CD]. La Jolla, CA: Philomel Records.

Deutsch, D. (2004). The octave illusion revisited again. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 355−364.

Deutsch, D. (2010). The paradox of pitch circularity. *Acoustics Today, July Issue*, 8−15.

Deutsch, D., Dooley, K., & Henthorn, T. (2008). Pitch circularity from tones comprising full harmonic series. *Journal of the Acoustical Society of America*, *124*, 589−597.

Deutsch, D., & Feroe, J. (1981). The internal representation of pitch sequences in tonal music. *Psychological Review*, *88*, 503−522.

Deutsch, D., Hamaoui, K., & Henthorn, T. (2007). The glissando illusion and handedness. *Neuropsychologia*, *45*, 2981−2988.

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.

Deutsch, J. A., & Deutsch, D. (1963). Attention: some theoretical considerations. *Psychological Review*, *70*, 80−90.

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

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

Dowling, W. J., Lung, K. M., & Herrbold, S. (1987). Aiming attention in pitch and time in the perception of interleaved melodies. *Perception & Psychophysics*, *41*, 642−656.

Elhalali, M., Xiang, J., Shamma, S. A., & Simon, J. Z. (2009). Interaction between attention and bottom-up saliency mediates the representation of foreground and background in an auditory scene. *Public Library of Science: Biology*, *7*, 1−14.

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

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

Fay, R. R. (1998). Auditory stream segregation in goldfish (*Carassius auratus*). *Hearing Research*, *120*, 69−76.

Ferrier, C. H., Huiskamp, G .J. M., Alpherts, W. C. P., Henthorn, T., & Deutsch, D. (in preparation). The octave illusion: A noninvasive tool for presurgical assessment of language lateralization.

Fishman, Y. I., Arezzo, J. C., & Steinschneider, M. (2004). Auditory stream segregation in monkey auditory cortex: effects of frequency separation, presentation rate, and tone duration. *Journal of the Acoustical Society of America*, *116*, 1656−1670.

Fishman, Y. I., Reser, D. H., Arezzo, J. C., & Steinschneider, M. (2001). Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. *Hearing Research*, *151*, 167−187.

Frankland, B. W., & Cohen, A. J. (2004). Parsing of melody: quantification and testing of the local grouping rules of Lerdahl and Jackendoff's 'A Generative Theory of Tonal Music'. *Music Perception*, *21*, 499−543.

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.

Gomes, H., Bernstein, R., Ritter, W., Vaughan, H. G., & Miller, J. (1997). Storage of feature conjunctions in transient auditory memory. *Psychophysiology*, *34*, 712−716.

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

Grimault, N., Bacon, S. P., & Micheyl, C. (2002). Auditory stream segregation on the basis amplitude-modulation rate. *Journal of the Acoustical Society of America*, *111*, 1340−1348.

Grimault, N., Micheyl, C., Carlyon, R. P., Arthaud, P., & Collett, L. (2000). Influence of peripheral resolvability on the perceptual segregation of harmonic complex tones differing in fundamental frequency. *Journal of the Acoustical Society of America*, *108*, 263−271.

Gutschalk, A., Micheyl, C., Melcher, J. R., Rupp, A., Scherg, M., & Oxenham, A. J. (2005). Neuromagnetic correlates of streaming in human auditory cortex. *Journal of Neuroscience*, *25*, 5382−5388.

Gutschalk, A., Oxenham, A. J., Micheyl, C., Wilson, E. C., & Melcher, J. R. (2007). Human cortical activity during streaming without spectral cues suggest a general neural substrate for auditory stream segregation. *Journal of Neuroscience*, *27*, 13074−13081.

Hall, M. D., Pastore, R. E., Acker, B. E., & Huang, W. (2000). Evidence for auditory feature integration with spatially distributed items. *Perception & Psychophysics*, *62*, 1243−1257.

Hamaoui, K., & Deutsch, D. (2010). The perceptual grouping of musical sequences: Pitch and timing as competing cues. In S. M. Demorest, S. J. Morrison, & P. S. Campbell (Eds.), *Proceedings of the 11th International Conference on Music Perception and Cognition*, Seattle, Washington (pp. 81−87).

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

Hari, R. (1990). The neuromagnetic method in the study of the human auditory cortex. In F. Grandori, M. Hoke, & G. L. Romani (Eds.), *Auditory evoked magnetic fields and electric potentials: Advances in audiology* (pp. 222−282). Basel, Switzerland: S. Karger.

Hartmann, W. M., & Goupell, M. J. (2006). Enhancing and unmasking the harmonics of a complex tone. *Journal of the Acoustical Society of America*, *120*, 2142−2157.

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, NY: Dover.

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.

Houtgast, T. (1972). Psychophysical evidence for lateral inhibition in hearing. *Journal of the Acoustical Society of America*, *51*, 1885−1894.

Houtsma, A. J. M., Rossing, T. D., & Wagenaars, W. M. (1987). *Auditory demonstrations*. Eindhoven, The Netherlands, and the Acoustical Society of America.

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.

Huron, D. (2001). Tone and voice: A derivation of the rules of voice-leading from perceptual principles. *Music Perception*, *19*, 1−64.

Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation*. Cambridge, MA: MIT Press.

Isaacs, K. L., Barr, W. B., Nelson, P. K., & Devinsky, O. (2006). Degree of handedness and cerebral dominance. *Neurology*, *66*, 1855−1858.

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.

Izumi, A. (2002). Auditory stream segregation in Japanese monkeys. *Cognition*, *82*, B113−B122.

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.

Kaas, J. H., & Hackett, T. A. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proceedings of the National Academy of Sciences USA*, *97*, 11793−11799.

Knecht, S., Drager, B., Deppe, M., Bobe, L., Lohmann, H., & Floel, A., et al. (2000). Handedness and hemispheric language dominance in healthy humans. *Brain*, *123*, 2512−2518.

Kondo, H. M., & Kashino, M. (2009). Involvement of the thalamocortical loop in the spontaneous switching of percepts in auditory streaming. *Journal of Neuroscience*, *29*, 12695−12701.

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.

Lamminmaki, S., & Hari, R. (2000). Auditory cortex activation associated with octave illusion. *Neuroreport*, *11*, 1469−1472.

Lamminmaki, S., Mandel, A., Parkkonen, L., & Hari, R. (in press). Binaural interaction and pitch perception as contributors to the octave illusion. *Journal of the Acoustical Society of America.*

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.

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

MacDougall-Shackleton, S. A., Hulse, S. H., Gentner, T. Q., & White, W. (1998). Auditory scene analysis by European starlings (*Sturnus vulgaris*): perceptual segregation of tone sequences. *Journal of the Acoustical Society of America*, *103*, 3581−3587.

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

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.

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, The Netherlands: 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.

McAdams, S., & Bertoncini, J. (1997). Organization and discrimination of repeating sound sequences by newborn infants. *Journal of the Acoustical Society of America*, *102*, 2945−2953.

McClurkin, H., & Hall, J. W. (1981). Pitch and timbre in a two-tone dichotic auditory illusion. *Journal of the Acoustical Society of America*, *69*, 592−594.

McDermott, J. H., & Oxenham, A. J. (2008). Spectral completion of partially masked sounds. *Proceedings of the National Academy of Sciences USA*, *105*, 5939−5955.

McDermott, J. H., Wrobleski, D., & Oxenham, A. J. (2011). Recovering sound sources from embedded repetition. *Proceedings of the National Academy of Sciences USA*, *108*, 1188−1193.

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

Meyer, L. B. (1956). *Emotion and meaning in music*. Chicago, IL: University of Chicago Press.

Meyer, L. B. (1973). *Explaining music: Essays and explorations*. Berkeley, CA: University of California Press.

Micheyl, C., Tian, B., Carlyon, R. P., & Rauschecker, J. P. (2005). Perceptual organization of tone sequences in the auditory cortex of awake macaques. *Neuron*, *48*, 139−148.

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.

Milner, B., Branch, C., & Rasmussen, T. (1966). Evidence for bilateral speech representation in some nonrighthanders. *Transactions of the American Neurological Association*, *91*, 306−308.

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.

Narmour, E. (1990). *The analysis and cognition of basic melodic structures: The implication-realization model*. Chicago, IL: University of Chicago Press.

Narmour, E. (1999). Hierarchical expectation and musical style. In D. Deutsch (Ed.), *The psychology of music* (2nd ed., pp. 441−472). San Diego, CA: Academic Press.

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.

Patterson, R. D., Milroy, R., & Allerhand, M. (1993). What is the octave of a harmonically rich note? *Contemporary Music Review*, *9*, 69−81.

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

Petkov, C. I., O'Connor, K. N., & Sutter, M. L. (2003). Illusory sound perception in macaque monkeys. *Journal of Neuroscience*, *23*, 9155−9161.

Petkov, C. I., O'Connor, K. N., & Sutter, M. L. (2007). Encoding of illusory continuity in primary auditory cortex. *Neuron*, *54*, 153−165.

Petkov, C. I., & Sutter, M. L. (2011). Evolutionary conservation and neuronal mechanisms of auditory perceptual restoration. *Hearing Research*, *271*, 54−65.

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

Pressnitzer, D., Sayles, M., Micheyl, C., & Winter, I. M. (2008). Perceptual organization of sound begins in the auditory periphery. *Current Biology*, *18*, 1124−1128.

Pujol, J., Deus, J., Losilla, J. M., & Capdevila, A. (1999). Cerebral lateralization of language in normal left-handed people studied by functional MRI. *Neurology*, *52*, 1038−1043.

Rahne, T., Bockmann, M., von Specht, H., & Sussman, E. S. (2007). Visual cues can modulate integration and segregation of objects in auditory scene analysis. *Brain Research*, *1144*, 127−135.

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, improvisation, and composition* (pp. 71−90). Oxford, U.K.: Oxford University Press.

Rauschecker, J. P. (1998). Parallel processing in the auditory cortex of primates. *Audiology and Neurootology*, *3*, 86−103.

Recanzone, G. H., Guard, D. C., Phan, M. L., & Su, T. K. (2000). Correlation between the activity of single auditory cortical neurons and sound localization behavior in the macaque monkey. *Journal of Neurophysiology*, *83*, 2723−2739.

Recanzone, G. H., & Sutter, M. L. (2008). The biological basis of audition. *Annual Review of Psychology*, *59*, 119−142.

Remijn, G. B., Nakajima, Y., & Tanaka, S. (2007). Perceptual completion of a sound with a short silent gap. *Perception*, *36*, 898−917.

Riecke, L., Mendelsohn, D., Schreiner, C., & Formisano, E. (2009). The continuity illusion adapts to the auditory scene. *Hearing Research*, *247*, 71−77.

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

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." *Journal of the Acoustical Society of America*, *80*, 961−962.

Roberts, B., Glasberg, B. R., & Moore, B. C. J. (2002). Primitive stream segregation of tone sequences without differences in fundamental frequency or passband. *Journal of the Acoustical Society of America*, *112*, 2074−2085.

Roberts, B., Glasberg, B. R., & Moore, B. C. J. (2008). Effects of the build-up and resetting of auditory stream segregation on temporal discrimination. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 992−1006.

Rogers, W. L., & Bregman, A. S. (1993). An experimental evaluation of three theories of auditory stream regulation. *Perception & Psychophysics*, *53*, 179−189.

Rogers, W. L., & Bregman, A. S. (1998). Cumulation of the tendency to segregate auditory streams: Resetting by changes in location and loudness. *Perception & Psychophysics*, *60*, 1216−1227.

Ross, J., Tervaniemi, M., & Näätänen, R. (1996). Neural mechanisms of the octave illusion: electrophysiological evidence for central origin. *Neuroreport*, *8*, 303−306.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*, 27−52.

Salzer, F. (1962). *Structural hearing*. New York, NY: Dover.

Sandell, G. J., & Darwin, C. J. (1996). Recognition of concurrently-sounding instruments with different fundamental frequencies. *Journal of the Acoustical Society of America*, *100*, 2683.

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

Saupe, K., Koelsch, S., & Rubsamen, R. (2010). Spatial selective attention in a complex auditory environment such as polyphonic music. *Journal of the Acoustical Society of America*, *127*, 472−480.

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

Schenker, H. (1956). *Neue musikalische theorien and phantasien: Der freie satz.* Vienna, Austria: Universal Edition.

Schenker, H. (1973). *Harmony* (O. Jonas, Ed. and annotator; E. M. Borgese, trans.). Cambridge, MA: MIT Press.

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

Seeba, F., & Klump, G. M. (2009). Stimulus familiarity affects perceptual restoration in the European starling (*Sturnus vulgaris*). *PLoS One*, *4*, e5974.

Shamma, S. A., & Micheyl, C. (2010). Behind the scenes of auditory perception. *Current Opinion in Neurobiology*, *20*, 361−366.

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, NY: 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.

Snyder, J. S., Alain, C., & Picton, T. W. (2006). Effects of attention on neuroelectric correlates of auditory stream segregation. *Journal of Cognitive Neuroscience*, *18*, 1−13.

Snyder, J. S., Carter, O. L., Lee, S.-K., Hannon, E. E., & Alain, C. (2008). Effects of context on auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 1007−1016.

Sonnadara, R. R., & Trainor, L. J. (2005). Perceived intensity effects in the octave illusion. *Perception & Psychophysics*, *67*, 648−658.

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

Suga, N., & Ma, X. (2003). Multiparametric corticofugal modulation and plasticity in the auditory system. *Neuroscience*, *4*, 783−794.

Sugita, Y. (1997). Neuronal correlates of auditory induction in the cat cortex. *Neuroreport*, *8*, 1155−1159.

Sussman, E., Gomes, H., Manette, J., Nousak, K., Ritter, W., & Vaughan, H. G. (1998). Feature conjunctions in auditory sensory memory. *Brain Research*, *793*, 95−102.

Sussman, E., Ritter, W., & Vaughan, J. H. G. (1999). An investigation of auditory streaming effect using event-related brain potentials. *Psychophysiology*, *36*, 22−34.

Sussman, E., Horvath, J., Winkler, I., & Orr, M. (2007). The role of attention in the formation of auditory streams. *Perception & Psychophysics*, *69*, 136−152.

Takegata, R., Huotilainen, M., Rinne, T., Näätänen, R., & Winkler, I. (2001). Changes in acoustic features and their conjunctions are processed by separate neuronal populations. *Neuroreport*, *12*, 525−529.

Tan, N., Aiello, R., & Bever, T. G. (1981). Harmonic structure as a determinant of melodic organization. *Memory and Cognition*, *9*, 533−539.

Temperley, D. (2001). *The cognition of basic musical structures*. Cambridge, MA: MIT Press.

Tenny, J., & Polansky, L. (1980). Temporal Gestalt perception in music. *Journal of Music Theory*, *24*, 205−241.

Teranishi, R. (1982). Endlessly ascending/descending chords performable on a piano. *Reports of the Acoustical Society of Japan, H*62−68.

Thomson, W. (1999). *Tonality in music: A general theory*. San Marino, CA: Everett Books.

Thompson, W. F., Hall, M. D., & Pressing, J. (2001). Illusory conjunctions of pitch and duration in unfamiliar tone sequences. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 128−140.

Tian, B., Reser, D., Durham, A., Kustov, A., & Rauschecker, J. P. (2001). Functional specialization in rhesus monkey auditory cortex. *Science*, *292*, 290−293.

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.

Treisman, A., & Gelade, A. (1980). A feature integration theory of attention. *Cognitive Psychology*, *12*, 97−136.

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, Italy: 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.

Vliegen, J., Moore, B. C. J., & Oxenham, A. J. (1999). The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task. *Journal of the Acoustical Society of America*, *106*, 938−945.

Vliegen, J., & Oxenham, A. J. (1999). Sequential stream segregation in the absence of spectral cues. *Journal of the Acoustical Society of America*, *105*, 339−346.

Warren, J. D., Uppenkamp, S., Patterson, R. D., & Griffiths, T. D. (2003). Separating pitch chroma and pitch height in the human brain. *Proceedings of the National Academy of Sciences USA*, *100*, 10038−10042.

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., 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.

Wilson, E. C., Melcher, J. R., Micheyl, C., Gutschalk, A., & Oxenham, A. J. (2007). Cortical fMRI activation to sequences of tones alternating in frequency: relationship to perceived rate and streaming. *Journal of Neurophysiology*, *97*, 2230−2238.

Winer, J. A. (2006). Decoding the auditory corticofugal systems. *Hearing Research*, *212*, 1−8.

Winkler, I., Czigler, I., Sussman, E., Horváth, J., & Balazs, L. (2005). Preattentive binding of auditory and visual stimulus features. *Journal of Cognitive Neuroscience*, *17*, 320−339.

Winkler, I., Kushnerenko, E., Horváth, J., Ceponiene, R., Fellman, V., & Huotilainen, M., et al. (2003). Newborn infants can organize the auditory world. *Proceedings of the National Academy of Sciences*, *100*, 11812−11815.

Zwicker, T. (1984). Experimente zur dichotischen Oktav- Tauschung [Experiments on the dichotic octave illusion]. *Acustica*, *55*, 128−136.