7 The Processing of Pitch Combinations

Diana Deutsch

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

I. Introduction

In this chapter, we examine ways in which pitch combinations are processed by the perceptual system. We first inquire into the types of abstraction that give rise to the perception of local features, such as intervals, chords, and pitch classes. We also explore low-level abstractions that result in the perception of global features, such as contour. We next consider how combinations of features are further abstracted so as to give rise to perceptual equivalences and similarities. We discuss the roles played by basic, and probably universal, organizational principles in the perception of musical patterns, and the contributions made by stored knowledge concerning the statistical properties of music. We argue for the view that music is represented in the mind of the listener as coherent patterns that are linked together so as to form hierarchical structures.

Other sections of the chapter are concerned with memory. We show how different aspects of musical tones are retained in parallel in separate memory systems, and that the output from these different systems is combined to determine memory judgments. We also consider the involvement of short-term memory for individual tones in our perception of tonal patterns. The final sections of the chapter concern a group of illusions that are produced by certain combinations of tones. These illusions have implications for individual differences in the perception of music, and for relationships between music and speech.

II. Feature Abstraction

A. Octave Equivalence

A strong perceptual similarity exists between tones that are related by octaves; that is, whose fundamental frequencies stand in a ratio of 2:1. Octave equivalence is implied in the music of many different cultures (cf. Nettl, 1956). In the Western

The Psychology of Music. DOI: http://dx.doi.org/10.1016/B978-0-12-381460-9.00007-9

musical scale, tones that stand in octave relation are given the same name, so that a tone is specified first by its position within the octave and then by the octave in which it occurs (D_2, $F\sharp_3$, and so on). In one version of Indian musical notation, a tone is represented by a letter to designate its position within the octave, together with a dot or dots to designate its octave placement.

Various observations related to octave equivalence have been reported. For example, listeners with absolute pitch may sometimes place a note in the wrong octave, even though they name it correctly (Bachem, 1955; Lockhead & Byrd, 1981; Miyazaki, 1989). Generalization of response to tones standing in octave relation has been found in human adults (Humphreys, 1939) and infants (Demany & Armand, 1984), as well as in animals (Blackwell & Schlosberg, 1943). Further, interference and consolidation effects in memory for pitch exhibit octave generalization (Deutsch, 1973b; Deutsch & Lapidis, in preparation).

Given that tones standing in octave relation are in a sense perceptually equivalent, it has been suggested that pitch should be treated as a bidimensional attribute; the first dimension representing overall pitch level (*pitch height*) and the second dimension defining the position of the tone within the octave (*tone chroma* or *pitch class*) (Bachem, 1955; Deutsch, 1969, 1973b; Deutsch, Dooley, & Henthorn, 2008; Deutsch, Kuyper & Fisher, 1987; Patterson, 1986; Pickler, 1966; Risset, 1969; Ruckmick, 1929; Shepard, 1964, 1982; Ueda & Ohgushi, 1987; Warren, Uppenkamp, Patterson, & Griffiths, 2003). This is discussed in detail later.

B. Perceptual Equivalence of Intervals and Chords

When two tones are presented either simultaneously or in succession, there results the perception of a musical interval, and intervals are perceived as the same in size when the fundamental frequencies of their component tones stand in the same ratio. This principle forms a basis of the traditional musical scale. The smallest unit of this scale is the semitone, which corresponds to a frequency ratio of approximately 1:1.06. Tone pairs that are separated by the same number of semitones are given the same name, such as major third, minor sixth, and so on.

Chords consisting of three or more tones are also classified in part by the ratios formed by their components. However, a simple listing of these ratios is not sufficient to define a chord. For instance, major and minor triads are perceptually quite distinct, yet they are both composed of a major third (five semitones), a minor third (four semitones), and a perfect fifth (seven semitones). So it is of perceptual importance that the minor third lies above the major third in the major triad, and below it in the minor triad; this needs to be taken into account in considering how chords might be abstracted by the nervous system.

Given the principles of octave and interval equivalence, one might hypothesize that the perceptual equivalence of intervals would persist if their component tones were placed in different octaves. This assumption has frequently been made by contemporary music theorists, who describe such intervals as in the same *interval class*. Traditional music theory assumes that such equivalence holds for simultaneous intervals. Those whose components have reversed their positions along the height dimension are treated as harmonically equivalent (Piston, 1948/1987), and we easily recognize root progressions of chords in their different instantiations. Plomp, Wagenaar, and Mimpen (1973) and Deutsch and Roll (1974) have provided evidence for the perceptual similarity of harmonic intervals that are related by inversion. For successive intervals, however, it appears that interval class is not perceived directly, but rather through a process of hypothesis confirmation, in which the features that are directly apprehended are pitch class and interval (Deutsch, 1972c).

Deutsch (1969) proposed a neural network that would accomplish the abstraction of low-level pitch relationships so as to produce basic equivalences found in music perception. The model is based on findings concerning the abstraction of low-level features in vision, such as orientation and angle size (Hubel & Wiesel, 1962).

The hypothesized neural network consists of two parallel channels, along each of which information is abstracted in two stages. An outline of this model is shown in Figure 1. The first channel mediates the perceptual equivalence of intervals and chords under transposition. In the first stage of abstraction along this channel, first-order units that respond to tones of specific pitch project in groups of two or three onto second-order units, which in consequence respond to specific intervals and chords, such as (C_4, E_4, G_4) or (D_5, G_5). It is assumed that such linkages occur only between units underlying pitches that are separated by an octave or less. In the second stage of abstraction along this channel, second-order units project onto third-order units in such a way that second-order units activated by tones standing in the same relationship project onto the same unit. So, for example, all units activated by an ascending interval of four semitones (a major third) project onto one unit, all those activated by a descending interval of seven semitones (a perfect fifth)

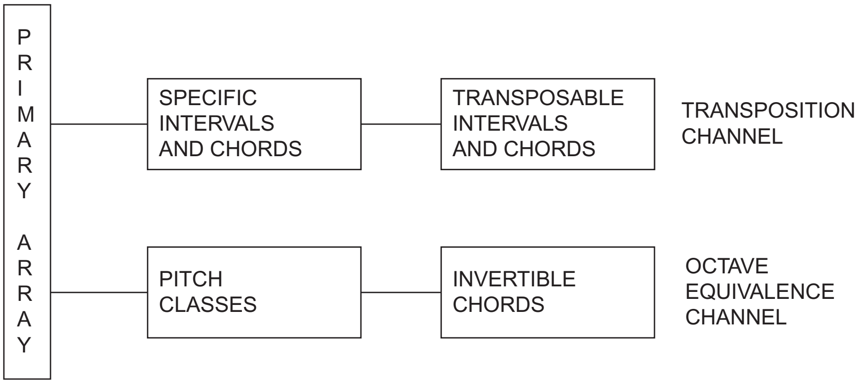

Figure 1 Model for the abstraction of pitch relationships. Pitch information is abstracted along two parallel channels; one mediating transposition and the other mediating octave equivalence.
Adapted from Deutsch (1969). ©1969 by the American Psychological Association. Adapted with permission.

project onto a different unit, all those activated by a major triad project onto yet a different unit, and so on (Figure 2).

The second channel mediates the perceptual equivalence of tones that stand in octave relation. In the first stage of abstraction along this channel, first-order units that respond to tones of specific pitch project onto second-order units in such a way that those standing in octave relation project onto the same unit. These second-order units then respond to tones in a given pitch class, regardless of the octave in which they occur, so can be termed "pitch class units." In the second stage of abstraction along this channel, second-order units project in groups of two or three onto third-order units, which in consequence respond to combinations of pitch classes. Such units therefore mediate the perceptual similarity of intervals and chords that are related by inversion (Figure 3). This level of convergence is assumed to occur only for units that are activated by simultaneously presented tones.

The general type of architecture proposed by this model has been found in mammalian auditory systems. Neurons have been found that act as AND gates, as hypothesized for the transposition channel, and others as OR gates, as hypothesized for the pitch class channel. In addition, the physiological evidence has shown that

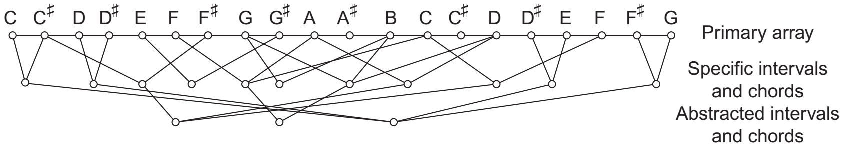

Figure 2 Two stages of abstraction along the transposition channel.
Adapted from Deutsch (1969). ©1969 by the American Psychological Association. Adapted with permission.

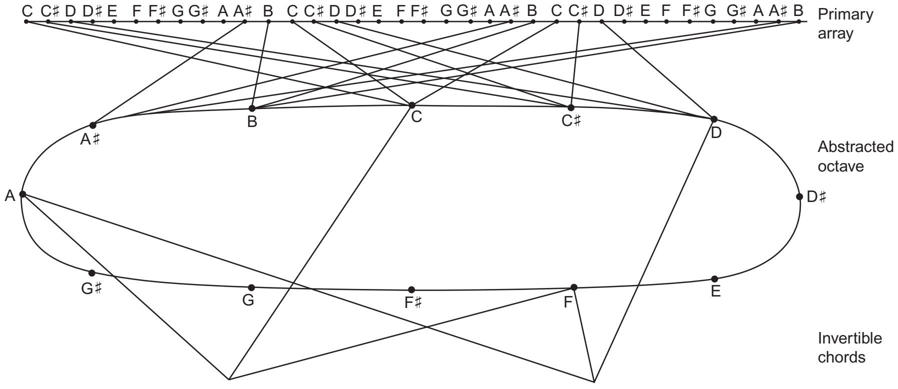

Figure 3 Two stages of abstraction along the octave-equivalence channel.
Adapted from Deutsch (1969). ©1969 by the American Psychological Association. Adapted with permission.

many auditory analyses are carried out in parallel subsystems, each of which is organized in hierarchical fashion (Knudsen, du Lac, & Esterly, 1987; Patterson, Uppenkamp, Johnsrude, & Griffiths, 2002; Schreiner, 1992; Suga, 1990; Sutter & Schreiner, 1991; Wessinger, VanMeter, Tian, Van Lare, Pekar, & Rauschecker, 2001).

With respect specifically to interval identification, Suga, O'Neil, and Manabe (1979) have described neurons in the auditory cortex of the bat that showed facilitation when the second harmonic of a tone was delivered simultaneously with the third harmonic, so that the combination formed a perfect fifth. Other units showed facilitation when the second and fourth harmonics were simultaneously presented, so that the combination formed an octave; yet others showed facilitation when the third and fourth harmonics were simultaneously presented, so that the combination formed a perfect fourth. Such units often responded poorly to single tones in isolation, but strongly and consistently when the appropriate tonal combination was presented. On the present model, units with such characteristics are hypothesized to occur at the first stage of abstraction along the transposition channel. With respect to the pitch class channel, Evans (1974) found neurons in the auditory cortex of the cat that exhibited peaks of sensitivity at more than one band of frequencies, and peaks spaced at octave intervals were commonly found. Also Suga and Jen (1976) noted the presence of neurons in the bat auditory cortex that showed two peaks of sensitivity that were approximately harmonically related.

Ross, Choi, and Purves (2007) hypothesized that the intervals of Western tonal music have a special status, resulting from our constant exposure to speech sounds. The authors analyzed a database of spoken English vowels and found that, expressed as ratios, the frequency relationships between the first two formants in vowel phones represent all 12 intervals in the chromatic scale. It is intriguing to hypothesize, therefore, that through extensive exposure to speech sounds, higher-order connections are formed between lower-order units in such a way as to emphasize those units that feature the 12 chromatic intervals.

Bharucha (1987, 1999) has hypothesized a more elaborate neural network, whose basic architecture has features similar to those proposed by Deutsch (1969). The model assumes that such feature detectors develop as a result of passive exposure to the music of our tradition, and it is discussed further in Chapter 8.

C. Interval Class

When different two-tone combinations form the same interval by appropriate octave displacement, these combinations are held to be in the same interval class. For example, C_3 paired with D_5, form the same interval class as G_2 paired with F_6. As noted earlier, the conditions under which interval class forms a basis for perceptual equivalence are complex ones. Experimental evidence for such equivalence has been obtained for simultaneous intervals, as mentioned earlier (Deutsch & Roll, 1974; Plomp et al., 1973). For successive intervals, however, the issue is complicated. If interval class were indeed a perceptual invariant, we should have no difficulty in recognizing a melody when its component tones are placed haphazardly in different octaves. As a test of this prediction, Deutsch (1972c) generated the first half of the tune "Yankee Doodle" in different versions. First, it was produced without transformation in each of three adjacent octaves. Second, it was generated in such a way that each tone was in its correct position within the octave (i.e., the interval classes were preserved) but the octave placement of the tones varied haphazardly across the same three octaves. Third, the tune was generated as a series of clicks, so that the pitch information was removed entirely but the rhythm remained.

The different versions of the tune were played to separate groups of subjects, who were given no clues as to its identity other than being assured that it was well known. Although the untransformed melody was universally recognized, the scrambled-octaves version was recognized no better than the version in which the pitch information was removed entirely. However, when the subjects were later given the name of the tune, and so knew what to listen for, they were able to follow the scrambled-octaves version to a large extent. This shows that the subjects were able to use pitch class to confirm the identity of the tune, although they had been unable to recognize it in the absence of cues on which to base a hypothesis. (This brief experiment is presented on the CD by Deutsch, 1995).

This experiment shows that perception of interval class, where successions of tones are concerned, requires the involvement of an active, "top-down" process, in which the listener matches each tone as it arrives with his or her image of the expected tone. On this line of reasoning, the extent to which listeners perceive interval class depends critically on their knowledge and expectations.

Other experimental findings have further indicated that interval class is not directly apprehended where successions of tones are concerned. Deutsch (1979) presented listeners with a standard six-tone melody, followed by a comparison melody. The comparison melody was always transposed four semitones up from the standard. On half the trials, the transposition was exact, and on the other half, two of the tones in the transposed melody were permuted, while the melodic contour was unchanged.

There were four conditions in the experiment. In the first, the standard melody was played once, followed by the comparison melody. In the second, the standard melody was repeated six times before presentation of the comparison melody. In the third condition, the standard melody was again repeated six times, but now on half of the repetitions it was transposed intact an octave higher, and on the other half it was transposed intact an octave lower, so that the intervals forming the melody were preserved. In the fourth condition, the standard melody was again repeated six times, but now on each repetition the individual tones were placed alternately in the higher and lower octaves, so that the interval classes were preserved, but the intervals themselves were altered.

Exact repetition of the standard melody resulted in a substantial improvement in recognition performance, and an improvement also occurred when the standard melody was repeated intact in the higher and lower octaves. However, when the standard melody was repeated in such a way that its tones alternated between the higher and lower octaves, performance was significantly poorer than when it was not repeated at all. This experiment provides further evidence that interval class cannot be considered a first-order perceptual feature. Repeating a set of intervals resulted in memory consolidation for these intervals; however, repeating a set of interval classes did not do so.

Deutsch and Boulanger (1984) further addressed this issue by presenting musically trained subjects with novel melodic patterns, which they recalled in musical notation. As shown in the examples in Figure 4, each pattern consisted of a haphazard ordering of the first six notes of the C-major scale. In the first condition, all the tones were taken from a higher octave; in the second, they were all taken from a lower octave. In the third condition, the individual tones alternated between these two octaves, so that roughly two thirds of the intervals formed by successive tones spanned more than an octave. The percentages of tones that were correctly notated in the correct serial positions in these different conditions are also shown in Figure 4, and it can be seen that performance in the third condition was substantially poorer than in the other two.

The findings from these three experiments are in accordance with the two-channel model of Deutsch (1969), which assumes that neural linkages underlying the abstraction of successive intervals occur only between units responding to pitches that are separated by no more than an octave. It is interesting in this regard to consider the use of octave jumps in traditional music. On the present line of reasoning, such jumps can be made with impunity, provided the musical setting is such that the octave-displaced tone is anticipated by the listener. We should therefore expect that octave jumps would tend to be limited to such situations. Indeed, this appears to be the case. For example, a melodic line may be presented several times without transformation. A clear set of expectations having been established, a jump to a different octave occurs. The passage in Figure 5a, for instance, occurs after the melody has been presented several times without octave jumps. Interval class can also be successfully invoked when the harmonic structure is clear and unambiguous, so that again the displaced tones are highly probable. This is illustrated in the segment in Figure 5b.

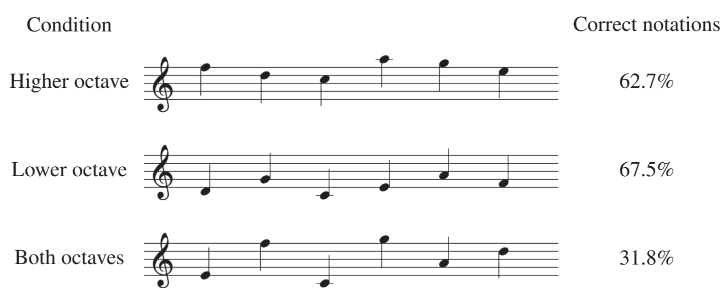

Figure 4 Examples of sequences used in different conditions of the experiment on the effect of octave jumps on recall of melodic patterns. At the right are shown the percentages of tones that were correctly recalled in the correct serial positions in the different conditions. Adapted from Deutsch and Boulanger (1984). ©1984 by the Regents of the University of California.

Figure 5 Two examples of octave jumps in traditional Western music. Here the jumps are readily processed. (a) From Beethoven, Rondo in C, Op. 5, No. 1; (b) from Beethoven, Sonata in C minor, Op. 10, No. 1.

The technique of 12-tone composition uses very frequent octave jumps, and this raises the question of whether the listener does indeed identify as equivalent different instantiations of the same tone row under octave displacement. Given the evidence and arguments outlined earlier, such identification should be possible in principle, but only if the listener is very familiar with the material, or if its structure is such as to give rise strongly to the appropriate expectations (see also Meyer, 1973; Thomson, 1991).

D. Contour

We use global as well as specific cues in recognizing music. Such cues include, for example, overall pitch range, the distribution of interval sizes, and the relative proportions of ascending and descending intervals. Melodic contour plays a particularly important role here. As shown in Figure 6, melodies can be represented by their distinctive contours, even when their interval sizes are altered.

One line of experimentation involving contour was initiated by Werner (1925). He reported that melodies could be recognized when they were transformed onto scales in which the octave was replaced by a different ratio, such as a fifth or two octaves, with these micro- or macro-octaves being divided into 12 equal intervals, so producing micro- or macro-scales. Later, Vicario (1983) carried out a study to determine how well listeners were able to recognize well-known melodies that had been transformed in this fashion. The results of this study are shown in Figure 7. As can be seen, although listeners were able to recognize such distorted melodies to some extent, the distortions nevertheless impaired melody recognition, with the amount of impairment being a function of the degree of expansion or compression of the octave.

In another experiment, White (1960) found that listeners could recognize melodies to some extent when all the intervals were set to one semitone, so that only the sequence of directions of pitch change remained. Performance was enhanced when

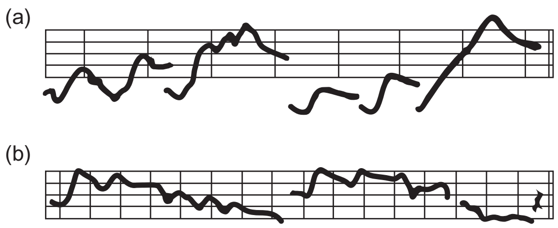

Figure 6 Contours from Beethoven piano sonatas as represented by Schoenberg: (a) from Sonata in C minor, Op. 10/I-III; (b) from Sonata in D, Op.10/3-III, mm. 1−16. From Schoenberg (1967).

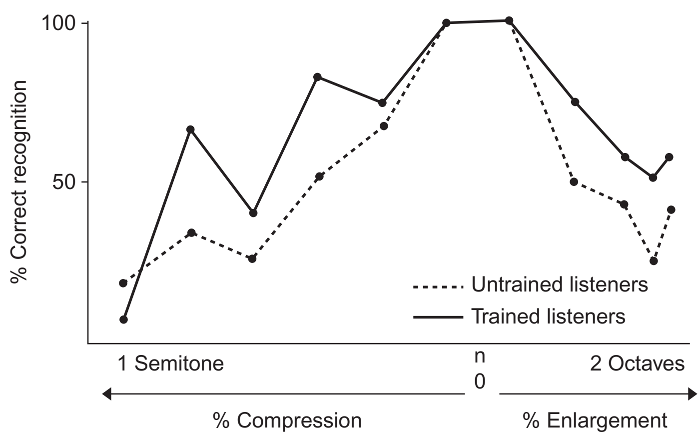

Figure 7 Percent correct recognition of melodies that have been transformed by compressing or enlarging the octave to differing extents.
Adapted from Vicario (1983).

the relative sizes of the intervals were retained, but their absolute sizes were altered. Further studies have confirmed that contour can serve as a salient cue to melody recognition (see, e.g., Croonen, 1994; Dowling, 1978; Dowling & Fujitani, 1971; Edworthy, 1985; Idson & Massaro, 1978; and Kallman & Massaro, 1979).

Further research has examined the cues that we use in judging similarity of contour. In much early work, contour was defined simply as the pattern of rises and falls in pitch, considering only temporally adjacent notes (cf. Dowling, 1978; Idson & Massaro, 1978). However, recent theoretical work has been concerned both with relationships between temporally adjacent notes and also with larger-scale features of contour (Marvin & LaPrade, 1987; Polansky & Bassein, 1992; Quinn, 1997). In an investigation of the relative salience of these two aspects of contour, Quinn (1999) constructed pairs of melodies that were either equivalent in note-to-note contour but not in the relationships between each note and the other notes in the melody, equivalent according to both criteria, or not equivalent according to either criterion. The subjects rated the degree of similarity between the members of each pair of melodies. The ratings indicated that note-to-note equivalence of contour played a primary role in similarity judgment, but that relationships between nonadjacent notes also had an influence.

Schmuckler (1999, 2004, 2009) adopted an alternative approach to contour perception. He characterized contour in terms of the relative degrees of strength of its cyclic information, as quantified by Fourier analysis. Schmuckler (2010) produced some interesting experimental support for this approach, though more findings are needed to evaluate it in detail.

E. Pitch Organization in Melody

We now turn to the related question of how listeners organize pitches so as to perceive coherent melodic phrases. As described in Chapter 6, pitch proximity is a powerful organizing principle in melody: We tend to group together tones that are close in pitch, and to separate out those that are further apart. When tones are presented at a rapid tempo, and these are drawn from two different pitch ranges, the listener perceives two melodic streams in parallel, one corresponding to the lower tones and the other to the higher ones—a phenomenon termed *stream segregation* (Bregman, 1990). However, pitch proximity also operates to group together tones when stream segregation does not occur. Hamaoui and Deutsch (2010) presented subjects with sequences of tones at interonset intervals of roughly 300 ms. The basic pattern consisted of a sequence of 12 tones that ascended or descended in semitone steps. Pitch distances of 2, 5, and 11 semitones were inserted between every three or four tones, and the subjects reported whether they heard the sequence as grouped into units of three or four tones each. When the sequences were isochronous, grouping by pitch proximity always occurred with the insertion of 5- or 11-semitone distances between successive tones, and such grouping even occurred to a statistically significant extent with the insertion of 2-semitone distances.

Grouping by pitch proximity is associated with substantial processing advantages. In a study by Deutsch (1978a), listeners compared the pitches of two tones that were separated by a sequence of intervening tones. As shown later in Figure 23, the smaller the average interval size formed by the intervening tones, the lower the error rate in judging whether the test tones were the same or different in pitch (see also Deutsch, 1974). Using a different paradigm, Aarden (2003) had subjects listen to folksong melodies. When each tone was played, subjects responded whether it was higher, lower, or identical in pitch to the previous tone. It was found that the closer the successive tones were in pitch, the faster were the subjects' reaction times.

The cognitive advantage conferred by smaller melodic intervals may account, at least in part, for the finding that in many cultures the frequency of occurrence of a melodic interval decreases as a function of its size. This has been shown in melodies from Africa, America, Asia, and Europe (Dowling, 1967; Huron, 2001; Merriam, Whinery & Fred, 1956; Ortmann, 1926). Further, in an analysis of melodic intervals in more than 4,000 folk songs, the average interval size formed by tones within phrases was 2.0 semitones, whereas that between tones at the end of one phrase and the beginning of the next was 2.9 semitones (Huron, 2006). This last finding indicates that smaller intervals serve to produce coherent relationships between tones within a phrase, and that larger intervals serve to separate out tones that cross phrase boundaries.

Repetition is also an important factor. We can note that there is a cross-cultural tendency for musical phrases to contain one or more tones that are repeated more often than others. From an experimental perspective, Deutsch (1970a, 1972a, 1975a) had listeners compare the pitches of two tones that were separated by a sequence of intervening tones, and found that repetition of the first test tone resulted in considerable memory enhancement for that tone (see also Section IV). Given these findings, we should expect that phrases containing repeated tones would be better remembered, and that the more often a tone is repeated, the better this tone would be remembered, so the greater would be its influence on the organization of the entire phrase.

So when we consider these two low-order effects together (i.e., grouping by pitch proximity and memory enhancement through repetition), we can see that there a considerable processing advantage is to be gained from a system in which there are a limited number of anchor tones—which are well remembered through repetition—surrounded by satellite tones that are linked to these anchor tones by pitch proximity. As argued by Deutsch (1982b), these two low-order effects acting together may well have influenced the development of musical systems across cultures. Erickson (1984) and Kim (2011) have also argued that such a principle—which Erickson termed "melodic tonal centering"—is a universal and possibly innate characteristic of tonal organization, which is not bound to any particular musical culture or historical period. A similar argument has been made by Thomson (1999, 2006) who proposed that melodies in different cultures share a type of organization that he termed "tonality frames," in which certain pitches serve as anchors in defining the pitch ranges of tones within melodies.

Another cross-cultural tendency was documented by Vos and Troost (1989) in an analysis of samples of music from Western classical composers, and from European and African-American folk songs. These authors found that large melodic intervals were more likely to ascend and small intervals to descend. Huron (2006) later extended these findings to samples of music from Australia, Asia, and Africa.

Meyer (1956) and Narmour (1990) have proposed that when presented with a melodic interval of small size, listeners expect to hear a further melodic interval that moves in the same direction. Evidence for this conjecture was obtained by Von Hippel (2002) in a study of anticipation judgments, and by Aarden (2003) in a reaction time study. For the case of large intervals, music theorists have observed that these generally tend to be followed by a change in direction—a tendency referred to as "post-skip reversal." Watt (1924), in analyses of Schubert lieder and Ojibway songs, found that as the size of an interval increased, the probability increased that the next interval would move in the opposite direction. Later, Von Hippel and Huron (2000) extended Watt's finding to traditional European, Chinese, South African, and Native American folk songs. Interestingly, Han, Sundararajan, Bowling, Lake, and Purves (2011) found that changes in pitch direction occurred more frequently, and intervals tended to be larger, in samples of music from tone language cultures than from nontone language cultures. And as expected, there were also more frequent changes in pitch direction and larger intervals in speech samples from tone language cultures. However, the general cross-cultural findings of post-skip reversal, and the prevalence of steps rather than skips in melodies, still held in this study.

The reason for the tendency for post-skip reversal has been a matter of debate. Meyer (1973) proposed that this occurs because listeners "want" to hear the gap produced by the large interval as filled with pitches lying within the gap. Von Hippel (2000) and Von Hippel and Huron (2000) later advanced an alternative explanation in terms of regression to the mean. Sampling a large number of melodies, they observed that pitches in most melodies formed a normal distribution, so that those in the center of a melody's range occurred most frequently, and the probability that a particular pitch would occur decreased with an increase in its distance from the center of the range. They argued, therefore, that most large intervals take a melody to an extreme of its range, creating the likelihood that the next pitch would be closer to the center. They obtained evidence for this view in a study of several hundred melodies from different cultures and periods. Interestingly, though, they also found—in line with Meyer's conjecture—that listeners expected large intervals to be followed by a change in direction, regardless of the location of the pitches relative to the center of the distribution.

The perceptual tendencies explored so far are related to Narmour's (1990, 1992) *implication-realization model* of musical expectations. Narmour proposed that listeners bring to their perception of melodies a number of expectations based on universal, and possibly innate, principles of music perception and cognition. One basic principle proposed by Narmour is that listeners expect small intervals to be followed by continuations in the same direction, and large intervals to be followed by a directional change. As another basic principle, Narmour proposed that listeners expect a small interval to be followed by one that is similar in size, and a large interval to be followed by one of smaller size. Narmour's principles have been the subject of substantial investigation (Cuddy & Lunny, 1995; Pearce & Wiggins, 2006; Schellenberg, 1996, 1997; Schmuckler, 1989; Thompson & Stainton, 1998), and considerable supporting evidence for them has been obtained. Variations of Narmour's model have also been proposed. For example, Schellenberg (1997) proposed a two-factor model of musical expectations involving pitch proximity and pitch reversal; further, to account for more global expectations, he extended the principle of proximity to noncontiguous tones.

Another important principle is the involvement of tonal schemata (Gjerdingen, 1988, 2007; Meyer, 1973). Certain musical patterns are prominent in works composed in particular musical styles, and these musical schemata and archetypes influence memory and perception of music in listeners who are familiar with the appropriate style (see also Kim, 2011).

In considering overall pitch relationships within phrases, two types of structure appear to occur quite commonly. Sachs (1962) has noted that in certain cultures and contexts, melodies are dominated by phrases that begin with a large ascending interval, and continue with a series of tones that descend in stepwise fashion. He termed these melodies "tumbling strains," and noted that they tend to occur, for example, in East European laments. A tendency has also been noted for phrases to rise and then fall in pitch, producing an arch-shaped contour. Huron (1996), in an analysis of phrases taken from more than 6,000 European folk songs, found that more than 40% of the analyzed phrases followed this pattern.

It is interesting to relate tumbling strains and melodic arch patterns to paralinguistic utterances, and to pitch patterns in exaggerated speech. Tumbling strains in laments bear a resemblance to wails that are produced in extreme distress, and may well derive in part from these. Also, both these contours bear strong resemblances to the exaggerated pitch patterns employed by mothers in communicating with preverbal infants—a form of speech termed "motherese." For example, mothers use falling pitches to sooth distressed infants, and they use steep arch-shaped contours to express approval or praise, as in saying "Go-o-od' girl!" Interestingly, these particular speech patterns occur in many different languages and cultures. Babies tend to respond appropriately even though they do not yet understand speech, even to phrases that are spoken in a foreign language (Fernald, 1993). We may then surmise that arch-shaped and falling pitch contours in music are related to a primitive and perhaps universal desire to produce such patterns in appropriate situations, and to a primitive impulse to respond to them.

III. Abstraction of Higher-Order Shapes

We next inquire into how higher-order abstractions are derived so as to lead to perceptual equivalences and similarities. We recognize visual shapes when these differ in size, position in the visual field, and to some extent in orientation. What transformations result in analogous equivalences in music?

Theorists have long drawn analogies between perception of pitch relationships and relationships in visual space (Helmholtz, 1859/1954; Koffka, 1935; Mach, 1906/1959). In contrast to visual space, however, pitch was conceived as represented along one dimension only. As Mach (1906/1959) wrote:

> *A tonal series is something which is an analogue of space, but is a space of one dimension limited in both directions and exhibiting no symmetry like that, for instance of a straight line running from right to left in a direction perpendicular to the median plane. It more resembles a vertical right line…*

Several investigators have shown that auditory analogues of visual grouping phenomena may be created by mapping one dimension of visual space into log frequency and the other into time (Bregman, 1990; Deutsch, 1975b; Van Noorden, 1975). The principle of proximity emerges clearly, for example, in the visual representation of the sequence shown in Figure 4 of Chapter 6. We may therefore inquire whether other perceptual equivalences in vision have analogues in the perception of music.

A. Transposition

Von Ehrenfels (1890), in his influential paper on form perception, pointed out that when a melody is transposed it retains its essential form, the *Gestaltqualitat*, provided the relations among the individual tones are preserved. In this respect, he argued, melodies are similar to visual shapes; these retain their perceptual identities when they are translated to different locations in the visual field.

A number of factors influence the extent to which a transposed and slightly altered melody is judged as similar to the original one. For example, when the original and transposed melodies can be interpreted as in the same key, and the successive tones comprising the melodies form the same number of steps along the diatonic scale, the melodies are generally judged as very similar to each other. This holds true whether or not the intervals forming the melodies are the same (Bartlett & Dowling, 1980; Dewitt & Crowder, 1986; Dowling, 1978, 1986; Takeuchi & Hulse, 1992; Van Egmond & Povel, 1994a, b), and can be taken to reflect the projection of pitch information onto overlearned alphabets, as proposed in the model of Deutsch and Feroe (1981) to be described later, and illustrated later in Figures 10 and 11.

Several researchers have hypothesized that the extent to which a transposed melody is perceived as related to the original one is influenced by the key distance between them. Key distance is defined in terms of distance along the cycle of fifths. So, for example, a melody that has been transposed from C major to G major is held to be more related to the original melody than one that has been transposed from C major to F♯ major (see, e.g., Bartlett & Dowling, 1980; Cuddy, Cohen, & Mewhort, 1981; Dowling, 1991; Dowling & Bartlett, 1981; Takeuchi & Hulse, 1992; Trainor & Trehub, 1993; Van Egmond & Povel, 1994a, 1994b).

Key distance has been found to affect melody recognition in complex ways (Dowling, 1991; Takeuchi & Hulse, 1992; Van Egmond & Povel, 1994b), and explanations for the obtained findings have been strongly debated (Dowling, 1991; Takeuchi & Hulse, 1992; Takeuchi, 1994; Van Egmond & Povel, 1994a). An important point here is that the closer two keys stand along the cycle of fifths, the larger the overlap of their pitch classes. For example, the C-major scale consists of pitch classes (C, D, E, F, G, A, B) and the G-major scale consists of pitch classes (G, A, B, C, D, E, F♯); these two scales therefore share six out of seven pitch classes. However, the F♯-major scale consists of (F♯, G♯, A♯, B, C♯, D♯, F); so the C-major and F♯-major scales share only two out of seven pitch classes. As described in Section IV, repetition of a pitch or pitch class strongly enhances its representation in short-term memory (Deutsch, 1970a, 1972a, 1975a). So when two melodies are presented in a short-term setting, and these are related by transposition, the salience of the tones in near-key transpositions should be considerably enhanced relative to those in far-key transpositions.

As a further short-term memory effect, when two tones are compared for pitch, and these are separated by a sequence of intervening tones, including in the intervening sequence a tone that is a semitone removed from the first test tone produces an increase in errors. Further, presenting two tones in the intervening sequence, one a semitone higher than the first test tone and the other a semitone lower, produces a substantial increase in errors (Deutsch, 1973a, 1973b, 1975c; Deutsch & Feroe, 1975; see also Section IV). Now when the C-major scale is presented followed by the G-major scale (a near-key transposition), only one of the seven tones of the G-major scale is preceded by tones that are both a semitone above and a semitone below it—namely, the tone F♯. However, when the C-major scale is presented followed by the F♯-major scale (a far-key transposition), five of the seven tones of the F♯-major scale are preceded by tones that are both a semitone above and a semitone below them—namely, the tones F♯, G♯, A♯, C♯, and D♯. So for far-key transpositions, tones are subject to a larger amount of short-term memory interference than are near-key transpositions. This difference in amount of interference should differentially affect comparison judgments of melodies that are related by near and far keys.

Key distance effects have also been invoked for triads; for example, the C-major triad is considered more related to the G-major triad than to the F♯-major triad. Experiments exploring these effects have generally employed the following paradigm: A prime context consisting of a chord or a sequence of chords is followed by a target chord, and subjects make a perceptual judgment on the target chord—such as an intonation or temporal asynchrony judgment. Targets have been found to be better processed when they were preceded by a harmonically related prime than when they were preceded by a less related prime (Bharucha & Stoeckig, 1986, 1987; Bigand, Tillmann, Poulin-Charronat, & Manderlier, 2005; Justus & Bharucha, 2002; Tillmann & Bharucha, 2002; Tillmann, Bigand & Pineau, 1998; Tillmann & Lebrun-Guillaud, 2006).

These findings are also equivocal in their interpretation. Although they have generally been attributed to acquired knowledge concerning chord progressions, short-term effects of repetition and interference could have played a role. Some complex effects of repetition have been found (see, e.g., Tekman & Bharucha, 1998); however, such effects have frequently not been controlled for, and there has been no control for specific effects of memory interference. For example, the C-major (C, E, G) and G-major (G, B, D) triads—which are considered closely related—have a tone in common—namely, G; further, only one pair of tones across these triads stand in semitone relation—namely, C and B. On the other hand, the C-major (C, E, G) and B-major (B, D♯, F♯) triads—which are considered unrelated—have no tones in common, and all three pairs of tones across these triads stand in semitone relation—namely, C and B, E and D♯, and G and F♯. So although it is reasonable to hypothesize that harmonic priming effects could be based on acquired knowledge of abstract relationships in tonal music, it is unclear to what extent these effects result from such acquired knowledge, and to what extent short-term memory effects are responsible.

Other factors have also been found to influence the similarity of transposed melodies. For example, several researchers have observed that the closer two melodies are in pitch range, the greater their perceived similarity (Francès, 1958/1988; Hershman, 1994; Van Egmond & Povel, 1994b, Van Egmond, Povel, & Maris, 1996). In addition, the coding model of Deutsch and Feroe (1981) has been used successfully as a predictor of perceived similarity between transposed melodies (Van Egmond & Povel, 1996), as described in Section III,D.

B. Inversion and Retrogression

We may next inquire whether further equivalences can be demonstrated for musical shapes that are analogous to their visuospatial counterparts. Schoenberg (1951) argued that transformations similar to rotation and reflection in vision result in perceptual equivalences in music also. He wrote:

> *The unity of musical space demands an absolute and unitary perception. In this space . . . there is no absolute down, no right or left, forward or backward . . . Just as our mind always recognizes, for instance, a knife, a bottle or a watch, regardless of its position, and can reproduce it in the imagination in every possible position, even so a musical creator's mind can operate subconsciously with a row of tones, regardless of their direction, regardless of the way in which a mirror might show the mutual relations, which remain a given quantity.*

This statement may be compared with Helmholtz's (1844) description of imagined visuospatial transformations:

> *Equipped with an awareness of the physical form of an object, we can clearly imagine all the perspective images which we may expect upon viewing it from this or that side. (see Warren & Warren, 1968, p. 252)*

On this basis, Schoenberg proposed that a row of tones may be recognized as equivalent when it is transformed in such a way that all ascending intervals become descending ones, and vice versa ("inversion"), when it is presented in reverse order ("retrogression"), or when it is transformed by both these operations ("retrograde-inversion"). Figure 8 illustrates Schoenberg's use of his theory in compositional practice. As Schoenberg (1951) wrote:

> *The employment of these mirror forms corresponds to the principle of the absolute and unitary perception of musical space.*

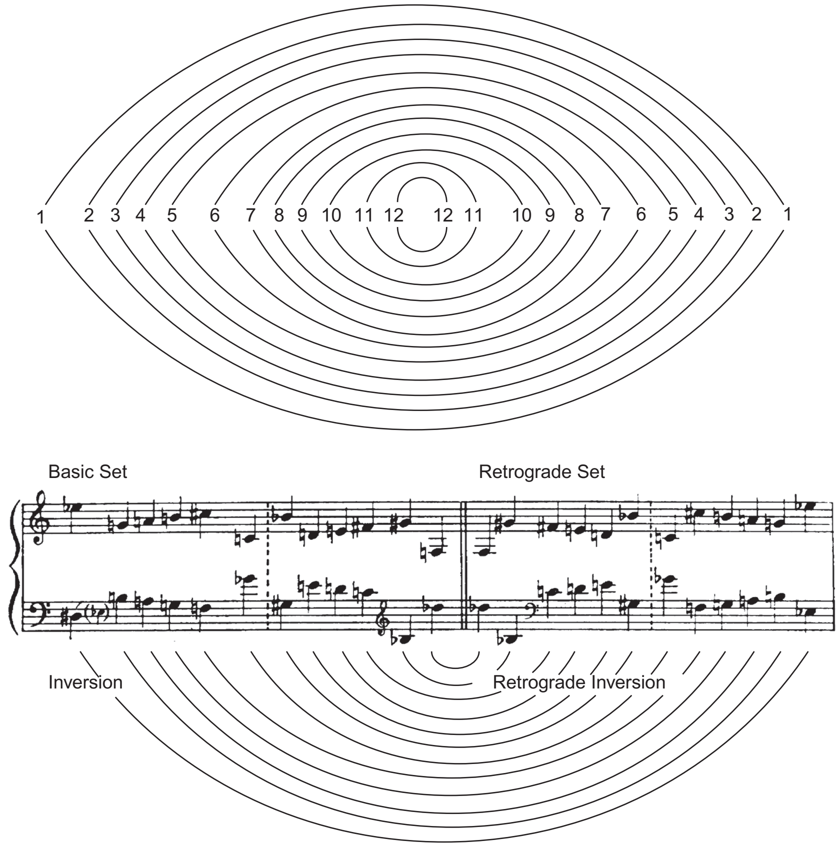

Figure 8 Schoenberg's illustration of his theory of equivalence relations between pitch structures, taken from his Wind Quartet, Op. 26.
From Schoenberg (1951).

Schoenberg did not conceive of the vertical dimension of musical space simply as pitch, but rather as pitch class. His assumptions of perceptual equivalence under transposition, retrogression, inversion, and octave displacement are fundamental to 12-tone composition (Babbitt, 1960). In this procedure, a given ordering of the 12 tones within the octave is adopted. The tone row is repeatedly presented throughout the piece; however, the above transformations are allowed on each presentation, and it is assumed that the row is perceived as an abstraction in its different manifestations.

Whether such transformations indeed result in perceptual equivalence is debatable. In the visual case, we must have evolved mechanisms that preserve the perceptual identities of objects regardless of their orientation relative to the observer. An analogous ecological argument cannot be made for inversion and retrogression of sound patterns. A second doubt is based on general experience. Sound sequences often become unrecognizable when they are reversed in time, as we can confirm by attempting to decode a segment of speech when it is played backward. Furthermore, many inverted three-note combinations are perceptually very dissimilar to the combinations from which they are derived. For example, a minor triad is an inversion of a major triad, yet the two are perceptually quite distinct from each other. It would appear, therefore, that when inverted and retrograde patterns are recognized, this is accomplished at a level of abstraction that is equivalent to the one that allows us to recite a segment of the alphabet backwards or to invert a series of numbers (Deutsch & Feroe, 1981). For further discussions of the perceptual status of 12-tone compositions, see Krumhansl, Sandell, and Sergeant (1987), Francès (1958/1988), and in particular Thomson (1991).

C. Models of Pitch Space

Over the centuries, theorists have proposed representations of pitch and pitch relationships in terms of distances in multidimensional space. For example, in order to capture the close perceptual similarity between tones that stand in octave relation, it has been suggested that pitch be represented as a helix, with the vertical axis corresponding to pitch height and tones separated by octaves lying closest within each turn of the helix (Section V; see also Chapter 6). More elaborate representations have also been proposed that would capture the complex patterns of pitch relationship that are invoked in listening to tonal music. For example, Longuet-Higgins (1962a, 1962b) has suggested that "tonal space" be characterized as a three-dimensional array: Tones that are adjacent along the first dimension are separated by fifths, those adjacent along the second dimension by major thirds, and those adjacent along the third dimension by octaves. The intervals of tonal music then appear as vectors in this tonal space. On this model, closely related tones, such as form a given major scale, produce a compact group in this array, so that a key can be defined as a neighborhood in tonal space. Similar representations have been proposed by others, such as Hall (1974), Balzano (1980), and Shepard (1982).

The spatial modeling of pitch relationships in the context of keys has a long tradition among music theorists. In particular, 18th century theorists developed circular configurations that would capture degrees of modulation between keys. In these models, adjacent positions along such circles depict close modulations, and positions that are further removed depict more distant ones. Later theorists such as Weber (1824) and Schoenberg (1954/1969) have produced related spatial models (Werts, 1983).

Leonard Meyer (1956) has argued that the mental representation of pitch relationships in classical tonal music is strongly influenced by hierarchies of relative stability and rest between tones in an established key. As he wrote:

> *The term "tonality" refers to the relationships existing between tones or tonal spheres within the context of a particular style system . . . some of the tones of the system*

are active. They tend to move toward the more stable points in the system—the structural or substantive tones.

But activity and rest are relative terms because tonal systems are generally hierarchical: tones which are active tendency tones on one level may be focal substantive tones on another level and vice versa. Thus in the major mode in Western music the tonic tone is the tone of ultimate rest toward which all other tones tend to move. On the next higher level the third and fifth of the scale, though active melodic tones relative to the tonic, join the tonic as structural tones; and all other tones, whether diatonic or chromatic, tend toward one of these. Going still further in the system, the full complement of diatonic tones are structural focal points relative to the chromatic notes between them. And, finally, as we have seen, any of these twelve chromatic notes may be taken as substantive relative to slight expressive deviations from their normal pitches. (Meyer, 1956, pp. 214−215)

The concept of a hierarchy of prominence for tones within a key was explored by Krumhansl (1979) in a study in which subjects judged similarities between pairs of tones that were presented in a tonal context. Multidimensional scaling of similarity ratings produced a three-dimensional conical structure around which tones were ordered according to pitch height. The components of the major triad formed a closely related structure near the vertex of the cone; the other tones in the major diatonic scale formed a less closely related subset that was further from the vertex, and the remaining pitch classes were more widely dispersed and still further from the vertex. These layers were then hypothesized to represent different degrees of stability for the pitch classes within a key.

There is a problem, however, with a representation that assigns to each pitch class a fixed degree of stability within a key regardless of the short term context in which it is embedded; a tone that is heard as highly stable in one context is heard as less stable in others. As a further problem, such a representation does not explain how the different pitch classes within a key are connected so as to form a unified whole. We need to know how tones at each hierarchical level are connected so as to form coherent patterns, and how such patterns are connected across hierarchical levels. Gjerdingen (1988, 2007), Narmour (1990, 1992), and Kim (2011) have all stressed that hierarchies in tonal music are formed of perceptually stable and closed tonal-temporal patterns, rather than nontemporal pitch hierarchies.

Deutsch and Feroe (1981) proposed a model for the mental representation of pitch sequences in tonal music in terms of tonal-temporal patterns that are linked together as hierarchies. The model also assumes that there is a hierarchy of pitch alphabets within an established key, though the role of any given pitch class depends on the short-term context in which it occurs. Pitch sequences composed of such alphabets at any one level form structural units at that level. Further, at each level, tones are elaborated by further tones at the next-lower level. Conversely, structural units at any one level contain tones that serve as reference points that unite to form structural units at the next-higher level.

A representation of Deutsch and Feroe's hierarchy of embedded alphabets is shown in Figure 9. The model assumes that, through extensive exposure to Western tonal music, the listener acquires this repertoire of embedded alphabets,

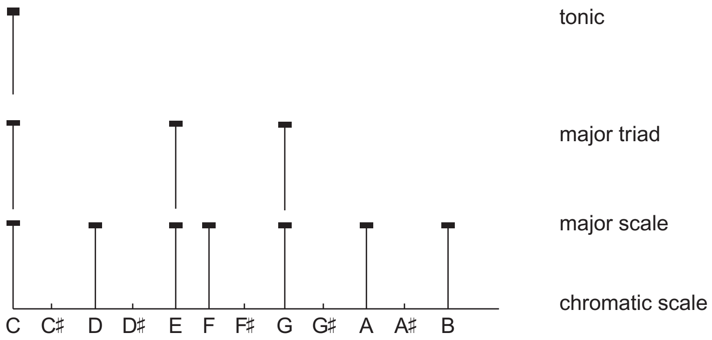

Figure 9 A hierarchy of embedded pitch alphabets.
Adapted from Deutsch and Feroe (1981). ©1981 by the American Psychological Association. Adapted with permission.

most prominently the chromatic scale, diatonic scales, and triads. At the lowest level, the chromatic alphabet serves as the parent alphabet from which families of subalphabets are derived. The major and minor scales are represented at the next-higher level; these can be expressed in terms of proximal distances along the chromatic alphabet. Triads are represented at the next-higher level; these can be expressed in terms of proximal distances along diatonic alphabets. Lerdahl (2001) has proposed an elaboration of Deutsch and Feroe's hierarchy of alphabets that also takes account of a number of other characteristics of tonal music, such as patterns of proximity between chords (see Lerdahl, 2001, p. 47.)

Compositional practice reflects our use of such overlearned alphabets. For example, in the short-term transposition of motives, the number of steps along an alphabet is often preserved, so that even when such transpositions result in alterations in interval size, they still appear appropriate to the listener. Figures 10 and 11 give two such examples. The first, from a Bach fugue, shows a motive that traverses the D-major scale four times in succession, each time beginning on a different position along the scale. The second, from a Schubert impromptu, shows a motive that traverses the A♭-minor triad five times in succession, each time beginning at different positions along the triad. In both cases, preservation of the pitch alphabet has the consequence that the intervals vary in the different instantiations of the motive (Deutsch, 1977, 1978d).

There is experimental evidence that pitch structures in Western tonal music are represented by listeners in terms of such embedded alphabets. Deutsch (1980) had subjects listen to sequences of tones that were drawn from such alphabets, and recall what they heard in musical notation. When errors in notation occurred, they rarely departed from the alphabet that had been presented. (So, for example, if a sequence consisted of tones in the G-major triad, erroneous notations would also be in the G-major triad.) In general, sequences were recalled very accurately when they could be simply represented as hierarchical structures, with different pitch alphabets at different levels of the hierarchy (see below).

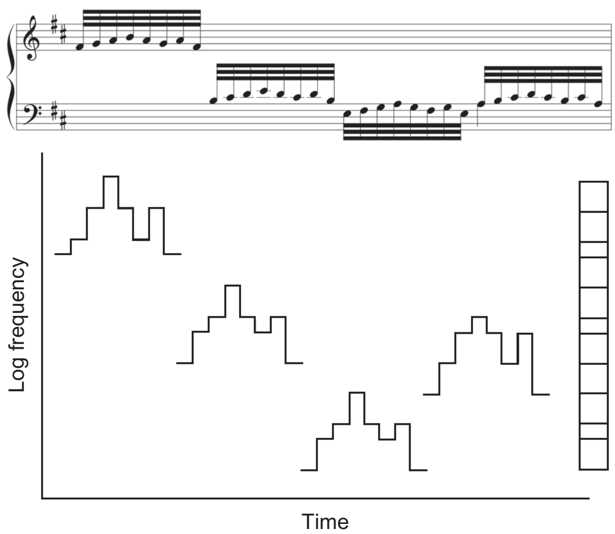

Figure 10 Transposition along the alphabet of the D-major scale. The same pattern is presented four times in succession at different positions along the scale. Because the major scale consists of unequal intervals, there result differences in the intervals comprising the pattern. The ladder at the right displays the scale. From J. S. Bach, *The Well-Tempered Clavier*, Book 1, Fugue V.
From Deutsch (1977).

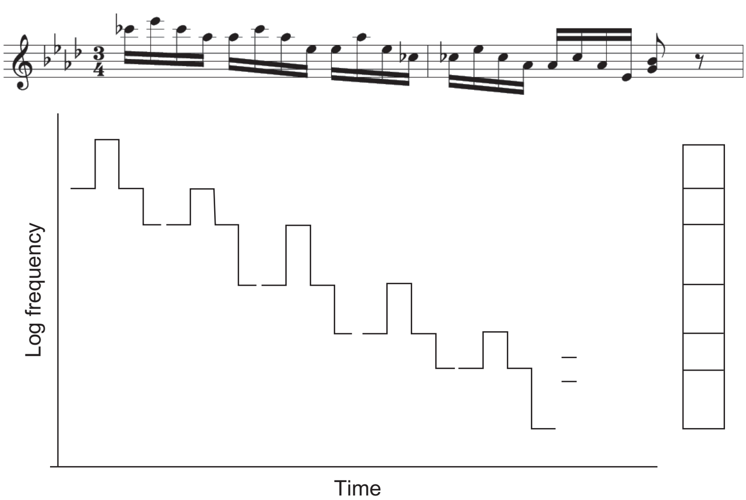

Figure 11 Transposition along the alphabet of the A♭-minor triad. The same pattern is presented five times in succession, at different positions along this triad. Because the triad consists of uneven intervals, there result differences in the intervals comprising the pattern. The ladder at the right displays the triad.
From F. Schubert, *Four Impromptus*, Op. 90, No. IV.

Further evidence comes from findings that melodies were better remembered when they were composed only of tones in a particular diatonic set than when they also contained tones outside the set (Cuddy et al., 1981; Dowling, 1991; Francès, 1958/1988). Presumably, adhering to a diatonic set increases the likelihood that the listener would invoke a key, and so use overlearned pitch alphabets as an aid to memory. It has also been reported that altering the context of a melody so as to suggest a different key rendered the melody more difficult to recognize (Dowling, 1986). Yet other studies have found that transposed melodies that did not involve a change in key were judged as very similar to the original ones, regardless of whether or not the intervals were preserved (Bartlett & Dowling, 1980; Dewitt & Crowder, 1986; Dowling, 1978, 1986; Takeuchi & Hulse, 1992; Van Egmond & Povel, 1994b). In addition, an alteration in a melody has been found easier to detect when it could be interpreted as a departure from its key and so as departing from the alphabets appropriate to the key (Francès, 1958/1988, Dewar, Cuddy, & Mewhort, 1977; Dowling, 1978).

D. The Deutsch/Feroe Model

The model proposed by Deutsch and Feroe (1981) (hereafter termed D&F) describes how pitch sequences in tonal music are encoded and represented in memory. Music theorists have argued that Western tonal music is composed of segments that are organized in hierarchical fashion (Lerdahl & Jackendoff, 1983; Meyer, 1956, 1973; Narmour, 1990, 1992; Schenker, 1956), and it is reasonable to suppose that this form of organization reflects the ways in which musical information is encoded and retained. As Greeno and Simon (1974) point out, we appear to retain many different types of information as hierarchies. We also appear to retain hierarchies of rules (Scandura, 1970), of programs (Miller, Galanter, & Pribram, 1960), and of goals in problem solving (Ernst & Newell, 1969). Visual scenes appear to be retained as hierarchies of subscenes (Palmer, 2002). The phrase structure of a sentence lends itself readily to hierarchical interpretations (Miller & Chomsky, 1963).

Restle (1970) and Restle and Brown (1970) have provided evidence that we readily acquire serial patterns as hierarchies that reflect the structure of these patterns. Parallel theoretical developments by Simon and his colleagues (Simon, 1972; Simon & Kotovsky, 1963; Simon & Sumner, 1968) and by others (Jones, 1978; Leewenberg, 1971; Vitz & Todd, 1969) have addressed the ways in which we acquire and retain serials patterns in terms of hierarchies of operators.

The D&F model is in the coded-element tradition, but it differs fundamentally from others in its basic architecture. The structural units of the model are sequences that are organized in accordance with universal grouping principles, such as proximity and good continuation. Structural units can also be based on schemata that have been acquired through exposure to the music of a particular tradition. These structural units combine to form a hierarchical network, in which elements that are present at any given level are elaborated by further elements so as to form structural units at the next-lower level, until the lowest level is reached. It should be

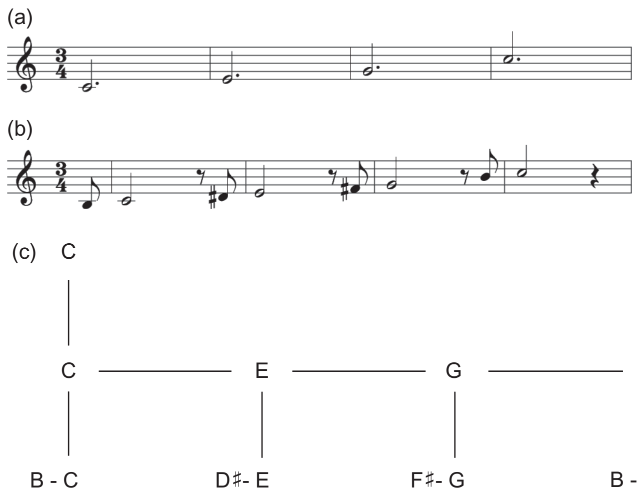

Figure 12 A series of pitches represented on two hierarchical levels. (a) At the higher level, there is an arpeggiation of the C-major triad. (b) At the lower level, each note of the triad is preceded by one a semitone lower, so forming a two-note pattern. (c) The hierarchical structure as a tree diagram.
Adapted from Deutsch and Feroe (1981). ©1981 by the American Psychological Association. Adapted with permission.

emphasized that although the model focuses on Western tonal music of the common practice era, it can equally well be applied to the music of other periods and cultures, and it assumes only that, through long-term exposure to music in a given style, listeners have become familiar with the pitch alphabets of the music in that style.

The model is introduced by a musical example. The pitch sequence shown in Figure 12b can, in principle, be represented in terms of steps along the chromatic scale: A basic subsequence consisting of a step up this scale is presented four times in succession, the second instantiation being four steps up from the first, the third being three steps up from the second, and the fourth being five steps up from the third. This analysis assigns prominence to the basic subsequence and does not relate its different instantiations in a meaningful way. A musical analysis of this pattern would instead describe it as on the two structural levels shown in Figures 12a and 12b. The basic relationship expressed here is that of the elaboration of a higher-level subsequence by a lower-level subsequence. The higher level, shown in Figure 12a, consists of an arpeggiation that ascends through the C major triad (C-E-G-C). At the lower level, each note of the triad is preceded by a neighbor embellishment, so that the two-note patterns (B-C), D♯-E), (F♯-G), (B-C) are formed. Figure 12c represents this hierarchical structure in tree form.

Specifically, a simplified version of the D&F model is as follows:

1. A structure is notated as $(A_1, A_2, \ldots, A_{l-2}, A_{l-1}, *, A_{l+1}, A_{l+2}, \ldots, A_n)$, where A_j is one of the operators n, p, s, n^i, or p^i. The asterisk (*) provides a reference point for the other operators, and appears exactly once in the structure.

2. Each structure $(A_1, A_2, \ldots, *, \ldots, A_n)$ has associated with it an alphabet, α. The combination of a structure and an alphabet is called a *sequence* (or *subsequence*). This, together with the reference element r, produces a *sequence of notes*.

3. The effect of each operator in a structure is determined by that of the operator closest to it, but on the same side as the asterisk. Thus the operator n refers to traversing one step up the alphabet associated with the structure. The operator p refers to traversing one step down this alphabet. The operator s refers to remaining in the same position. The two operators n^i and p^i refer to traversing up or down i steps along the alphabet, respectively.

4. The values of the sequence of notes $(A_1, A_2, \ldots, *, \ldots, A_n)$, α, r, where α is the alphabet and r the reference element, are obtained by taking the value of the asterisk to be that of r.

5. To produce another sequence from the two sequences $A = (A_1, A_2, \ldots, *, \ldots, A_m)\ \alpha$, and $B = (B_1, B_2, \ldots, *, \ldots, B_n)$, β, where α and β are two alphabets, we define the compound operator pr (prime). *A*[pr]*B*;r, where r is the reference element, refers to assigning values to the notes produced from $(B_1, B_2, \ldots, *, \ldots, B_n)$ such that the value of the asterisk is the same as the value of A_1, when the sequence *A* is applied to the reference element r. Values are then assigned to the notes produced from $(B_1, B_2, \ldots, *, \ldots, B_n)$ such that the value of the asterisk is the same as the value of A_2, and so on. This gives a sequence of length $m \times n$. Other compound operators such as inv (inversion) and ret (retrograde) are analogously defined.

So according to the formalism just outlined, the pattern shown in Figure 12 can be represented as:

$$A = (*,\ 3n)C_{tr}$$
$$B = (p,\ *)Cr$$
$$S = A[pr]B,\ C_4$$

where C_{tr} represents the C-major triad, Cr the chromatic scale, and C_4 the reference element.

In other words, sequence *A* consists of a reference point followed by two successive steps along the C-major triad. Sequence *B* represents an ascending half-step that ends on a reference point. To combine these two sequences so as to produce the full sequence, the reference element C_4 replaces the reference point in sequence *A*; this produces the sequence of notes $(C_4{-}E_4{-}G_4{-}C_5)$. The sequence *B* is then applied to each note of sequence *A*, taking each note of sequence *A* as the reference point. This produces the entire sequence of notes (B_3-C_4-$D\sharp_4$-E_4-$F\sharp_4$-G-B_4-C_5).

In many other hierarchical representations of music, such as proposed by Schenker (1956) and the coded element models referred to earlier, elements that are present at all but the lowest level are rule systems rather than actual notes.

In contrast, in the D&F model, an actual sequence of notes is realized at each structural level. This confers the advantage that notes that are present at any given level are also present at all levels below it. In consequence, the higher the level at which a note is represented, the more often and so the more firmly it is represented. This has the consequence that higher-level subsequences serve to cement lower level subsequences together. As a further advantage, by repeatedly invoking the same structure, the model enables long sequences to be encoded in parsimonious fashion—essentially acting as a compression algorithm. A related processing advantage is that the model enables subsequences at different structural levels to be encoded as chunks of a few items each; this in turn is conducive to optimal memory performance (Anderson, Reder, & Lebiere, 1996; Estes, 1972; Wickelgren, 1967).

As another processing advantage, the D&F model enables the encoding of subsequences in terms of laws of figural goodness, such as proximity and good continuation, and also enables the invocation of melodic schemata and archetypes in the representation of subsequences. This has the effect of binding the tones within subsequences together, and so also helps the listener to apprehend and remember the full sequence. As yet a further advantage, the model enables different pitch alphabets to be invoked at different hierarchical levels. The use of multiple alphabets here has the benefit of helping to clarify and disambiguate the different levels of the hierarchy.

Experimental evidence indicates that listeners process pitch sequences in accordance with the D&F model when given the opportunity to do so. One hypothesis that arises from the model is that a sequence of notes should be processed more easily when it can be parsimoniously represented in accordance with its rules. In an experiment to test this hypothesis, Deutsch (1980) presented musically trained listeners with sequences of notes, which they recalled in musical notation. Examples of these sequences are shown in Figure 13. The passage in Figure 13a (a "structured sequence") consists of a higher-level subsequence of four elements that acts

Figure 13 Examples of sequences used in the experiment to study utilization of pitch structure in recall. Sequence (a) can be represented parsimoniously as a higher-level subsequence of four elements (an arpeggiation of the G-major triad) that acts on a lower-level subsequence of three elements (a step down and then up the chromatic scale). Sequence (b) consists of a haphazard reordering of the notes in sequence (a) and cannot be parsimoniously represented.
Adapted from Deutsch (1980).

Figure 14 Types of temporal structure used in the experiment to study the utilization of pitch structure in recall. (a) Sequence unsegmented. (b) Sequence segmented in groups of three, so that segmentation is in accordance with pitch structure. (c) Sequence segmented in groups of four, so that segmentation is in conflict with pitch structure.

on a lower-level subsequence of three elements. The passage in Figure 13b (an "unstructured sequence") consists of a haphazard reordering of the passage in Figure 13a, and does not lend itself to a parsimonious representation. It was predicted, on the basis of the model, that the structured sequences would be notated more accurately than the unstructured ones.

Another factor was also examined in this experiment. It has been found in studies using strings of verbal materials that we tend to recall such strings in accordance with their temporal grouping (Bower & Winzenz, 1969; McLean & Gregg, 1967; Mueller & Schumann, 1894). This effect was found to be so powerful as to offset grouping by meaning (Bower & Springston, 1970). Analogous results have also been obtained using nonverbal materials (Dowling, 1973; Handel, 1973; Restle, 1972). It was predicted, therefore, that temporal grouping would affect ease of recall of the present tonal sequences in analogous fashion. In particular, temporal grouping in accordance with pitch structure was expected to enhance performance, whereas grouping in conflict with pitch structure was expected to result in performance decrements. See London (2012) for an excellent discussion of the effects of timing on perception of pitch structures.

Given these considerations, sequences such as these were presented in three temporal configurations (Figure 14). In the first, the tones were spaced at equal intervals; in the second, they were spaced in four groups of three, so that they were segmented in accordance with pitch structure; in the third, they were spaced in three groups of four, so that they were segmented in conflict with pitch structure.

Large effects of both pitch structure and temporal segmentation were obtained. For structured sequences that were segmented in accordance with pitch structure, performance levels were very high. For structured sequences that were unsegmented, performance levels were still very high, though slightly lower. However, for structured sequences that were segmented in conflict with pitch structure, performance levels were much lower. For unstructured sequences, performance levels were considerably lower than for structured sequences that were segmented in accordance with their structure or that were unsegmented; instead, they were in the same range as for structured sequences that were segmented in conflict with pitch structure.

Figure 15 shows the percentages of tones that were correctly recalled in their correct serial positions in the different conditions of the experiment. Typical bow-shaped curves are apparent, and in addition, discontinuities occur at the boundaries between temporal groupings. This pattern of results indicates that the subjects encoded the temporal groupings as chunks, which were retained or lost independently of each other. This pattern is very similar to that found by others with the use of verbal materials (Bower & Winzenz, 1969).

The transition shift probability (TSP) provides a further measure of interitem association. This is defined as the joint probability of either an error following a correct response on the previous item, or of a correct response following an error on the previous item (Bower & Springston, 1970). If groups of elements tend to be retained or lost as chunks, we should expect the TSP values to be smaller for transitions within a chunk, and larger for the transition into the first element of a chunk. It was indeed found that TSPs were larger on the first element of each temporal grouping than on other elements. This is as expected on the hypothesis that temporal groupings serve to define subjective chunks that are retained or lost independently of each other.

In general, the findings of Deutsch (1980) provide strong evidence that listeners perceive hierarchical structures that are present in tonal sequences, and that they use such structures in recall. For the structured sequences used here, the listener needed only to retain two chunks of three or four items each; however, for the unstructured sequences, no such parsimonious encoding was possible. The error rates for the unstructured sequences were much higher than for the structured sequences, in accordance with the hypothesis that they imposed a much heavier memory load.

Another study was carried out by Van Egmond and Povel (1996). A paired comparison paradigm was employed to investigate perceived similarities between melodies and their transpositions, when the latter had been altered in various ways. The D&F model was used as a qualitative predictor of the degree of perceived similarity between the original and transposed melodies. The authors hypothesized that the larger the number of items by which the codes for the original and transposed melodies differed, the more dissimilar the two melodies would appear.

More specifically, Van Egmond and Povel predicted that an exact transposition would be judged as most similar to the original melody, because its code would differ only in terms of one item; i.e., the key. For a transposition that was chromatically altered, the prediction concerning perceived similarity would depend on whether the transposed melody could be represented parsimoniously in the same key as the original. If it could be so represented, then its code would differ in terms of only one item—the reference element. If it could not be so represented, then its code would differ in terms of two items—the key and the reference element. Finally, a transposition that was diatonically altered would be judged as most

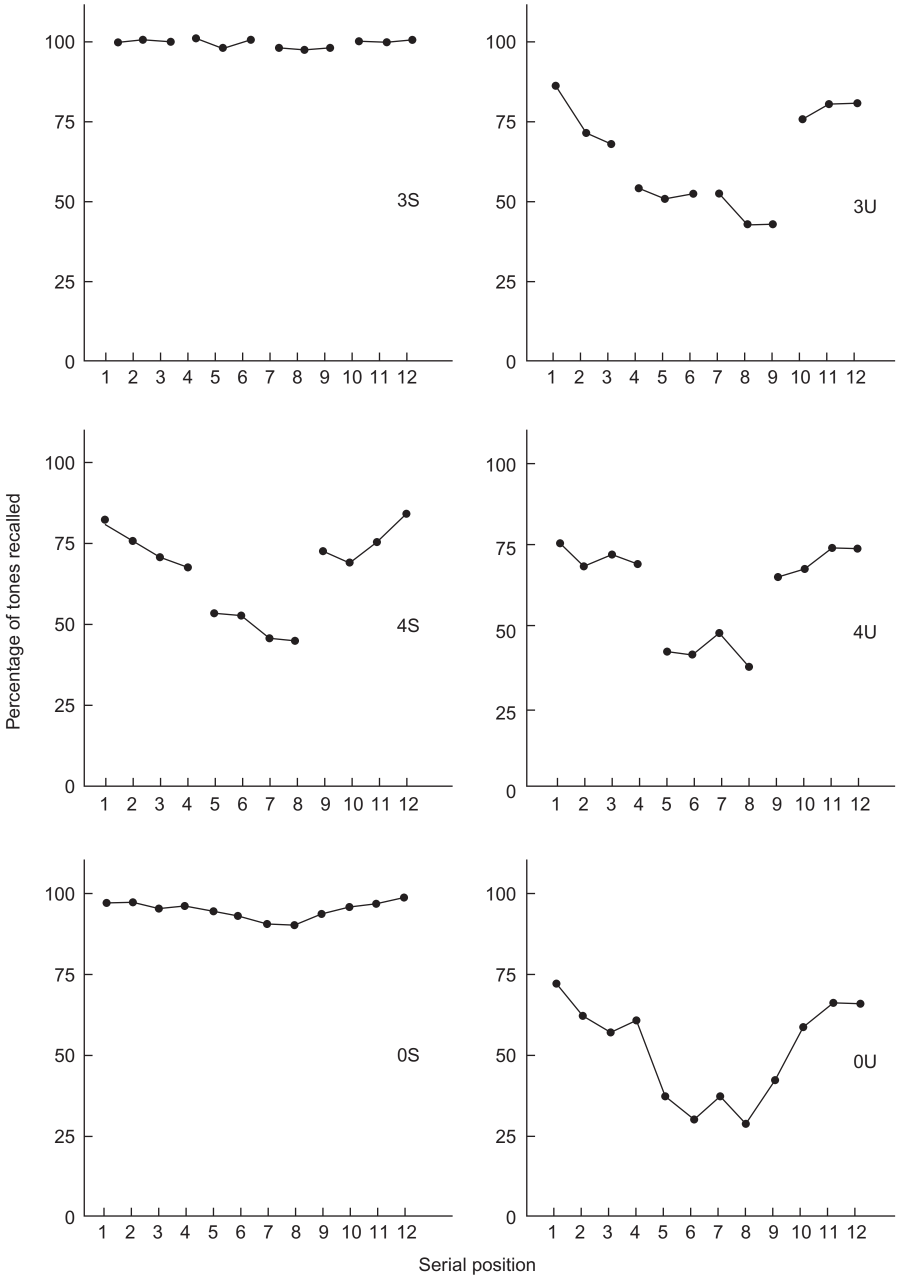

Figure 15 Serial position curves for the different conditions of the experiment to study the utilization of pitch structure in recall. 3: Temporal segmentation in groups of three. 4: Temporal segmentation in groups of four. 0: No temporal segmentation. S: Structured sequence. U: Unstructured sequence.
From Deutsch (1980).

dissimilar to the original melody, because its code would differ in terms of six items—the key and five structure operators.

The experimental findings confirmed the hypothesis. Exact transpositions were judged to be most similar to the original melodies. Chromatically altered transpositions that could be interpreted as in the same key as the original melodies were judged to be more similar than were those that could not be so interpreted. Transpositions that were diatonically altered were judged to be more dissimilar than were chromatically altered transpositions.

In a further set of experiments, Hamaoui and Deutsch (2010) constructed two groups of sequences. Those in one group could be parsimoniously represented in hierarchical fashion according to the D&F rules. Those in the other group were unstructured, but they matched the structured sequences in terms of starting pitch, number of changes in pitch direction, overall pitch movement, and interval size content. The effect of grouping by hierarchical structure—as measured by the duration of conflicting temporal gaps required to overrule it—was found to be remarkably strong.

In yet another study, Oura (1991) presented subjects with a melody, which they recalled in musical notation. Tones that were represented at higher structural levels were recalled better than were those that were represented at lower levels. Further, Dibben (1994) had subjects listen to a musical segment, and then to a pair of reductions, and they judged which reduction best matched the full segment. She found that the subjects chose the version that matched the full segment at higher structural levels. The findings from both these studies are in accordance with the prediction from the D&F model, that the higher in a tonal-temporal hierarchy a note or sequence of notes is represented, the more often it is represented, and so the more firmly it should be embedded in memory (see also Wang & Sogin, 1990).

So far we have been considering the processing of a single melodic line. However, tonal music generally involves several such lines, and even where only one is presented, a harmonic progression is generally implied. We can assume that such progressions are also encoded in hierarchical fashion. In addition, the use of parallel linear sequences, which must also combine to form an acceptable harmonic sequence, places constraints on the choice of elements in each sequence; this in turn serves to reduce the processing load.

E. Acquisition of a Representation

We now consider how the D&F model addresses the process by which the listener acquires a representation of a passage. The model assumes that an initial set of subsequences is formed on the basis of simple organizational principles, such as proximity and good continuation. We can also assume that the listener's choice of a dominant note in a subsequence—which then serves as a reference point—is also initially guided by low-level factors, such as an increase in loudness or duration, metrical stress, and the temporal position of the note in the subsequence.

We can consider, as an example, the sequence in Figure 16, which was derived from Figure 1 of Deutsch and Feroe (1981). This pattern can be described as an arpeggiation that ascends through the C-major triad (E-G-C) with each note of the triad preceded by a neighbor embellishment. In other words, the notes E, G, and C are targeted for representation at a higher level, as shown in the associated tree diagram. As a result, the key of C major is clearly attributed, even though two of the notes in the sequence (D♯ and F♯) are outside the C-major scale. However, when the identical sequence of notes is played in reverse order, as shown in Figure 17, it is no longer heard as in C major, but instead as in E minor. We target the notes B, F♯, and D♯ so as to form the subsequence (B-F♯-D♯) at the next-higher level, as shown in the associated tree diagram. So we hear an arpeggiation that descends through the B-major triad, and we hear it as implying the dominant of E minor, leading us to attribute the key of E minor instead.

Deutsch (1984) suggested that this paradoxical outcome is based on the following process: Using primitive organizational principles, the listener forms low-level groupings from the two-note patterns that are related by proximity, and then assigns prominence to the second note of each two-note pattern. This leads to the assignment of the subsequence (E-G-C) at the higher level when the sequence is played forward, and (B-F♯-D♯) when the sequence is played backward.

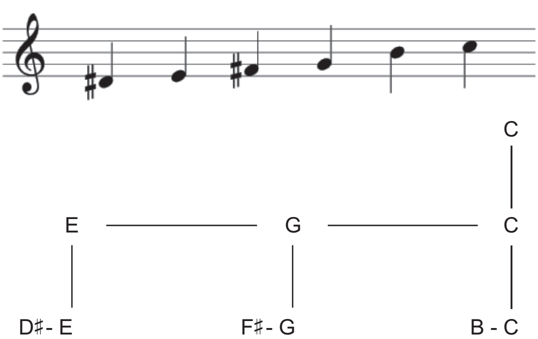

Figure 16 Pitch sequence to illustrate the effect of a particular temporal ordering on a given collection of tones. This sequence is heard as in C major although two tones are outside the C-major scale. The tree diagram illustrates the hypothesized mental representation of this sequence.
Adapted from Deutsch (1984).

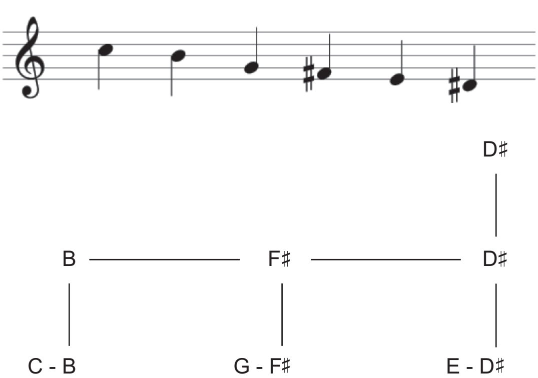

Figure 17 The identical pitch sequence as in Figure 16, but presented in reverse order. The tree diagram illustrates the hypothesized mental representation of this sequence, which is heard as in E minor.
Adapted from Deutsch (1984).

As another example of the strong influence of ordering, we can consider the passages shown in Figure 13. Passage (a) (G-F♯-G-D-C♯-D-B-A♯-B-G-F♯-G) clearly invokes the key of G major, even though two of the notes (C♯ and A♯) are outside the G-major scale. Again, the listener forms low-level groupings based on pitch proximity (G-F♯-G, and so on), and targets the notes (G-D-B-G) to form a subsequence at the next-higher level. However, when the same set of notes is played in haphazard order, as in Passage (b), the listener cannot form a parsimonious hierarchical representation of the passage, so the key becomes ambiguous. So the D&F model and the associated experiments clarify that (1) key assignments can be readily made for passages that include tones outside the scale for the assigned key, (2) they are strongly dependent on the ordering of the notes in the passage, and (3) listeners can use simple organizational principles based on ordering to create a hierarchical structure from these notes, and so to assign a key to the passage.

Kim (2011) has addressed the important question of why the listener chooses the second of each pair of notes in the examples in Figures 16 and 17 as the dominant note. He pointed out that melodic steps have been proposed by several music theorists to have an inhibitory effect. For example, Komar (1971) described the second note of a linear pair as the "stepwise displacement" of the first note. Further, Larson (1997) observed that this concept relies on the distinction between steps and leaps: In a melodic step (defined as a distance of one or two semitones), the second note tends to displace the trace of the first note in memory, so that it becomes the more prominent note. Kim proposed, therefore, that resulting from stepwise displacement, the listener perceives the second note of each two-note grouping as more prominent, and so targets this note for representation at a higher structural level.

Bharucha (1984a, 1984b) has advanced the alternative proposal that the listener needs to assign a key in order to generate a hierarchy of prominence of notes within a passage. In other words, he proposed that the decision as to which notes assume prominence is driven by the internalized knowledge of hierarchies of prominence within a key (see also Krumhansl, 1990). In contrast, Kim (2011), while

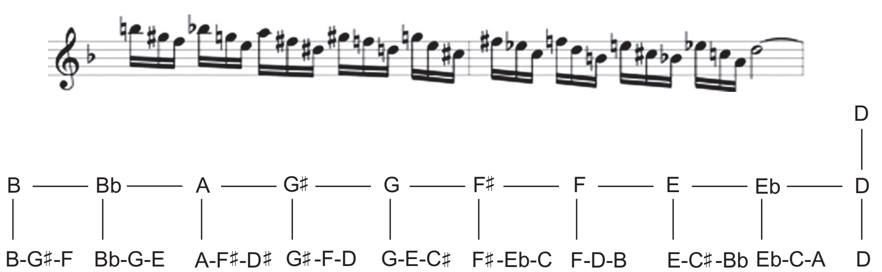

Figure 18 Passage with a clear hierarchical structure independent of key. The higher-level subsequence consists of a descending chromatic scale, and the lower-level subsequences are all diminished triads. See text for details. From Prelude VI in D minor, by J. S. Bach.

acknowledging that top-down processing is also invoked, including making reference to an established key, contended that bottom-up processes are heavily involved in establishing hierarchies of prominence.

In this context, we can observe that the D&F model does not require that listeners first attribute a key in order to acquire a hierarchical representation of a passage. The passage in Figure 18, taken from Bach's Prelude in D Minor, consists of a higher-level subsequence that traverses the chromatic scale from B_5 down to D_5. Each note in this subsequence is elaborated by an arpeggiation that descends through the diminished triad. The full sequence so produced can be notated as:

$$A = (9n,\ *)\ \mathrm{Cr}$$
$$B = (*,\ 2p)\ \mathrm{dim}_{tr}$$
$$S = A\ [\mathrm{pr}]B,\ D_5$$

where Cr indicates the chromatic alphabet, and dim_{tr} indicates the diminished triad. The sequence ends on the note D (the tonic) but could in principle have ended on any note in the chromatic set. So rather than relying on an established key, these hierarchical representations play a large role in the process of key identification itself, through an elaborate bootstrapping operation in which different cues feed back on each other.

F. Other Approaches to Key Identification

A number of approaches to key identification have been taken, and these fall into several categories (see also Temperley, Chapter 8). One approach holds that listeners possess a template that represents the distribution of pitch classes for each of the 12 major and minor keys. When a piece is heard, its pitch class distribution is compared with that in each of the templates, and the template that provides the best match wins. This view assumes that the ordering of the different pitch classes in a piece is ignored, with only the statistical distribution of the pitch classes remaining.

An early model in this category was proposed by Longuet-Higgins and Steedman (1971). As a first pass, the model assumes that as a note is sounded, it eliminates all keys whose scales do not contain that note. This process continues until only one candidate key remains. A problem with this model is that it cannot account for correct key identifications of passages containing notes that are outside the scale for that key, as in the examples given in Figures 12 and 13a.

Krumhansl and Schmuckler (1986; see also Krumhansl, 1990) proposed a distributional model based on a set of "key profiles," which were derived from a study by Krumhansl and Kessler (1982) (hereafter termed K&K). To generate the profiles, musically trained subjects were presented with a musical context (a scale, chord, or cadence) that was followed by a probe tone, and they judged how well the probe tone fit in the context provided. Probe-tone ratings were obtained for all 12 pitch classes in each context. The ratings from the different keys and contexts were then averaged so as to generate a single major-key profile and a single minor-key profile.

The procedure used to generate the K&K profiles has been criticized on a number of grounds. In particular, in averaging across the contexts provided to the subjects, taking the major and minor keys separately (the procedure used by K&K), one obtains distributions of the number of repetitions of each pitch class that correspond remarkably well to the profiles obtained from the subjects' rating judgments (Butler, 1989). The profiles could, therefore, simply reflect enhancement by repetition in short-term memory. Later, Huron and Parncutt (1993) and Leman (2000) produced models that simulated K&K's probe tone data, but were based on short-term memory effects. Further, Deutsch (1970a, 1972a, 1975a) observed that repetition of the pitch of a tone in an atonal setting resulted in memory enhancement for that tone; these findings produced direct evidence that a probe tone should be heard as more salient as a function of its repeated presentation in the context provided (see Section IV). In other work, Oram and Cuddy (1995) and Creel and Newport (2002) carried out probe-tone studies employing context melodies that were generated from artificial pitch class distributions designed to be very dissimilar to those in major or minor scales. The subjects' judgments correlated with the pitch class distributions in the context melodies, so that those pitches that occurred more often in the context were given higher ratings—findings that are again attributable to repetition effects in short-term memory. In sum, since probe tone ratings are strongly influenced by short-term contexts, they cannot be assumed by default to reflect long-term exposure to music of the listener's tradition.

Another argument has been advanced by Temperley and Marvin (2008), Aarden (2003), and Huron (2006), based on statistical analyses of large samples of Western tonal music. These authors found that although the K&K profiles correlated with the distributions of pitch classes within keys, the correlations were imperfect, and for certain scale degrees there were substantial discrepancies between the profiles and the actual distributions.

At all events, the Krumhansl and Schmuckler algorithm adds information about note duration to the K&K profiles, and then determines the key of a passage (or piece) by comparing its pitch class distribution with the amended K&K profiles for each of the 12 major and minor keys and choosing the one with best fit. Other models based on the distribution of pitch classes in a passage or a piece have been proposed by Chew (2002), Vos and Van Geenen (1996), Yoshino and Abe (2004), and Temperley (2007).

The distributional approach to key finding has been criticized on the grounds that it neglects the effect of temporal ordering of the pitches in a passage. Several alternative approaches that emphasize temporal ordering have been proposed. Most prominently, Butler, Brown, and colleagues (Brown, 1988; Brown & Butler, 1981; Brown, Butler, & Jones, 1994; Browne, 1981; Butler, 1989; Butler & Brown, 1984; Van Egmond & Butler, 1997) have contended that key identification is strongly influenced by the presence of rare intervals within a key; in particular, minor seconds and tritones. Their work has focused on the tritone, which in the major scale only occurs between two scale degrees ($\hat{4}$ and $\hat{7}$). Even considering the tritone, ordering is important: for example, F-B implies the key of C whereas B-F implies the key of F♯. Vos (1999) also emphasized the importance of certain melodic intervals for key identification. Specifically, he proposed that a rising fifth or a descending fourth at the beginning of a melody provides important cues. In addition, Callender, Quinn, and Tymoczko (2008) have proposed a substantial model of voice leading that emphasizes the ordering of chord progressions.

Evidence for the strong influence of ordering was provided in the study by Deutsch (1980) discussed earlier. It was shown that a set of pitches that were ordered in such a way that they could be encoded parsimoniously as phrases in tonal music were easily processed, whereas the same set of pitches that were reordered haphazardly were processed only poorly (Figure 13). Also, as described in Deutsch (1984), the sequence shown in Figures 16 and 17 can be heard either as in C major or as in E minor, depending on whether it is played forward or backward.

Further experimental evidence for the importance of ordering and intervallic information was obtained by Brown (1988). Subjects were presented with pitch class sets that were ordered either to evoke a particular tonal center, or to evoke a different tonal center, or to be tonally ambiguous. The subjects' key judgments were strongly influenced by these manipulations (see also Brown et al., 1994). Matsunaga and Abe (2005) also found that subjects' choices of tonal centers for passages were influenced by the orderings of the presented tones. In another experiment, West and Fryer (1990) presented subjects with quasi-random orderings of the tones in a diatonic scale, in each case followed by a probe tone, and the subjects judged the suitability of the probe tone as a tonic in the context of the sequence they had just heard. It was found that the actual tonic was not judged as uniquely suitable as the tonal center; instead scale degrees $\hat{1}$, $\hat{3}$, $\hat{4}$, and $\hat{5}$ were rated as equally suitable.

Smith and Schmuckler (2004) created sequences in which the K&K profiles (or variants of these) were used to create distributions of the durations and frequencies of occurrence of the different pitch classes, which were then randomly ordered. Subjects were presented with these sequences, and they produced probe-tone profiles that were used by the authors to draw inferences concerning perceptions of key for these sequences. The tone profiles that the subjects produced were found to be similar to the original K&K profiles from which the sequences were derived. The authors interpreted this result to reflect the subjects' use of long-term knowledge of pitch class distributions within keys in making their judgments. However, since very similar distributional contexts were employed to generate both the original K&K profiles and the profiles obtained in their experiment, the results could instead have reflected the similarity of these two short-term contexts, rather than reflecting the use of long-term mental templates.

Based in part on this reasoning, Temperley and Marvin (2008) argued that, rather than drawing inferences from probe tone responses—which are equivocal in their interpretation—a better procedure would be to have subjects identify the key of a passage explicitly. They also argued that subjects' judgments should be compared against pitch class distributions that are found in actual music, because probe-tone profiles correlate only imperfectly with these distributions.

Reasoning along these lines, Temperley and Marvin presented subjects with melodies that were generated quasi-randomly from scale-degree distributions. The distributions were created from the first eight measures of each of the string quartet movements by Haydn and Mozart. The authors then created a profile displaying the proportion of events of each scale degree for each passage. The profiles from all major-key passages were averaged to create a major-key profile, and the analogous procedure was used to create a minor-key profile. The profiles were then employed to generate scale degrees in a stochastic fashion, so as to produce the presented melodies.

The subjects, who were musically trained, listened to each passage, and then made explicit key judgments by locating the tonic on a keyboard—a task that is easy for musically trained listeners to accomplish. It was found that only slightly more than half of the subjects' judgments of the presented melodies matched the generating key. In a further analysis, the authors determined for each melody the key that was chosen by the largest number of subjects, and they found that judgments of this key accounted for only 56.1% of the key judgments, showing that the subjects disagreed among themselves substantially in their choice of key. From these findings, Temperley and Marvin concluded that listeners perform poorly in determining the key of a melody when it is generated from pitch class distributions alone, and that structural cues produced by the ordering of the tones in the sequence must also be employed in the process of key identification.

IV. The Organization of Short-Term Memory for Tones

We here present a theoretical framework for the representation of tones in short-term memory (otherwise known as working memory). This issue is fundamental to our understanding of music perception and cognition, because tones form the basic units from which musical structures are derived. Indeed, as we have argued, certain characteristics of higher-level tonal organization can be attributed to interactions between tones at this basic level.

It is evident from general considerations that memory for music must be the function of a heterogeneous system, whose various subdivisions differ in the persistence with which they retain information. For example, the system that subserves memory for pitch relationships must be capable of retaining information over very long periods of time, whereas this is not true of the system that retains absolute pitch values. Similarly, the system that retains temporal patterns must preserve information for considerably longer than the system that retains absolute values of duration. Based on such considerations, we can assume that when memory for a musical pattern is tested after various time periods have elapsed, differences in its form of encoding would emerge.

More specifically, the model assumes that musical tones are initially subjected to a set of perceptual analyses, which are carried out in different subdivisions of the auditory system. Such analyses result in the attribution of values of pitch, loudness, duration, and so on, as well as values resulting from higher-level analyses, such as intervals, chords, rhythms, and timbres. It is further assumed that in many of these subsystems, information is represented along arrays that are systematically organized with respect to a simple dimension, such as pitch, loudness, or duration, or some higher-level dimension such as interval size, or in a multidimensional space, such as timbre.

The model further assumes that the outputs of these analyses are projected onto arrays in corresponding subdivisions of the auditory memory system. So, for example, one subdivision retains values of pitch, and others retain values of duration, loudness, interval size, timbre; and so on. Information is retained in parallel in these different subdivisions; however, the time constants of retention in these subdivisions vary considerably. It is further assumed that specific interactions take place within these subdivisions that are analogous to those that occur in systems processing auditory information at the incoming level. The outputs of these different subdivisions then combine during retrieval of information from memory.

Neurophysiological findings support the hypothesis of multiple auditory memory stores that subserve different stimulus attributes. When a listener is presented with a series of identical tones followed by a new tone, the new tone elicits an event-related brain potential called the ‘‘mismatch negativity’’ or MMN, which is assumed to reflect the detection of a difference between the incoming stimulus and the stimuli that have been stored in memory. Giard et al. (1995) analyzed the MMNs elicited by pure tones that deviated from standard tones in frequency, intensity, or duration. They found that the scalp topographies of the MMNs varied according to type of stimulus deviance, and they concluded that the frequency, intensity, and duration of a sound have separate neural representations in memory. In addition, MMNs obtained from tones that differed in terms of two features have been found to be roughly equal to the sum of the MMNs obtained from tones that differed in terms of a single feature—indicating that the standard tones leave multiple representations in the brain (Levanen, Hari, McEvoy, & Sams, 1993; Schroger, 1995). Within this framework of multiple parallel stores, we first focus on memory for pitch, and examine how values of this attribute are represented in storage and how they are accessed during retrieval. We then consider how other attributes of tone are represented in memory.

A. The System That Retains Absolute Pitch Values

In considering the characteristics of the system that retains absolute pitch values, a number of hypotheses may be advanced. For example, such memory might simply deteriorate with the passage of time. Another possibility is that pitch information is retained in a general store that is limited in terms of the number of items it can retain, so that memory loss results from a general information overload. As a third possibility, memory for pitch might be the function of an organized system whose elements interact in specific ways.

We can begin with the following observations. When a tone is presented, and this is followed immediately by another tone that is either identical in pitch to the first or that differs by a semitone, most listeners find it very easy to determine whether the two tones are the same or different in pitch. The task continues to be very easy when a silent interval of 6 s intervenes between the tones to be compared. Although memory for pitch has been shown to fade gradually with the passage of time (Bachem, 1954; Clément, Demany, & Semal, 1999; Harris, 1952; Kaernbach & Schlemmer, 2008; Rakowski, 1994; Wickelgren, 1966, 1969), the amount of fading during a silent retention interval of 6 s is so small that it is barely apparent in this situation. However, when eight extra tones intervene during the retention interval, the task becomes strikingly difficult, and this is true even when the listener is instructed to ignore the intervening tones. Deutsch (1970b) found that listeners who made no errors in comparing such tone pairs when they were separated by 6 s of silence made 40% errors when eight tones intervened during the retention interval. In a companion experiment, either four, six, or eight tones intervened during a retention interval of constant duration, and the error rate increased with an increase in the number of intervening tones.

We can conclude that memory for pitch is subject to a small amount of decay with time, and also to a large interference effect produced by other tones. What, then, is the basis of this interference effect? One possibility is that the intervening tones produce attention distraction, and that attention to the tone to be remembered is necessary for memory to be preserved. If this were the case, then other intervening materials would also result in memory loss, provided that these, too, distracted the listener's attention. As another hypothesis, pitch information might be held in a general store of limited capacity, along with other types of material. Further materials that enter this store would then also impair pitch recognition. As a third hypothesis, pitch information might be retained in a specialized system, and memory loss might result from interactions that occur specifically within this system.

In an experiment to examine these different hypotheses, Deutsch (1970c) had subjects compare the pitches of two tones that were separated by a 5-s retention interval. The test tones were either identical in pitch or they differed by a semitone.

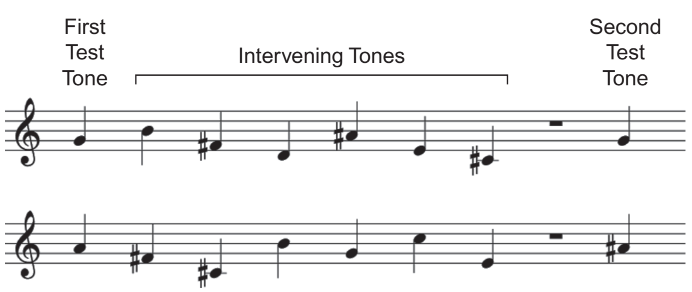

Figure 19 Examples of tone series employed in experiments to examine the effects of intervening tones on memory for the pitch of a tone.

In the first condition, six tones intervened during the retention interval (Figure 19). In the second condition, six spoken numbers intervened instead. In both these conditions, listeners were asked to ignore the intervening materials, and simply to judge whether the test tones were the same or different in pitch. A third condition was identical to the second, except that the listeners were asked to recall the numbers in addition to comparing the pitches of the test tones; this ensured that the numbers were attended to and entered memory. In a fourth condition, the subjects were asked to ignore the test tones and simply to recall the numbers.

It was found that the intervening tones produced substantial impairment in memory for the pitch of the first test tone. However, the intervening spoken numbers produced only minimal impairment, even when the subjects were asked to recall them. In addition, the error rate in number recall was no higher when the subjects were simultaneously performing the pitch-recognition task than when they could ignore the test tones. This experiment indicated, therefore, that decrements in pitch memory resulting from intervening tones are due to interactions that take place within a specialized system. More specifically, the hypothesis was proposed that *the system that retains pitch information is organized in many ways along the same principles as the system that receives it* (Deutsch, 1975c).

B. Further Evidence for a Separate Pitch Memory System

There is further evidence that pitch memory is the function of a specialized system. Deutsch (1974) had subjects compare the pitches of two test tones that were separated by a retention interval that contained eight intervening tones. In one condition, the intervening tones were all drawn from the same octave as the test tones; in a second, they were all drawn from the octave above; in a third, they were all drawn from the octave below. The intervening tones produced substantial interference in all conditions; however, the amount of interference varied depending on the octave in which the intervening tones were placed. The largest performance decrement in these three conditions occurred when the intervening tones were in the same octave as the test tones, the next largest when they were in the octave above, and the smallest when they were in the octave below. This experiment indicated, therefore, that the amount of interference produced by intervening tones depends on the pitch relationships between these and the test tones.

Semal, Demany, and colleagues have shown that interference in pitch memory results from interactions that occur within a system that is sensitive to pitch relationships, but insensitive to other attributes of sound (Demany & Semal, 2008). In one study, Semal and Demany (1991) had subjects compare the pitches of two test tones that were separated by a retention interval containing six intervening tones. The test tones were sine waves, and in some conditions the intervening tones were also sine waves, whereas in others they were of complex spectral composition. Substantial decrements in pitch recognition occurred when the intervening tones were close in pitch to the test tones, regardless of their spectra. However, when the intervening tones were remote in pitch from the test tones, the amount of memory impairment they produced was substantially smaller, again regardless of their spectra.

In a further experiment, Semal and Demany (1991) studied the effect of intervening tones that were composed of several harmonics of a missing fundamental. Again, memory performance depended essentially on the pitches of the intervening tones, and not on their spectral composition. Intervening tones that were close in pitch to the test tones were associated with poor performance, regardless of their spectra. Performance levels were higher when the intervening tones were remote in pitch from the test tones, again regardless of their spectra.

In yet another experiment, Semal and Demany (1993) found that differences in the amplitudes of the intervening tones had remarkably little effect on performance. The amount of memory impairment produced by the intervening tones was not a monotonically increasing function of their amplitudes, neither did maximal interference occur when the amplitudes of the test and intervening tones were identical. The authors also found that performance on pitch memory tasks was affected very little by whether or not the test and intervening tones had the same time-varying envelopes.

These experiments lead to the conclusion that the system that underlies memory for pitch is insensitive to other stimulus attributes. We can then ask whether the lack of interference in memory for the pitch of a tone when spoken words intervene during the retention interval is due to the storage of verbal and nonverbal stimuli in separate systems, or to the storage of pitch information in a separate system, regardless of whether the pitches are produced by tones or by words. This issue was addressed by Semal, Demany, Ueda, and Hallé (1996), who had subjects make memory judgments concerning words that were spoken at different pitches. The test words were separated by a retention interval that contained intervening materials that were either words spoken at different pitches, or complex tones presented at different pitches. The amount of memory impairment produced by the intervening materials was greater when their pitches were close to those of the test words than when they were remote from them, regardless of whether the intervening materials were words or tones. The authors concluded that the pitches of spoken words are not processed in a specialized “speech module,” but rather in a system that is responsible for retaining pitch information, which disregards other attributes of sound.

Other studies have explored the effects on pitch memory of a difference between the perceived spatial locations of the test and intervening tones. Deutsch (1978c) obtained slightly better memory performance when the test and intervening tones were presented to different ears rather than to the same ear; however, the difference between these two conditions was fairly small. Kallman, Cameron, Beckstead, and Joyce (1987) confirmed the small advantage produced by delivering the test and intervening tones to different ears; however, this advantage was present only when the ear of input for the intervening tones was fixed within a block—the procedure used by Deutsch (1978c). When the ear of input for the intervening tones varied unpredictably from trial to trial, the advantage produced by differences in ear of input disappeared. In yet a further experiment, Ries and DiGiovanni (2007) found a small yet significant improvement in the threshold for discriminating between the test tones when the intervening tones were made to differ in perceived spatial location through the use of interaural phase and amplitude cues. Taking these findings together, we can conclude that differences in perceived spatial location can reduce the disruptive effect of interpolated tones, though only to a small extent.

C. Specific Interactions within the Pitch Memory System

We next inquire more specifically into the types of interaction that occur within the system that retains pitch information. If, as hypothesized by Deutsch (1975c), the system is organized in ways that are similar to the system that processes incoming pitch information, we should expect facilitatory and inhibitory interactions to take place within this system that are a function of the relationships between the elements involved. This hypothesis was tested in a series of experiments.

Deutsch (1972b) had subjects compare the pitches of two test tones that were separated by a sequence of six intervening tones. The test tones were either identical in pitch or they differed by a semitone. The effects were explored of placing a tone whose pitch bore a critical relationship to the pitch of the first test tone (the "critical tone") in the second serial position of the intervening sequence. This distance varied in steps of ⅙ tone between identity and a whole-tone separation. As shown in Figure 20, when the first test tone and the critical tone were identical in pitch, memory facilitation was produced. As the pitch distance between these two tones increased, errors in pitch recognition also increased. Errors peaked at ⅔-tone separation and then decreased, returning to baseline at roughly a whole-tone separation.

Based on these findings, it was conjectured that pitch memory is the function of an array whose elements are activated by tones of specific pitch. These elements

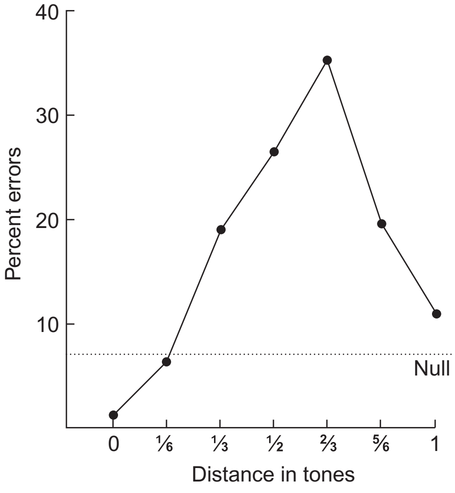

Figure 20 Percentage of errors in pitch recognition as a function of the pitch distance between the first test tone and a critical intervening tone. The line labeled Null shows the error rate in a control condition in which all intervening tones were at least 1½ tones removed in pitch from the first test tone. The maximal error rate occurred when the critical intervening tone was ⅔ tone removed from the first test tone.
Reprinted with permission from Deutsch (1972b). ©1972 American Association for the Advancement of Science.

are organized tonotopically on a log frequency continuum, and inhibitory interactions take place along this array that are a function of the distance between the interacting elements. It was further hypothesized that these interactions are analogous to recurrent lateral inhibitory interactions in systems processing sensory information at the incoming level (Ratliff, 1965). It was further conjectured that when these elements are inhibited, they produce weaker signals, so that increased errors in memory judgment result.

A number of considerations further support this conjecture. Error rates in pitch memory judgment cumulate when two critical tones that are a semitone removed from the first test tone are interpolated, placed one on either side of the first test tone along the pitch continuum (Deutsch, 1973a). This is analogous to the finding that lateral inhibitory effects cumulate when two inhibitory stimuli are placed, one on either side of the test stimulus along a visuospatial continuum (Ratliff, 1965). Furthermore, the parameters obtained here are consistent with tuning curves recorded from neurons in human auditory cortex (Bitterman, Mukamel, Malach, Fried, & Nelken, 2008) and extrapolated from psychophysical experiments (McLachlan, 2011). Just as lateral inhibition serves to sharpen the tuning response of auditory neurons at the incoming level (Yang, Pollack, & Resler, 1992), so we can assume that such sharpening must exist to maintain the fine-grained resolution of pitch information in memory.

If the pitch memory system were indeed organized as a recurrent lateral inhibitory network, one might also expect to find evidence for disinhibition: If a tone that was inhibiting memory for another tone were itself inhibited by a third tone, memory for the first tone should return. Specifically, in sequences where the test tones are identical in pitch, if two critical tones were placed in the intervening sequence, one always ⅔-tone removed from the test tone and the other further removed along the pitch continuum, errors should vary as a function of the pitch relationship between the two critical tones: The error rate should be highest when these two tones are identical in pitch, decline as the second critical tone moves away from the first, dip maximally at a ⅔-tone separation, and then return to baseline. In other words, the curve produced should be roughly the inverse of the curve plotting the original disruptive effect.

To test this prediction, Deutsch and Feroe (1975) performed an experiment in which subjects compared the pitches of two test tones when these were separated by a sequence of six intervening tones. A tone was always placed in the second serial position of the intervening sequence, whose pitch was ⅔ tone removed from that of the first test tone; that is, in a relationship expected to produce maximal inhibition. Errors were plotted as a function of the pitch of a second critical tone, which was placed in the fourth serial position, whose pitch relationship to the first critical tone varied in ⅙-tone steps between identity and a whole-tone separation.

As can be seen in Figure 21, a systematic return of memory was indeed obtained. The error rate in sequences where the second critical tone was identical in pitch to the first was significantly higher than baseline, and the error rate where the two critical tones were separated by ⅔ tone was significantly lower than baseline.

A first-order inhibitory function was obtained in a companion experiment, and this was used to calculate the theoretical disinhibition function, assuming that the error rate was determined simply by the strength of the signal produced by the element underlying the first test tone. As also shown in Figure 21, there was a good correspondence between the disinhibition function obtained experimentally and the one derived theoretically on the lateral inhibition model. This experiment therefore provided strong evidence that pitch memory elements are arranged as a lateral inhibitory network, analogous to those handling sensory information at the incoming level. As described earlier, this network could in turn affect patterns of perceived salience of

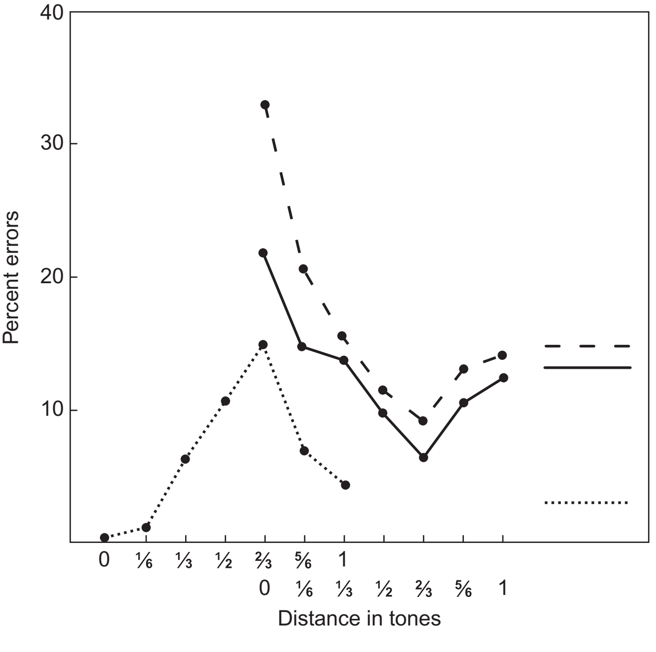

Figure 21 Percentages of errors in pitch recognition obtained experimentally and predicted theoretically. Dotted line displays percentage of errors in a baseline experiment that varied the pitch relationship between the first test tone and a critical intervening tone. (Dotted line at right displays percentage of errors where no tone was interpolated in the critical range.) Solid line displays percentage of errors in an experiment where a tone that was ⅔ tone removed from the first test tone was always interpolated. Errors are plotted as a function of the pitch relationship between this tone and a second critical intervening tone that was further removed along the pitch continuum. Dashed line displays percentage of errors for the same experimental conditions predicted theoretically from the lateral inhibition model. (Solid and dashed lines at right display percentages of errors obtained experimentally and predicted theoretically where no further critical tone was interpolated.)
Adapted from Deutsch and Feroe (1975).

tones so as to influence, for example, which tone is targeted as the most salient in a sequence; judgments involving key distance between melodies; and so on.

D. Item and Order Information

Another factor has also been found to cause impairment in pitch memory. When two test tones differ in pitch, and the intervening sequence includes a critical tone whose pitch is identical to that of the second test tone, there results a substantial increase in errors of misrecognition. This increase in errors is greater when the critical tone is placed early in the intervening sequence rather than late (Deutsch, 1970a, 1972a).

In order to explain this effect, Deutsch (1972a) proposed a model in which pitch information is retained along both a pitch continuum and a temporal continuum. When a tone is presented, its pitch can be represented as a bell-shaped distribution along these two continua, such as shown in Figure 22. With the passage of time, this distribution spreads along both continua, but particularly along the temporal continuum. When a further tone that is close to the first in time and pitch is then presented, the overlapping portions of the distributions underlying the two tones sum. As a result, in our present paradigm, when the first and second test tones differ in pitch, and a tone of the same pitch as the second test tone is included in the intervening sequence, the subject recognizes that it had occurred in the sequence, but is uncertain *when* it had occurred and so sometimes erroneously assumes that it had been the first test tone. This effect of misrecognition should be greater when the critical tone is placed early in the intervening sequence rather than late, as was found in the experiments of Deutsch (1970a, 1972a).

An experiment by Deutsch (1975c) lent further support to this model. Here the pitch difference between the first and second test tones was varied, and errors were again plotted as a function of the pitch relationship between the first test tone and the critical intervening tone. It was found, as predicted, that in sequences where the critical intervening tone and the second test tone were on the same side of the first test tone along the pitch continuum, then as the pitch of the second test tone shifted along this continuum, the peak of errors produced by the critical intervening tone

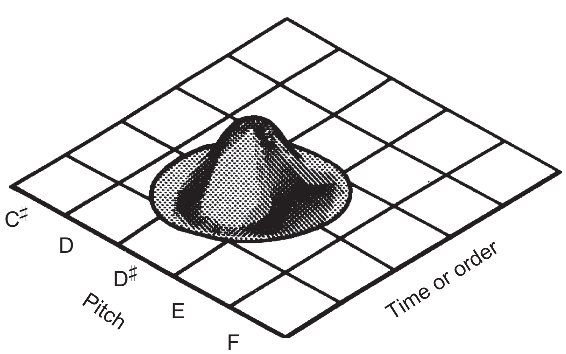

Figure 22 Distribution hypothesized to underlie memory for the pitch of a tone. See text for details.
Adapted from Deutsch (1972a).

shifted in parallel. Further support for the model was provided by Ries and DiGiovanni (2009), who determined the threshold for detecting a pitch difference between two test tones that were separated by a sequence of intervening tones. These authors found that the threshold for discriminating the test tones rose considerably when these differed and a tone of the same pitch as the second test tone was included in the intervening sequence.

Interestingly, this effect of temporal smearing has been shown to occur with proactively presented tones also. Ruusurvita (2000) had subjects listen to two test tones that were separated by a silent retention interval, and judge whether the second test tone was higher or lower than the first. Three interfering tones were presented before the first test tone, and the subjects were instructed to ignore them. It was found that errors were enhanced when the interfering tones and the second test tone were on the same side of the first test tone along the pitch continuum, and that errors were reduced when these tones were on the opposite side of the first test tone along the pitch continuum. This finding is as expected from the effects of retroactively presented tones obtained by Deutsch (1972a, 1972b) described earlier.

E. Memory Enhancement Through Repetition

The model proposed by Deutsch (1972a) gives rise to a further prediction: When a tone that is identical in pitch to the first test tone is included in the intervening sequence, the memory distributions for these tones should sum, and this should result in a stronger memory trace for the first test tone. This in turn should lead to an enhancement in pitch recognition and discrimination, which should be greater when the critical tone is placed early in the intervening sequence rather than late. Such a pattern of results was obtained in a number of experiments (e.g., Deutsch, 1970a, 1972a, 1975c). Indeed, the error rate in comparing the pitches of the test tones was considerably lower when the intervening sequence contained six tones, one of which was identical to the first test tone, compared with a sequence containing only four intervening tones, none of which was identical to the first test tone (Deutsch, 1975a, 1975c). Further, Ries and DiGiovanni (2009) found that including in the intervening sequence a tone of identical pitch to the first test tone produced a substantial lowering of threshold in detecting whether the test tones were the same or different. As was discussed earlier, this enhancement of the memory trace through repetition has important implications for how we process tonal structures.

F. Octave Generalization

As described earlier, there is considerable evidence that pitch is represented along both a linear array of pitch height and also a circular array of pitch class. This leads us to inquire whether interference in pitch memory takes place along a pitch height array, or a pitch class array, or whether both such arrays are involved.

In one experiment, Deutsch (1973b) had subjects compare the pitches of two tones that were separated by six intervening tones. The experiment explored the effects of intervening tones that bore the same relationship to the test tones as had been found earlier to produce memory disruption, but that were further displaced by an octave. In sequences where the test tones were identical, the effects were studied of including two critical tones, one a semitone higher than the first test tone and the other a semitone lower, but that were further displaced by an octave. In sequences where the test tones differed, the effects were studied of including a critical tone that was identical in pitch to the second test tone, but again displaced by an octave.

Substantial generalization of the disruptive effect of including tones a semitone removed from the first test tone occurred when such tones were displaced an octave higher, and a weaker effect occurred when such tones were displaced an octave lower. However, the disruptive effect was largest from tones that were placed in the middle octave, in which both pitch height and pitch class were involved. It was concluded that disruptive effects in pitch memory take place along both a pitch height and a pitch class array.

Memory facilitation through repetition also generalizes across octaves. Deutsch and Lapidis (in preparation) had subjects compare the pitches of two tones that were separated by a sequence of six intervening tones. As found earlier (Deutsch, 1975a), including in the intervening sequence a tone that was identical in pitch to the first test tone resulted in an enhancement of performance, both when the test tones were the same in pitch and also when they differed. This effect of enhancement also occurred when the critical intervening tone was an octave higher than the first test tone. Although this enhancement effect was somewhat reduced by the octave displacement, the difference depending on octave placement was not statistically significant. We can conclude, therefore, that enhancement of memory for pitch through repetition also generalizes across octaves, and so also takes place along a pitch class array.

G. Pitch Proximity and Pitch Memory

In listening to sequences such as we have been describing, the listener processes not only the individual tones but also the melodic intervals between them. These intervals then provide a framework of pitch relationships to which the test tones can be anchored. So intervening sequences that form melodic patterns that are easily processed should be associated with enhanced performance on this memory task.

As described in Chapter 6, there is considerable evidence that melodic patterns are processed more effectively when they are composed of small rather than large intervals, in accordance with the principle of proximity. One might then hypothesize that in our present situation also, intervening sequences that are composed of small melodic intervals would be associated with higher performance levels than those composed of larger intervals. In an experiment to test this hypothesis, Deutsch (1978a) had subjects compare the pitches of two test tones that were separated by a sequence of six intervening tones. There were four conditions in the experiment. In the first, the intervening tones were chosen at random from within

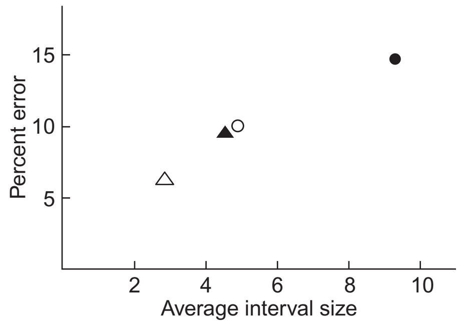

Figure 23 Percentage of errors in pitch recognition as a function of the average size of the melodic interval in the sequence. Open triangle: Intervening tones span a one-octave range and are ordered monotonically. Filled triangle: Intervening tones span a one-octave range and are ordered at random. Open circle: Intervening tones span a two-octave range and are ordered monotonically. Filled circle: Intervening tones span a two-octave range and are ordered at random.
From Deutsch (1978a).

a one-octave range, and they were also ordered at random. The second condition was identical to the first, except that the intervening tones were arranged in monotonically ascending or descending order, so that the average size of the melodic intervals was reduced. In the third condition, the intervening tones were chosen at random from within a two-octave range, and they were also ordered at random. The fourth condition was identical to the third, except that the intervening tones were arranged in monotonically ascending or descending order. (The only restriction on this random ordering was that, to avoid the specific effects we have been describing, no sequence contained repeated pitch classes, and no intervening sequence contained a tone that was a semitone removed from the first test tone.)

As shown in Figure 23, the error rate was found to increase with an increase in the average size of the melodic intervals formed by the intervening tones. There was no evidence that monotonic ordering of the intervening tones had an effect, beyond that of producing a smaller average interval size. As described earlier, there is a striking cross-cultural tendency for the frequency of occurrence of a melodic interval to be inversely correlated with its size (Dowling, 1967; Huron, 2001; Merriam et al., 1956; Ortmann, 1926). One might hypothesize that this tendency is based on an increasing difficulty in processing melodic intervals as interval size increases. As indicated in the present experiment, this should in turn result in decreased accuracy in pitch recognition judgment.

In a related study, Deutsch (1974) compared the effects on pitch recognition of placing the intervening tones in different octaves. In the condition where the intervening tones were drawn from the same octave as the test tones, the error rate was higher than in conditions where they were drawn consistently from the octave above or the octave below. However, the error rate was highest when the intervening tones were drawn from both the higher and the lower octaves. In this last condition, the intervening tones formed very large intervals, so that listeners were unable to make use of frameworks of pitch relationships to which the test tones could be anchored. Olson and Hanson (1977) also found that increased distances between the test and intervening tones were associated with increased errors in pitch recognition. The experiment by Deutsch and Boulanger (1984), described earlier, also led to the same conclusion.

H. Memory for Timbre

Continuing to examine the hypothesis that the different attributes of musical tones are retained in separate memory subsystems, we now consider memory for timbre. Starr and Pitt (1997) employed tones of different spectral composition, each of which consisted of a fundamental together with three adjacent harmonics. So one type of tone consisted of the fundamental together with harmonics 2, 3, and 4; another consisted of the fundamental together with harmonics 3, 4, and 5; and so on. Eight values of timbre were generated in accordance with this algorithm, and in this way a "spectral similarity dimension" was created. Subjects compared the timbres of two test tones when these were separated by a 5-s retention interval. Performance was substantially disrupted by intervening tones that were very similar in timbre to the first test tone. Further, performance improved as the difference in timbre between the first test tone and the intervening tones increased—analogous to the similarity-based interference found in memory for pitch (Deutsch, 1972b). When the pitch distance between the test and intervening tones was also varied, the effect of timbre similarity persisted regardless of this pitch distance, and there was only a negligible effect of pitch variations on memory for timbre.

In another study, Mercer and McKeown (2010) had subjects compare two test tones that were separated by a 10-s retention interval. The test tones comprised six harmonics, one of which (the "critical feature") was increased in amplitude so as to produce an alteration in timbre. A single distractor tone, consisting of two adjacent harmonics, was presented either before or after the first test tone. Performance was not significantly impaired by distractors that consisted of features that were contained in the first test tone. However, performance was substantially impaired when the distractor contained two new features. Performance was even more impaired when the test tones differed and the distractor and the second test tone contained the same critical feature. These findings are again analogous to those obtained in memory for pitch, where disruption occurred from intervening tones that differed in pitch from the test tones, and even greater disruption occurred when the test tones differed and a tone in the intervening sequence was identical in pitch to the second test tone.

I. Memory for Loudness

Considering the hypothesis of a separate memory store for loudness, Clément et al. (1999) reasoned that in the absence of intervening stimuli, the trace of a pitch sensation may not decay at the same rate as that of a loudness sensation. To examine this hypothesis, the authors presented subjects with test tones that were separated by retention intervals of different durations, and they compared memory for pitch and loudness at these different intervals. Intensity discrimination declined markedly as retention intervals increased from 0.5 s to 2 s, and remained fairly constant thereafter for intervals of at least 10 s. In comparison, frequency discrimination declined in a way that was significantly less abrupt. Jump and Ries (2008) examined further the hypothesis of a separate loudness memory store, by presenting test tones that were separated by a retention interval of 4800 ms. Relative loudness judgments were substantially disrupted when four tones of varying intensity intervened during the retention interval, and did not improve when the intervening tones were made to differ in apparent spatial location through the use of interaural phase and intensity cues.

J. Memory for Spatial Location

Ries, Hamilton, and Grossmann (2010) addressed the same issue with respect to memory for the spatial location of a tone. These authors produced differences in apparent spatial location through the use of interaural phase differences. Subjects compared the apparent spatial locations of two test tones that were separated by retention intervals of up to 15 s. Similarly to the findings described above for other attributes, memory for location decayed slowly during a silent retention interval, but was substantially disrupted when four intervening tones with differing apparent locations were interpolated during this interval.

K. Memory for Duration

The system underlying memory for duration has characteristics that are similar to those underlying memory for pitch. Deutsch (1986b) presented subjects with a pair of blips that defined a first test duration, followed by another pair of blips that defined a second test duration. The subjects judged whether the second test duration was identical to the first, or longer, or shorter. In one set of conditions, a series of additional blips was interpolated during the retention interval; these blips defined durations that were in the same range as the test durations. When the intervening durations were identical to the first test duration, performance levels were higher than when no blips were interpolated. This effect was analogous to the enhancement of pitch memory that occurred when a tone of identical pitch to the first test tone was included in the intervening sequence. In addition, the error rate was considerably higher when the interpolated blips defined durations that were slightly removed from the first test duration, and judgments reflected distortions in memory for the first test duration in the direction of the intervening durations. This effect was analogous to the large misrecognition effect that occurred in memory for pitch when a tone of identical pitch to the second test tone was included in the intervening sequence (Deutsch, 1970a, 1972a).

L. Memory for Pitch Relationships

There is evidence that the system underlying memory for pitch relationships is organized in ways that are similar to the organization of memory for absolute pitch values. As described earlier (Section I), Deutsch (1969) suggested that interval information is place-coded in the nervous system. Deutsch (1975c) further proposed that this information is projected onto a memory array, such that memory for intervals is the function of a continuum whose elements are activated by the simultaneous or successive presentation of pairs of tones. Tone pairs standing in the same ratio project onto the same elements, and so onto the same point along the continuum; tone pairs standing in closely similar ratios project onto adjacent points along the continuum; and so on. It was further proposed that interactive effects take place along this memory continuum that are analogous to those occurring in the system that retains absolute pitch values. Such effects include memory enhancement through repetition and similarity-based interference.

In one experiment, Deutsch and Roll (1974) had subjects make pitch-recognition judgments when the test tones were separated by six intervening tones. It was found that these judgments were strongly influenced by the relational context in which the test tones were placed. The subjects tended to judge test tones as identical in pitch when these were presented in the context of identical harmonic intervals and to judge them as different when they were presented in the context of different intervals.

Deutsch (1982a) obtained a similar effect in memory for melodic intervals. Here subjects compared two test tones that were each preceded by tones of lower pitch, and the subjects were instructed to ignore the preceding tones. The test tone combinations were separated by a retention interval during which six extra tones were interpolated. When the first and second test tone combinations formed the identical melodic interval, there was an increased tendency to judge the test tones as identical. Further, when the test tone combinations formed different melodic intervals, there was an increased tendency to judge the test tones as different. This pattern occurred both when the test tones were the same in pitch and also when they differed.

Given these findings, an experiment was performed to examine indirectly whether memory for intervals was subject to specific interactive effects similar to those found in memory for absolute pitch values (Deutsch, 1978b). Specifically, memory for harmonic intervals was tested. Subjects compared the pitches of two test tones when these were both accompanied by tones of lower pitch. The test tones were either identical in pitch or they differed by a semitone. However, the tone accompanying the first test tone was always identical in pitch to the tone accompanying the second test tone. So when the test tones were identical, the intervals formed by the test-tone combinations were also identical. Similarly, when the test tones differed, the intervals formed by the test-tone combinations also differed.

The test-tone combinations were separated by a sequence of six intervening tones. The tones in the second and fourth serial positions of the intervening sequence were also accompanied by tones of lower pitch. It was found that when the intervals formed by the intervening combinations were identical in size to the interval formed by the first test combination, the error rate was lower than when the intervals formed by the intervening combinations were chosen at random. Furthermore, when the intervals formed by the intervening combinations differed in size by a semitone from the interval formed by the first test combination, the error rate was higher than when the sizes of the intervals formed by the intervening combinations were chosen at random. This experiment indicated that effects analogous to those in the system retaining absolute pitch information also occur in the system retaining abstracted pitch values; namely, memory enhancement through repetition and similarity-based interference.

V. Paradoxes Based on Pitch Class

As described earlier, the pitch of a tone is held to vary along two dimensions: The monotonic dimension of pitch height defines its position along a continuum from low to high, and the circular dimension of pitch class defines its position within the octave (Babbitt, 1960; Bachem, 1955; Charbonneau & Risset, 1973; Deutsch, 1969, 1972c, 1973b, 1986a; Deutsch & Boulanger, 1984; Deutsch et al., 2008; Forte, 1973; Révész, 1913, Risset, 1969, 1971; Ruckmick, 1929; Shepard, 1964, 1982; Ueda & Ohgushi, 1987). In order to accommodate the dimensions of pitch class and height in a single spatial representation, it has been suggested that pitch be depicted as a geometrically regular helix in which the entire structure maps into itself under transposition (Drobisch, 1855; Shepard, 1964, 1982). Such a representation is shown in Figure 24, and it can be seen that tones that are separated by octaves are depicted as in close spatial proximity. This geometric model assumes that the dimensions of pitch class and pitch height are orthogonal, so that the pitch class of a tone would not influence its perceived height.

Shepard (1964) noted that the helical model of pitch has an intriguing consequence. If one could suppress the monotonic component of pitch height, leaving only the circular component of pitch class, all tones that are related by octaves could be mapped onto the same tone, which would then have a clearly defined

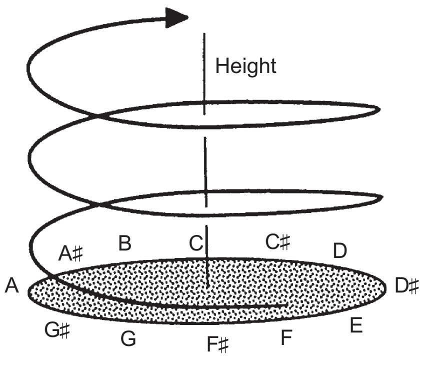

Figure 24 Pitch as a geometrically regular helix.
Adapted from Shepard (1965).

pitch class but an indeterminate height. In this way, the tonal helix would be collapsed into a circle, and judgments of pitch would become completely circular.

In an associated experiment, Shepard generated a set of tones, each of which consisted of 10 sinusoidal components that were separated by octaves, and whose amplitudes were determined by a fixed, bell-shaped, spectral envelope. The pitch classes of the tones were varied by shifting the components up or down in log frequency, holding the position and shape of the envelope constant. Shepard argued that because the spectral envelope remained fixed, the perceived heights of these tones would remain constant as their pitch classes were varied.

Subjects listened to ordered pairs of such tones, and they reported in each case whether they heard an ascending or a descending pattern. When the tones within a pair were separated by one or two steps along the pitch class circle (Figure 25), judgments were determined almost entirely by proximity. For example, the tone pair C-C♯ was heard as ascending, and the pair C-B was heard as descending. When the tones were separated by larger distances along the circle, the tendency to follow by proximity was gradually reduced, and when they were separated by exactly a half-octave, ascending and descending judgments occurred equally often.

Shepard (1964) concluded from these findings that the dimensions of pitch class and height were indeed orthogonal, arguing that such a view would at all events be expected on common-sense grounds:

> *tonality [i.e., pitch class] seems quite analogous to the attribute of being clockwise or counterclockwise. One of two nearby points on a circle can be said to be clockwise from the other; but it makes no sense to say how clockwise a single point is absolutely.*

However, this conclusion does not necessarily follow from Shepard's findings. Where judgments were heavily influenced by proximity, any effect of pitch class on perceived height could have been overwhelmed by this factor. Furthermore, because the data were averaged across pitch classes, any effect of pitch class on perceived height would have been lost in the averaging process. The issue of

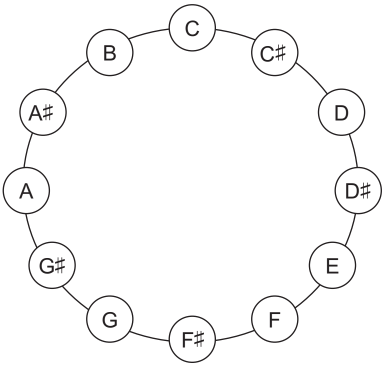

Figure 25 The pitch class circle.

whether the pitch class of a tone might influence its perceived height therefore remained unresolved in Shepard's study.

A. The Tritone Paradox

Given such considerations, Deutsch (1986a) had subjects listen to ordered pairs of tones that were related by a half-octave (or tritone), and they reported in each case whether they heard an ascending or a descending pattern. Each tone consisted of six octave-related components whose amplitudes were scaled by a bell-shaped spectral envelope. In order to control for possible effects of the relative amplitudes of the components of the tones, and also to examine the effects of varying their overall heights, the tone pairs were generated under envelopes that were placed at six different positions along the spectrum, which were spaced at half-octave intervals.

It was reasoned that because the tones within each pair were in opposite positions along the pitch class circle, proximity could not here be used as a cue in making judgments of relative height. So if the assumption of orthogonality were correct, these judgments would not be influenced by the pitch classes of the tones. But it was also reasoned that an interaction between the two dimensions might emerge: The listener might perceive tones in one region of the pitch class circle as higher and tones in the opposite region as lower.

More specifically, it was conjectured that listeners might arrange pitch classes as a circular map, similar to a clock face. This map might have a particular orientation with respect to height. For example, C could be in the 12:00 position and F♯ in the 6:00 position, so that the listener would perceive the tone pairs C-F♯ (and B-F and C♯-G) as descending, and tone pairs F♯-C (and F-B and G-C♯) as ascending. If, on the other hand, this map were oriented so that F♯ stood in the 12:00 position and C in the 6:00 position, the listener would instead perceive the tone pair C-F♯ as ascending and the pair F♯-C as descending.

The hypothesis of an effect of pitch class on perceived height was strikingly confirmed: The judgments of most subjects showed that tones in one region of the pitch class circle were perceived as higher, and those in the opposite region as lower. Another striking finding also emerged: The relationship between pitch class and perceived height differed radically from one subject to another. Figure 26 presents, as an example, the judgments of two subjects who showed particularly clear and consistent relationships between pitch class and perceived height. (The judgments were averaged over tones generated under all six spectral envelopes.)

The first subject heard tone pairs C♯-G, D-G♯, D♯-A and E-A♯ as ascending, and tone pairs F♯-C, G-C♯, G♯-D, A-D♯, A♯-E, and B-F as descending. In contrast, the second subject heard tone pairs B-F, C-F♯, C♯-G, D-G♯, D♯-A, and E-A♯ as descending, and pairs F♯-C, G-C♯, G♯-D, and A-D♯ as ascending. So for the most part, when the first subject heard an ascending pattern, the second subject heard a descending one; and vice versa. In consequence, also as shown in Figure 26, extended patterns formed of such tone pairs were heard by these two subjects as producing entirely different melodies. Figure 27 shows the perceptual orientations of the pitch

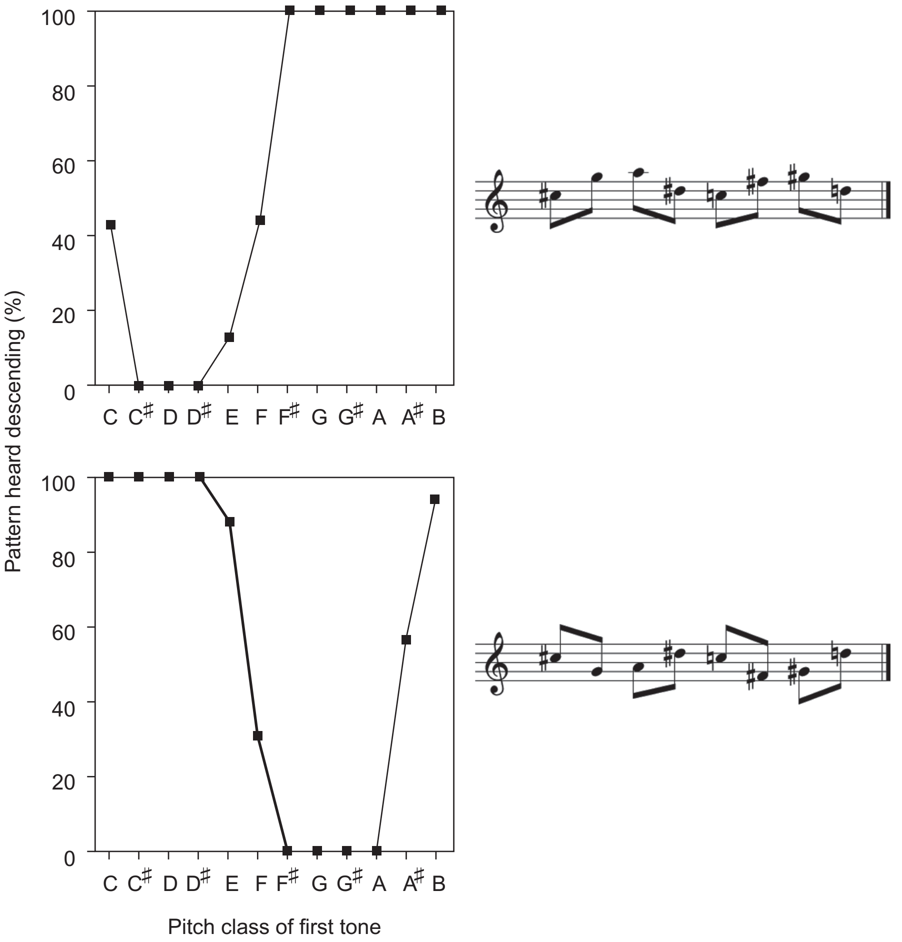

Figure 26 The tritone paradox as perceived by two different subjects. The graphs show the percentages of judgments that a tone pair formed a descending pattern, plotted as a function of the pitch class of the first tone of the pair. Notations on the right show how the identical series of tone pairs was perceived by these two subjects.
Data from Deutsch (1986a).

class circle that were derived from the judgments of these two subjects. For the first subject, the peak pitch classes (i.e., those that stood at the highest position along the pitch class circle) were G♯ and A; however, for the second subject, the peak pitch classes were C♯ and D instead.

Figure 28 shows the judgments of four more subjects whose patterns were less pronounced than were those shown in Figure 26. These data were taken from experiments in which four spectral envelopes were used, which were spaced at half-octave intervals,

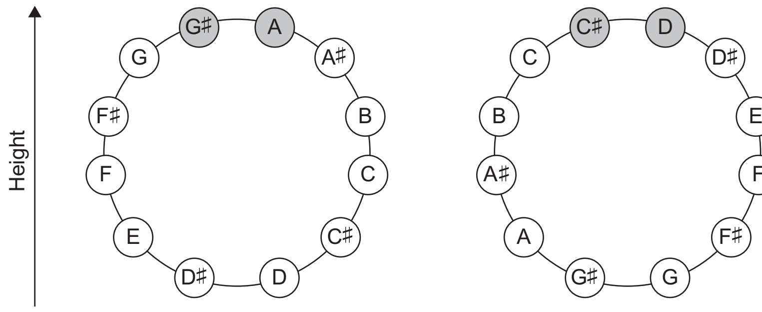

Figure 27 Perceptual orientations of the pitch class circle, derived from the judgments of the two subjects whose data are displayed in Figure 26. The circle on the left is derived from the graph shown in the upper portion of Figure 26, and the circle on the right is derived from the graph shown in the lower portion. The pitch classes that mark the highest position along the circle are termed *peak pitch classes*.

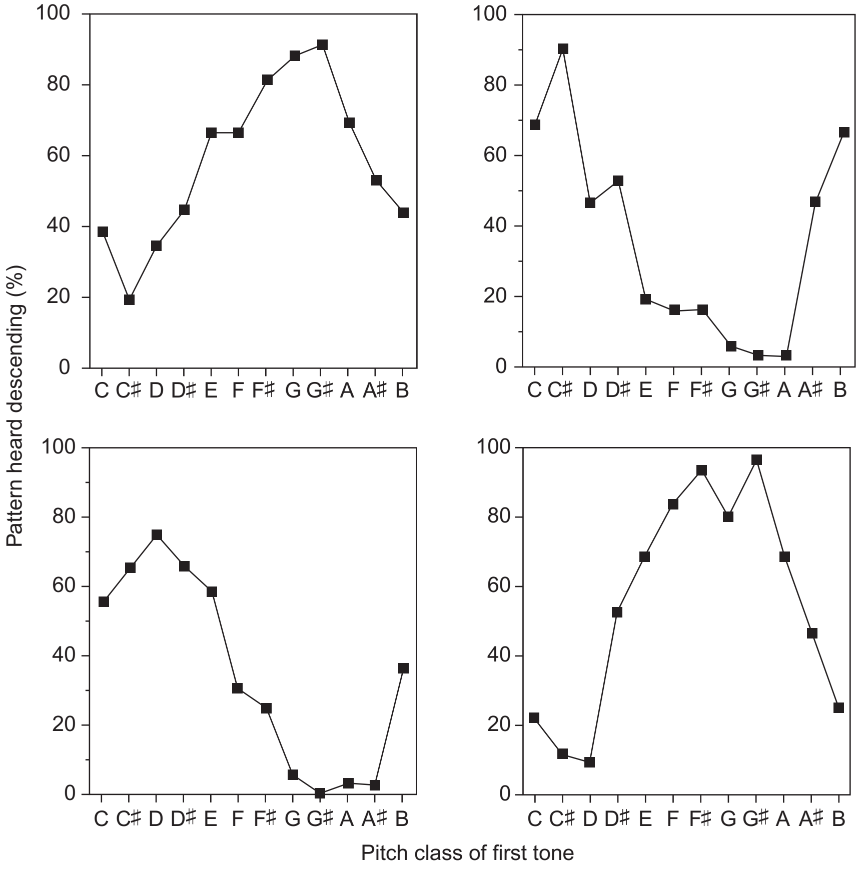

Figure 28 The tritone paradox as perceived by four more subjects.

and the judgments under these four envelopes were averaged. It can be seen that all four subjects showed clear relationships between pitch class and perceived height; however, the form of this relationship varied from one subject to another.

Deutsch et al. (1987) performed a study to examine perception of the tritone paradox in a general population. A group of subjects were selected on the only criteria that they were UCSD undergraduates, had normal hearing, and could judge reliably whether pairs of sine-wave tones that were related by a tritone formed ascending or descending patterns. The judgments of most subjects reflected clear relationships between pitch class and perceived height. Furthermore, computer simulations showed that the patterns obtained in this experiment were extremely unlikely to have occurred by chance. It was concluded that the tritone paradox exists to highly significant extent in a general population. (The sound patterns comprising a full experiment on the tritone paradox, together with instructions for analyzing the results, are published in the CD by Deutsch, 1995.)

B. The Semitone Paradox

We next inquire into what happens when more than one tone is presented at a time. Deutsch (1988b) examined this issue using a basic pattern that consisted of two sequentially presented tone pairs, which were diametrically opposed along the pitch class circle. An example of such a pattern is shown in Figure 29. On one side of the circle the second tone was higher than the first (in this example, G♯ was followed by A), and on the other side the second tone was lower than the first (in this example, D♯ was followed by D). In general, subjects linked the tones sequentially in accordance with pitch proximity, so that they perceived the pattern as two stepwise lines that moved in contrary motion. However, the higher line could be heard as ascending and the lower line as descending, or vice versa.

Subjects were presented with such sequential tone pairs, and they judged in each case whether the line that was higher in pitch formed an ascending or a descending pattern. From these judgments it was inferred which pitch classes were heard as higher and which as lower. Taking the tone pairs in Figure 29, for example, if the subject heard the higher line as ascending, this indicated that he or she perceived G♯ and A as higher and D♯ and D as lower [as in Percept (a)]. However, if the subject heard the higher line as descending, this indicated that he or she perceived D♯ and D as higher and G♯ and A as lower [as in Percept (b)].

Just as with the tritone paradox, subjects' judgments here reflected orderly relationships between the pitch classes of the tones and their perceived heights. Also as with the tritone paradox, the form of this relationship varied radically from one subject to another. This is illustrated in the judgments of two subjects shown in Figure 30. For the first subject, tones F, F♯, G, G♯, A, and A♯ were heard as higher and C, C♯, D, and D♯ were heard as lower. In contrast, for the second subject, C♯, D, and D♯ were heard as higher and F, F♯, G, G♯, A, A♯, and B were heard as lower. In consequence, also as shown in Figure 30, musical passages produced by series of such tone pairs were heard by these two subjects in entirely different ways.

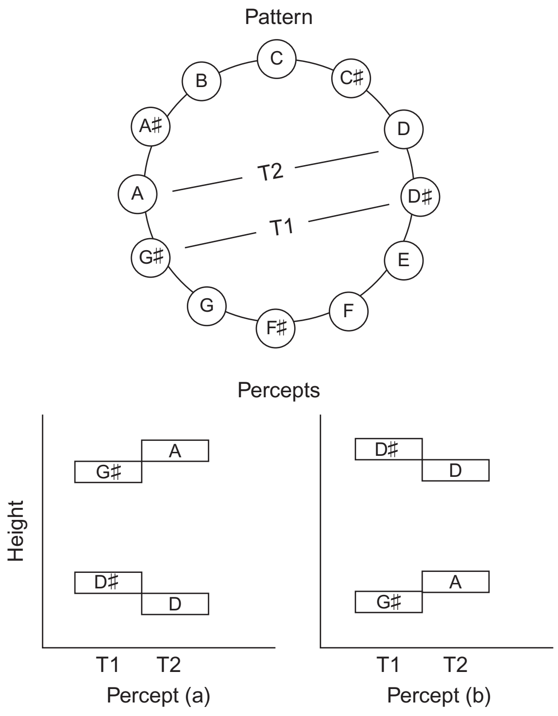

Figure 29 Example of pattern giving rise to the semitone paradox, together with two alternative perceptual organizations. Tones G♯ and D♯ are presented simultaneously at T1, and tones A and D at T2. Listeners organize this pattern as two stepwise lines that move in contrary motion; that is, they hear the ascending line G♯-A together with the descending line D♯-D. However, some listeners hear the ascending line as higher [Percept (a)] while other listeners hear the descending line as higher [Percept (b)].
Adapted with permission from Deutsch (1988b). ©1988 by The Regents of the University of California.

C. *The Melodic Paradox*

We can further ask what happens when more complex patterns are presented. Deutsch, Moore, and Dolson (1986) investigated this question using patterns consisting of three sequentially presented tone pairs. Specifically, the pattern shown in Figure 31 was played in two different keys. In C major, the pattern consisted of the succession of tones D-E-F played together with B-A-G. In F♯ major, the tones G♯-A♯-B were played together with E♯-D♯-C♯.

When this pattern was heard unambiguously, listeners always organized the tones sequentially in accordance with pitch proximity. So they heard one melodic line that ascended by a minor third, together with another that descended by a major third. However, also as shown in Figure 31, the descending line could be heard as higher and the ascending line as lower [as in Percept (a)] or the ascending line could be heard as higher and the descending line as lower [as in Percept (b)].

Analogous effects were found to occur here also: When the pattern was played in one key, it was perceived with the higher line ascending. However, when the pattern was played in the other key, it was heard with the higher line descending instead. So transposing the pattern from one key to the other led to a perceived interchange of voices. Furthermore, when the pattern was played in any one key, it was heard with the higher line as ascending by some listeners, but as descending by others.

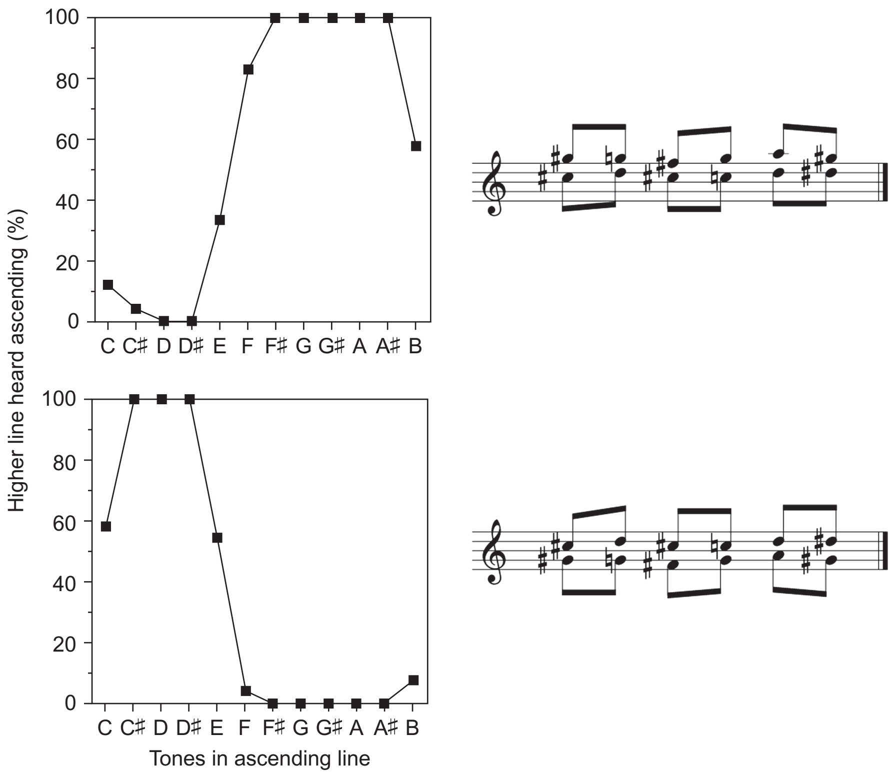

Figure 30 The semitone paradox as perceived by two different subjects. The graphs show the percentages of trials in which a tone was heard as part of the higher line, plotted as a function of the pitch class of the tone. For both subjects, when the pattern was transposed, the ascending and descending lines appeared to interchange positions. Moreover, in general, when the first subject heard the higher line ascending, the second subject heard it descending, and vice versa. Notations on the right show how the identical series of patterns was perceived by these two subjects.

Deutsch (1988a) performed a further experiment, in which this two-part pattern was played in six different keys: C, D, E, F♯, G♯, and A♯ major. The judgments of four subjects were examined, and they all showed orderly effects of key and also differed radically in the direction in which key influenced their judgments. In consequence, extended passages formed of such patterns were heard by these subjects in entirely different ways.

As illustration, Figure 32 displays the percepts of two of the subjects. The first subject heard the pattern in the keys of C and D with the higher line ascending, yet in the keys of E, F♯, and G♯ with the higher line descending. The second subject,

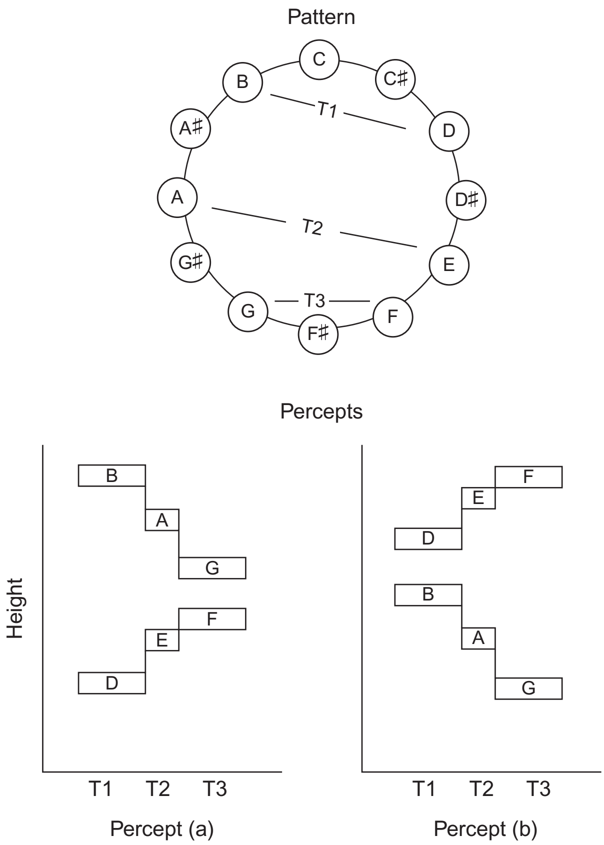

Figure 31 Pattern giving rise to the melodic paradox, together with alternative perceptual organizations. Tones D and B are simultaneously presented at time T1, tones E and A are presented at T2, and tones F and G are presented at T3. This pattern is generally heard as two stepwise lines that move in contrary motion. However, some listeners hear the higher line descending and the lower line ascending [Percept (a)], whereas others hear the higher line ascending and the lower line descending [Percept (b)].
From experiment by Deutsch et al. (1986).

in contrast, heard the pattern in the keys of D, E, F♯, and G♯ with the higher line ascending, yet in the keys of C and A♯ with the higher line descending. Thus for the most part, when the first subject heard the higher line ascending, the second subject heard it descending; and vice versa. This is also illustrated in the notation on the right-hand part of the figure.

D. Implications of These Musical Paradoxes

The paradoxes described here show that pitch class and pitch height are not orthogonal dimensions; rather, the perceived height of a tone is systematically related to its position along the pitch class circle, when other factors are controlled for.

The paradoxes are surprising on a number of grounds. First, they provide striking violations of the principle of perceptual equivalence under transposition; a principle that had been assumed to be universal. In the case of the tritone paradox, transposing the pattern from one key to another can cause it to appear to change from an ascending pattern to a descending one, and vice versa. In the case of the

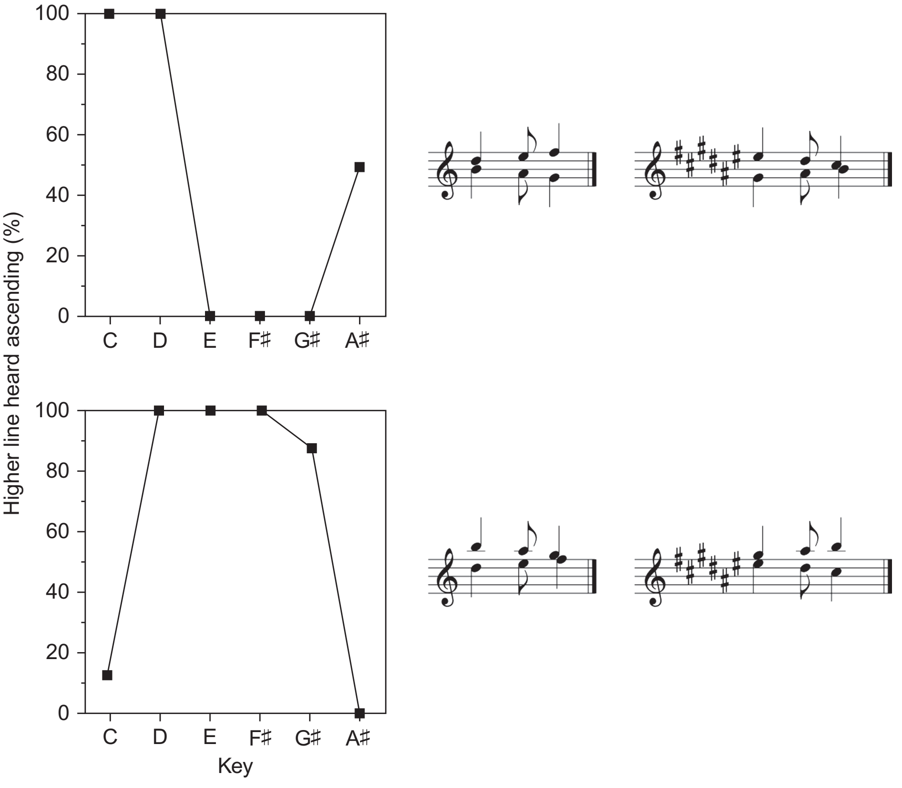

Figure 32 Melodic paradox as perceived by two different subjects, when presented in six different keys. In general, as the pattern was transposed, the ascending and descending lines appeared to interchange positions. Moreover, for the most part when the first subject heard the higher line ascending, the second subject heard it descending, and vice versa. Notations on the right show how the identical series of patterns was perceived by these two subjects. Adapted from Deutsch (1988a).

paradoxes involving two-part patterns, transposition can result in a perceived interchange of voices.

Another surprising implication concerns absolute pitch, a faculty that is generally assumed to be very rare. Because the majority of listeners experience these musical paradoxes, it follows that the majority of us have at least a partial form of absolute pitch, in that we hear tones as higher or lower depending simply on their note names, or pitch classes. Other studies, described in Chapter 5, have also indicated that most people possess an implicit form of absolute pitch, even though they are unable to name the notes they are judging (Halpern, 1989; Levitin, 1994; Schellenberg & Trehub, 2003; Terhardt & Seewann, 1983; Vitouch & Gaugusch, 2000).

A further unexpected outcome of this work concerns the striking differences between listeners in how these paradoxes are perceived. These differences are as pronounced among experienced musicians as among listeners without musical training. It is to the basis of such differences that we now turn.

E. Basis of the Tritone Paradox

Studies exploring the bases of these musical paradoxes have focused largely on the tritone paradox. A number of experiments have examined the involvement of spectral factors in this effect. Deutsch (1987) generated such tritone pairs under 12 different spectral envelopes, with peaks spaced at ¼-octave intervals, so that their positions varied over a three-octave range. Although the relationship between pitch class and perceived height was found sometimes to vary as a function of the overall height of the spectral envelope, and sometimes also as a function of the relative amplitudes of the components of the tones, such variations tended to be small in absolute terms (see also Dawe, Platt, & Welsh, 1998; Deutsch, 1994; Giangrande, 1998; Repp & Thompson, 2010).

A number of informal observations led the author to hypothesize that the tritone paradox might be related to the processing of speech sounds. It was conjectured that the listener develops a long-term representation of the overall pitch range of his or her speaking voice. Included in this representation is a specification of the octave band in which the largest proportion of pitch values occurs. The listener then takes the pitch classes delimiting the octave band for speech as marking the highest position along the pitch class circle; this in turn determines his or her orientation of the pitch class circle with respect to height.

In a test of this hypothesis, Deutsch, North, and Ray (1990) selected a group of subjects who showed clear relationships between pitch class and perceived height in making judgments of the tritone paradox. We took a 15-min recording of natural speech from each subject, and from this recording, we identified the octave band containing the largest number of pitch values. Comparing across subjects, we obtained a significant correspondence between the pitch classes defining this octave band, and those defining the highest position along the pitch class circle, as determined by judgments of the tritone paradox.

Two versions of the hypothesis relating the tritone paradox to the pitch range of speech may then be advanced. The first does not assume that the pitch range of an individual's speaking voice is itself determined by a learned template. The second, and broader, version assumes that we acquire such a template as a result of exposure to speech around us. This template is then used both to constrain our own speech output and to evaluate the speech of others. If the second hypothesis were correct, we would expect the orientation of the pitch class circle to be similar for people in a given linguistic subculture, but to vary among people in different subcultures.

In a review of the literature concerning the pitch ranges of speech in different linguistic communities, Dolson (1994) described evidence in support of the latter hypothesis. First, most people confine the pitch range of their speech to roughly an octave. Second, within a given linguistic community, in general the speech of females is close to an octave above that of males; for this reason, a template based on pitch class rather than pitch would enable the mapping of male and female speech onto a single mental representation. Further, the pitch ranges of speech differ remarkably little within a given linguistic community (except, of course, for the gender difference); however, there are considerable variations in the pitch ranges of speech across different linguistic communities. Moreover, there is a surprising lack of correlation between the pitch range of a person's speech and physiological parameters such as his or her height, weight, chest size, and laryngeal size. This indicates that the pitch range of a person's speaking voice is based on cultural consensus. Deutsch, Le, Shen, and Henthorn (2009) have provided detailed evidence for the hypothesis of a culturally acquired pitch range for speech that spans an octave.

Deutsch (1991) performed a further experiment to test the speech-related hypothesis. The judgments of two groups of subjects were compared: The first group had grown up in California, and the second group had grown up in the south of England. The two groups were found to differ statistically in their perceptions of the tritone paradox, so that frequently when a Californian subject heard the pattern as ascending, a subject from the south of England heard it as descending; and vice versa (Figure 33).

Other laboratories have obtained further evidence for a geographic association. Giangrande (1998) found that a group of subjects at Florida Atlantic University produced a distribution of peak pitch classes that was similar to the one found by Deutsch (1991) among Californians. Treptoe (1997) found a very similar distribution among subjects at the University of Wisconsin, Steven's Point. In contrast, Dawe et al. (1998) found that a group of students at McMaster University, Ontario, produced a distribution that was quite similar to the one found by Deutsch (1991) for subjects from the south of England.

Examining this correlate in greater detail, Ragozzine and Deutsch (1994) discovered a regional difference in perception of the tritone paradox within the United States. Among subjects who had grown up in the area of Youngstown, Ohio, the perceptions of those whose parents had also grown up in this region differed significantly from those whose parents had grown up elsewhere within the United States. These findings indicate that perception of the tritone paradox is influenced by a template that is acquired in childhood. Further evidence was provided by Deutsch (2007), who found a significant correlation between the way children and their mothers heard the tritone paradox. This correlation was obtained even though the children had all been born and raised in California, whereas their mothers had grown up in many different geographical regions, both within and outside the United States.

We can then ask what happens in the case of people who had been exposed to one language in infancy and later acquired a different language. Will such people hear the tritone paradox in accordance with their first language, or will they hear it in accordance with the language that they now speak? Deutsch, Henthorn, and Dolson (2004) addressed this question by testing subjects whose first language was

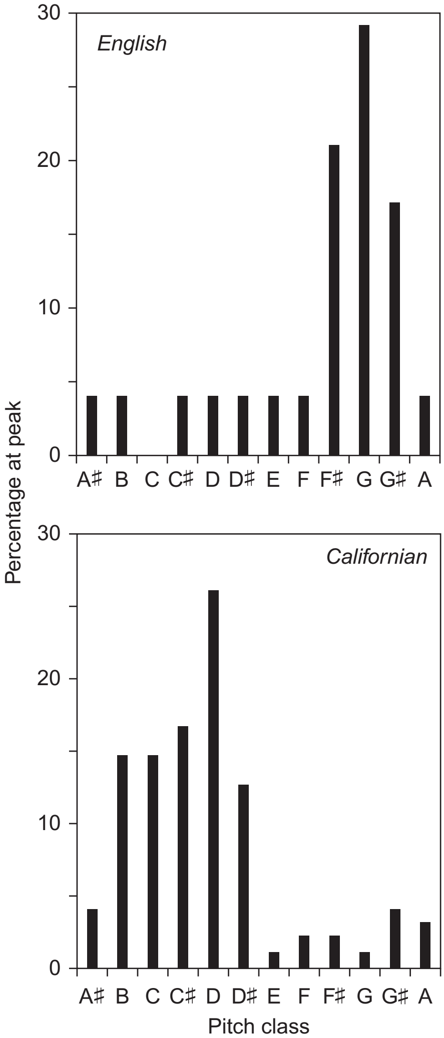

Figure 33 Distributions of peak pitch classes in two groups of subjects. The first group had grown up in the south of England and the second group had grown up in California. Reprinted with permission from Deutsch (1991). ©1991 by The Regents of the University of California.

Vietmanese, and who now live in California. The subjects were all from south or central Vietnam. The first, older, group had arrived in the United States as adults. They all spoke perfect Vietnamese, but little English. The second, younger, group had arrived in the United States as infants or children. They all spoke perfect English, and most were not fluent speakers of Vietnamese. In addition, we tested a third group, which consisted of native speakers of Californian English, both of whose parents were also native speakers of Californian English.

Figure 34 shows the distribution produced by the Vietnamese subjects, together with the one produced by the native speakers of Californian English. In addition,

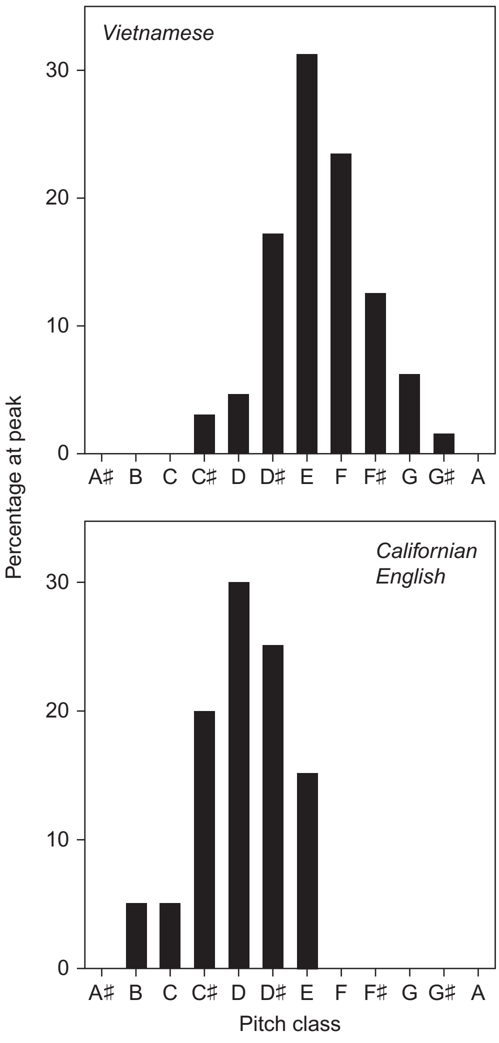

Figure 34 Distributions of peak pitch classes in two groups of subjects. The first group had been born in Vietnam, and the second group were native speakers of Californian English. Reprinted from Deutsch et al., 2004. ©2004 by The Regents of the University of California.

the distributions produced by the older and younger Vietnamese groups taken separately both differed highly significantly from that produced by the native speakers of Californian English. However, the difference between the distributions produced by the older and younger Vietnamese groups did not begin to approach significance. So these results are in accordance with the conjecture that perception of the tritone paradox reflects a speech-related template that develops early in life. In a further experiment, Vietnamese subjects read out a 5-min passage in Vietnamese, and from each recording, we identified the octave band containing the largest number of pitch values. As with the earlier study by Deutsch et al. (1990) on English-speaking subjects, there was a significant correspondence between the pitch classes defining this octave band and those defining the highest position along the pitch class circle, as determined by judgments of the tritone paradox. This again supports the conjecture that perception of the tritone paradox is determined by a pitch template that is derived from speech heard early in life.

We can then ask whether these perceptual paradoxes occur in natural musical situations. In unpublished experiments, the effects were found to persist with the use of tone complexes whose partials were stretched slightly so that they did not stand in octave relation. The effects also persisted when the sinusoidal components of the tones were replaced by sawtooth waves, so that the power spectrum of each complex was similar to one produced by several instruments playing simultaneously. Furthermore, the effects were not destroyed by imposing a vibrato, a tremolo, or a fast decay such as occurs with a plucked string. Most interestingly, the tritone paradox was produced, at least in some individuals, when each tone of a pair consisted of a single harmonic series, with the relative amplitudes of the odd and even harmonics adjusted so that the tones were roughly equivalent in perceived height. It appears, therefore, that effects such as these might well be found in music performed by live instruments, when the composer has introduced ambiguities of height, such as in orchestral pieces by Debussy and Ravel.

VI. Illusory Transformation from Speech to Song

Finally, we briefly discuss an illusion that I discovered while fine-tuning the spoken commentary on my CD *Musical Illusions and Paradoxes* (1995). I had the phrase "*sometimes behave so strangely*" on a loop, and noticed that after a number of repetitions, the phrase sounded as though sung rather than spoken, as notated in Figure 35. Later, I included this illusion in my CD *Phantom Words, and Other Curiosities* (Deutsch, 2003).

The illusion raises fundamental issues concerning relationships between speech and song. It is generally assumed that whether a phrase is heard as spoken or sung depends on its acoustical characteristics. In contrast, this illusion occurs without altering the signal in any way, without training, and without any context provided by other sounds, but simply as a result of repeating the phrase several times over.

Deutsch, Henthorn, and Lapidis (2011) carried out two experiments to explore this effect. The first experiment explored certain constraints governing the illusion, using a rating task as the measure. It was found that the illusion occurred when the repeated presentations of the spoken phrase were exact replicas of the original one. Yet when on repetition the phrase was transposed slightly, or the syllables were

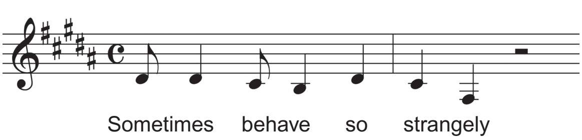

Figure 35 The spoken phrase, as it generally appears to be sung.
Reprinted from Deutsch (2003).

jumbled, the illusion did not occur. In a second experiment, subjects were asked to repeat back the phrase exactly as they heard it, either following a single presentation or following 10 presentations. The subjects' renditions following 10 presentations corresponded more closely to the pattern notated on Figure 35 than to the sequence of intervals formed by the original spoken phrase. It was hypothesized that during the process of repetition, the pitches forming the phrase increase in perceptual salience, and that they are also perceptually distorted so as to conform to a well-formed tonal melody. It appears, therefore, that the neural circuitries underlying the perception of speech and song can accept the same input, but process it differently, so as to produce different outputs.

VII. Conclusion

In the foregoing pages, we have considered the principles whereby the listener forms abstractions based on pitch and retains pitch information at different levels of abstraction. Where appropriate, we have considered underlying neurophysiological mechanisms, and we have also drawn on insights provided by music theorists. We have argued for the view that music is represented in the mind of the listener in the form of coherent patterns that are linked together so as to form hierarchical structures. We have also examined the system underlying memory for tones, and have explored a number of paradoxical illusions together with their implications. The system that we are dealing with is very complex, but an understanding of its characteristics is slowly emerging.

Acknowledgments

I am grateful to Trevor Henthorn for help with the illustrations. Preparation of this chapter was supported in part by an Interdisciplinary Research Award to the author from the University of California, San Diego.

References

Aarden, B. (2003). *Dynamic melodic expectancy* (Unpublished doctoral dissertation). Columbus, Ohio: Ohio State University.

Anderson, J. R., Reder, L. M., & Lebiere, C. (1996). Working memory: Activation limitations on retrieval. *Cognitive Psychology*, *30*, 221−256.

Babbitt, M. (1960). Twelve-tone invariants as compositional determinants. *The Musical Quarterly*, *46*, 246−259.

Bachem, A. (1954). Time factors in relative and absolute pitch determination. *Journal of the Acoustical Society of America*, *26*, 751−753.

Bachem, A. (1955). Absolute pitch. *Journal of the Acoustical Society of America*, *27*, 1180−1185.

Balzano, G. (1980). The group-theoretic description of 12-fold and microtonal pitch systems. *Computer Music Journal*, *4*, 66–84.

Bartlett, J. C., & Dowling, W. J. (1980). Recognition of transposed melodies: a key-distance effect in developmental perspective. *Journal of Experimental Psychology: Human Perception and Performance*, *6*, 501–515.

Bharucha, J. J. (1984a). Anchoring effects in music: The resolution of dissonance. *Cognitive Psychology*, *16*, 485–518.

Bharucha, J. J. (1984b). Event hierarchies, tonal hierarchies, and assimilation: a reply to Deutsch and Dowling. *Journal of Experimental Psychology: General*, *113*, 421–425.

Bharucha, J. J. (1987). Music cognition and perceptual facilitation: a connectionist approach. *Music Perception*, *5*, 1–30.

Bharucha, J. J. (1999). Neural nets, temporal composites, and tonality. In D. Deutsch (Ed.), *The psychology of music* (pp. 413–441). San Diego, CA: Academic Press.

Bharucha, J. J., & Stoeckig, K. (1986). Reaction time and musical expectancy: priming of chords. *Journal of Experimental Psychology: Human Perception and Performance*, *12*, 403–410.

Bharucha, J. J., & Stoeckig, K. (1987). Priming of chords: spreading activation or overlapping frequency spectra? *Perception & Psychophysics*, *41*, 519–524.

Bigand, E., Tillmann, B., Poulin-Charronat, B., & Manderlier, D. (2005). Repetition priming: Is music special? *Quarterly Journal of Experimental Psychology*, *58A*, 1347–1375.

Bitterman, Y., Mukamel, R., Malach, R., Fried, I., & Nelken, I. (2008). Ultra-fine frequency tuning revealed in single neurons of human auditory cortex. *Nature*, *451*, 197–201.

Blackwell, H. R., & Schlosberg, H. (1943). Octave generalization, pitch discrimination, and loudness thresholds in the white rat. *Journal of Experimental Psychology*, *33*, 407–419.

Bower, G. H., & Springston, F. (1970). Pauses as recoding points in letter series. *Journal of Experimental Psychology*, *83*, 421–430.

Bower, G., & Winzenz, D. (1969). Group structure, coding and memory for digit series. *Journal of Experimental Psychology*, *80*, 1–17.

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

Brown, H. (1988). The interplay of set content and temporal context in a functional theory of tonality perception. *Music Perception*, *5*, 219–250.

Brown, H., & Butler, D. (1981). Diatonic trichords as minimal cue-cells. *In Theory Only*, *5*, 39–55.

Brown, H., Butler, D., & Jones, M. R. (1994). Musical and temporal influences on key discovery. *Music Perception*, *11*, 371–407.

Browne, R. (1981). Tonal implications of the diatonic set. *In Theory Only*, *5*, 3–21.

Butler, D. (1989). Describing the perception of tonality in music: a critique of the tonal hierarchy theory and a proposal for a theory of intervallic rivalry. *Music Perception*, *6*, 219–241.

Butler, D., & Brown, H. (1984). Tonal structure versus function: studies of the recognition of harmonic motion. *Music Perception*, *2*, 6–24.

Callender, C., Quinn, I., & Tymoczko, D. (2008). Generalized voice leading spaces. *Science*, *320*, 346–348.

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

Chew, E. (2002). The spiral array: An algorithm for determining key boundaries. In C. Anagnostopoulou, M. Ferrand, & A. Smaill (Eds.), *Music and artificial intelligence* (pp. 18–31). Berlin, Germany: Springer.

Clément, S., Demany, L., & Semal, C. (1999). Memory for pitch versus memory for loudness. *Journal of the Acoustical Society of America*, *106*, 2805−2811.

Creel, S. C., & Newport, E. L. (2002). Tonal profiles of artificial scales: Implications for music learning. In C. Stevens, D. Burnham, G. McPherson, E. Schubert, & J. Renwick (Eds.), *Proceedings of the 7th International Conference on Music Perception and Cognition*, Sydney, Australia.

Croonen, W. L. (1994). Effects of length, tonal structure, and contour in the recognition of tone series. *Perception & Psychophysics*, *55*, 623−632.

Cuddy, L. L., Cohen, A. L., & Mewhort, D. J. (1981). Perception of structure in short melodic sequences. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 869−883.

Cuddy, L. L., & Lunney, C. A. (1995). Expectancies generated by melodic intervals: perceptual judgments of melodic continuity. *Perception & Psychophysics*, *57*, 451−462.

Dawe, L. A., Platt, J. R., & Welsh, E. (1998). Spectral-motion aftereffects and the tritone paradox among Canadian subjects. *Perception & Psychophysics*, *60*, 209−220.

Demany, L., & Armand, P. (1984). The perceptual reality of tone chroma in early infancy. *Journal of the Acoustical Society of America*, *76*, 57−66.

Demany, L., & Semal, C. (2008). The role of memory in auditory perception. In W. A. Yost, A. N. Popper, & R. R. Fay (Eds.), *Auditory perception of sound sources* (pp. 77−113). New York, NY: Springer.

Deutsch, D. (1969). Music recognition. *Psychological Review*, *76*, 300−307.

Deutsch, D. (1970a). Dislocation of tones in a musical sequence: a memory illusion. *Nature*, *226*, 286.

Deutsch, D. (1970b). *The deterioration of pitch information in memory* (Unpublished doctoral dissertation). University of California at San Diego, La Jolla.

Deutsch, D. (1970c). Tones and numbers: specificity of interference in short-term memory. *Science*, *168*, 1604−1605.

Deutsch, D. (1972a). Effect of repetition of standard and comparison tones on recognition memory for pitch. *Journal of Experimental Psychology*, *93*, 156−162.

Deutsch, D. (1972b). Mapping of interactions in the pitch memory store. *Science*, *175*, 1020−1022.

Deutsch, D. (1972c). Octave generalization and tune recognition. *Perception & Psychophysics*, *11*, 411−412.

Deutsch, D. (1973a). Interference in memory between tones adjacent in the musical scale. *Journal of Experimental Psychology*, *100*, 228−231.

Deutsch, D. (1973b). Octave generalization of specific interference effects in memory for tonal pitch. *Perception & Psychophysics*, *13*, 271−275.

Deutsch, D. (1974). Generality of interference by tonal stimuli in recognition memory for pitch. *Quarterly Journal of Experimental Psychology*, *26*, 229−234.

Deutsch, D. (1975a). Facilitation by repetition in recognition memory for tonal pitch. *Memory & Cognition*, *3*, 263−266.

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

Deutsch, D. (1975c). The organization of short-term memory for a single acoustic attribute. In D. Deutsch, & J. A. Deutsch (Eds.), *Short-term memory* (pp. 107−151). New York, NY: Academic Press.

Deutsch, D. (1977). Memory and attention in music. In M. Critchley, & R. A. Henson (Eds.), *Music and the brain* (pp. 95−130). London, England: Heinemann.

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

Deutsch, D. (1978b). Interactive effects in memory for harmonic intervals. *Perception & Psychophysics*, *24*, 7–10.

Deutsch, D. (1978c). Interference in pitch memory as a function of ear of input. *Quarterly Journal of Experimental Psychology*, *30*, 283–287.

Deutsch, D. (1978d). The psychology of music. In E. C. Carterette, & M. P. Friedman (Eds.), *Handbook of perception* (Vol. X, pp. 191–218). New York, NY: Academic Press.

Deutsch, D. (1979). Octave generalization and the consolidation of melodic information. *Canadian Journal of Psychology*, *33*, 201–205.

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

Deutsch, D. (1982a). The influence of melodic context on pitch recognition judgment. *Perception & Psychophysics*, *31*, 407–410.

Deutsch, D. (1982b). The processing of pitch combinations. In D. Deutsch (Ed.), *The psychology of music* (1st ed., pp. 271–316). New York, NY: Academic Press.

Deutsch, D. (1984). Two issues concerning tonal hierarchies: comment on Castellano, Bharucha, and Krumhansl. *Journal of Experimental Psychology: General*, *113*, 413–416.

Deutsch, D. (1986a). A musical paradox. *Music Perception*, *3*, 275–280.

Deutsch, D. (1986b). Recognition of durations embedded in temporal patterns. *Perception & Psychophysics*, *39*, 179–187.

Deutsch, D. (1987). The tritone paradox: effects of spectral variables. *Perception & Psychophysics*, *42*, 563–575.

Deutsch, D. (1988a). Pitch class and perceived height: Some paradoxes and their implications. In E. Narmour, & R. Solie (Eds.), *Explorations in music, the arts, and ideas: Essays in honor of Leonard B. Meyer*. Stuyvesant, NY: Pendragon Press.

Deutsch, D. (1988b). The semitone paradox. *Music Perception*, *6*, 115–134.

Deutsch, D. (1991). The tritone paradox: an influence of language on music perception. *Music Perception*, *8*, 335–347.

Deutsch, D. (1994). The tritone paradox: some further geographical correlates. *Music Perception*, *12*, 125–136.

Deutsch, D. (1995). *Musical illusions and paradoxes* [CD]. La Jolla, CA: Philomel Records.

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

Deutsch, D. (2007). Mothers and their offspring perceive the tritone paradox in closely similar ways. *Archives of Acoustics*, *32*, 3–14.

Deutsch, D., & Boulanger, R. C. (1984). Octave equivalence and the processing of tonal sequences. *Music Perception*, *3*, 40–51.

Deutsch, D., & Feroe, J. (1975). Disinhibition in pitch memory. *Perception & Psychophysics*, *17*, 320–324.

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

Deutsch, D., Henthorn, T., & Dolson, M. (2004). Speech patterns heard early in life influence later perception of the tritone paradox. *Music Perception*, *21*, 357–372.

Deutsch, D., Kuyper, W. L., & Fisher, Y. (1987). The tritone paradox: its presence and form of distribution in a general population. *Music Perception*, *5*, 79–92.

Deutsch, D., & Lapidis, R. (in preparation). Octave generalization of consolidation effects in memory for pitch.

Deutsch, D., Moore, F. R., & Dolson, M. (1986). The perceived height of octave-related complexes. *Journal of the Acoustical Society of America*, *80*, 1346−1353.

Deutsch, D., North, T., & Ray, L. (1990). The tritone paradox: correlate with the listener's vocal range for speech. *Music Perception*, *7*, 371−384.

Deutsch, D., & Roll, P. L. (1974). Error patterns in delayed pitch comparison as a function of relational context. *Journal of Experimental Psychology*, *103*, 1027−1034.

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., Henthorn, T., & Lapidis, R. (2011). Illusory transformation from speech to song. *Journal of the Acoustical Society of America*, *129*, 2245−2252.

Deutsch, D., Le, J., Shen, J., & Henthorn, T. (2009). The pitch levels of female speech in two Chinese villages. *JASA Express Letters*, *125*, 208−213.

Dewar, K. M., Cuddy, C. L., & Mewhort, D. J. K. (1977). Recognition memory for single tones with and without context. *Journal of Experimental Psychology: Human Learning & Memory*, *3*, 60−67.

Dewitt, L. A., & Crowder, R. G. (1986). Recognition of novel melodies after brief delays. *Music Perception*, *3*, 259−274.

Dibben, N. (1994). The cognitive reality of hierarchic structure in tonal and atonal music. *Music Perception*, *12*, 1−25.

Dolson, M. (1994). The pitch of speech as function of linguistic community. *Music Perception*, *11*, 321−331.

Dowling, W. J. (1967). *Rhythmic fission and the perceptual organization of tone sequences* (Unpublished doctoral dissertation). Harvard University, Cambridge, MA.

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

Dowling, W. J. (1978). Scale and contour: two components of a theory of memory for melodies. *Psychological Review*, *85*, 342−354.

Dowling, W. J. (1986). Context effects on melody recognition: scale-step and interval representation. *Music Perception*, *3*, 281−296.

Dowling, W. J. (1991). Tonal strength and melody recognition after long and short delays. *Perception & Psychophysics*, *50*, 305−313.

Dowling, W. J., & Bartlett, J. C. (1981). The importance of interval information in long-term memory for melodies. *Psychomusicology*, *1*, 30−41.

Dowling, W. J., & Fujitani, D. S. (1971). Contour, interval and pitch recognition in memory for melodies. *Journal of the Acoustical Society of America*, *49*, 524−531.

Drobisch, M. (1855). Uber musikalische Tonbestimmung und Temperatur. In *Abhandlungen der Koniglich sachsischen Gesellschaft der Wissenschaften zu Leipzig. Vierter Band: Abhandlungen der mathematisch-physischen Classe. Zweiter Band* (pp. 3−121). Leipzig, Germany: S. Hirzel.

Edworthy, J. (1985). Melodic contour and musical structure. In P. Howell, I. Cross, & R. J. West (Eds.), *Musical structure and cognition* (pp. 169−188). Orlando, FL: Academic Press.

Erickson, R. (1984). A perceptual substrate for tonal centering? *Music Perception*, *2*, 1−5.

Ernst, G. W., & Newell, A. (1969). *GPS: A case study in generality and problem solving*. New York, NY: Academic Press.

Estes, W. K. (1972). An associative basis for coding and organization in memory. In A. W. Melton, & E. Martin (Eds.), *Coding processes in human memory*. Washington, DC: Winston.

Evans, E. F. (1974). Neural processes for the detection of acoustic patterns and for sound localization. In F. O. Schmitt, & F. T. Worden (Eds.), *The neuroscience, third study program* (pp. 131−147). Cambridge, MA: MIT Press.

Fernald, A. (1993). Approval and disapproval: infant responsiveness to vocal affect in familiar and unfamiliar languages. *Child Development*, *64*, 657−674.

Forte, A. (1973). *The structure of atonal music*. New Haven, CT: Yale University Press.

Francès, R. (1988). *The perception of music* (W. J. Dowling, Trans.). Hillsdale, NJ: Erlbaum. (Original work published 1958)

Giangrande, J. (1998). The tritone paradox: effects of pitch class and position of the spectral envelope. *Music Perception*, *13*, 253−264.

Giard, M. H., Lavikainen, J., Reinikainen, K., Perrin, F., Bertrand, O., Perrier, J., & Naatanen, R. (1995). Separate representation of stimulus frequency, intensity, and duration in auditory sensory memory: an event-related potential and dipole-model analysis. *Journal of Cognitive Neuroscience*, *7*, 133−143.

Gjerdingen, R. O. (1988). *A classic turn of phrase: Music and the psychology of convention*. Philadelphia, PA: University of Pennsylvania Press.

Gjerdingen, R. O. (2007). *Music in the galant style*. New York, NY: Oxford University Press.

Greeno, J. G., & Simon, H. A. (1974). Processes for sequence production. *Psychological Review*, *81*, 187−196.

Hall, D. (1974). Quantitative evaluation of musical scale tunings. *American Journal of Physics*, *48*, 543−552.

Halpern, A. R. (1989). Memory for the absolute pitch of familiar songs. *Memory & Cognition*, *17*, 572−581.

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 (ICMPC11),* Seattle, Washington.

Han, S., Sundararajan, J., Bowling, D. L., Lake, J., & Purves, D. (2011). Co-variation of tonality in the music and speech of different cultures. *PLoS ONE*, *6*, e20160. doi:10.1371/journal.pone.0020160

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

Harris, J. D. (1952). The decline of pitch discrimination with time. *Journal of Experimental Psychology*, *43*, 96−99.

Helmholtz, H. von (1844). The origin of the correct interpretations of our sensory impressions. *Zeitchrift für Psychologie und Physiologie der Sinnesorgane*, *7*, 81−96.

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. (Original work published 1859)

Hershman, D. P. (1994). Key distance effects in ecological contexts. In I. Deliège (Ed.), *Proceedings of the 3rd International Conference on Music Perception and Cognition, Liège* (pp. 243−244). Liège, Belgium: ICMPC.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, *160*, 106−154.

Humphreys, L. F. (1939). Generalization as a function of method of reinforcement. *Journal of Experimental Psychology*, *25*, 361−372.

Huron, D. (1996). The melodic arch in Western folksongs. *Computing in Musicology*, *10*, 3−23.

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.

Huron, D., & Parncutt, R. (1993). An improved model of tonality perception incorporating pitch salience and echoic memory. *Psychomusicology*, *12*, 154−171.

Idson, W. L., & Massaro, D. W. (1978). A bidimensional model of pitch in the recognition of melodies. *Perception & Psychophysics*, *24*, 551−565.

Jones, M. R. (1978). Auditory patterns: Studies in the perception of structure. In E. C. Carterette, & M. P. Friedman (Eds.), *Handbook of Perception: Vol. VIII. Perceptual coding*. New York, NY: Academic Press.

Jump, R. J., & Ries, D. T. (2008). Effect of interaural level and phase cues on intervening interference in auditory working memory for loudness. *Hearing Research*, *240*, 112−115.

Justus, T. C., & Bharucha, J. J. (2002). Music perception and cognition. In S. Yantis, & H. Pashler (Eds.), *Stevens' Handbook of Experimental Psychology: Vol. 1. Sensation and Perception* (3rd ed., pp. 453−492). New York, NY: Wiley.

Kaernbach, C., & Schlemmer, K. (2008). The decay of pitch memory during rehearsal. *Journal of the Acoustical Society of America*, *123*, 1846−1849.

Kallman, H. J., Cameron, P. A., Beckstead, J. W., & Joyce, E. (1987). Ear of input as a determinant of pitch-memory interference. *Memory & Cognition*, *15*, 454−460.

Kallman, H. J., & Massaro, D. W. (1979). Tone chroma is functional in melody recognition. *Perception & Psychophysics*, *26*, 32−36.

Kim, J. C. (2011). *Tonality in music arises from perceptual organization* (Unpublished doctoral dissertation). Northwestern University, Evanston, IL.

Knudsen, E. I., du Lac, S., & Esterly, S. D. (1987). Computational maps in the brain. *Annual Review of Neuroscience*, *10*, 41−65.

Komar, A. J. (1971). *Theory of suspensions: A study of metrical and pitch relations in tonal music*. Princeton, NJ: Princeton University Press.

Koffka, K. (1935). *Principles of Gestalt psychology*. New York, NY: Harcourt.

Krumhansl, C. L. (1979). The psychological representation of musical pitch in a tonal context. *Cognitive Psychology*, *11*, 346−374.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. New York, NY: Oxford University Press.

Krumhansl, C. L., & Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review*, *89*, 334−368.

Krumhansl, C. L., Sandell, G. J., & Sergeant, D. C. (1987). The perception of tone hierarchies and mirror forms in twelve-tone serial music. *Music Perception*, *5*, 31−78.

Krumhansl, C. L. & Schmuckler, M. A. (1986). *Key finding in music: An algorithm based on pattern matching to tonal hierarchies.* Paper presented at the 19th annual Mathematical Psychology meeting, Cambridge, MA.

Larson, S. (1997). The problem of prolongation in tonal music: terminology, perception, and expressive meaning. *Journal of Music Theory*, *41*, 101−136.

Leewenberg, E. L. (1971). A perceptual coding language for visual and auditory patterns. *American Journal of Psychology*, *84*, 307−349.

Leman, M. (2000). An auditory model of the role of short-term memory in probe-tone ratings. *Music Perception*, *17*, 481−509.

Lerdahl, F. (2001). *Tonal pitch space*. Oxford, UK: Oxford University Press.

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

Levanen, S., Hari, R., McEvoy, L., & Sams, M. (1993). Responses of the human auditory cortex to changes in one versus two stimulus features. *Journal of Experimental Brain Research*, *97*, 177–183.

Levitin, D. J. (1994). Absolute memory for musical pitch: evidence for the production of learned melodies. *Perception, & Psychophysics*, *56*, 414–423.

Lockhead, G. R., & Byrd, R. (1981). Practically perfect pitch. *Journal of the Acoustical Society of America*, *70*, 387–389.

London, J. (2012). *Hearing in time: Psychological aspects of musical meter* (2nd ed.). Oxford, U.K.: Oxford University Press.

Longuet-Higgins, H. C. (1962a). Letter to a musical friend. *Music Review*, *23*, 244–248.

Longuet-Higgins, H. C. (1962b). Second letter to a musical friend. *Music Review*, *23*, 271–280.

Longuet-Higgins, H. C., & Steedman, M. J. (1971). On interpreting Bach. *Machine Intelligence*, *6*, 221–241.

Mach, E. (1959). *The analysis of sensations and the relation of the physical to the psychical* (C. M. Williams, Trans.; W. Waterlow, review and supplement). New York, NY: Dover. (Original work published 1906)

Marvin, E. W., & Laprade, P. A. (1987). Relating musical contours: extensions of a theory for contour. *Journal of Music Theory*, *31*, 225–267.

Matsunaga, A., & Abe, J. (2005). Cues for key perception of a melody: pitch set alone? *Music Perception*, *23*, 153–164.

McLachlan, N. (2011). A neurocognitive model of recognition and pitch segregation. *Journal of the Acoustical Society of America*, *130*, 2845–2854.

McLean, R. S., & Gregg, L. W. (1967). Effects of induced chunking on temporal aspects of serial retention. *Journal of Experimental Psychology*, *74*, 455–459.

Mercer, T., & McKeown, D. (2010). Updating and feature overwriting in short-term memory for timbre. *Attention, Perception, & Psychophysics*, *72*, 2289–2303.

Merriam, A. P., Whinery, S., & Fred, B. G. (1956). Songs of a Rada community in Trinidad. *Anthropos*, *51*, 157–174.

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.

Miller, G. A., & Chomsky, N. (1963). Finitary models of language users. *Handbook of Mathematical Psychology*, *2*, 419–493.

Miller, G. A., Galanter, E. H., & Pribram, K. H. (1960). *Plans and the structure of behavior*. New York, NY: Holt.

Miyazaki, K. (1989). Absolute pitch identification: effects of timbre and pitch region. *Music Perception*, *7*, 1–14.

Mueller, G. E., & Schumann, F. (1894). Experimentelle Beitrage zur Untersuchung des Gedächtnisses. *Zeitschrift fur Psychologie und Physiologie der Sinnesorgane*, *6*, 81–190, 257–339

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

Narmour, E. (1992). *The analysis and cognition of melodic complexity: The implication-realization model*. Chicago, IL: University of Chicago Press.

Nettl, B. (1956). *Music in primitive culture*. Cambridge, MA: Harvard University Press.

Olson, R. K., & Hanson, V. (1977). Interference effects in tone memory. *Memory & Cognition*, *5*, 32−40.

Oram, N., & Cuddy, L. L. (1995). Responsiveness of Western adults to pitch-distributional information in melodic sequences. *Psychological Research*, *57*, 103−118.

Ortmann, O. (1926). On the melodic relativity of tones [Special issue]. *Psychological Monographs*, *35*. (Whole No. 162).

Oura, Y. (1991). Constructing a representation of a melody: transforming melodic segments into reduced pitch patterns operated on by modifiers. *Music Perception*, *9*, 251−266.

Palmer, S. E. (2002). Perceptual organization in vision. In H. Pashler, & S. Yantis (Eds.), *Stevens' Handbook of Experimental Psychology*: 3rd ed., Vol. 1, *Sensation and Perception* (pp. 177−234). New York, NY: Wiley.

Patterson, R. D. (1986). Spiral detection of periodicity and the spiral form of musical scales. *Psychology of Music*, *14*, 44−61.

Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., & Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron*, *36*, 767−776.

Pearce, M. T., & Wiggins, G. A. (2006). Expectation in melody: the influence of context and learning. *Music Perception*, *23*, 377−405.

Pickler, A. G. (1966). Logarithmic frequency systems. *Journal of the Acoustical Society of America*, *39*, 1102−1110.

Piston, W. (1987). *Harmony* (2nd ed.). London, England: Norton. (Original work published 1948)

Plomp, R., Wagenaar, W. A., & Mimpen, A. M. (1973). Musical interval recognition with simultaneous tones. *Acustica*, *29*, 101−109.

Polansky, L., & Bassein, R. S. (1992). Possible and impossible melodies: Some formal aspects of contour. *Journal of Music Theory*, *36*, 259−284.

Quinn, I. (1997). Fuzzy extensions to the theory of contour. *Music Theory Spectrum*, *19*, 232−263.

Quinn, I. (1999). The combinatorial model of pitch contour. *Music Perception*, *16*, 439−456.

Ragozzine, F., & Deutsch, D. (1994). A regional difference in perception of the tritone paradox within the United States. *Music Perception*, *12*, 213−225.

Rakowski, A. (1994). Investigating short term auditory memory with the method of vocal pitch control. In *Proceedings of the Stockholm Music Acoustics Conference* (pp. 53−57). Stockholm, Sweden: Royal Institute of Technology.

Ratliff, F. (1965). *Mach bands: Quantitative studies of neural networks in the retina*. San Francisco, CA: Holden Day.

Repp, B. H., & Thompson, J. M. (2010). Context sensitivity and invariance in perception of octave-ambiguous tones. *Psychological Research*, *74*, 437−456.

Restle, F. (1970). Theory of serial pattern learning: Structural trees. *Psychological Review*, *77*, 481−495.

Restle, F. (1972). Serial patterns: the role of phrasing. *Journal of Experimental Psychology*, *92*, 385−390.

Restle, F., & Brown, E. (1970). Organization of serial pattern learning. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 4, pp. 249−331). New York, NY: Academic Press.

Révész, G. (1913). *Zur grundlegguncy der tonpsychologie*. Leipzig, Germany: Feit.

Ries, D., & DiGiovanni, J. (2007). Release from interference in auditory working memory for pitch. *Hearing Research*, *230*, 64−72.

Ries, D. T., & DiGiovanni, J. J. (2009). Effects of recurrent tonal information on auditory working memory for pitch. *Hearing Research*, *255*, 14−21.

Ries, D. T., Hamilton, T. R., & Grossmann, A. J. (2010). The effects of intervening interference on working memory for sound location as a function of inter-comparison interval. *Hearing Research*, *268*, 227−233.

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

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

Ross, D., Choi, J., & Purves, D. (2007). Musical intervals in speech. *Proceedings of the National Academy of Sciences*, *104*, 9852−9857.

Ruckmick, C. A. (1929). A new classification of tonal qualities. *Psychological Review*, *36*, 172−180.

Ruusuvirta, T. (2000). Proactive interference of a sequence of tones in a two-tone pitch comparison task. *Psychonomic Bulletin & Review*, *7*, 327−331.

Sachs, C. (1962). *The wellsprings of music*. The Hague, The Netherlands: Martinus Nijhoff.

Scandura, J. M. (1970). Role of rules in behavior: toward an operational definition of what (rule) is learned. *Psychological Review*, *77*, 516−533.

Schellenberg, E. G. (1996). Expectancy in melody: tests of the implication-realization model. *Cognition*, *58*, 75−125.

Schellenberg, E. G. (1997). Simplifying the implication-realization model of musical expectancy. *Music Perception*, *14*, 295−318.

Schellenberg, E. G., & Trehub, S. E. (2003). Good pitch memory is widespread. *Psychological Science*, *14*, 262−266.

Schenker, H. (1956). *Neue musikalische theorien und phantasien: Der Freie Satz*. Vienna, Austria: Universal Edition.

Schmuckler, M. A. (1989). Expectation in music: Investigations of melodic and harmonic processes. *Music Perception*, *7*, 109−150.

Schmuckler, M. A. (1999). Testing models of melodic contour similarity. *Music Perception*, *16*, 295−326.

Schmuckler, M. A. (2004). Pitch and pitch structures. In J. Neuhoff (Ed.), *Ecological psychoacoustics* (pp. 271−315). San Diego, CA: Academic Press.

Schmuckler, M. A. (2009). Components of melodic processing. In S. Hallam, I. Cross, & M. Thaut (Eds.), *Oxford handbook of music psychology* (pp. 93−106). Oxford, UK: Oxford University Press.

Schmuckler, M. A. (2010). Melodic contour similarity using folk melodies. *Music Perception*, *28*, 169−194.

Schoenberg, A. (1951). *Style and idea*. London, England: Williams & Norgate.

Schoenberg, A. (1967). In G. Strong (Ed.), *Fundamentals of musical composition*. New York, NY: St. Martin's Press.

Schoenberg, A. (1969). *Structural functions of harmony* (rev. ed.). New York, NY: Norton. (Original work published 1954)

Schreiner, C. E. (1992). Functional organization of the auditory cortex: maps and mechanisms. *Current Opinion in Neurobiology*, *2*, 516−521.

Schroger, E. (1995). Processing of auditory deviants with changes in one versus two stimulus dimensions. *Psychophysiology*, *32*, 55−65.

Semal, C., & Demany, L. (1991). Dissociation of pitch from timbre in auditory short-term memory. *Journal of the Acoustical Society of America*, *89*, 2404−2410.

Semal, C., & Demany, L. (1993). Further evidence for an autonomous processing of pitch in auditory short-term memory. *Journal of the Acoustical Society of America*, *94*, 1315−1322.

Semal, C., Demany, L., Ueda, K., & Hallé, P. A. (1996). Speech versus nonspeech in pitch memory. *Journal of the Acoustical Society of America*, *100*, 1132−1140.

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

Shepard, R. N. (1965). Approximation to uniform gradients of generalization by monotone transformations of scale. In D. L. Mostofsky (Ed.), *Stimulus generalization*. Stanford, CA: Stanford University Press.

Shepard, R. N. (1982). Structural representations of musical pitch. In D. Deutsch (Ed.), *The psychology of music* (1st ed., pp. 343−390). New York, NY: Academic Press.

Simon, H. A. (1972). Complexity and the representation of patterned sequences of symbols. *Psychological Review*, *79*, 369−382.

Simon, H. A., & Kotovsky, K. (1963). Human acquisition of concepts for sequential patterns. *Psychological Review*, *70*, 534−546.

Simon, H. A., & Sumner, R. K. (1968). Pattern in music. In B. Kleinmuntz (Ed.), *Formal representation of human judgment*. New York, NY: Wiley.

Smith, N. A., & Schmuckler, M. A. (2004). The perception of tonal structure through the differentiation and organization of pitches. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 268−286.

Starr, G. E., & Pitt, M. A. (1997). Interference effects in short-term memory for timbre. *Journal of the Acoustical Society of America*, *102*, 486−494.

Suga, N. (1990). Cortical computational maps for auditory imaging. *Neural Networks*, *3*, 3−21.

Suga, N., O'Neill, W. F., & Manabe, T. (1979). Harmonic-sensitive neurons in the auditory cortex of the mustache bat. *Science*, *203*, 270−274.

Suga, N., & Jen, P. H-S. (1976). Disproportionate tonotopic representation for processing CF-FM sonar signals in the mustache bat auditory cortex. *Science*, *194*, 542−544.

Sutter, M. I., & Schreiner, C. E. (1991). Physiology and topography of neurons with multipeaked tuning curves in cat primary auditory cortex. *Journal of Neurophysiology*, *65*, 1207−1226.

Takeuchi, A. H. (1994). More on key-distance effects in melody recognition: a response to van Egmond and Povel. *Music Perception*, *12*, 143−146.

Takeuchi, A. H., & Hulse, S. H. (1992). Key-distance effects in melody recognition reexamined. *Music Perception*, *10*, 1−24.

Tekman, H. G., & Bharucha, J. J. (1998). Implicit knowledge versus psychoacoustic similarity in priming of chords. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 252−260.

Temperley, D. (2007). *Music and probability*. Cambridge, MA: MIT Press.

Temperley, D., & Marvin, E. W. (2008). Pitch-class distribution and the identification of key. *Music Perception*, *25*, 193−212.

Terhardt, E., & Seewann, M. (1983). Aural key identification and its relationship to absolute pitch. *Music Perception*, *1*, 63−83.

Thomson, W. (1991). *Schoenberg's error*. Philadelphia, PA: University of Pennsylvania Press.

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

Thomson, W. (2006). Pitch frames as melodic archetypes. *Empirical Musicology Review*, *1*, 85−102.

Thompson, W. F., & Stainton, M. (1998). Expectancy in bohemian folk song melodies: Evaluation of implicative principles for implicative and closural intervals. *Music Perception*, *15*, 231–252.

Tillmann, B., & Bharucha, J. J. (2002). Effect of harmonic relatedness on the detection of temporal asynchronies. *Perception & Psychophysics*, *64*, 640–649.

Tillmann, B., Bigand, E., & Pineau, M. (1998). Effects of global and local contexts on harmonic expectancy. *Music Perception*, *16*, 99–117.

Tillmann, B., & Lebrun-Guillaud, G. (2006). Influence of tonal and temporal expectations on chord processing and on completion judgments of chord sequences. *Psychological Research*, *70*, 345–358.

Trainor, L. J., & Trehub, S. E. (1993). Musical context effects in infants and adults: key distance. *Journal of Experimental Human Perception and Performance*, *19*, 615–626.

Treptoe, R. (1997). *Further study of the tritone paradox* (Unpublished manuscript). University of Wisconsin–Stevens Point.

Ueda, K., & Ohgushi, K. (1987). Perceptual components of pitch: spatial representation using multidimensional scaling technique. *Journal of the Acoustical Society of America*, *82*, 1193–1200.

Van Egmond, R., & Butler, D. (1997). Diatonic connotations of pitch-class sets. *Music Perception*, *15*, 1–29.

Van Egmond, R., & Povel, D. J. (1994a). Factors in the recognition of transposed melodies: a comment on Takeuchi and Hulse. *Music Perception*, *12*, 137–142.

Van Egmond, R., & Povel, D. J. (1994b). Similarity judgments of transposed melodies as a function of overlap and key distance. In I. Deliège (Ed.), *Proceedings of the 3rd International Conference on Music Perception and Cognition (ICMPC), Liège* (pp. 219–220). Liège: ICMPC.

Van Egmond, R., & Povel, D. J. (1996). Perceived similarity of exact and inexact transpositions. *Acta Psychologica*, *92*, 283–295.

Van Egmond, R., Povel, D. J., & Maris, E. (1996). The influence of height and key on the perceptual similarity of transposed melodies. *Perception & Psychophysics*, *58*, 1252–1259.

Van Noorden, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences* (Unpublished doctoral thesis). Technische Hogeschool, Eindhoven, Holland.

Vicario, G. B. (1983). *Micro- and macromelodies*. Paper presented at the Fourth Workshop on the Physical and Neuropsychological Foundations of Music, Ossiach, Austria.

Vitouch, O., & Gaugusch, A. (2000). Absolute recognition of musical keys in non-absolute-pitch possessors. In C. Woods, G. Luck, R. Brochard, F. Seddon, & J. A. Sloboda (Eds.), *Proceedings of the 6th International Conference on Music Perception and Cognition*. Dept. of Psychology, Keele University, United Kingdom.

Vitz, P. C., & Todd, T. C. (1969). A coded element model of the perceptual processing of sequential stimuli. *Psychological Review*, *76*, 433–449.

Von Ehrenfels, C. (1890). Uber Gestaltqualitäten. *Vierteljahrschrift fur Wissenschaftliche Philosophie*, *14*, 249–292.

Von Hippel, P. (2000). Redefining pitch proximity: Tessitura and mobility as constraints on melodic intervals. *Music Perception*, *17*, 315–327.

Von Hippel, P. (2002). Melodic-expectation rules as learned heuristics. In C. Stevens, D. Burnham, G. McPherson, E. Schubert, & J. Renwick (Eds.), *Proceedings of the 7th International Conference on Music Perception and Cognition,* Sydney, Australia.

Von Hippel, P., & Huron, D. (2000). Why do skips precede reversals? The effect of tessitura on melodic structure. *Music Perception*, *18*, 59–85.

Vos, P. G. (1999). Key implications of ascending fourth and descending fifth openings. *Psychology of Music*, *27*, 4−18.

Vos, P. G., & Troost, J. M. (1989). Ascending and descending melodic intervals: Statistical findings and their perceptual relevance. *Music Perception*, *6*, 383−396.

Vos, P. G., & Van Geenen, E. W. (1996). A parallel processing key-finding model. *Music Perception*, *14*, 185−224.

Wang, C. C., & Sogin, D. W. (1990). The recognition of melodic fragments as components of tonal patterns. *Psychology of Music*, *18*, 140−149.

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*, *100*, 10038−10042.

Warren, R. M., & Warren, R. P. (1968). *Helmholtz on perception: Its physiology and development*. New York, NY: Wiley.

Watt, H. J. (1924). Functions of the size of interval in the songs of Schubert and of the Chippewa [i.e., Ojibway] and Teton Sioux [i.e., Lakota] Indians. *British Journal of Psychology*, *14*, 370−386.

Weber, G. (1824). *Versuch einer Heordeneten Theorie*. Mainz, Germany: B. Schotts Sohne.

Werner, H. (1925). Uber Mikromelodik und Mikroharmonik. *Zeitschrift fur Psychologie*, *98*, 74−89.

Werts, D. (1983). *A theory of scale references* (Unpublished doctoral dissertation). Princeton University, Princeton, NJ.

Wessinger, C. M., Van Meter, J., Tian, B., Van Lare, J., Pekar, J., & Rauschecker, J. P. (2001). Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. *Journal of Cognitive Neuroscience*, *13*, 1−7.

West, R. J., & Fryer, R. (1990). Ratings of suitability of probe tones as tonics after random orderings of notes of the diatonic scale. *Music Perception*, *7*, 253−258.

White, B. (1960). Recognition of distorted melodies. *American Journal of Psychology*, *73*, 100−107.

Wickelgren, W. A. (1966). Consolidation and retroactive interference in short-term recognition memory for pitch. *Journal of Experimental Psychology*, *72*, 250−259.

Wickelgren, W. A. (1967). Rehearsal grouping and the hierarchical organization of serial position cues in short-term memory. *Quarterly Journal of Experimental Psychology*, *19*, 97−102.

Wickelgren, W. A. (1969). Associative strength theory of recognition memory for pitch. *Journal of Mathematical Psychology*, *6*, 13−61.

Yang, L., Pollack, G. D., & Resler, C. (1992). GABergic circuits sharpen tuning curves and modify response properties in the mustache bat inferior colliculus. *Journal of Neurophysiology*, *68*, 1760−1774.

Yoshino, I., & Abe, J. (2004). Cognitive modeling of key interpretation in melody perception. *Japanese Psychological Research*, *46*, 283−297.