

Tonal Cognition

CAROL L. KRUMHANS� AND PETRI TOIVIAINEN

Department of Psychology, Cornell University, Ithaca, New York 14853, USA

Department of Music, University of Jyväskylä, Jyväskylä, Finland

ABSTRACT: This article presents a self-organizing map (SOM) neural network model of tonality based on experimentally quantified tonal hierarchies. A toroidal representation of key distances is recovered in which keys are located near their neighbors on the circle of fifths, and both parallel and relative major/minor key pairs are proximal. The map is used to represent dynamic changes in the sense of key as cues to key become more or less clear and modulations occur. Two models, one using tone distributions and the other using tone transitions, are proposed for key-finding. The tone transition model takes both pitch and temporal distance between tones into account. Both models produce results highly comparable to those of musically trained listeners, who performed a probe tone task for ten nine-chord sequences. A distributed mapping of tonality is used to visualize activation patterns that change over time. The location and spread of this activation pattern is similar for experimental results and the key-finding model.

KEY WORDS: Music; Tonality; Cognition; Probe tone

INTRODUCTION

Tonality induction refers to the process through which the listener develops a sense of the key of a piece of music. The concept of tonality is central to Western music, but eludes definition. From the point of view of musical structure, tonality is related to a cluster of features, including musical scale (usually major or minor), chords, the conventional use of sequences of chords in cadences, and the tendencies for certain tones and chords to suggest or be “resolved” to others. From the point of view of experimental research on music cognition, tonality has implications for establishing hierarchies of tones and chords, and for inducing certain expectations in listeners about how melodic and harmonic sequences will continue. One method for studying the perception of tonality is the probe tone method, which quantifies the tonal hierarchy. When applied to unambiguous key-defining contexts, it provides a standard for determining key strengths when more ambiguous and complex musical materials are presented. In addition to experimental studies, considerable effort has been spent developing computational models. This effort has produced various symbolic and neural network models, including a number that take musical input and return a key identification, sometimes called key-finding models.

Address for correspondence: Professor Carol L. Krumhansl, Department of Psychology, Uris Hall, Cornell University, Ithaca, NY 14853. Voice: 607-255-6351; fax: 607-255-8433.
clk4@cornell.edu

PROBE TONE METHODOLOGY

An experimental method introduced to study tonality is sometimes referred to as the probe tone method.¹ It is best illustrated with a concrete example. Suppose you hear the tones of the ascending C major scale: C D E F G A B. There is a strong expectation that the next tone will be the tonic, C, first, because it is the next logical tone in the series and, second, because it is the tonic of the key. In the experiment, the incomplete scale context was followed by the tone C (the probe tone), and listeners were asked to judge how well it completed the scale on a numerical scale (1 = very bad, 7 = very good). As expected, the C received the maximal rating. Other probe tones, however, also received fairly high ratings, and they were not necessarily those that are close to the tonic C in pitch. For example, the most musically trained listeners also gave high ratings to the dominant, G, and the mediant, E. In general, the tones of the scale received higher ratings than the nonscale tones, C# D# F# G# A#. This suggested that it is possible to get quantitative judgments of the degree to which different tones are perceived as stable, final tones in tonal contexts.

A subsequent study² used this method with a variety of musical contexts at the beginning of the trials. They were chosen because they are clear indicators of a key. They included the scale, the tonic triad chord, and chord cadences in both major and minor keys. These were followed by all possible probe tones in the chromatic scale, which musically trained listeners were instructed to judge in terms of how well they fit with the preceding context in a musical sense. Different major keys were used, as were different minor keys. The results for contexts of the same mode were similar when transposed to a common tonic. Also, the results were similar independent of which particular type of context was used. Consequently, the data were averaged over these factors. We call the resulting values the K-K profiles, which can be expressed as vectors. The vector for major keys is: K-K major profile = <6.35, 2.23, 3.48, 2.33, 4.38, 4.09, 2.52, 5.19, 2.39, 3.66, 2.29, 2.88>. The vector for minor keys is: K-K minor profile = <6.33, 2.68, 3.52, 5.38, 2.60, 3.53, 2.54, 4.75, 3.98, 2.69, 3.34, 3.17>.

We can generate K-K profiles for 12 major keys and 12 minor keys from these. If we adopt the convention that the first entry in the vector corresponds to the tone C, the second to C#/Db, the third to D, and so on, then the vector for C major is: <6.35, 2.23, 3.48, 2.33, 4.38, 4.09, 2.52, 5.19, 2.39, 3.66, 2.29, 2.88>, the vector for Db major is: <2.88, 6.35, 2.23, 3.48, 2.33, 4.38, 4.09, 2.52, 5.19, 2.39, 3.66, 2.29>, and so on. The vectors for the different keys result from shifting the entries the appropriate number of places to the tonic of the key.

TRACING THE DEVELOPING AND CHANGING SENSE OF KEY

The probe tone method was then used to study how the sense of key develops and changes over time.² Ten nine-chord sequences were constructed, some of which contained modulations between keys. Musically trained listeners did the probe tone task after the first chord, then after the first two chords, then after the first three chords, and continued until the full sequence was heard. This meant that 12 (probe tones) \times 9 (chord positions) \times 10 (sequences) = 1080 judgments were made by each listener. Each of the 90 sets of probe tone ratings was compared with the ratings made for the

unambiguous key-defining contexts. That is, each set of probe tone ratings was correlated with the K-K profiles for the 24 major and minor keys. For some of the sets of probe tone ratings (some probe positions in some of the chord sequences), a high correlation was found, indicating a strong sense of key. For other sets of probe tone ratings, no key was highly correlated, which was interpreted as an ambiguous or weak sense of key.

As should be obvious from the above, the probe tone task requires an intensive empirical effort to trace how the sense of key develops and changes, even for short sequences. In addition, the sequence needs to be interrupted, and the judgment is made after the sequence has been interrupted. For these reasons, the judgments may not faithfully mirror the experience of music in time. Hence, we³ are currently testing an alternative form of the probe tone methodology. In this method, which we call the concurrent probe tone task, the probe tone is presented continuously while the music is played. The complete passage is sounded together with a probe tone. Then the passage is sounded again, this time with another probe tone. This process is continued until all probe tones have been sounded. Preliminary results suggest this methodology produces interpretable results, at least for musically trained participants. Our focus here, however, is on how the sense of key can be represented, whether the input to the representation is from a probe tone task or from a model of key-finding as described later.

A GEOMETRIC MAP OF KEY DISTANCES FROM THE TONAL HIERARCHIES

The K-K profiles generated a highly regular and interpretable geometric representation of musical keys.² The basic assumption underlying this approach was that two keys are closely related to each other if they have similar tonal hierarchies. That is, keys were assumed to be closely related if tones that are stable in one key are also relatively stable in the other key. To measure the similarity of the profiles, a product-moment correlation was used. It was computed for all possible pairs of major and minor keys, giving a 24×24 matrix of correlations showing how similar the tonal hierarchy of each key was to every other key. To give some examples, C major correlated relatively strongly with A minor (0.651), with G major and F major (both 0.591), and with C minor (0.511). C minor correlated relatively strongly with E \flat major (0.651), with C major (0.511), with A \flat major (0.536), and less strongly with F minor and G minor (both 0.339).

A technique called multidimensional scaling was then used to create a geometric representation of the key similarities. The algorithm locates 24 points (corresponding to the 24 major and minor keys) in a spatial representation to best represent their similarities. It searches for an arrangement such that points that are close correspond to keys with similar K-K profiles (as measured by the correlations). In particular, non-metric multidimensional scaling seeks a solution such that distances between points are (inversely) related by a monotonic function to the correlations. A measure (called "stress") measures the amount of deviation from the best-fitting monotonic function. The algorithm can search for a solution in any specified number of dimensions. In this case, a good fit to the data was found in four dimensions.

The four-dimensional solution located the 24 keys on the surface of a torus (generated by one circle in dimensions 1 and 2, and another circle in dimensions 3 and 4). Because of this, any key can be specified by two values: its angle on the first circle and its angle on the second circle. Thus, the result can be depicted in two dimensions as a rectangle where it is understood that the left edge is connected to the right edge, and the bottom edge is connected to the top edge. The locations of the 24 keys were interpretable in terms of music theory. There was one circle of fifths for major keys (...F \sharp /G \flat , D \flat , A \flat , E \flat , B \flat , F, C, G, D, A, E, B, F \sharp /G \flat ...) and one circle of fifths for minor keys (...f \sharp , c \sharp , g \sharp , d \sharp /e \flat , b \flat , f, c, g, d, a, e, b, f \sharp ...). These wrap diagonally around the torus such that each major key is located near both its relative minor (for example, C major and A minor) and its parallel minor (for example, C major and C minor). (See FIGS. 1 and 2 below.)

REPRESENTING THE SENSE OF KEY ON THE TORUS

The continuous spatial medium containing the 24 major and minor keys affords representing the sense of key in a graphical form,² using a technique called multidimensional unfolding. It is a method that is closely related to multidimensional scaling. Multidimensional unfolding begins with a multidimensional scaling solution, in this case the torus representation of the 24 major and minor keys. This solution is considered fixed. The algorithm then finds a point in the multidimensional scaling solution to best represent the sense of key at each point in time. Let P_1 be the probe tone ratings after the first chord in a sequence; it is a 12-dimensional vector of ratings for each tone of the chromatic scale. This vector was correlated with each of the 24 K-K vectors, giving a 24-dimensional vector of correlations. The unfolding algorithm finds a point to best represent these correlations. Suppose P_1 correlates highly with the K-K profile for F major and fairly highly with the K-K profile for D minor. Then the unfolding algorithm will produce a point near these keys and far from the keys with low correlations. Then the vector of correlations was computed for P_2 , giving a second point. This process continues until the end of the sequence.

In this manner, each of the ten nine-chord sequences² generated a series of nine points on the torus representation of keys. For nonmodulating sequences, the points remained in the neighborhood of the intended key. For the modulating sequences, the first points were near the initial intended key, then shifted to the region of the second intended key. Modulations to closely related keys appeared to be assimilated more rapidly than those to distantly related keys. That is, the points shifted to the region of the new key earlier in sequences containing close modulations than in sequences containing distant modulations.

MEASUREMENT ASSUMPTIONS OF THE MULTIDIMENSIONAL SCALING AND UNFOLDING METHODS

The above methods make a number of assumptions about measurement, only some of which will be noted here. The torus representation is based on the assumption that correlations between the K-K profiles are appropriate measures of interkey distance. It further assumes that these distances can be represented in a relatively

low-dimensional space (four dimensions). This latter assumption was supported by the low stress value (high goodness-of-fit value) of the multidimensional scaling solution. It was further supported by a subsidiary Fourier analysis of the K-K major and minor profiles, which found two relatively strong harmonics.⁴ In fact, plotting the phases of the two Fourier components for the 24 key profiles was virtually identical to the multidimensional scaling solution. This supports the torus representation, which consists of two orthogonal circular components. Nonetheless, it would seem desirable to see whether an alternative method with completely different assumptions recovers the same toroidal representation of key distances.

The unfolding method also adopts correlation as a measure of distances from keys, this time using the ratings for each probe position in the chord sequences and the K-K vectors for the 24 major and minor keys. The unfolding technique finds the best-fitting point in the four-dimensional space containing the torus. It does not provide a way of representing cases in which no key is strongly heard, because it cannot generate points outside the space containing the torus. Thus, an important limitation of the unfolding method is that it does not provide a representation of the strength of the key or keys heard at each point in time. For this reason, we sought a method that is able to represent both the region of the key or keys that are heard, together with their strengths.

In addition, the unfolding approach assumes that spatial locations between the 24 points for major and minor keys are meaningful. An intermediate position would result from a blend of tonal hierarchies of nearby keys. However, other sets of probe tone ratings might also map to the same position. Thus, the identification of points between keys is not necessarily unique. This motivated an alternative model which explicitly specifies the meaning of positions between keys.

THE SELF-ORGANIZING MAP (SOM) OF KEYS

The self-organizing map (SOM)⁵ is an artificial neural network that simulates the formation of ordered feature maps. The SOM consists of a two-dimensional grid of units, each of which is associated with a reference vector. Through repeated exposure to a set of input vectors, the SOM settles into a configuration in which the reference vectors approximate the set of input vectors according to some similarity measure; the most commonly used similarity measures are the Euclidean distance and the direction cosine. The direction cosine between an input vector \mathbf{x} and a reference vector \mathbf{m} is defined by

$$\cos\theta = \frac{\sum_i x_i m_i}{\sqrt{\sum_i x_i^2} \sqrt{\sum_i m_i^2}} = \frac{\mathbf{x} \cdot \mathbf{m}}{\|\mathbf{x}\| \|\mathbf{m}\|}. \quad (1)$$

Another important feature of the SOM is that its configuration is organized in the sense that neighboring units have similar reference vectors. For a trained SOM, a mapping from the input space onto the two-dimensional grid of units can be defined by associating any given input vector with the unit whose reference vector is most similar to it. Because of the organization of the reference vectors, this mapping is

smooth in the sense that similar vectors are mapped onto adjacent regions. Conceptually, the mapping can be thought of as a projection onto a nonlinear surface determined by the reference vectors.

We trained the SOM with the 24 K-K profiles. The SOM was specified in advance to have a toroidal configuration, that is, the left and the right edges of the map were connected to each other as were the top and the bottom edges. Euclidean distance and direction cosine, when used as similarity measures in training the SOM, yielded identical maps. (The resulting map is displayed in Figs. 1 and 2 below.) The map shows the units with reference vectors that correspond to the K-K profiles. The SOM configuration is highly similar to the multidimensional scaling solution² and the Fourier-analysis-based projection⁴ obtained with the same set of vectors. Unlike those maps, however, all locations in the map are explicitly associated with a reference vector so that they are uniquely identified.

REPRESENTING THE SENSE OF KEY ON THE SOM

A distributed mapping of tonality can be defined by associating each unit with an activation value. For each unit, this activation value depends on the similarity between the input vector and the reference vector of the unit. Specifically, the units whose reference vectors are highly similar to the input vector have a high activation, and vice versa. The activation value of each unit can be calculated, for instance, using the direction cosine of Equation 1. Dynamically changing data from either probe tone experiments or key-finding models can be visualized as an activation pattern that changes over time. The location and spread of this activation pattern provides information about the perceived key and its strength. More specifically, a focused activation pattern implies a strong sense of key and vice versa.

KEY-FINDING MODELS

A variety of key-finding algorithms have been proposed. The objective of these models is to assign a key to an input sample of music (such as the fugue subjects of J. S. Bach's *Well-Tempered Clavier*). Symbolic models, reviewed elsewhere,⁴ have taken into consideration a number of factors in assigning key, including scales that contain the tones of the sample, and the presence of such cues to key as the tonic-fifth, tonic-third, and tonic-leading-tone intervals, characteristic tone sequences, and cadences. Some effort has been made to take phrasing, melodic accent, and rhythm into account. More recently proposed symbolic models^{6,7} have made advances in both computational and music-analytic sophistication.

Concurrently, neural network models have provided an alternative subsymbolic approach.^{8,9} In these, the input sample typically gives rise to activation levels of units associated with different keys. Thus, these models return graded measures of key strength. The problem of representing these in a way that takes into account the distances between keys, however, remains unsolved. In addition, little has been done to compare the output of key-finding algorithms to perceptual judgments, which may change over time as the cues to key become more or less clear and as the music

may modulate to other keys. For the most part, the model's output has been compared with the composer's key signature.

Both these issues were addressed by a key-finding algorithm developed using the K-K profiles.⁴ The input to this algorithm was a 12-dimensional vector specifying the total durations of the twelve chromatic scale tones in the musical selection to which the key was to be assigned. One application was to Bach's *C Minor Prelude* from Book II of the *Well-Tempered Clavier*. It was treated on a measure-by-measure basis. Each input vector was projected onto the toroidal map of keys using the phases of the two strong Fourier components. Two musical experts gave quantitative judgments of key strength in each measure of the piece, which were also projected onto the map. The key-finding algorithm followed the same pattern of modulations found by the experts, suggesting the approach lends itself to tracing dynamic changes in key as modulations occur. One obvious limitation of this approach is that it ignores the order of the tones in the input sample. This potentially important cue for key led to the development of the tone transition model described below.

STONE TRANSITIONS AND KEY-FINDING

The order in which tones are played may provide additional information that is useful for key-finding. This is supported by studies on both tone transition probabilities^{10,11} and perceived stability of tone pairs in tonal contexts.^{4,13} In samples of compositions by Bach, Beethoven, and Webern, only a small fraction of all the possible tone transitions were actually used (the fractions were 23, 16, and 24 percent, respectively).¹⁰ Furthermore, in a sample of 20 songs by Schubert, Mendelssohn, and Schumann, there was an asymmetry in the transition frequencies in the sense that certain tone transitions were used more often than the same tones in the reverse temporal order.¹¹ For instance, the transition B-C was used 93 times, whereas the transition C-B was used only 66 times. A similar asymmetry was found in studies on perceived stability of tone pairs in a tonal context.^{4,13} After the presentation of a tonal context, tone pairs that ended with a tone that was high in the tonal hierarchy were given higher ratings than the reverse temporal orders. For instance, in the context of C major, the ratings for the transitions B-C and C-B were 6.42 and 3.67, respectively.

Determining tone transitions in a piece of polyphonic music is not a trivial task, especially if one aims at a representation that corresponds to perceptual reality. Even in a monophonic piece, the transitions can be ambiguous in the sense that their perceived strengths may depend on the tempo. Consider, for example, the tone sequence C4-G3-D4-G3-E4, where all the tones have equal durations. When played slowly, this sequence is heard as a succession of tones oscillating up and down in pitch. With increasing tempi, however, the subsequence C4-D4-E4 becomes increasingly prominent. This is because these tones segregate into one stream due to the temporal and pitch proximity of its members, separate from G3-G3. With polyphonic music, the ambiguity of tone transitions becomes even more obvious. Consider, for instance, the sequence consisting of a C major chord followed by a D major chord, where the tones of each chord are played simultaneously. In principle, this passage contains nine different tone transitions. Some of these transitions are, however, perceived as

stronger than others. For instance, the transition G-A is, due to pitch proximity, perceived as stronger than the transition G-D.

It seems, thus, that the analysis of tone transitions in polyphonic music should take into account principles of auditory stream segregation.¹⁴ Furthermore, it may be necessary to code the presence of transitions on a continuous instead of a discrete scale. In other words, each transition should be associated with a strength value instead of just coding whether that particular transition is present or not. Below, a dynamic system that embraces these principles is described. In regard to the evaluation of transition strength, the system bears a resemblance to a proposed model applying the concept of apparent motion to music.¹⁵

TONE TRANSITION MODEL

Let the piece of music under examination be represented as a sequence of tones, where each tone is associated with pitch, onset time, and duration. The main idea of the model is the following: given any tone in the sequence, there is a transition from that tone to all the tones following that particular tone. The strength of each transition depends on three factors: pitch proximity, temporal proximity, and the duration of the tones. More specifically, a transition between two tones has the highest strength when the tones are proximal in both pitch and time and have long durations. These three factors are included in the following dynamic model.

Representation of Input

The pitches of the chromatic scale are numbered consecutively. The onset times of tones having pitch k are denoted by t_{ki}^n , $i = 1, \dots, n_k$, and the offset times by t_{ki}^f , $i = 1, \dots, n_k$, where n_k is the total number of times the k^{th} pitch occurs.

Pitch Vector $\mathbf{p}(t) = (p_k(t))_k$

Each component of the pitch vector has nonzero value whenever a tone with the respective pitch is sounding. It has the value of 1 at each onset at the respective pitch, decays exponentially after that, and is set to zero at the tone offset. The time evolution of \mathbf{p} is governed by the equation

$$\dot{p}_k = -p_k/\tau_p + \sum_{i=1}^{n_k} \delta(t - t_{ki}^n) - p_k \sum_{i=1}^{n_k} \delta(t - t_{ki}^f), \quad (2)$$

where \dot{p}_k denotes the time derivative of p_k and $\delta(\cdot)$ the Dirac delta function (unit impulse function). The time constant τ_p has the value of $\tau_p = 0.5$ sec. With this value, the integral of p_k saturates at about 1 sec after tone onset, thus approximating the durational accent as a function of tone duration.¹⁶

Pitch Memory Vector $\mathbf{m}(t) = (m_k(t))_k$

The pitch memory vector provides a measure of both the perceived durational accent and the recency of notes played at each pitch. In other words, a high value of m_k

indicates that a tone with pitch k and a long duration has been played recently. The dynamics of \mathbf{m} are governed by the equation

$$\dot{\mathbf{m}} = \mathbf{p} - \mathbf{m}/\tau_m, \quad (3)$$

The time constant τ_m determines the dependence of transition strength on the temporal distance between the tones. In the simulations, the value of $\tau_m = 3$ sec has been used, corresponding to typical estimates of the length of the auditory sensory input.^{17–19}

Transition Strength Matrix $\mathbf{S}(t) = (s_{kl}(t))_{kl}$

The transition strength matrix provides a measure of the instantaneous strengths of transitions between all pitch pairs. More specifically, a high value of s_{kl} indicates that a long tone with pitch k has been played recently and a tone with pitch l is currently sounding. The temporal evolution of \mathbf{S} is governed by the equation

$$s_{kl} = m_k \frac{1 + \text{sgn}(p_l - p_k)}{2} p_l e^{-(k-l)^2/\alpha^2}, \quad (4)$$

In this equation, the nonlinear term $(1 + \text{sgn}(p_l - p_k))/2$ is used for distinguishing between simultaneously and sequentially sounding pitches. This term is nonzero only when $p_l > p_k$, that is, when the most recent onset of pitch l has occurred more recently than that of pitch k . The term $e^{-(k-l)^2/\alpha^2}$ weights the transitions according to the interval size. For the parameter α , the value $\alpha = 6$ has been used. With this value a perfect fifth gets a weight of about 0.37 times the weight of a minor second.

Dynamic Tone Transition Matrix $\mathbf{N}(t) = (n_{kl}(t))_{kl}$

The dynamic tone transition matrix is obtained by temporal integration of the transition strength matrix. At a given point of time, it provides a measure of the strength and recency of each possible tone transition. The time evolution of \mathbf{N} is governed by the equation

$$\dot{\mathbf{N}} = \mathbf{S} - \mathbf{N}/\tau_N, \quad (5)$$

where the time constant τ_N is equal to τ_m , that is, $\tau_N = 3$ sec.

To examine the role of tone transitions in key-finding, we developed two key-finding models. Model 1 is based on pitch class distributions. Model 2 is based on tone transition distributions. Below, a brief description of the models is given.

KEY-FINDING MODEL 1

Model 1 is based on pitch class distributions only. Like the earlier algorithm based on the K-K profiles,⁴ it does not take tone transitions into account. However, it has a dynamic character in that both the pitch vector and the pitch memory vector depend on time. It uses a pitch class vector $\mathbf{p}_c(t)$, which is similar to the pitch vector $\mathbf{p}(t)$ used in the dynamic tone transition matrix, except that it ignores octave information. Consequently, the vector has 12 components that represent the pitch classes.

The pitch class memory vector $\mathbf{m}_c(t)$ is obtained by temporal integration of the pitch class vector according to the equation

$$\dot{\mathbf{m}}_c = \mathbf{p}_c - \mathbf{m}_c / \tau_d. \quad (6)$$

Again, the time constant has the value $\tau_d = 3$ seconds. To obtain estimates for the key, vector $\mathbf{m}_c(t)$ is correlated with the K-K profiles for each key. Alternatively, or in addition, the vectors $\mathbf{m}_c(t)$ can be projected onto the toroidal key representation using activation values as described earlier. Both approaches will be taken here.

KEY-FINDING MODEL 2

Model 2 is based on tone transitions. Using the dynamic transition matrix \mathbf{N} , it calculates the octave-equivalent transition matrix $\mathbf{N}' = (n'_{ij})_{ij}$ according to

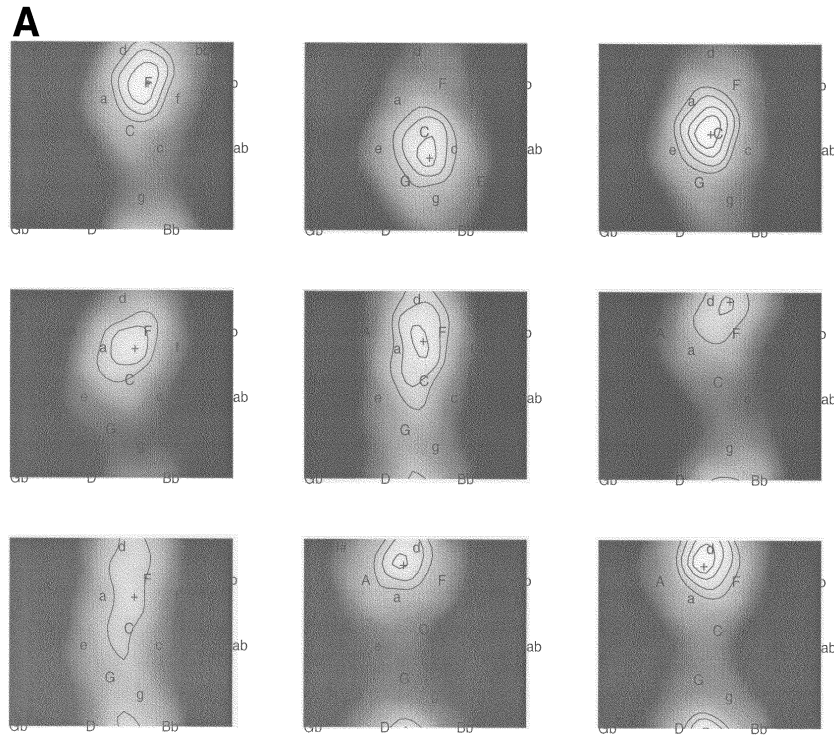
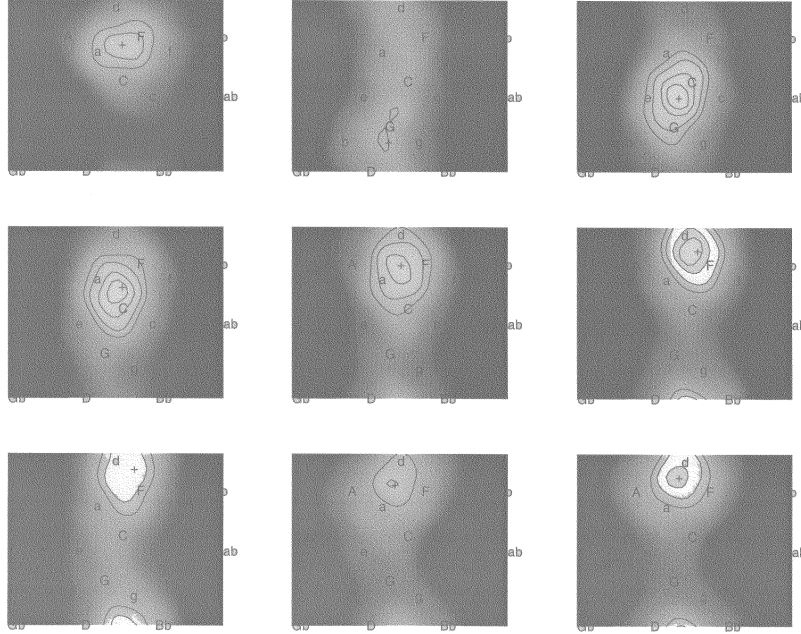


FIGURE 1. (A) Distributed mapping of tonality on the SOM for probe tone judgments for the nine-chord sequence modulating from C major to D minor. The *top row* shows the results for chords 1, 2, and 3 (*left to right*), the *second row* for 4, 5, and 6 (*left to right*), and the *bottom row* for chords 7, 8, and 9 (*left to right*).

B**FIGURE 1. (B)** Distributed mapping of model 1 for the same sequence.

$$n_{ij}'(t) = \sum_{\substack{i = p \bmod 12 \\ j = p \bmod 12}} n_{pq}(t). \quad (7)$$

In other words, transitions whose first and second tones have identical pitch classes are considered equivalent, and their strengths are added. Consequently, the melodic direction of the tone transition is not taken into account. To obtain estimates for the key, the pitch class transition matrix is correlated with the matrices representing the perceived stability of two-tone transitions for each key.⁴

APPLICATION OF MODELS 1 AND 2 TO THE CHORD SEQUENCE DATA

The probe tone data for the ten chord sequences described earlier² were compared with the results of models 1 and 2 applied to those sequences. An interchord onset time of one second was used. The perceptual judgments were correlated with the K-K profiles, and these were compared with the key correlations of the two models. Model 1 results correlated highly ($r(2158) = 0.87, p < 0.0001$) with the experimental data. Model 2 also correlated highly ($r(1918) = 0.86, p < 0.0001$). (Model 2 produces

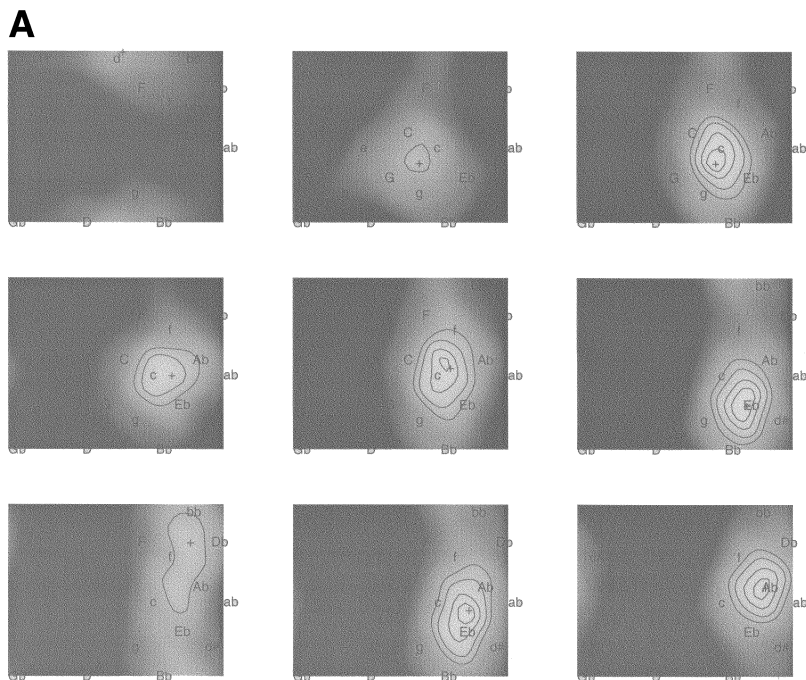


FIGURE 2. (A) Distributed mapping of tonality on the SOM for probe tone judgments for the nine-chord sequence modulating from C minor to A♭ minor.

an output only after the first two chords.) Model 2 contributed additional precision when combined with model 1, as shown by a multiple regression predicting the experimental data from both models ($R(2, 1917) = 0.89$, $p < 0.0001$, with both models contributing significantly at $p < 0.0001$; the standard coefficient was 0.50 for model 1 and 0.43 for model 2). Thus, the two models generally matched well the experimental results, and modeled slightly different aspects of the listeners' responses.

FIGURES 1 and 2 show the distributed mapping of tonality on the SOM for the listeners (FIGS. 1a and 2a) and model 1 (FIGS. 1b and 2b) for two illustrative chord sequences. (Because of issues about how best to visualize the results of model 2, we show only model 1 here.) In these representations, a single contour outlines correlations 0.70 or above, double contours outline correlations 0.80 or above, triple contours outline correlations 0.90 or above, and quadruple contours outline correlations 0.95 or above.

The sequence depicted in FIGURE 1 consists of the chords F G C F d B♭ e° A d, containing a relatively distant modulation from C major to D minor. After the first chord, F, appeared, the listeners and the model had a clear focus on the key of F major in which the chord is I. When the second chord, G, was sounded, listeners apparently interpreted this as a IV–V progression in C major, resulting in a focus near that

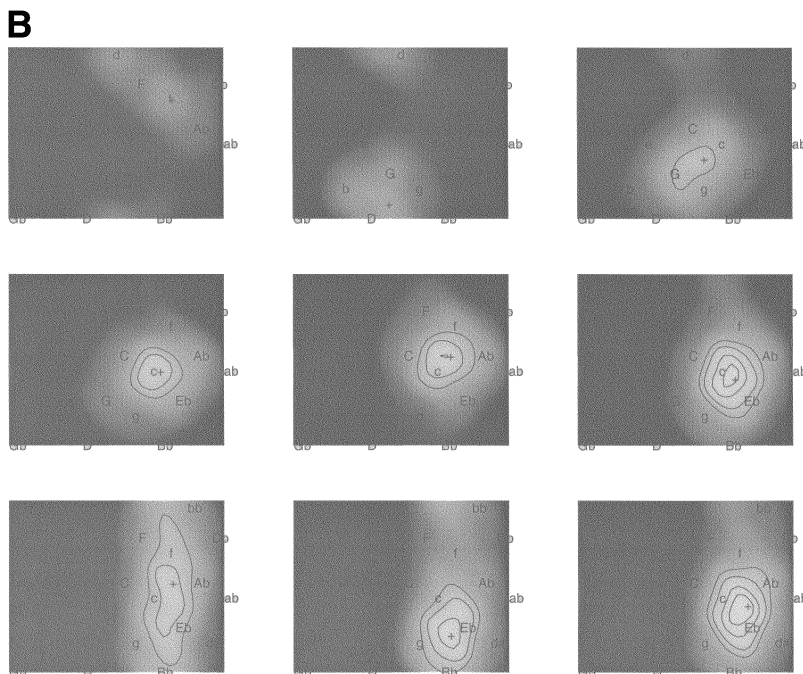


FIGURE 2. (B) Distributing mapping of model 1 for the same sequence.

key. The model, however, did not find this focus until the tonic triad, C, in the third position. For the fourth and fifth chords, F and d, the focus of both listeners and model remained in the region of C major but was shifted somewhat toward F major and D minor. The sixth chord, B \flat , which is contained in D minor but not C major, shifted the focus farther toward D minor. The seventh chord, e $^\circ$, greatly weakened the sense of key for both listeners and model; a diffuse elongated region of weak activation was found. With the last two chords, A and d, the final focus on D minor (in which these are V–I) was arrived at.

The sequence depicted in FIGURE 2 consists of the chords d $^\circ$ G c A b F D b b b E b A b, containing a modulation from C minor to A b major, a modulation to what is considered a relatively close key. Both listeners and model yielded an extremely weak activation pattern for the first two chords, d $^\circ$ and G, although the focus for listeners after the second chord was weakly near C minor. The third chord, c, clarified this focus near C minor for both listeners and model, where it remained for the fourth and fifth chords, A b and f. The sixth chord, E b, which is the tonic triad in the relative major of C minor (E b major) produced a strong focus on E b major for listeners and a shift in that direction for the model. The seventh chord, B b, diffused this focus for both listeners and model before the new key of A b major clarified with the E b and A b chords in the last two positions (which are V–I in the A b major key).

CONCLUSION

Experimental studies suggest that the sense of tonality undergoes dynamic and subtle changes when a listener hears music. The sense of key develops and strengthens as certain cues appear, then may weaken or shift to a new key as subsequent events are sounded. An important step in understanding this process is a suitable means of representing such changes. Toward this end, we have developed a spatial representation based on psychological data. The distributed map of the SOM provided a visually accessible representation of these subtle dynamic changes. The key-finding models described here suggested that listeners' sense of key can be modeled quite well using tone distributions and tone transitions. Two models were formulated to incorporate various psychological phenomena, including durational accent as a function of tone duration, the duration of sensory memory, temporal-order asymmetries, and pitch streaming as a function of the distances between tones in pitch and time. The models accounted well for the results of an experiment with relatively simple and short chord sequences. Further tests of such models with more complex, extended, and musically realistic materials may point to additional factors, such as rhythm, meter, phrasing, and form, that may also influence the sense of tonality.

REFERENCES

1. KRUMHANSL, C.L. & R.N. SHEPARD. 1979. Quantification of the hierarchy of tonal functions within a diatonic context. *J. Exp. Psychol. Hum. Percept. Perform.* **5**: 579–594.
2. KRUMHANSL, C.L. & E.J. KESSLER. 1982. Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychol. Rev.* **89**: 334–368.
3. KRUMHANSL, C.L. & P. TOIVAINEN. 2000. Dynamics of tonality induction: a new method and a new model. *In* Proceedings of the Sixth International Conference on Music Perception and Cognition. C. Woods, G. Luck, R. Brochard, F. Seddon & J.A. Sloboda, Eds.: 1504–1513. Keele University. Keele, UK.
4. KRUMHANSL, C.L. 1990. *Cognitive Foundations of Musical Pitch*. Oxford University Press. New York.
5. KOHONEN, T. 1997. *Self-organizing Maps*. Springer-Verlag. Berlin.
6. VOS, P.G. & E.W. VAN GEENEN. 1996. A parallel-processing key-finding model. *Music Percept.* **14**: 185–223.
7. VOS, P.G. & M. LEMAN, Eds. 2000. *Tonality Induction*. Special issue of *Music Percept.*
8. LEMAN, M. 1995. A model of retroactive tone-center perception. *Music Percept.* **12**: 439–471.
9. BHARUCHA, J.J. 1999. Neural nets, temporal composites, and tonality. *In* *The Psychology of Music*. 2nd edit. D. Deutsch, Ed.: 413–440. Academic Press. San Diego, CA.
10. FUCKS, W. 1962. Mathematical analysis of the formal structure of music. *IRE Trans. Information Theory* **8**: 225–228.
11. YOUNGBLOOD, J.E. 1958. Style as information. *J. Music Theory* **2**: 24–35.
12. KNOPOFF, L. & W. HUTCHINSON. 1978. An index of melodic activity. *Interface* **7**: 205–229.
13. KRUMHANSL, C.L. 1979. The psychological representation of musical pitch in a tonal context. *Cognit. Psychol.* **11**: 346–374.
14. BREGMAN, A.S. 1990. *Auditory Scene Analysis*. M.I.T. Press. Cambridge, MA.
15. GJERDINGEN, R.O. 1994. Apparent motion in music? *Music Percept.* **11**: 335–370.
16. PARNCUTT, R. 1994. A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Percept.* **11**: 409–464.

17. DARWIN, C.J., M.T. TURVEY & R.G. CROWDER. 1972. An auditory analogue of the Sperling partial report procedure: evidence for brief auditory storage. *Cognit. Psychol.* **3**: 255–267.
18. FRAISSE, P. 1982. Rhythm and tempo. *In* *The Psychology of Music*. 1st edit. D. Deutsch, Ed.: 149–180. Academic Press. San Diego, CA.
19. TREISMAN, A.M. 1964. Verbal cues, language, and meaning in selective attention. *Am. J. Psychol.* **77**: 206–219.