

Psychological Review

VOLUME 89 NUMBER 4 JULY 1982

Geometrical Approximations to the Structure of Musical Pitch

Roger N. Shepard
Stanford University

Rectilinear scales of pitch can account for the similarity of tones close together in frequency but not for the heightened relations at special intervals, such as the octave or perfect fifth, that arise when the tones are interpreted musically. Increasingly adequate accounts of musical pitch are provided by increasingly generalized, geometrically regular helical structures: a simple helix, a double helix, and a double helix wound around a torus in four dimensions or around a higher order helical cylinder in five dimensions. A two-dimensional "melodic map" of these double-helical structures provides for optimally compact representations of musical scales and melodies. A two-dimensional "harmonic map," obtained by an affine transformation of the melodic map, provides for optimally compact representations of chords and harmonic relations; moreover, it is isomorphic to the toroidal structure that Krumhansl and Kessler (1982) show to represent the psychological relations among musical keys.

A piece of music, just as any other acoustic stimulus, can be physically described in terms of two time-varying pressure waves, one incident at each ear. This level of analysis has, however, little correspondence to

the musical experience. Because the ear is responsive to frequencies up to 20 kHz or more, at a sampling rate of two pressure values per cycle per ear, the physical specification of a half-hour symphony requires well in excess of a hundred million numbers. Clearly, our response to the music is based on a much smaller set of psychological attributes abstracted from this physical stimulus. In this respect the perception of music is like the perception of other stimuli such as colors or speech sounds, where the vast number of physical values needed to specify the complete power spectrum of a stimulus is reduced to a small number of psychological values, corresponding, say, to locations on red-green, blue-yellow, and black-white dimensions for homogeneous colors (Hurvich & Jameson, 1957) or on high-low and front-back dimensions for steady-state vowels (Peterson & Barney, 1952; Shepard, 1972). But what, exactly, are the basic perceptual attributes of music?

Just as continuous signals of speech are perceptually mapped into discrete internal representations of phonemes or syllables,

I first described the double helical representation of pitch and its toroidal extensions in 1978 (Shepard, Note 1; also see Shepard, 1978b, p. 183, 1981a, 1981c, p. 320). The present, more complete report, originally drafted before I went on sabbatical leave in 1979, has been slightly revised to take account of a related, elegant development by my colleagues Krumhansl and Kessler (1982).

The work reported here was supported by National Science Foundation Grant BNS-75-02806. It owes much to the innovative researches of Gerald Balzano and Carol Krumhansl, with whom I have been fortunate enough to share the excitement of various attempts to bridge the gap between cognitive psychology and the perception of music. Less directly, the work has been influenced by the writings of Fred Attneave and Jay Dowling. Finally, I am indebted to Shelley Hurwitz for assistance in the collection and analysis of the data and to Michael Kubovy for his many helpful suggestions on the manuscript.

Requests for reprints should be sent to Roger N. Shepard, Department of Psychology, Jordan Hall, Building 420, Stanford University, Stanford, California 94305.

continuous signals of music are perceptually mapped into discrete internal representations of tones and chords; just as each speech sound can be characterized by a small number of distinctive features, each musical tone can be characterized by a small number of perceptual dimensions of pitch, loudness, timbre, vibrato, tremolo, attack, decay, duration, and spatial location. In addition, much as the internal representations of speech sounds are organized into higher level internal representations of meaningful words, phrases, and sentences, the internal representations of musical tones and chords are organized into higher level internal representations of meaningful melodies, progressions, and cadences.

The Fundamental Roles of Pitch and Time in Music

The perceptually salient attributes of the tones making up the musically significant chords and melodies are not equally important for the determination of those higher order units. In fact, for the music of all human cultures, it is the relations specifically of pitch and time that appear to be crucial for the recognition of a familiar piece of music. Other attributes, for example, loudness, timbre, vibrato, attack, decay, and apparent spatial location, although contributing to audibility, comfort, and aesthetic quality, can be varied widely without disrupting recognition or even musical appreciation.

The reasons for the primacy of pitch and time in music are both musical and extramusical. From the extramusical standpoint there are compelling arguments, recently advanced by Kubovy (1981), that pitch and time alone are the attributes that are "indispensable" for the perceptual segregation of the auditory ensemble into discrete tones. From the musical standpoint a case can be made that the richness and power of music depends on the listener's interpretation of the tones in terms of a discrete structure that is endowed with particular group-theoretic properties (Balzano, 1980, in press, Note 2). In the case of the human auditory system, moreover, the requisite properties appear to be fully available only in the dimensions of

pitch (Balzano, 1980; Krumhansl & Shepard, 1979; Shepard, 1982) and time (Jones, 1976; Pressing, in press), the dimensions within which higher order musical units such as melodies and chords are capable of structure-preserving transformations.

In this paper I confine myself to the case of pitch and to the question of how a single psychological attribute corresponding, in the case of pure sinusoidal tones, to a simple one-dimensional physical continuum of frequency affords the structural richness requisite for tonal music. One can perhaps readily conceive how structural complexity is achievable in the dimension of time, through overlapping patterns of rhythm and stress, but the structural properties of pitch seem to be manifested even in purely melodic sequences of purely sinusoidal tones (Krumhansl & Shepard, 1979). In the absence of physical overlap of upper harmonics of the sort considered by Helmholtz (1862/1954) and by Plomp & Levelt (1965), wherein does this structure reside?

Cognitive Versus Psychoacoustic Approaches to Pitch

Until recently, attempts to bring scientific methods to bear on the perception of musical stimuli have mostly adopted a psychoacoustic approach. The goal has been to determine the dependence of psychological attributes, such as pitch, loudness, and perceived duration, on physical variables of frequency, amplitude, and physical duration (Stevens, 1955; Stevens & Volkman, 1940) or on more complex combinations of physical variables (de Boer, 1976; J. Goldstein, 1973; Plomp, 1976; Terhardt, 1974; Wightman, 1973).

By contrast, the cognitive psychological approach looks for structural relations within a set of perceived pitches independently of the correspondence that these structural relations may bear to physical variables. This approach is particularly appropriate when such structural relations reside not in the stimulus but in the perceiver—a circumstance that is well known to students of music theory, who recognize that an interval defined by a given physical difference in log frequency may be heard very differently in different musical contexts. To elaborate on

an example mentioned by Risset (1978, p. 526), in a C-major context the interval B-F is heard as having a strong tendency to resolve by contraction into the smaller interval C-E, whereas in an F⁺-major context the physically identical interval (now called B-E⁺, however) is heard as having a similar strong tendency to resolve by expansion into the larger interval A⁺-F⁺. Moreover, Krumhansl (1979) has now provided systematic, quantitative evidence that the perceived relations between the various tones within an octave do indeed depend on the context-induced musical key with respect to which those tones are interpreted. How, then, are we to represent the relations of pitch between tones as those relations are perceived by a listener who is interpreting the tones musically?

Previous Representations of Pitch Rectilinear Representations

The simplest representations that have been proposed for pitch have been unidimensional scales based on judgments made in nonmusical contexts. Examples are the "mel" scale that Stevens, Volkman, and Newman (1937) and Stevens and Volkman (1940) based on a method of fractionation or the similar scale that Beck and Shaw (1961) later based on a method of magnitude estimation. Pitches are represented in such scales by locations along a one-dimensional line. In these scales, moreover, the location of each pure tone is related to the logarithm of its frequency in a nonlinear manner dictated by the psychoacoustic fact that pairs of low-frequency tones are less discriminable than are pairs of high-frequency tones separated equally in log frequency. Such scales thus deviate from musical scales (or from the spacing of keys on a piano keyboard) in which position is essentially logarithmic with frequency. From a musical standpoint such scales are in fact anomalous in two respects.

First, because of the resulting nonlinear relation between pitch and log frequency, the distance between tones separated by the same musical interval, such as an octave or a fifth, is not invariant under transposition up or down the scale. The failure of musi-

cally significant relations of pitch to emerge as invariant in these scales seems to be a direct consequence of the assiduous avoidance, by the psychoacoustic investigators, of any musical context or tonal framework within which the listeners might interpret the stimuli musically. Atneave and Olson (1971) showed that when familiar melodies, as opposed to arbitrarily selected nonmusical tones, were to be transposed in pitch, listeners required that the separations between the tones be preserved on the musically relevant scale of log frequency—not on a nonlinearly related scale such as the mel scale. That the scale underlying judgments of musical pitch must be logarithmic with frequency is in fact now supported by several kinds of empirical evidence (Dowling, 1978b; Null, 1974; Ward, 1954, 1970).

Second, because of the unidimensionality of scales of pitch such as the mel scale, perceived similarity must decrease monotonically with increasing separation between tones on the scale. There is, therefore, no provision for a particularly significant musical interval, such as the octave, may be perceived as having more in common than tones separated by a somewhat smaller but musically less significant interval, such as the major seventh. Indeed, even the musically more relevant log-frequency scale, also being unidimensional, is subject to this same limitation. Yet, in the case of the octave, which corresponds to an approximately two-to-one ratio of frequencies, the phenomenon of augmented perceptual similarity at that particular interval (a) has long been anticipated (Boring, 1942, pp. 376, 380; Licklider, 1951, pp. 1003-1004; Kucknick, 1929), (b) was in fact empirically observed at about the time that the unidimensional mel scale was being perfected (Blackwell & Schlosberg, 1943; Humphreys, 1939), (c) has since been much more securely established (Allen, 1967; Bachem, 1954; Balzano, 1977; Dowling, 1978a; Dowling & Hollombe, 1977; Idson & Massaro, 1978; Kellman & Massaro, 1979; Krumhansl & Shepard, 1979; Thurlow & Erchul, 1977), and (d) probably underlies the remarkable precision and cross-cultural consistency with which listeners are able to adjust a variable tone so that it stands

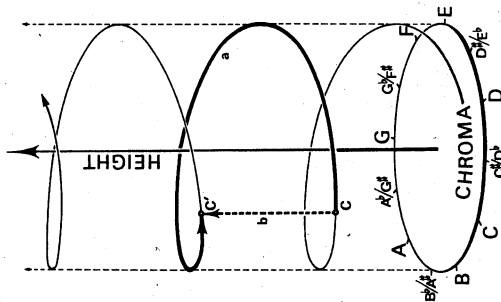


Figure 1. A simple regular helix to account for the increased similarity between tones separated by an octave. (From "Approximation to Uniform Gradients of Generalization by Monotone Transformations of Scale" by Roger N. Shepard. In D. I. Mostofsky (Ed.), *Stimulus Generalization*. Stanford, Calif.: Stanford University Press, 1965, p. 105. Copyright 1965 by Stanford University Press. Reprinted by permission.)

in an octave relation to a given fixed tone (Burns, 1974; Dowling, 1978b; Sundberg & Lindqvist, 1973; and, originally, Ward, 1954).¹

Simple Helical Representations

We can obtain geometrical representations that are consistent with an increased similarity at the octave by deforming a rectilinear representation of pitch into a higher dimensional embedding space to form a helix, as proposed for this purpose by Drobisch as early as 1846, or a spiral, as proposed by Donkin in 1874 (see Piker, 1966; Ruckmick, 1929; and for a recently proposed planar spiral, Hahn & Jones, 1981). For, unlike a straight line, a helix or spiral that completes one turn per octave achieves the desired increase in spatial proximity between points an octave apart—at least if the slope of the curve is not too steep. (Compare Paths a and

b between the tones C and C', an octave apart, in Figure 1.) Moreover, this is true whether the curve is embedded in a cylinder, as proposed by Drobisch, a flat plane, as suggested by Donkin, or a bell-shaped surface of revolution, as advocated by Ruckmick (1929). Despite their differences, the representations proposed by these authors were alike in having adjacent turns more closely spaced toward the low-frequency end, where given differences in log frequency are less discriminable. (See the figures reproduced in Piker, 1966; Ruckmick, 1929.) These representations were analogous, in this respect, to the unidimensional psychophysical representations of Stevens and his colleagues (Stevens et al., 1937; Stevens & Volkman, 1940).

In 1954, before learning of these early proposals, I had attempted to accommodate a heightened similarity at the octave by means of a tonal helix (Note 3) that differed, however, from the ones proposed by Drobisch, Donkin, and Ruckmick in being geometrically uniform or regular. (See Figure 1, which, except for relabeling, is reproduced from Shepard, Note 3—as it later appeared in Shepard, 1965.) Because it is geometrically regular, this is the helical analog of the unidimensional scale having the musically more relevant logarithmic structure advocated by Attneave and Olson (1971); in it the distance corresponding to any particular musical interval is invariant throughout the representation. The uniform helix also possesses an additional advantage over curves embedded in variously shaped surfaces of revolution, such as Ruckmick's (1929) "tonal bell." Only when the helix is regular, and hence embeddable in a cylindrical surface, will tones standing in the octave relation, in addition to coming into closer proximity with each other, fall on a common straight line parallel to the axis of the helix. Such lines

¹ Incidentally, these studies agree in showing that the ratio of physical frequencies of pure tones that subjects set in the octave relationship is about 2.02:1 and not exactly 2:1. This small "stretched-octave" effect (Balzano, 1977; Burns, 1974; Dowling, 1973a; Sundberg & Lindqvist, 1973; Ward, 1954) still leads to an approximately logarithmic relation between pitch and frequency and does not vitiate any of the conclusions to be drawn here.

can be thought of as projecting all tones with the same musical name but differing by octaves (e.g., the tones C, C', C'', etc.) down into a single corresponding point in a "chroma circle" on a plane orthogonal to the axis of the helix (see Figure 1). Moreover, this projective property, unlike the property of augmented proximity, holds regardless of the slope of the helix.

Physical Realization of the Chroma Circle

It is, in fact, the projective property of the regular helix that subsequently led me to a method of physically separating the two underlying components of pitch implicit in the helical representation, namely, the rectilinear component called *pitch height*, corresponding to the axis of the helix (or of the cylinder in which it lies), and the circular component variously called *tone quality* (Revesz, 1954) or *tone chroma* (Bachem, 1950, 1954), corresponding to the circumference of that cylinder (see Shepard, 1964b). What was required was the specification of two physical operations corresponding to the geometrical projections of the entire helix onto the central axis, in the one case, and onto an orthogonal plane, in the other. The auditory realization of the required operations was greatly facilitated by the development, at just this time, of computer techniques for the additive synthesis of arbitrarily specified tones (Mathews, 1963).

For the first operation I proposed a broadening of the band of energy around the center frequency of each tone until the resulting narrow-band noise encompassed about an octave. Because the different sounds generated in this way have different center frequencies, they still differ over the whole range of pitch height. But, because they have all been spread alike around the chroma circle, they can no longer be discriminated with respect to chroma. This operation thus corresponds to collapsing the helix onto its central axis.

For the second operation I proposed, instead, a harmonic elaboration of each tone until it included, alike, all multiples and submultiples of the original frequency (i.e., all tones standing in octave and multiple-octave relations to that original tone), with the am-

plitudes of the component frequencies determined by a fixed bell-shaped spectral envelope that was at its maximum near the middle of the standard musical range and that gradually fell away in both directions to below-threshold levels for very low and very high frequencies. The different sounds generated in this way remain fully distinct in chroma but are all equivalent in height. Thus, in shifting through chroma, from C to C' to D to D' and so on to B, the next step (though still heard as a step up in pitch), instead of leaving one an octave higher at C, leaves one back at the original starting tone C (Shepard, 1964b).² Indeed, application of multidimensional scaling to measures of similarity derived from judgments of relative pitch between tones varied in this manner (Shepard, 1964b) yielded the almost perfectly circular solution displayed in Figure 2 (a) (see Shepard, 1978a). (A similarly circular representation for tones generated in this way has also been reported by Charbonneau and Risset, 1973.)

Although this circular component, chroma, emerges most compellingly when the rectilinear component, height, is artificially suppressed, as with these special, computer-generated tones, the claim is that this circular component is necessarily present in all musical tones for which tones separated by an octave are perceived as more closely related than tones separated by a somewhat smaller interval. Circular multidimensional scaling solutions have in fact been obtained for ordinary musical tones. Figure 2 (b), for example, reproduces a similarly circular pattern subsequently obtained by Balzano (1977, 1982) from a multidimensional scaling analysis of his own discriminative reaction time data for melodic intervals.³

² The illusion of circular or "endlessly ascending" tones can be heard on a commercial record ("Shepard's Tones," 1970) or on a short 16-mm film (Shepard & Zajac, 1965), which we believe to be the first film in which both the sound and the animation were generated by computer. I demonstrated the independent variation of linear pitch height and circular tone chroma at the 1978 meeting of the Western Psychological Association (Shepard, Note 1).

³ One should, however, exercise caution in basing the inference of circularity solely on a multidimensional scaling solution. The curvature evident in some obtained solutions (e.g., the one reported by Levitt, Van de Geer,

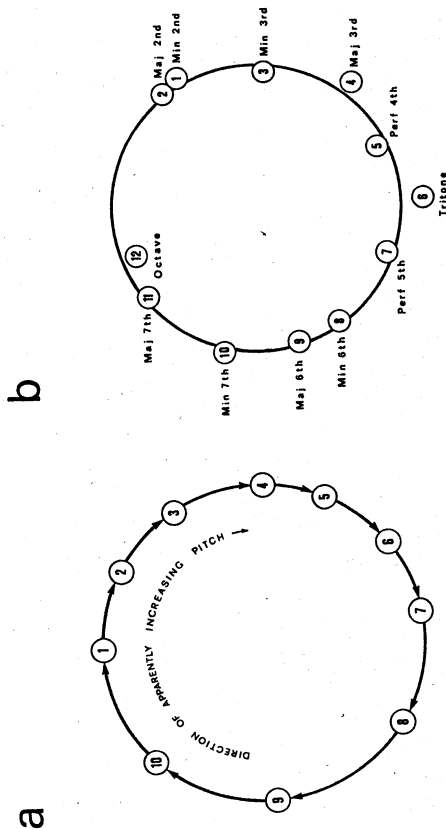


Figure 2. Chroma circles recovered by multidimensional scaling (a) for 10 computer-generated tones especially designed to eliminate differences in pitch height and (b) for twelve ordinary musical tones. (Panel a is from "The Circumplex and Related Topological Manifolds in the Study of Perception" by R. N. Shepard. In S. Shye (Ed.), *Theory Construction and Data Analysis in the Behavioral Sciences*. San Francisco, Calif.: Jossey-Bass, 1978. Copyright 1978 by Jossey-Bass, Inc. Reprinted by permission. Panel b is from *Chronometric Studies of the Musical Interval Sense* by G. J. Balzano. Unpublished doctoral dissertation, Stanford University, 1977. Reprinted by permission.)

By now the possibility of analyzing perceived pitch into the rectilinear and circular components of height and chroma has been accepted by a number of researchers (e.g., Bachem, 1950, 1954; Balzano, 1977; Deutsch, 1972, 1973; Jones, 1976; Kallman & Marsaro, 1979; Pikler, 1955; Revesz, 1954; Risset, 1978). Even among advocates of a helical representation, though, opinions may still differ concerning the relative merits of a geometrically regular structure such as I proposed versus a more or less distorted variant such as Ruckmick (1929) advocated. Here, again, one's predilection may depend on whether one takes a more psychoacoustic or a more cognitive and musical point of view.

Argument for a Geometrically Regular Structure

From the psychoacoustic standpoint it seems natural to suggest that the spacing

& Plomp, 1966) may simply reflect an artifact that almost always arises when basically one-dimensional data are fit in a higher dimensional space (Shepard, 1974, pp. 386-388).

frequencies, amplitudes, or durations that can be adequately transduced by a particular ear or whether or not a preceding context has been provided that is sufficient to activate and to orient or tune the internal structure required for a musical interpretation.

From this cognitive standpoint the already cited evidence for invariance under transposition (Attneave & Olson, 1971; Dowling, 1978b) would require not only that the structure be helical (rather than spiral, say) but also that the helix be geometrically regular and, so, be carried into itself by a rigid screwlike motion under the musical transposition that maps any tone into any other. Only then will the geometrical structure preserve the property, essential to music, that all pairs of tones separated by a given interval, such as an octave or a fifth, have the same musical relation regardless of their overall pitch height. Indeed, from this standpoint the reason that this structure must take a helical form (just as much as the reason that the DNA molecule must take a helical form) is a consequence of the fundamental geometrical fact that the most general rigid motion of space into itself is a combination of a rotation and a translation along the axis of the rotation, that is, a screw displacement (Coxeter, 1961, pp. 101, 321; H. Goldstein, 1950, p. 124; Greenwood, 1965, p. 318; Hilbert & Cohn-Vossen, 1932/1952, pp. 82, 285).

One should not suppose, here, that the octave-related tones that project onto the same point on the chroma circle share some absolutely identifiable quality of chroma any more than the projection of any tone onto the central axis of the helix corresponds to an absolutely identifiable quality of pitch height. Only those rare individuals possessing "absolute pitch" can recognize a chroma absolutely (Bachem, 1954; Siegel, 1972; Ward, 1963a, 1963b). For most individuals (including most musicians), in the case of pitch just as in the case of many other perceptual dimensions (loudness, brightness, size, distance, duration, etc.), it is the relations between presented values that have well-defined internal representations, not the values themselves (Shepard, 1978c, 1981b). In fact, as Attneave and Olson (1971) have so persuasively argued, pitch is perhaps

best thought of not as a stimulus but as a "medium" in which auditory patterns (chords or tunes) can move about while retaining their structural identity, just as visual space is a medium in which luminous patterns can move about while retaining their structural identity. We are not very good at recognizing the position of an isolated point of light in otherwise empty space, but we can say with great precision whether one such point is above or below another or whether three points form a straight line or a right triangle. Likewise, many of us are poor at identifying the pitch of an isolated tone but quite good at saying whether one tone is higher or lower than another or whether three tones stand in octave relations or form a major triad. In either the visual or the auditory case, any such pattern moved to a different location or pitch continues to be recognized as the same pattern.

I go beyond the formulations of Attneave and Olson (1971) and of Dowling (1978b), however, in saying that in the case of pitch, such a motion must be helical. For example, suppose we alternate one well-defined pattern, say a major triad, with exactly the same pattern but each time with the second pattern displaced farther away from the first in pitch height. As the two patterns are moved apart from their initially complete coincidence, each will retain its own structural identity. But the perceived degree of relation between the two patterns will first decrease and then increase again as all three components of one come into the octave relation with corresponding components of the other. (The perceived relation may also increase somewhat at certain intermediate positions as the two triads pass through other special tonal relations, just as the visual similarity to the original of a rotating copy of a triangle will increase at certain intermediate orientations before coming again into complete coincidence at 360°. See Shepard, 1982.) Such a phenomenon is not explicable solely in terms of increasing separation within a purely rectilinear medium; it implies a medium with a circular component. Motions in such a medium are thus auditory analogs of the operations of mental rotation and rotational apparent motion in the visual domain (see Shepard & Cooper, 1982).

Limitations of the Simple Helix

The structure of the simple regular helix pictured in Figure 1 was dictated by two considerations: invariance under transposition and increased similarity at the octave. In such a regular helix, moreover, the special octave relation is represented by unique collinearity or projectability as well as by augmented spatial proximity. As it stands, however, the helical structure does not provide either augmented proximity or collinearity for tones separated by any other special musical interval. Yet, beginning with Ebbinghaus, Drobisch's proposal of a helical representation has been criticized for its failure to account for the special status of the interval of a perfect fifth (see Ruckmick, 1929).

There are, indeed, a number of converging reasons for supposing that the fifth should, like the octave, have a unique status. (a) As has been known at least since Pythagoras, after the 2-to-1 ratio in the lengths of a vibrating string that corresponds to the octave (which as we now know determines a 1-to-2 ratio in the resulting frequency of vibration), the fifth corresponds to the next simplest, 3-to-2, ratio (Helmholtz, 1862/1954). (b) In the case of musical and, therefore, harmonically rich tones, those separated by a fifth also have, within the octave, the fewest upper harmonics that deviate from coincidence by an amount expected to produce noticeable beats (Helmholtz, 1862/1954; Plomp & Levelt, 1965). (c) Moreover, such beats can be subjectively experienced even when the harmonics that most contribute to them are not physically present (Mathews, Note 4; see also Mathews & Sims, 1981). (d) Correspondingly, simultaneously sounded tones differing by a fifth—even pure, sinusoidal tones—tend to be heard as particularly smooth, harmonious, or consonant, and the fifth, together with the similarly harmonic unison major third, completes the uniquely stable and tonally centered chord, the major triad (Meyer, 1956; Piston, 1941; Ratner, 1962; Schenker, 1906/1954, p. 252). (e) In addition, the interval of the fifth plays a pivotal role in tonal music, being the interval that separates musical keys that share the greatest number of tones and between which

the initial experiments in which listeners rated, in effect, the extent to which each individual test tone out of the 13 within one complete octave was substitutable for the tonic tone (*do*) that would normally have completed the major scale presented as context (Krumhansl & Shepard, 1979; also see Krumhansl & Kessler, 1982) and in further experiments in which listeners rated the similarities between the two test tones in all possible pairs selected from one complete octave (Krumhansl, 1979).

Only for listeners with little musical background did the obtained orderings of the musical intervals agree with previous psychoacoustic results in which similarity was determined primarily by proximity in pitch height between the two tones making up the interval, that is, in which the ranking of the intervals with respect to similarity or mutual substitutability of their two component tones was, from greatest to least, unison, minor second, major second, minor third, major third, and so on. For the more musically oriented listeners, the results tended, instead, toward the entirely different ranking: unison and octave (nearly equivalent to each other), followed by the fifth and sometimes the major third, followed by the other tones of the diatonic scale, followed by the remaining, nondiatonic tones (those corresponding in the key of C major to the sharps and flats or black keys of the piano).

In short, data collected from listeners who invest the test tones with a musical interpretation consistently reveal a whole hierarchy of tonal relations that cannot be accommodated within previously proposed geometrical representations of pitch whether rectilinear, helical, or spiral. Accordingly, it now appears justified to present some alternative, generalized helical structures together with the steps that led to their construction and some evidence that such generalized structures are indeed capable of accommodating the musically primary tonal relations. The following is intended, therefore, as the first full account of these new representations of pitch—first briefly described in 1978 (Shepard, Note 1; also see Shepard, 1981a, or, for a description concurrent with that presented here but following a different derivation, Shepard, 1982).

New Representations for Musical Pitch The Diatonic Scale as an Interpretive Schema

In their characteristic eschewal of musical context, psychoacoustically oriented investigators missed the essential musical aspect of pitch. By failing to elicit, within the listeners, the discrete tonal schema or "hierarchy of tonal functions" (Meyer, 1956, pp. 214-215; Piston, 1941; Ratner, 1962) associated with a particular musical key, these investigators left the listeners with no unique cognitive framework within which to interpret the test tones. Even musically sophisticated listeners therefore had little choice but to make their judgments on the basis of the simplest attribute of tones differing in frequency—pitch height.

Cognitively oriented researchers are now recognizing that interpretive schemata play an essential role in the perception of musical pitch, just as they do in perception generally. In the case of pitch, the primary schema seems to be the musical scale—usually, in the case of Western listeners, the familiar major diatonic scale (*do, re, mi, etc.*). As noted by Dowling (1978b, in press), even though the most commonly used musical scales differ somewhat from culture to culture, they all share certain basic properties. Regardless of the total number of tones permitted by each scale, most are organized around five to seven "focal pitches" per octave. Moreover, the steps between such pitches rather than being constant in log frequency are almost always arranged according to a particular asymmetric pattern that repeats exactly within every octave.

The cyclic repetition of the pattern from octave to octave can be explained in terms of the perceived equivalence of tones differing by an approximately 2-to-1 ratio of frequencies, which led to the proposed simple helix for pitch. The other structural universals of musical scales have been attributed to pervasive cognitive constraints on the number of absolutely identifiable categories per perceptual dimension (7 ± 2 , as enumerated by Miller, 1956; cf. Dowling, 1978b) and to the requirement that the scale have a structure that affords reference points or tonal centers to which a melody can move