

Modeling Tonality: Applications to Music Cognition

Elaine Chew (eniale@alum.mit.edu)

University of Southern California
Integrated Media Systems Center, and
Department of Industrial and Systems Engineering
Los Angeles, CA 90089-1450 USA

Abstract

Processing musical information is a task many of us perform effortlessly, and often, unconsciously. In order to gain a better understanding of this basic human cognitive ability, we propose a mathematical model for tonality, the underlying principles for tonal music. The model simultaneously incorporates pitch, interval, chord and key relations. It generates spatial counterparts for these musical entities by aggregating musical information. The model also serves as a framework on which to design algorithms that can mimic the human ability to organize musical input. One such skill is the ability to determine the key of a musical passage. This is equivalent to being able to pick out the most stable pitch in the passage, also known as “doh” in *solfege*. We propose a computational algorithm that mimics this human ability, and compare its performance to previous models. The algorithm is shown to predict the correct key with high accuracy. The proposed computational model serves as a research and pedagogical tool for putting forth and testing hypotheses about human perception and cognition in music. By designing efficient algorithms that mimic human cognitive abilities, we gain a better understanding of what it is that the human mind can do.

Introduction

Music cognition is a complex task requiring the integration of information at many different levels. Nevertheless, processing musical information is an act with which we are all familiar. The mind is so adept at organizing and extracting meaningful patterns when listening to music that we are often not even aware of what it is that we do when comprehending music. Some of this unconscious activity includes determining the tonal center¹, the rhythm, and the phrase structure of the piece.

I illustrate our unconscious ability to process music by a short anecdote from my own experiences. In my first semester as a pianolab² instructor at MIT, I encountered a few students who had no prior musical background. I asked one such student, after he carefully traced out the melodic line for Yankee Doodle, “What is the key³ of

this piece?” He responded with a reasonable question: “What do you mean by key?” I began singing the piece and stopped mid-stream. I then asked the student if he could sing me the note on which the piece should end. Without hesitation, he sang the correct pitch⁴, thereby successfully picking out the first degree, and most stable pitch, in the key. The success of this method raised more questions than it answered. What is it we know that causes us to hear one pitch as being more stable than others? How does the mind assess the function of this stable pitch over time as the music evolves?

Before we can study music cognition, we first need a representation for musical structure. In this paper, we propose a mathematical model for tonality, the underlying principles of tonal music. According to Bamberger (2000), “tonality and its internal logic frame the coherence among pitch relations in the music with which [we] are most familiar.” The model uses spatial proximity to represent perceived distances between musical entities. The model simultaneously incorporates representations for pitch, interval, chord and key relations.

Using this model, we design a computational algorithm to mimic human decisions in determining keys. The process of key-finding precedes the evaluation of melodic and harmonic structure, and is a fundamental problem in music cognition. We relate this new representation to previous models by Longuet-Higgins & Steedman (1971) and Krumhansl & Schmuckler (1986). The computational algorithm is shown to identify keys at a high level of accuracy, and its performance is compared to that of the two previous models.

The Representation

Western tonal music is governed by a system of rules called *tonality*. The first part of the paper proposes a geometric representation, the Spiral Array model, that captures this system of relations among tonal elements. The Spiral Array model offers a parsimonious description of the inter-relations among tonal elements, and suggests

ordered, the first degree of the scale gives the scale its name. This is also the most stable pitch, known as the *tonic*.

⁴A *pitch* is a sound of some frequency. High frequency sounds produce a high pitch, and low frequency sounds produce a low pitch. This is distinct from a *note*, which is a symbol that represents two properties, pitch and duration.

¹The *tonal center*, also called the *tonic* of the key, is the pitch that attains greatest stability in a musical passage.

²A keyboard skills class for students enrolled in Music Fundamentals and Composition courses.

³Excerpted from the Oxford Dictionary of Music: A *key* implies adherence, in any passage, to the note-material of one of the major or minor scales. When the pitches in a scale are

new ways to re-conceptualize and reorganize musical information. A hierarchical model, the Spiral Array generates representations for pitches, intervals, chords and keys within a single spatial framework, thus allowing comparisons among elements from different hierarchical levels. The basic idea behind the Spiral Array is the representation of higher level tonal elements as aggregates of their lower level components.

Spatial analogues of physical and psychological phenomena are known to be powerful tools for solving abstract intellectual problems (Shepard, 1982). Some have argued that problems in music perception can be reduced to that of finding an optimal data representation (Tanguine, 1993). Shepard (1982) determined that “the cognitive representation of musical pitch must have properties of great regularity, symmetry, and transformational invariance.” The model placed all twelve chromatic pitches equally over one full turn of a spiral, and highlighted pitch height relations. Further extensions to a double helix emphasized perfect fifth interval⁵ relations, but did not account for major and minor third relations.

Applying multi-dimensional scaling techniques to experimental data, Krumhansl (1978,1990) mapped listener ratings of perceived relationships between probe tones and their contexts into space. The resulting cone (1978) places pitches in the tonic triad closest to each other, confirming the psychological importance of fifth and third interval relations, which form triads. Parncutt (1988) has also presented a psychoacoustical basis for the perception of triadic units.

Another representation that incorporates spatial counterparts for both perfect fifth and major/minor third relations is the *tonnetz*, otherwise known as the Harmonic Network. This model has been used by music theorists since Riemann (see, for example, Lewin, 1987; Cohn, 1998), who posited that tonality derives from the establishing of significant tonal relationships through chord functions. Cohn (1998) has traced the earliest version of this network to the 18th century mathematician Euler, and used the *tonnetz* representation to characterize different compositional styles, focussing on preferred chord transitions in the development sections. More recently, Krumhansl (1998) presented experimental support for the psychological reality of these neo-Riemannian transformations.

Our proposed Spiral Array model derives from a three-dimensional realization of the Harmonic Network, and takes into account the inherent spiral structure of the pitch relations. It is distinct from the Harmonic Net-

⁵Excerpted from the Oxford Dictionary of Music, an *interval* is the distance between any two pitches expressed by a number. For example, C to G is a 5th, because if we proceed up the major scale of C, the fifth pitch is G. The 4th, 5th and octave are all called Perfect. The other intervals, measured from the first pitch, in the ascending major scale are all called Major. Any Major interval can be chromatically reduced by a semitone (distance of a half step) to become Minor. If any Perfect or Minor interval is so reduced, it becomes Diminished; if any Perfect or Major interval is increased by a semitone it becomes Augmented.

work in that it assigns spatial representations for higher level musical entities in the same structure. The representations for intervals, chords and keys are constructed as mathematical aggregates of spatial representations of their component parts.

Like the models derived from multi-dimensional scaling, the Spiral Array model uses proximity to incorporate information about perceived relationships between tonal elements. Distances between tonal entities as represented spatially in the model correspond to perceived distances among sounding entities. Perceptually close intervals are defined following the principles of music theory. In accordance with the Harmonic Network, the Spiral Array assigns greatest prominence to perfect fifth and major/minor third interval relations, placing elements related by these intervals in proximity to each other.

In the calibration of the model, the parameter values that affect proximity relations are prescribed based on a few perceived relations among pitches, intervals, chords and keys. These proximity relations will be described in a later section.

The Spiral Array Model

As the name suggests, in the Spiral Array Model, pitches are represented by points on a spiral. Adjacent pitches are related by intervals of perfect fifths. Pitches are indexed by their number of perfect fifths from C, which has been chosen arbitrarily as the reference pitch. For example, D has index two because C to G is a perfect fifth, and G to D is another. $P(k)$ denotes the point on the spiral representing a pitch of index k . Each pitch can be defined in terms of transformations from its previous neighbor - a rotation, and a vertical translation.

$$P(k+1) \stackrel{\text{def}}{=} R P(k) + h,$$

$$\text{where } R = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and } h = \begin{bmatrix} 0 \\ 0 \\ h \end{bmatrix}.$$

The pitch C is arbitrarily set at the point $[0,1,0]$.

Since the spiral makes one full turn every four pitches to line up vertically above the starting pitch position. Positions representing pitches four indices, or a major third, apart are related by a simple vertical translation:

$$P(k+4) = P(k) + 4h.$$

For example, C and E are a major third apart, and E is positioned vertically above C.

At this point, we diverge from the original *tonnetz* to define chord and key representations in the three-dimensional model. The added complexity of the three-dimensional realization allows one to define representations off the grid, and *within* the spiral. A chord is the composite result, or effect, of its component pitches. A key is the effect of its defining chords. We propose that this effect can be represented spatially by a convex combination of its components.

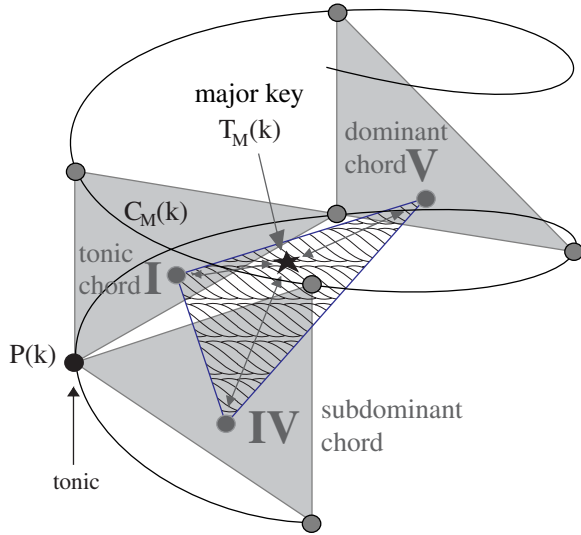


Figure 1: The Spiral Array Model.

Mathematically, the chord's representation is generated by a convex combination of its three component pitch positions. Geometrically, the chord representation resides strictly within the boundaries of the triangle outlined by the triad (see Figure 1). A chord is represented by a weighted average of its component pitch positions: the root $P(k)$, the fifth $P(k+1)$, and the third $P(k+4)$ for major triads, and $P(k-3)$ for minor triads:

The representation for a major triad is

$$C_M(k) \stackrel{\text{def}}{=} w_1 P(k) + w_2 P(k+1) + w_3 P(k+4),$$

where $w_1, w_2, w_3 > 0$ and $\sum_{i=1}^3 w_i = 1$.

The minor triad is generated by a similar combination,

$$C_m(k) \stackrel{\text{def}}{=} u_1 P(k) + u_2 P(k+1) + u_3 P(k-3),$$

where $u_1, u_2, u_3 > 0$ and $\sum_{i=1}^3 u_i = 1$.

The weights, w_i and u_i , on the pitch positions represent the importance of the pitch to the generated chord. For longstanding psychological, physical and theoretical reasons, the root is deemed the most important, followed by the fifth, then the third. Correspondingly, the weights are constrained to be monotonically decreasing from the root, to the fifth, to the third. In order that spatial distance mirrors these relations, there are additional constraints on the aspect ratio h/r . These constraints are described in Chew (2000).

An important property of the Spiral Array is that representations of pitches in a given key occupy a compact neighborhood. Each major chord, together with its right and left neighbor major chords, combine to produce the effect of a major key. In music terminology, these chords

are given names, with respect to the key, that reflect their function. The center chord is called the tonic chord (I)⁶, the one to its right the dominant (V), and the one to its left the subdominant (IV). Hence, we represent the major key as a combination of its I, V and IV chords. For example, the representation of the C major key is generated by the C major, G major and F major chord representations. See Figure 1 for an example of a major key representation.

Mathematically, the representation for a major key, $T_M(k)$ is the weighted average of its tonic triad ($C_M(k)$), dominant triad ($C_M(k+1)$) and subdominant triad ($C_M(k-1)$) representations. As before, the design objective is to have the weights correspond to each chord's significance in the key. Hence, the I chord is given the largest weight, followed by that of the V chord, then the IV chord:

$$T_M(k) \stackrel{\text{def}}{=} \omega_1 C_M(k) + \omega_2 C_M(k+1) + \omega_3 C_M(k-1),$$

where $\omega_1, \omega_2, \omega_3 > 0$ and $\sum_{i=1}^3 \omega_i = 1$.

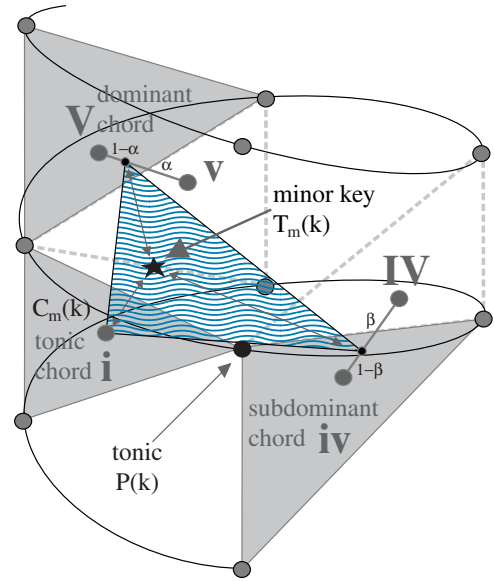


Figure 2: Geometric representation of a minor key, a composite of its tonic (i), dominants (V/v) and subdominant (iv/IV) chords.

The definition for the minor key is more complicated,

⁶We shall use roman numerals to denote chord function within a key. The number indicates the scale degree of the chord's root. For example, "I" represents the tonic chord. We adopt the convention of denoting major chords by upper case roman numerals, and minor chords by lower case ones. For example, a major chord with the tonic as root is "I" but a minor chord with the same root is "i".

but we will not go into the details at this time. It suffices to say that the center of effect for the minor key $T_m(k)$ is modeled as a combination of the tonic $C_m(k)$, the major and minor dominant triads $C_M(k+1)$ and $C_m(k+1)$, and the major and minor subdominant triad $C_m(k-1)$ and $C_M(k-1)$:

$$T_m(k) \stackrel{\text{def}}{=} v_1 C_m(k) + v_2 \alpha C_M(k-1) + (1-\alpha) C_m(k-1) + v_3 \beta C_m(k-1) + (1-\beta) C_M(k-1),$$

where $v_1, v_2, v_3 > 0$ and $v_1 + v_2 + v_3 = 1$,
and $0 \leq \alpha \leq 1, 0 \leq \beta \leq 1$.

See Figure 2 for the spatial representation of a minor key.

Properties of the Spiral Array Model

In the Spiral Array model, musical information is condensed, or aggregated, and represented by a single point. Proximity in the Spiral Array indicates some combination of the following: shared pitches, shared intervals, or tonal elements within a perfect fifth, major third or minor third interval of each other. This section summarizes the criteria for selecting the weights defined in the previous section so that relations between represented tonal entities have direct counterparts in the geometric structure. Details are given in Chew (2000). The criteria are summarized as follows:

1. Perceptually close intervals should be represented by shorter inter-pitch distances. For example, the closest distance between any two pitch positions denotes a perfect fifth relation; and, pitches a third apart are closer than those a second apart, etc.
2. Each chord representation is closest to its root, followed by the fifth, then the third; and, no other pitches are as close to the major chord center as its three constituent pitches.
3. The average position of two pitches an interval of a half step apart should be closest to the key related to the upper pitch; and, the average position of two pitches an interval of a perfect fourth apart should be closest to the key related to the upper pitch.

These preliminary criteria are subjective, and are by no means comprehensive. We found, through experiments, that by satisfying these few conditions, the model performed well when applied to the problem of key-finding (as described in the next sections). Thus, this could be reason to believe that with a few simple conditions, we might be able to capture the salient features in musical information in a way that concurs with listener perceptions.

Finding the Key of a Melody

The Spiral Array provided a framework on which to design viable and efficient computational algorithms for

problems in the analysis and manipulation of musical information. Because the model condenses musical information to a spatial point, it allows for efficient and dynamic tracking of a streams of musical signals. Using the model, an algorithm is designed to determine the key of musical passages. We illustrate how the algorithm works by an example, “Simple Gifts”. This algorithm is shown to perform better than existing ones when applied to the 24 fugue subjects in Book I of Bach’s “Well-Tempered Clavier” (henceforth, referred to as the WTC). This algorithm exemplifies the concept of mapping musical information onto the Spiral Array.

Analyzing the key of a melody poses many challenges. Given a melody, one must make informed decisions about its key based on little information. Furthermore, there could be more than one equally valid answer, in which case a list for the most likely candidates for key would be more appropriate than one definite key. This section introduces the key-finding algorithm (CEG) based on the Spiral Array that returns a ranked list of possible keys. CEG is an acronym for *Center of Effect Generator*. The CEG algorithm is fundamental to the Spiral Array model and uses the model to reframe the problem of key recognition as a computationally simple one of finding a distance-minimizing representation.

In the Spiral Array, the collection of pitches in a given key defines a compact space. As pitches in a melody are sounded in sequence, the geometric shape defined by the pitch positions becomes increasingly more complex. Instead of using this complex shape to identify the key, *the algorithm collapses the pitch information down to a single point*, the center of effect (c.e.). In this manner, the pitches combine to create an object in space — a point which is the composite sum of the pitch positions.

Since keys are also defined as points in space, it is then simple to compute the distance between the c.e. and the key, and nearby keys, to determine which key is closest to the c.e. Thus the mathematical sum of pitches affords parsimonious descriptions of, and comparisons between, different pitch collections.

However, the CEG algorithm more than simply compares pitch collections. By definition, the key representations favor triadic pitch configurations, and also tonic-dominant and tonic-subdominant relationships. These representations incorporate different levels of hierarchical structure and relationships. Not all pitches are weighted equally; and, the key representation is a structured but nonlinear combination of its pitch collection. By comparing the c.e.’s to these key representations, we expect certain pitch relations to prevail.

An Example

The algorithm is best explained by an example. Consider the Shaker tune, used in Copland’s symphonic suite “Appalachian Spring” (1945), shown in Figure 3.

At any point in time, the CEG method generates a c.e. from the given musical information that summarizes the tonal space generated by the pitches sounded. Define a step to be a pitch event. At each step, the pitches from the



Figure 3: “Simple Gifts”.

beginning to the present is weighted (multiplied) by its duration, and the c.e. is generated by *aggregating* these weighted pitch positions.

If the i -th note is represented in the Spiral Array by pitch position p_i and has duration d_i , then the *aggregate center* at the i -th pitch event is defined as:

$$c_i \stackrel{\text{def}}{=} \sum_{j=1}^i d_j p_j.$$

The CEG method updates its decision with each note or pitch event. The distance from the key representations to c_i is calculated and ranked. The key that is closest is ranked first, next closest second, and so on.

Figure 4 plots the exact distances from the four closest keys (F major, C major, F minor and C minor), at each successive pitch event. Observe, in the graph, that F major quickly establishes itself as the closest key. However, between pitch events $i = 22$ to 24, C major (the dominant of F) vies with F major for preeminence. The melody dwells on the dominant key area at $i = 19$ to 24, outlining the C major triad from $i = 21$ to 24. This behavior in the model concurs with listener perception.

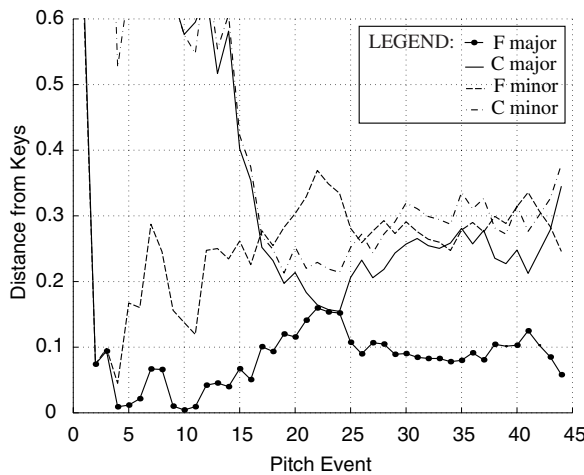


Figure 4: Distance to various keys as “Simple Gifts” unfolds.

Comparison to other Key-Finding Models

To validate the model, it was compared to Longuet-Higgins & Steedman’s Shape-Matching Algorithm (SMA) (1971) and to Krumhansl & Schmuckler’s Probe

Tone Profile Method (PTPM) (1986). Detailed discussions of each test run is documented in Chew (2000).

The tonic-dominant rule was devised for cases when the SMA algorithm failed to reach the desired conclusion by the end of the fugue theme. In such cases, the tonic-dominant rule derives the key from the first pitch which is assumed to be either the tonic or the dominant of the intended key. The † denotes cases when the tonic-dominant rule was applied. Numbers in brackets denote the average when considering only the fugue subjects in which the tonic-dominant rule was not applied by any of the three methods.

Table 1: Applying key-finding algorithm to Bach’s fugue subjects in the WTC. (Numbers generated using $h = \begin{bmatrix} 2 & 15 \end{bmatrix}$ ($r = 1$), and weights across all hierarchies set to $[0.516, 0.315, 0.168]$.)

Book I Fugue subj	Steps to key		
	CEG	PTPM	SMA
C major	2	2	16† ⁷
C minor	5	5	5
C major	6	7	16
C minor	3	3	4
D major	2	2	15†
D minor	3	3	8
E major	2	6	11†
D minor	2	6	12†
E major	14	12†	11
E minor	3	2	7†
F major	4	10	6
F minor	3	15	4†
F major	3	2	8
F minor	7	18	5†
G major	2	2	15
G minor	3	3	4
A major	3	2	7†
G minor	5	5	5
A major	2	4	7
A minor	5	5	5
B major	4	4	14
B minor	2	3	6†
B major	2	11	11
B minor	3	3	7
Average	3.75 (3.57)	5.25 (4.79)	8.71 (8.21)

For the fugue subjects in Book I of the WTC, the CEG required on average 3.75 pitch events, the PTPM 5.25, and the SMA 8.71 to determine the correct key. Given a melody, a hypothesis of its key based on its first pitch is not a meaningful one. The reliability of a hypothesis based on two pitch events is still questionable. Hence, on average, the absolute minimum number of pitches re-

quired to form an opinion of the key is 3. The CEG algorithm required, on average, 3.75 steps to determine the key of the 24 fugue subjects. Based on the reasons stated, we claimed that the key-finding algorithm using the Spiral Array has an average performance that is close to optimal.

Comments

The approach detailed in this paper is computational, and mimics the manifestation of human music cognitive abilities. It proposes mathematical ways to aggregate and organize musical information. However, this does not imply that the computational algorithm describes how the human mind processes musical information. The fact that it performs well suggests that it should be considered as a method of modeling human cognition in music.

A computational model serves as a research and pedagogical tool for putting forth and testing hypotheses about human perception and cognition in music. For example, one can generate hypotheses about how humans perceive musical groupings, and implement this theory using the model.

In the melodies used for model validation, and in the “Simple Gifts” example, the aggregate points (c.e.’s) were generated cumulatively as the melodies unfolded. For lengthier examples, some decay of the information over time should be incorporated into the c.e.’s. This would be a way to model short-term memory in listening to music.

At present, the model ignores the dimension of pitch height. Clearly, pitches from different registers will generate different perceptions of relatedness. Future modifications could take into consideration the modeling of pitch height by weighting pitches from different registers differently.

The CEG algorithm currently proceeds sequentially forward through time, and cumulatively aggregates the information to produce a representation for the c.e. Cognitively, a human listener makes judgements about the key not only sequentially forward in time as the melody unfolds, but also retroactively after having gained some future information. A harmonic analysis algorithm proposed by Winograd (1968) proceeds backwards from the end of the piece; and, Temperley’s (1999) extension of the Krumhansl-Schmuckler model employs dynamic programming, which also works backwards algorithmically. Future extensions of the Spiral Array Model might incorporate elements of both forward and retroactive decision-making.

By designing efficient algorithms that mimic human cognitive abilities, we gain a better understanding of what it is that the human mind can do. By studying the shortcomings of the algorithms, we can modify them, and in so doing, learn about the extent of human cognitive abilities. In the examples we discussed, the information was processed sequentially forward through time. In actual fact, the human listener can often retroactively change his or her decision about structural properties of the piece after having listened to more of the music. At-

tempts to model this would yield further insight as to the temporal nature of music cognition.

Acknowledgments

Jeanne Bamberger’s cogent advice and unflagging support has made this research possible. This work was funded in part by the Josephine de Karman Dissertation Fellowship administered through the Massachusetts Institute of Technology.

References

- Bamberger, Jeanne (2000). *Developing Musical Intuition*. New York, NY: Oxford University Press.
- Chew, Elaine (2000). *Towards a Mathematical Model of Tonality*. Doctoral dissertation, Department of Operations Research, Massachusetts Institute of Technology, Cambridge, MA.
- Cohn, Richard (1998). Introduction to Neo-Riemannian Theory: A Survey and a Historical Perspective. *Journal of Music Theory*, 42 (2), 167–180.
- Cohn, Richard (1997). Neo-Riemannian Operations, Parsimonious Trichords, and their Tonnetz Representations. *Journal of Music Theory*, 41 (1), 1–66.
- Longuet-Higgins, H. C. & Steedman, M. J. (1971). On Interpreting Bach. *Machine Intelligence*, 6, 221.
- Krumhansl, Carol L. (1990). *Cognitive Foundations of Musical Pitch*. New York, NY: Oxford University Press.
- Krumhansl, C. L. & Schmuckler, M. A. (1986). The Petroushka chord: A perceptual investigation. *Music Perception*, 4, 153–184.
- Krumhansl, C. L. (1998). Perceived Triad Distance: Evidence Supporting the Psychological Reality of Neo-Riemannian Transformations. *Journal of Music Theory*, 42 (2), 265–281.
- Krumhansl, C. L. (1978). *The Psychological Representation of Musical Pitch in a Tonal Context*. Doctoral dissertation, Department of Psychology, Stanford University, Stanford, CA.
- Lewin, David (1987). *Generalized Musical Intervals and Transformations*. New Haven, CT: Yale University Press.
- Shepard, Roger N. (1982). Structural representations of musical pitch. In D. Deutsch (Ed.), *The Psychology of Music*. New York, NY: Academic Press.
- Tanguine, Andranick S. (1993). *Artificial Perception and Music Recognition*. Lecture Notes in Artificial Intelligence. New York, NY: Springer-Verlag.
- Temperley, David (1999). What’s Key for Key? The Krumhansl-Schmuckler Key-Finding Algorithm Reconsidered. *Music Perception*, 17 (1), 65–100.
- Winograd, Terry (1968). Linguistics and the Computer Analysis of Tonal Harmony. *Journal of Music Theory*, 12 (1), 2–49.