Perception, 2003, volume 32, pages 741 – 766

DOI:10.1068/p3312

Measuring and modeling real-time responses to music: The dynamics of tonality induction

Petri Toiviainen
Department of Music, University of Jyväskylä, PO Box 35, Jyväskylä 40351, Finland;
e-mail: ptoiviai@jyu.fi

Carol L Krumhansl
Department of Psychology, Cornell University, 214 Uris Hall, Ithaca, NY 14853-7601, USA
Received 22 November 2001, in revised form 30 October 2002; published online 11 June 2003

Abstract. We examined a variety of real-time responses evoked by a single piece of music, the organ *Duetto* BWV 805 by J S Bach. The primary data came from a concurrent probe-tone method in which the probe tone is sounded continuously with the music. Listeners judged how well the probe tone fit with the music at each point in time. The process was repeated for all probe tones of the chromatic scale. A self-organizing map (SOM) [Kohonen 1997 *Self-organizing Maps* (Berlin: Springer)] was used to represent the developing and changing sense of key reflected in these judgments. The SOM was trained on the probe-tone profiles for 24 major and minor keys (Krumhansl and Kessler 1982 *Psychological Review* **89** 334 – 368). Projecting the concurrent probe-tone data onto the map showed changes both in the perceived keys and in their strengths. Two dynamic models of tonality induction were tested. Model 1 is based on pitch class distributions. Model 2 is based on the tone-transition distributions; it tested the idea that the order of tones might provide additional information about tonality. Both models contained dynamic components for characterizing pitch strength and creating pitch memory representations. Both models produced results closely matching those of the concurrent probe-tone data. Finally real-time judgments of tension were measured. Tension correlated with distance away from the predominant key in the direction of keys built on the dominant and supertonic tones, and also correlated with dissonance.

A great deal of experimental work on music perception has been carried out in which retrospective judgments have been used. This method consists of presenting subjects with musical stimuli, after which they are requested to rate the stimulus according to some feature. Retrospective judgments have been applied to studies on, for instance, pitch memory (Deutsch 1972, 1975, 1978; Cuddy et al 1981), musical expectancy (Schmuckler 1989; Cuddy and Lunney 1995; Krumhansl 1995a, 1995b; Krumhansl et al 1999, 2000), tonality (Krumhansl and Shepard 1979; Krumhansl and Kessler 1982), melodic similarity (Monahan and Carterette 1985; Schmuckler 1999), and timbre (Grey 1977; Iverson and Krumhansl 1993; Toiviainen et al 1995; Toiviainen et al 1998).

Collecting data with retrospective judgments has the advantage that it allows a relatively simple experimental setting. This method, however, has some limitations. First, judgments made by listeners after an excerpt has ended may not reflect the experience while the music is played. Second, during long musical excerpts the quality of music, and thus the responses to it, may change over time. These dynamic qualities are not well represented by a single retrospective judgment. The latter limitation can be overcome by presenting excerpts of varying length, taken from a long musical sequence, and collecting responses after each excerpt. For example, Krumhansl and Kessler (1982) used this method in a study on how the sense of key develops and changes over time. Although this methodology enables tracking dynamic features of the responses, the collection of data may be time-consuming, especially if long musical stimuli are used. Therefore, experimental psychologists have shown an increased interest in measuring responses to music in real-time as the music is played. This method provides a relatively efficient way of generating a rich set of data.

One class of real-time responses asks the listener to adjust the position of a physical device or computer indicator. Some examples of real-time measures from the literature are: judgments of tension (Nielsen 1983; Madson and Fredrickson 1993; Fredrickson 1995; Krumhansl 1996, 1997; Krumhansl and Schenck 1997), judgments of memorability and openness (Krumhansl 1998), and judgments of the amount and quality of emotions (Krumhansl 1997, 1998; Krumhansl and Schenck 1997; Schubert 1999). Validation for this approach comes from two sources. First, intersubject agreement in these methods has been found to be reasonably strong, and subjects make similar responses on repeated trials. Second, and more important for the interpretation of the responses, the dynamic ratings can be understood as responses to identifiable structural features of the music.

The research reported in this article introduces a concurrent probe-tone method and shows how it can be used to create dynamic representations of the sense of key. The concurrent probe-tone method was used in conjunction with an organ *Duetto* by J S Bach (1685–1750). It is the last of four organ *Duettos* for solo performer, BWV 805 (Williams 1980). The set of *Duettos*, BWV 802–805, was published as part of the *Clavier-Übung* (keyboard practice), part III. It lasts ~3 min played at 75 half-note beats per minute. It is in 2/2 meter and in the key of A minor, with 108 total measures, mostly grouped in 4-measure phrases. It consists of a single melodic line in the right hand (RH), and a single melodic line in the left hand (LH). Tonally the music goes through multiple keys related to the key of A minor, with widely varying amounts of tonal clarity and chromaticism. Thus, BWV 805 was convenient for studying tonality induction—the process through which the sense of key arises and changes over time. Musical notation of BWV 805 is available at http://www.perceptionweb.com/misc/p3312.

1 Background

Krumhansl and Shepard (1979) introduced the probe-tone technique to investigate one aspect of how a tonal context influences the perception of pitch. Music theorists describe a hierarchy of tones, with some tones more stable, structurally significant, or final sounding than others. To quantify this hierarchy, Krumhansl and Shepard (1979) played an incomplete major scale (omitting the final tonic). This context was followed on successive trials by each of the chromatic scale pitches in an octave (the probe tones). Listeners rated how well each probe tone completed the scale. Krumhansl and Kessler (1982) extended this method to a variety of different contexts (chords and chord cadences, as well as scales) both in major and in minor keys.

The results of these studies confirmed music-theoretic predictions, with the tonic highest in the hierarchy, followed by the third and fifth scale tones, followed by the remaining scale tones, and finally the nondiatonic tones. The results for contexts of the same mode were similar when transposed to a common tonic. Also, the results were similar independently of which particular type of key-defining 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 $= \langle 6.35, 2.23, 3.48, 2.33, 4.38, 4.09, 2.52, 5.19, 2.39, 3.66, 2.29, 2.88 \rangle$. The vector for minor keys is: K-K minor profile $= \langle 6.33, 2.68, 3.52, 5.38, 2.60, 3.53, 2.54, 4.75, 3.98, 2.69, 3.34, 3.17 \rangle$. The K-K profiles are depicted in figure 1.

In the context of Western tonal-harmonic music, the hierarchy has been found in a variety of different kinds of measures. Diverse subject populations have been studied to explore effects of development, musical training, and neurological condition. The same basic method has been used with contexts from 20th-century Western music (Krumhansl and Schmuckler 1986; Krumhansl et al 1987), North Indian classical music (Castellano et al 1984), and Balinese music (Kessler et al 1984) demonstrating that the hierarchy depends on the musical style. (For a recent review see Krumhansl and Cuddy, in press.)

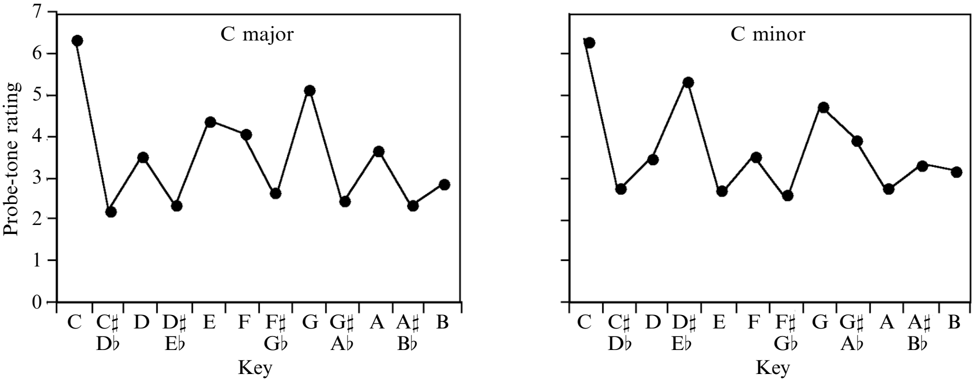


Figure 1. Probe-tone ratings for the keys of C major and C minor (Krumhansl and Kessler 1982).

The Krumhansl and Kessler (1982) probe-tone data also generated a musically interpretable geometric representation of musical keys. The basic assumption underlying this approach was that 2 keys are closely related to each other if they have similar tonal hierarchies, as measured by correlation. That is, keys were assumed to be closely related if tones that are relatively stable in one key are also relatively stable in the other key. Multidimensional scaling of the matrix of correlations achieved a good fit in four dimensions. It located the 24 keys on the surface of a torus. There was one circle of fifths for major keys (...F♯/G♭, D♭, A♭, E♭, B♭, F, C, G, D, A, E, B, F♯/G♭...) and one circle of fifths for minor keys (...f♯, c♯, g♯, d♯/e♭, b♭, f, c, g, d, a, e, b, f♯...). These wrap diagonally around the torus such that each major key is located both near its relative minor (for example, C major and A minor) and its parallel minor (for example, C major and C minor).

In an initial attempt to characterize how the sense of key changes over time, Krumhansl and Kessler (1982) used 10 nine-chord sequences, some of which contained modulations between keys. Listeners did the probe-tone task after the first chord, then after the first 2 chords, then after the first 3 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.

The continuous spatial medium of multidimensional scaling affords representing the changing sense of key in a graphical form. To do this, Krumhansl and Kessler (1982) used multidimensional unfolding, 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. Each set of probe-tone ratings was correlated with each of the 24 key probe-tone profiles. The unfolding algorithm finds a point to best represent these correlations. The advantage of this representation is that it makes the results visually accessible.

In this manner, each of the 10 nine-chord sequences in the Krumhansl and Kessler (1982) study generated a series of 9 points on the torus representation of keys. For non-modulating 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 the points 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.

Considerable modeling effort has been devoted to the problem of how to characterize the process of tonality induction (cf Vos and Leman 2000). Only those models most directly related to tonal hierarchies will be reviewed here. One of the first, if not the first, key-finding algorithms to be implemented on a computer was that of Longuet-Higgins and Steedman (1971). This algorithm uses a 2-D array of tones. It matches the tones of the input to box-shaped regions containing the scale tones of a key; there is one region for each major and minor key (the harmonic minor scale is used). The algorithm works by eliminating musical keys as the music progresses. The first tone eliminates keys in which the first tone is not contained; the second tone eliminates from the remaining keys those in which the second tone is not contained, and so on, until only 1 key remains. Two kinds of problems can arise. First, at some step all remaining keys may be eliminated at once. Then, an additional rule, the tonic – dominant preference rule, is invoked. The algorithm returns to the previous step, and preference is given to the key whose tonic is the first tone of the piece if it was one of the remaining candidates. If not, then preference is given to the dominant. The second kind of problem is that more than one possible key can remain at the end of the input segment. At this point, the tonic – dominant preference rule is also invoked.

The key-finding algorithm of Krumhansl and Schmuckler (Krumhansl 1990) was motivated by the idea that weighting tones according to the tonal hierarchy might result in a more accurate and efficient algorithm than simply assessing whether or not tones are scale members. The input to the algorithm is the distribution of tones in the input segment weighted according to their duration. That is, it is the total duration of each of the chromatic scale tones in the segment. The algorithm simply correlates this input with the 24 major and minor Krumhansl and Kessler (1982) key profiles. The correlations give a measure of the strength of each possible key. Note that more than one key may have a strong correlation, for example in the region around a pivot chord of a modulation. Also, it may be that no key has a strong correlation in a tonally weak segment of music.

The algorithm was tested with different kinds of input. One was a step-by-step input done for comparison with the Longuet-Higgins and Steedman (1971) algorithm, which had been tested with the fugue subjects of J S Bach's *Well-tempered Clavier*. Adding the tonal hierarchy information generally improved key-finding performance. Another test used the short initial segments of Bach's preludes for comparison with Cohen's (1991) listeners' judgments of key; the algorithm performed somewhat more accurately than human subjects. The third was a set of nonoverlapping 1-measure windows from the beginning to the end of a single Bach prelude. The results compared well with judgments by music-theory experts, and were projected onto the toroidal representation of musical keys described earlier.

Three variants of the Krumhansl and Schmuckler algorithm have been proposed recently. Temperley (1999) suggested a number of modifications. One modification, based on music-theoretic considerations, is to change the weights of the tones. The most notable change is to increase the weight of the 7th scale degree in major and the raised 7th degree in minor. Second, the input codes the music for whether or not a tone is present, ignoring duration. Third, a matching formula other than correlation is used. Fourth, a penalty is imposed for changing key from one input segment to the next.

The purpose of the penalty was to reduce the kind of instability sometimes observed, especially during modulations and when the input segment is relatively short. Finally, a retrospective re-evaluation of key is permitted. In addition to comparing the model with that of Longuet-Higgins, Temperley carried out the most extensive test to date with a music-theory textbook's key analysis of a large number of pieces. This was done on a measure-by-measure basis. The modified model performed well in these applications.

Shmulevich and Yli-Harja (2000) proposed an algorithm directed at smoothing local oscillations in key assignments by using neighboring key assignments. The input is a set of overlapping sliding windows of a fixed number of tones. There is one window for each tone of the music, so the model traces changes of key in detail. The key assignments of the windows were then smoothed by a graph-based method. The graph contains 24 nodes, corresponding to the major and minor keys. The edges are assigned lengths equal to the distances between keys in the multidimensional-scaling solution of Krumhansl and Kessler (1982). The final key assignment is the norm of the keys, that is, the key that is closest to all of the possible key assignments. The results compared well with judgments of music-theory experts in the test of the Krumhansl–Schmuckler algorithm (Krumhansl 1990).

Another approach to modeling key-finding is Leman's (1995) neural network model that uses acoustic information as input. It begins with an auditory component that processes the acoustic signal and forms pitch patterns. These patterns self-organize into a stable perceptual schema during perceptual learning. The resulting information then enters a self-organizing network that consists of cognitive schemas for pitch structures. These include a schema for tone-center perception (or key-finding) that results from long-term learning. The model outputs a sense of the tone center (major or minor key) based on a combination of the current information and preceding tone-center perceptions. The model allows for retrospective (retroactive) re-evaluation of tone centers. In addition, it yields a measure of key distances that is very similar to that derived from the Krumhansl and Kessler (1982) probe-tone profiles.

2 Objectives of the present study

As should be obvious, the retrospective probe-tone method 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. Thus, the judgments may not faithfully mirror the experience of music in time. For these reasons, we were motivated to develop an alternative form of the probe-tone methodology.

In this method, which we call the concurrent probe-tone judgment, the probe tone is presented continuously while the music is played. The complete passage is first sounded together with 1 probe tone. Then the passage is sounded again, this time with another probe tone. This process is continued until all 12 probe tones have been sounded. A time slice can then be taken at any point in the piece, giving the ratings of the 12 chromatic scale tones at that time. Because this was the initial application of the method, and the task appeared quite demanding for the relatively long and complex musical context which had to be repeated twelve times in its entirety, only musically trained subjects were tested.

Another objective in the present study was to consider an alternative approach to representing the sense of key reflected in probe-tone judgments. The above methods make a number of assumptions about measurement. 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). Support for this came from a subsidiary Fourier analysis of the K-K major and minor profiles (Krumhansl 1990), which found two relatively strong harmonics. Plotting the phases of the two Fourier components for the 24 key profiles was virtually identical to the multidimensional-scaling solution. Nonetheless, it would be 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 with 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 only 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.

Lastly, 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 meaning of points between keys is not well specified. This motivated an alternative model that explicitly identifies the meaning of positions between keys. To meet these objects, we used a self-organizing map (SOM, Kohonen 1997).

The third and final objective of the present study was to test the concurrent probe-tone judgments against two models of tonality induction. One model addresses the obvious limitation of the Krumhansl – Schmuckler (Krumhansl 1990) algorithm that it ignores the order of the tones in the input sample. The order in which tones are played may provide additional information that is useful for tonality induction. This is supported by studies both on tone transition probabilities (Fucks 1962; Youngblood 1958) and perceived stability of tone pairs in tonal contexts (Krumhansl 1979, 1990). 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%, respectively (Fucks 1962)]. 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 (Youngblood 1958). 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 (Krumhansl 1979, 1990). 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. The algorithm developed here creates a dynamic tone-transition matrix based on principles of auditory stream segregation (Bregman 1990). The tone-transition matrix is compared with the data on perceived similarity of pairs of tones in major and minor key contexts (Krumhansl 1990). This yields a measure of the strength of each major and minor key. A dynamic version of the Krumhansl – Schmuckler algorithm was also developed. Both models incorporate assumptions about the time-course of pitch extraction and memory decay. They were tested against the concurrent probe-tone data for the Bach organ *Duetto*. Based on the results presented above, it was expected that the tone-transition-based model would predict the probe-tone ratings better than the dynamic version of the Krumhansl – Schmuckler algorithm.

3 Experiment: The concurrent probe-tone method

3.1 *Method*

3.1.1 *Subjects*. Eight highly trained musicians participated in the experiment; they were paid $5 for their participation. Four were undergraduate students and four were graduate students at Cornell University; the graduate students were all enrolled in the doctoral program in music. All eight participants were currently active studying and performing music. These subjects (four males and four females) had taken instruction on an average of 3.25 different musical instruments (range 2 – 5), with an average of 20.6 total years of instruction summed across instruments (range 16 – 31.5 years). They had performed on most of these instruments, with an average of 21.9 total years of performance summed across instruments (range 12.5 – 38 years). They had studied music theory at a university level for an average of 2.6 years (range 0 – 5.5 years), and taken other university-level music courses for an average of 3.8 years (range 0 – 7 years). On a scale from 1 to 7, subjects reported a mean familiarity with the music of J S Bach of 6.0 (range 5 – 7). However, on another seven-point scale, subjects reported a mean familiarity with BWV 805 of 1.9 (range 1 – 4).

3.1.2 *Materials and stimuli.* Subjects controlled the experiment themselves using a MAX 3.5.9-9 object-oriented interface on a 350 MHz Macintosh G4 computer with a 17-inch Mitsubishi Diamond Plus 71 monitor. Additionally, the interface played all stimuli by sending the musical instrument digital interface (MIDI) information to Unity DS-1 1.2 software via Open Music System. Unity then converted the MIDI information into audio information, passing it out of the computer to a NAD 3125 stereo amplifier. Subjects listened to the stimuli over AKG K141 headphones.

The stimuli were based on a MIDI file of BWV 805, Bach's fourth organ *Duetto*, downloaded from the Classical MIDI Archives (http://www.prs.net/bach.html#800). This website lists the creator of the MIDI file as D. Jao. Before creating the stimuli, the original MIDI file was adapted by using Digital Performer 2.61-MTS in the following ways. First, we set the tempo to a half-note interbeat interval (IBI) of 800 ms, or 75 beats per minute. Second, we made the MIDI file metronomic such that all IBIs were equal throughout the piece. Third, we set all MIDI note-on velocities (which determine acoustic amplitude) to 110. Lastly, we played the MIDI file with the church organ timbre from the Unity DS-1 library.

For the concurrent probe-tone task, we created twelve versions by adding a probe tone to the MIDI file. The probe tones were all of the tones of the chromatic scale. Each probe tone was sounded in 6 octaves (in the range C1 – B6) in the same church organ timbre as the music. It was sounded at the beginning of each measure with a duration equal to the length of a measure minus 33 ms. This created a slight pulsing effect intended to prevent the probe tone from blending in with the music. Also toward this end the music was played to the right ear and the probe tone was played to the left ear through the headphones. This also eliminated peripheral sensory dissonance. The MAX interface displayed instructions to subjects for the task.

3.1.3 *Procedure.* Each subject ran in a single experimental session that lasted somewhat longer than 1 h. During the experiment, the experimenter was in a room adjacent to the subject. To start the music, the subjects pressed an on-screen button using the mouse and, after a 1000 ms delay, the music began playing. Subjects were instructed to judge how well the probe tone fit with the music in a musical sense. They made their judgments continuously during the music by dragging an on-screen slider (controlled by subjects with the computer mouse) to the right as the degree of fit increased and to the left as the degree of fit decreased. The position of the slider was recorded every 200 ms and was set at the midpoint of the scale at the beginning of each trial. The order in which the probe tones appeared across trials was determined randomly and was different for each subject.

After signing an experimental consent form, subjects practiced the task with a 45 s excerpt of another Bach organ *Duetto*, BWV 803. The excerpt was played first without a probe tone, so that they could become familiar with it. This was followed by three trials with 3 different probe tones. After the practice session, each subject heard the organ *Duetto* BWV 805 once without a probe tone. They then performed the task with each of the 12 probe tones. Prior to debriefing, subjects filled out a questionnaire concerning their music experience, their familiarity with the stimulus materials, and general demographic information.

3.2 *Results*

3.2.1 *Pre-processing of the data and intersubject correlations.* The raw data were the position of the slider recorded every 200 ms for each of the 12 probe tones. The data were smoothed by averaging over half-measure segments (800 ms). This was done out of a concern that the data used in subsequent analyses were sampled at time intervals long enough to allow for independent responses by the listeners. In other words, at the interval of 800 ms, motor limitations would not produce artifactual correlations between data at successive time points. This gives a total of 216 values for each of the 12 probe tones. The issue of serial correlation in time-series data has been discussed by Schubert (2001, 2001 – 2002), who recommends the use of a different transformation. We have not taken this approach because of the greater interpretability of the untransformed fitness judgments, and the fact that our objective is to develop a model for the data that explicitly takes tonal memory into account. In addition, it should be noted that the data for each of the 12 probe tones are independent of one another given that they were acquired on successive presentations of the piece of music. In passing, it might be mentioned that another promising approach to the analysis of time-series data is functional data analysis (Ramsay and Silverman 1997).

To assess the consistency across subjects, we computed intersubject correlations. These averaged $r_{2590} = 0.52$, $p < 0.0001$. Although significant, owing to the large number of degrees of freedom, the correlations suggest the possibility of differences between subjects. However, the matrix of intersubject correlations did not suggest the existence of identifiable subgroups within the group. One factor that may have contributed to the modest correlations is that subjects had to listen to the piece of music twelve times, each time with a different probe tone, and they may have had difficulties maintaining a constant criterion for using the response scale. When each subject's data were correlated with the group average, these correlations averaged $r_{2590} = 0.75$, $p < 0.0001$. This indicates that the group average is reasonably representative of the individual subjects. Consequently, we averaged across subjects, and the following analyses are based on the average data at the half-measure level.

3.2.2 *Representing the concurrent probe-tone data on a SOM*

A SOM 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 $\boldsymbol{x}$ and a reference vector $\boldsymbol{m}$ is defined by

$$\cos\theta = \frac{\sum_i x_i m_i}{\left(\sum_i x_i^2\right)^{1/2}\left(\sum_i m_i^2\right)^{1/2}} = \frac{\boldsymbol{x}\cdot\boldsymbol{m}}{||\boldsymbol{x}||\,||\boldsymbol{m}||}\,. \tag{1}$$

Another important factor 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 figure 2.) The map shows the units with reference vectors that correspond to the K-K profiles. The SOM configuration is virtually identical to the multidimensional-scaling solution (Krumhansl and Kessler 1982) and the Fourier-analysis-based projection (Krumhansl 1990) obtained with the same set of vectors. Unlike those maps, however, every location in the map is explicitly associated with a reference vector so that they are uniquely identified. SOMs with similar configurations have been obtained with training sets consisting of temporally integrated autocorrelation images taken from an auditory model (eg Leman and Carreras 1996).

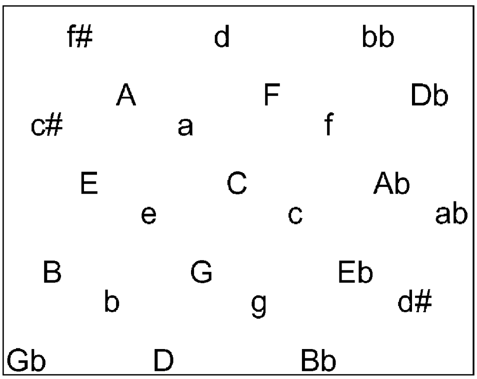


Figure 2. The configuration of a self-organizing map trained with the 24 Krumhansl – Kessler profiles. The map has a toroidal configuration, in other words, the left and the right edges are connected to each other as are the top and the bottom edges.

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, by using the direction cosine of equation 1. Dynamically changing data from either probe-tone experiments or tonality induction 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.

3.2.3 *Projection of concurrent probe-tone judgments onto the SOM.* The probe-tone judgments were projected onto the SOM by calculating the activation level of each unit, as described earlier. The activation was converted into a gray-scale such that lighter areas correspond with units that have higher activation. Four representative cases are shown in figure 3; the dynamic representation (in color and with the music sounded simultaneously) can be found on the Internet at http://www.cc.jyu.fi/~ptoiviai/bwv805/index.html. Figure 3a shows the activation pattern at the beginning of measure 11 with a clear tonal center at A minor. Figure 3b shows that by the beginning of measure 18 the

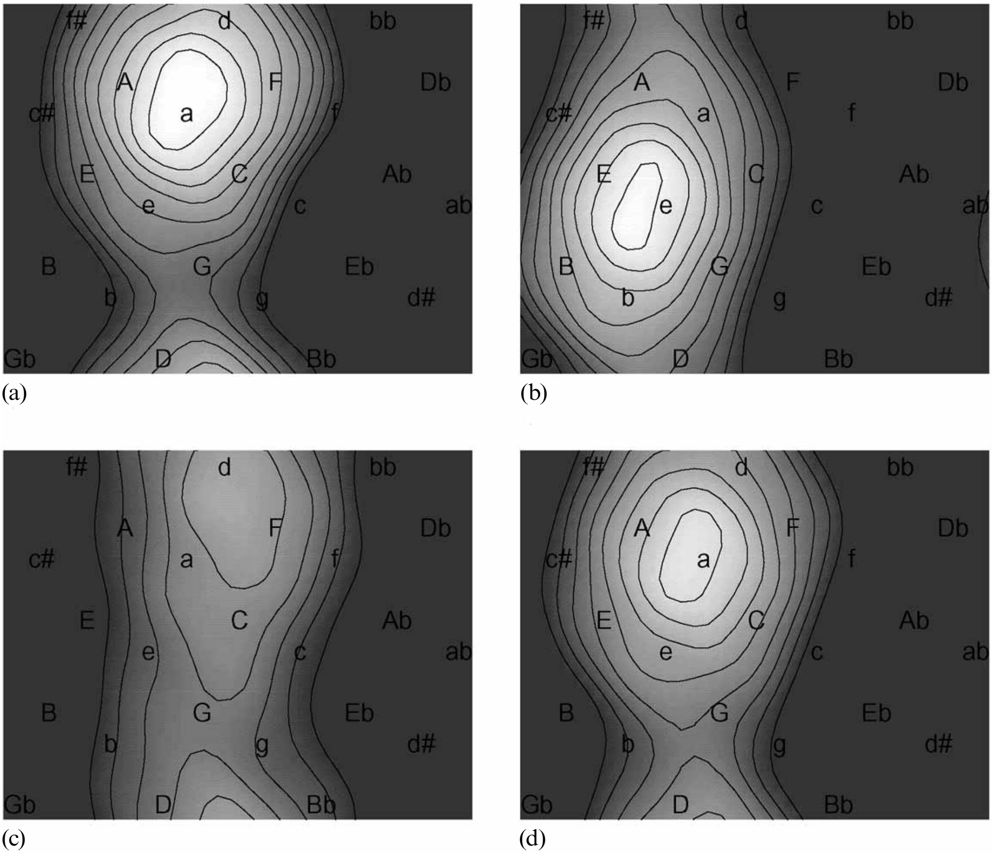


Figure 3. Projections of concurrent probe-tone judgments onto a self-organizing map trained with the 24 Krumhansl – Kessler profiles. Light areas correspond to units with high activation and vice versa. The projections are shown at the beginning of (a) measure 11; (b) measure 18; (c) measure 25; and (d) measure 34. The contour lines start at zero and are spaced at intervals of 0.1.

tonal center has shifted to E minor. Figure 3c shows the results at the beginning of measure 25 after the long chromatic passage. No units are strongly activated; instead, there is an elongated region of weak activation that includes C and F major and A and D minor. Figure 3d shows the return to the key of A minor at the beginning of measure 34. These examples illustrate typical patterns that appear in the SOM activation, but the changes that occur in the sense of key over time can best be appreciated by viewing the dynamic representation.

4 Dynamic models of tonality induction

In order to develop a model of tonality induction by using tone order, it is necessary to find a way of summarizing the tone transitions. However, this is not a trivial task for polyphonic music, 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, sub-sequence C4 – D4 – E4 becomes increasingly prominent. This is because these tones segregate into one stream owing 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 9 different tone transitions. Some of these transitions are, however, perceived as stronger than the others. For instance, the transition G–A is, because of 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 (Bregman 1990). 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. 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 (Gjerdingen 1994). A dynamic system that embraces these principles is described next.

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 2 tones has the highest strength when the tones are proximal both in pitch and in time, and have long durations. These three factors are included in the following dynamic model.

4.1 *Representation of input*

The pitches of the chromatic scale are numbered consecutively. The onset times of tones having pitch k are denoted by t^n_{ki}, $i = 1, ..., n_k$, and the offset times by $t^f_{ki} = 1, ..., n_k$, where n_k is the total number of times the kth pitch occurs.

4.2 *Pitch vector* $\boldsymbol{p}(t) = [p_k(t)]_k$

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

$$\dot{p}_k = -\frac{p_k}{\tau_p} + \sum_{i=1}^{n_k} \delta(t - t^n_{ki}) - p_k \sum_{i=1}^{n_k} \delta(t - t^f_{ki}), \tag{2}$$

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 $\tau_p = 0.5$ s. With this value, the integral of p_k saturates at about 1 s after tone onset, thus approximating the durational accent as a function of tone duration (Parncutt 1994).

4.3 *Pitch memory vector* $\boldsymbol{m}(t) = [m_k(t)]_k$

The pitch memory vector provides a measure of both the perceived durational accent and the recency of tones 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 $\boldsymbol{m}$ are governed by the equation

$$\dot{\boldsymbol{m}} = \frac{\boldsymbol{p} - \boldsymbol{m}}{\tau_m}. \tag{3}$$

The time constant τ_m determines the dependence of transition strength (defined next) on the temporal distance between the tones. In the simulations, the value $\tau_m = 3$ s has been used, corresponding to typical estimates of the length of the auditory sensory memory (Treisman 1964; Darwin et al 1972; Fraisse 1982).

4.4 *Transition strength matrix* $\boldsymbol{S}(t) = [s_{kl}(t)]_{kl}$

The transition-strength matrix provides a measure of the instantaneous strength of transitions between all pitch pairs. More specifically, a higher 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 $\boldsymbol{S}$ is governed by the equation

$$s_{kl} = \tfrac{1}{2}[1 + \mathrm{sgn}(p_l - p_k)]\, m_k p_l \exp[-(k-l)^2/\alpha^2] \,. \tag{4}$$

In this equation, the nonlinear term $\frac{1}{2}[1 + \mathrm{sgn}(p_l - p_k)]$ 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 $\exp[-(k-l)^2/\alpha^2]$ 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.

4.5 *Dynamic tone transition matrix* $\boldsymbol{N}(t) = [n_{kl}(t)]_{kl}$

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 $\boldsymbol{N}$ is governed by the equation

$$\dot{\boldsymbol{N}} = \boldsymbol{S} - \frac{\boldsymbol{N}}{\tau_N} \,, \tag{5}$$

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

The octave-equivalent transition matrix $\boldsymbol{N}' = (n'_{ij})_{ij}$ is defined according to

$$n'_{ij}(t) = \sum_{\substack{i=p \bmod 12 \\ j=q \bmod 12}} n_{pq}(t) \,. \tag{6}$$

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.

Figure 4 displays the temporal evolution of the pitch class vector $\boldsymbol{p}_c$, the pitch class memory vector $\boldsymbol{m}_c$, and the octave-equivalent transition matrix $\boldsymbol{N}'$ during the first 4 measures of the musical stimulus used in the study.

5 Testing the tonality-induction models

For the present application, we developed two tonality-induction models. Model 1 is based on pitch class distributions. Model 2 is based on the tone-transition distributions just described. Below, a brief description of the models is given and their results compared with the concurrent probe-tone judgments.

5.1 *Model 1*

Model 1 is based on pitch class distributions only. Like the earlier Krumhansl–Schmuckler algorithm based on the K-K profiles (Krumhansl 1990), 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 $\boldsymbol{p}_c(t)$, which is similar to the pitch vector $\boldsymbol{p}(t)$ 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 $\boldsymbol{m}_c(t)$ is obtained by temporal integration of the pitch class vector according to the equation

$$\dot{\boldsymbol{m}}_c = \boldsymbol{p}_c - \frac{\boldsymbol{m}_c}{\tau_d} \,. \tag{7}$$

Again, the time constant has the value $\tau_d = 3$ s. To obtain estimates for the key, vector $\boldsymbol{m}_c(t)$ is correlated with the K-K profiles for each key. Alternatively, or in addition,

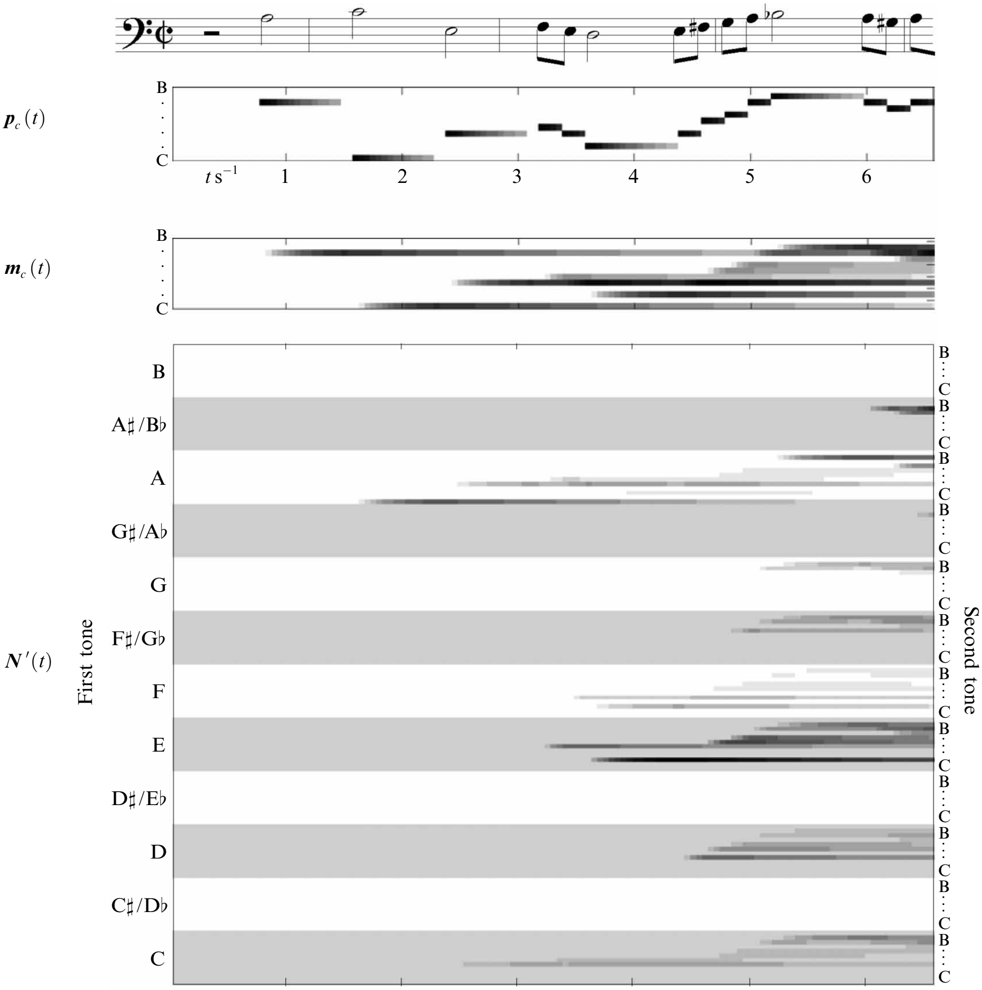


Figure 4. Temporal evolution of the pitch class vector $\boldsymbol{p}_c$, the pitch class memory vector $\boldsymbol{m}_c$, and the octave-equivalent transition matrix $\boldsymbol{N}'$ during the first four measures of the organ *Duetto* BWV 805. For the vectors, the pitch classes are ordered chromatically so that pitch class C is in the bottom. The components of matrix $\boldsymbol{N}'$ are ordered so that transitions having the same first tone are grouped together. This grouping is indicated by the alternating white and gray stripes. For example, the lowermost gray stripe contains the 12 transitions from pitch class C, with the transition C – C being the lowermost.

the vectors $\boldsymbol{m}_c(t)$ can be projected onto the toroidal key representation by using activation values as described earlier. Figure 5 shows the activation strengths for measures 11, 18, 25, and 34. Comparison with figure 3 shows that the model produced results similar to those of the listeners. The dynamic representation of the model can be found on the Internet (http://www.cc.jyu.fi/~ptoiviai/bwv805/index.html) for comparison with the concurrent probe-tone data as projected onto the SOM.

5.2 *Model 2*

Model 2 is based on tone transitions, and uses the dynamic octave-equivalent transition matrix $\boldsymbol{N}'$ defined in equation (6). To obtain estimates for the key, the pitch class transition matrix is correlated with the matrices representing the perceived relatedness of pairs of tones in major and minor key contexts (Krumhansl 1990, table 5.1, page 125). These matrices are graphically depicted in figure 6.

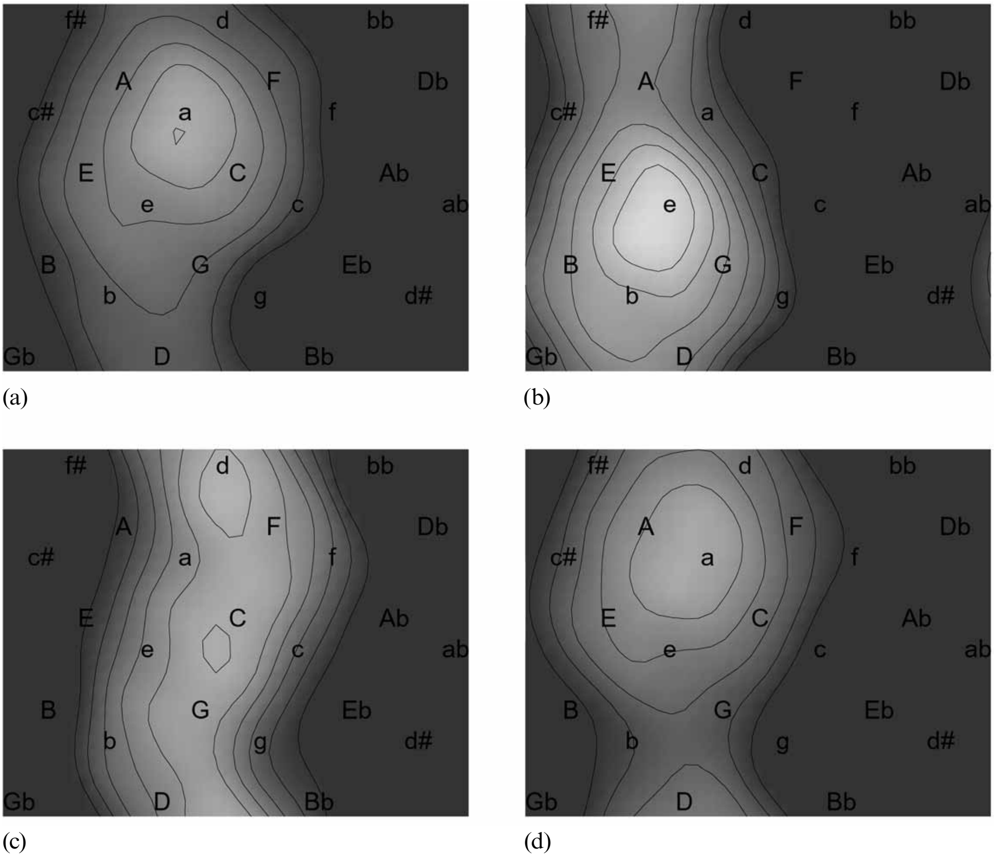


Figure 5. Projection of the pitch-class memory vector $\boldsymbol{m}_c(t)$ onto a self-organizing map trained with the 24 Krumhansl – Kessler profiles. Light areas correspond to units with high activation and vice versa. The projections are shown at the beginning of (a) measure 11; (b) measure 18; (c) measure 25; and (d) measure 34. The contour lines start at zero and are spaced at intervals of 0.1.

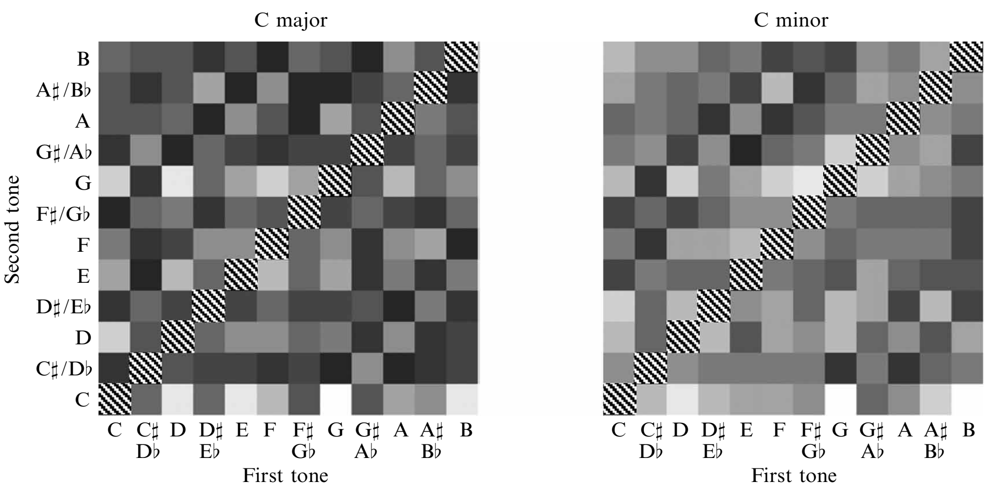


Figure 6. Graphical representation of relatedness ratings for all possible ordered pairs of tones presented after C major and C minor key-defining contexts (after Krumhansl 1990, page 125). The degree of darkness of each square displays the rating value, bright squares corresponding to high values and vice versa. For both the major and the minor contexts, the diagonal between bottom left and top right corners is missing because the experiment did not use repeated test tones.

5.3 *Comparing models with the concurrent probe-tone data*

Models 1 and 2 each yield a value measuring the strength of each key at each half-measure of the music. These were compared with those for the listeners' data (the correlation between the probe-tone judgments and each of the K-K vectors). Figures 7a and 7b display the listeners' data against each model's predictions for all major (figure 7a) and minor (figure 7b) keys. For these figures, the model's outputs have been scaled to have a standard deviation equal to that of the human data. This was necessary because the standard deviation of the output values of model 2 was, owing to the high number of degrees of freedom used to calculate the correlation values in the model, significantly lower than that of the human data.

The correlation between the listeners' data and model 1 was $r_{5182} = 0.89$, $p < 0.0001$. The correlation between the listeners' data and model 2 was also $r_{5182} = 0.89$, $p < 0.0001$. Thus, the two models were equally good at predicting the listeners' key judgments. To see whether the two models complement each other by providing slightly different results, a multiple regression was performed. The result, $R_{2,5183} = 0.90$, $p < 0.0001$, showed that the two models together only slightly improved the fit of either model alone. However, both models contributed significantly (at $p < 0.0001$) and a stepwise regression including both models indicated that they fit slightly different aspects of the listeners' judgments.

The performance of the models was further compared by calculating the Euclidean distance between the listeners' data and the predictions of each model, a common method for assessing the similarity between time series of equal length (Goldin and Kanellakis 1995). Before the calculation, the data were normalized to zero mean and unit variance. Furthermore, the Euclidean distance was divided by the number of components in each time series (5182). This was equivalent to calculating the root-mean-square error between the time series. The Euclidean distances between the listeners' data and the outputs of models 1 and 2 were 0.216 and 0.217, respectively. Thus, according to this measure, the models again performed about equally well.

6 A subsidiary experiment with other real-time measures of musical responses

In a subsidiary experiment, listeners made a number of other real-time judgments while listening to the Bach organ *Duetto*: tension, pulse clarity, predictability, and the relative salience of the right hand. Some of these continuous judgments were motivated by a study on pulse finding (Toiviainen and Snyder, in press). As part of the experiment, the subjects also tapped to the perceived pulse of the music. The focus of the present discussion is on tension, and its relationship to tonality and consonance. Therefore, the other measures recorded in this experiment will not be discussed here.

Music is often described as containing patterns of tension and relaxation (or release from tension). As noted in the introduction, a number of studies have obtained real-time measures of musical tension (Nielsen 1983; Madson and Fredrickson 1993; Fredrickson 1995; Krumhansl 1996, 1997; Krumhansl and Schenck 1997). Of psychological interest is the possible connection between musical tension and emotional or affective responses to music. Empirical studies have shown that judgments of musical tension correlate with judgments of the amount of emotion (Krumhansl and Schenk 1997), and with such basic emotions as happy, sad, and fearful (Krumhansl 1997). Musical tension has been found to correlate with changes in psychophysiological measures such as blood pressure and heart rate (Krumhansl 2000). In addition, tension ratings have been shown to correlate with key distances and consonance as theorized in Lerdahl's (1988, 1996, 2001) tonal pitch space model (Lerdahl et al 2000; Lerdahl and Krumhansl 2001a, 2001b). Thus, it seems important to try to understand the musical antecedents of tension.

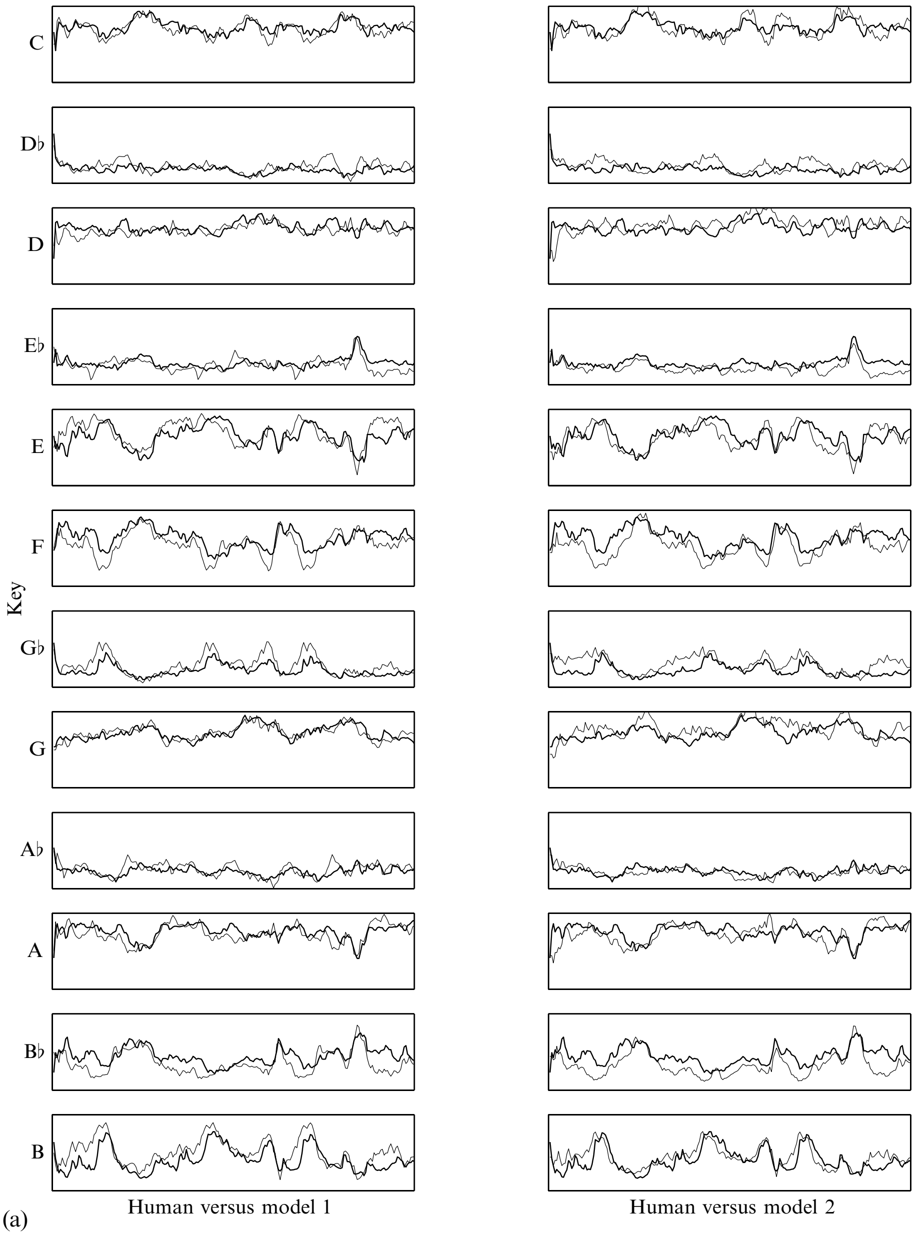
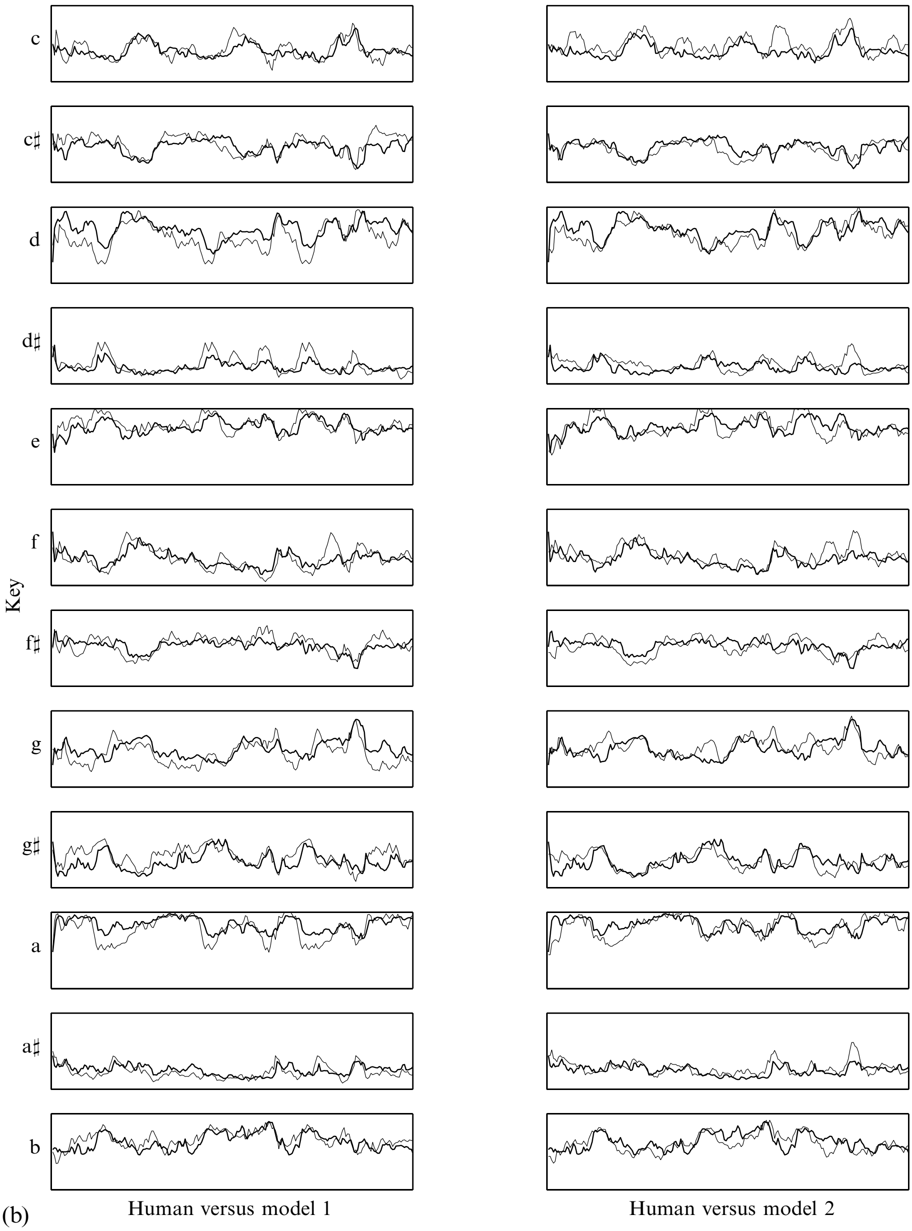


Figure 7. Comparison of listeners' data and the models' predictions about the strength of (a) each major key and (b) each minor key (facing page). The thick lines display the correlation between the probe-tone judgments and the K-K vector corresponding to the respective major key as a function of time. The thin lines show the models' predictions.

6.1 *Method*

6.1.1 *Subjects.* Twenty Cornell University undergraduate students participated for extra credit in psychology courses. Ten (six males and four females) had 8 or more years of musical experience with a mean of 11.3 years playing an instrument or singing. We refer to these subjects as *musicians*. The other ten subjects (five males and five females) had less than 8 years of musical experience with a mean of 3.2 years playing an instrument or singing. We refer to these subjects as *nonmusicians*. On a scale from 1 to 7,



(b)

Figure 7 (continued).

subjects reported an average familiarity with the music of J S Bach of 3.7. The musicians' mean familiarity was 4.8, while the nonmusicians' mean familiarity was 2.6. This difference was statistically significant ($t_{18} = 3.48$, $p < 0.005$). On another seven-point scale, subjects reported a mean familiarity with this particular piece, BWV 805, of 1.5, with means of 1.8 and 1.2 for musicians and nonmusicians, respectively. This difference was not statistically significant ($t_{18} = 1.70$, ns).

6.1.2 *Materials and stimuli.* In this experiment we used three versions of the Bach organ *Duetto*. The first was as described above, called the *whole* version. The second contained only the left-hand part of the whole version, called the LH version. The third contained only the right-hand part of the whole version, called the RH version.

Because the LH part begins at measure 1.5 (middle of measure 1) and the RH part begins at measure 9.5, the RH version is $\sim$13 s shorter than the whole and LH versions.

6.1.3 *Procedure.* Each subject participated in a single experimental session that lasted approximately 1 h. The instructions directed subjects to continuously judge the amount of tension in the music. Subjects made their judgments by dragging an on-screen slider (controlled by subjects with the computer mouse) to the right as tension increased and to the left as tension decreased. The position of the slider (on a scale from 0 to 100) was recorded every 250 ms. Subjects performed the task with all three versions (whole, LH, and RH; their order was counterbalanced across subjects).

At the start of the experiment, subjects practiced with a 45 s excerpt of BWV 803. At the end of the experiment, subjects filled out a questionnaire concerning their music and dance experience, their familiarity with the stimulus materials, and general demographic information.

6.2 *Results*

6.2.1 *Intergroup correlations.* We first computed the average of each subject group's judgments. The correlations between the averages for musicians and nonmusicians for the tasks were: LH tension, $r_{687} = 0.79$; RH tension, $r_{636} = 0.86$; and whole tension, $r_{687} = 0.79$. All these correlations are reasonably strong (and statistically significant at $p < 0.0001$) indicating that, as groups, the responses of musicians and nonmusicians were quite similar. Therefore, the following analyses will focus on the overall average data. However, all analyses were computed for musicians and nonmusicians separately, and observed differences will be noted.

6.2.2 *Treatment of the data.* The data were recorded at 250 ms intervals. To match the data points with the metrical structure of the music, the raw data were resampled at 200 ms intervals with linear interpolation. The resampled values were then averaged over half-measure segments as in the first experiment. This yielded a total of 216 data points for the tasks using the whole and LH versions, and 200 data points for the tasks using the RH version. These data will be used in the following analyses.

6.2.3 *Tension judgments.* Figure 8 shows the tension judgments for LH, RH, and whole versions. As can be seen, the general patterns were similar. The correlation between LH and whole tension judgments was $r_{214} = 0.67$, $p < 0.0001$; the correlation between RH and whole tension judgments was $r_{198} = 0.88$, $p < 0.0001$; and the correlation between LH and RH tension judgments was $r_{198} = 0.37$, $p < 0.0001$. Thus, the judgments for the LH and RH versions were least similar, as would be expected given the different music used in the two cases. To determine whether the whole tension judgments were a function of the judgments for the two separate voices (and to determine their relative strengths), a multiple correlation was computed. The resulting correlation, $R_{2,197} = 0.91$, $p < 0.0001$, supported the idea that the whole tension judgments were an additive combination of those for the two separate voices. The analysis showed that the RH contributed more to the whole tension judgments (standardized coefficient $= 0.79$, $p < 0.0001$) than the LH (standardized coefficient $= 0.23$, $p < 0.0001$). The same pattern was found when the analysis was applied to the data of the musicians and the nonmusicians separately, except that the RH predominated to an even greater extent for the nonmusicians (standardized coefficient for the RH $= 0.82$, $p < 0.0001$, standardized coefficient for the LH $= 0.08$, $p = 0.056$).

The question addressed by the next analysis was of central interest. It asks whether the tension ratings correspond in some way with the shifts that occur in key during this piece of music. The first step in this analysis was taking the probe-tone ratings at each half-measure segment and correlating them with the 24 K-K profiles. This generates a measure of the strength of each key throughout the piece. The correlations

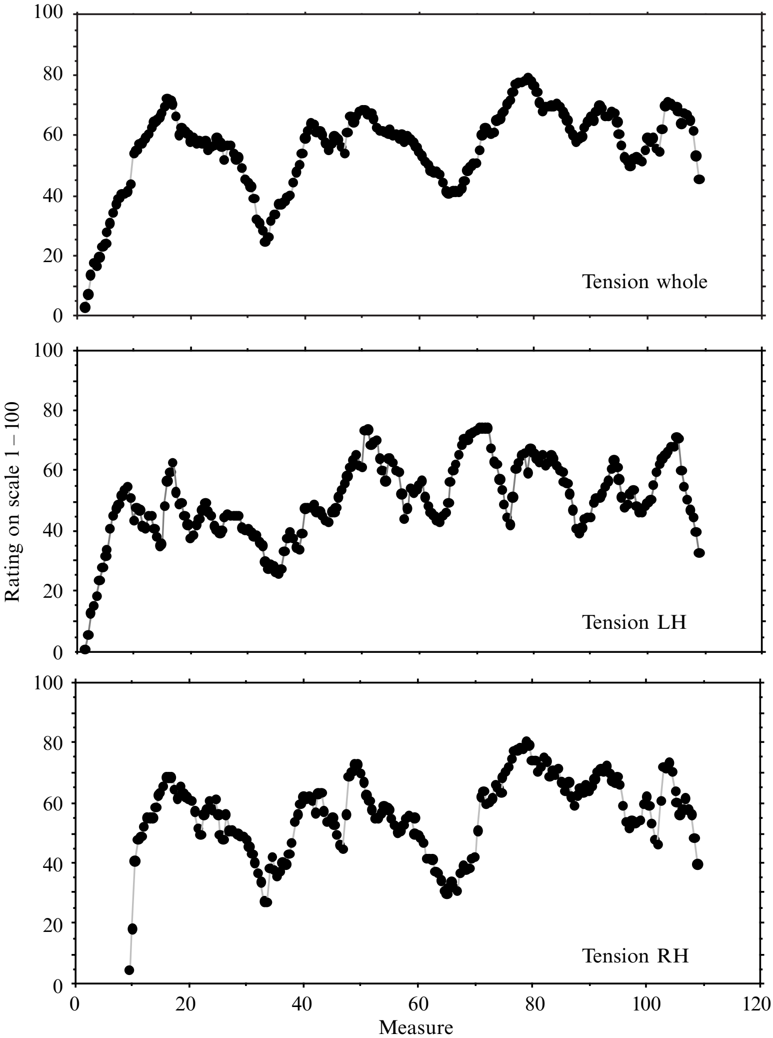


Figure 8. Continuous tension judgments from experiment 1 for whole version (top), LH version (middle), and RH version (bottom).

for each key were then correlated with the tension judgments. These correlations are shown in table 1. As can be seen, positive correlations were found for some keys, particularly E minor, E major, B major, G♯ minor, and B minor. This means that tension ratings increased when these keys were relatively strong. In contrast, negative correlations were found for other keys, particularly F minor, D♭ major, and F major. Tension ratings decreased when these keys were relatively strong. It should be noted that A minor, which is the key of the composition according to the musical score, was associated with a relatively large negative correlation: −0.26.

These correlations can be displayed on the SOM as shown in figure 9. The light region corresponds to keys associated with increased tension; the dark region corresponds to keys associated with decreased tension. According to this map, points where the subjects' tension ratings were high were usually associated with points where the perceived key was closest to the keys based on the V or the II degree of A minor. This is in line with the qualities of the chords built on the V and II degrees: chord successions such as V–I and II–V–I are usually regarded to form a tension–relaxation pattern.

Table 1. Correlations between tension judgments and key strengths.

Key	Correlation	Key	Correlation
C	−0.20	c	−0.20
D♭	−0.34	c♯	0.13
D	−0.10	d	−0.26
E♭	0.03	d♯	0.15
E	0.30	e	0.42
F	−0.33	f	−0.32
G♭	0.14	f♯	−0.24
G	0.14	g	−0.01
A♭	−0.26	g♯	0.37
A	−0.02	a	−0.26
B♭	−0.13	b♭	−0.17
B	0.30	b	0.30

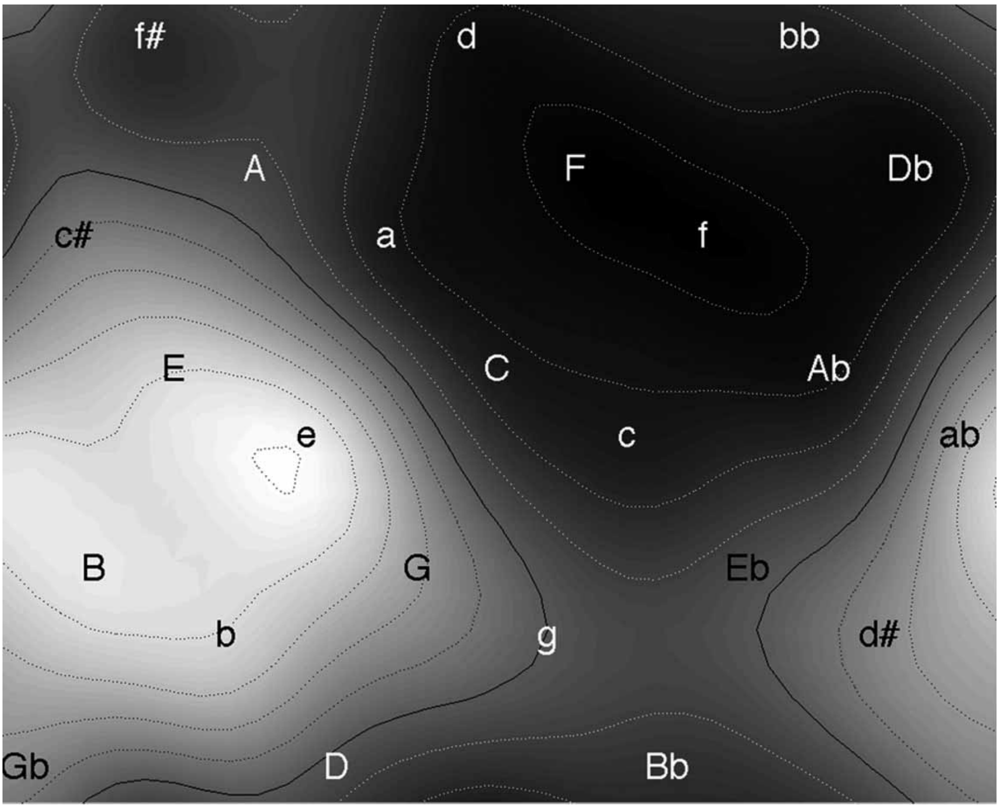


Figure 9. Correlations between the tension ratings and the strength of each key, displayed on the self-organizing map. The contour lines are spaced at intervals of 0.1, and the solid line shows the contour of zero correlation.

6.2.4 *Consonance.* The next analyses considered effects of consonance. The first compared the tension judgments with three measures of consonance: the consonance of the harmonic intervals formed by the RH and the LH, the consonance of the melodic intervals in the RH, and the consonance of the melodic intervals in the LH. The consonance values used were taken from Hutchinson and Knopoff (1978). The values were for tones with 10 harmonics and were smoothed over a 2-measure window (3.2 s) to avoid highly unstable values. They were then compared with the real-time judgments at the end of the window. On the whole, the correlations were weak, but generally significant.

For all listeners together, tension correlated (negatively) more strongly with the consonance of harmonic intervals ($r_{195} = -0.26$, $p = 0.0003$) than melodic intervals ($r_{195} = -0.14$, $p = 0.05$; and $r_{195} = -0.19$, $p = 0.007$, for RH and LH, respectively). However, when analyzed separately, musicians and nonmusicians showed a different pattern. For musicians, only the consonance of harmonic intervals was significant ($r_{185} = -0.34$, $p < 0.0001$), whereas for nonmusicians the consonance of the melodic intervals had a stronger influence ($r_{195} = -0.25$, $p = 0.0005$; and $r_{195} = -0.23$, $p = 0.001$ for RH and LH, respectively) than the consonance of the harmonic intervals ($r_{195} = -0.16$, $p = 0.025$). In general, although these correlations were significant, the effect size of consonance was not large.

Finally, the consonance of each probe tone with the LH and RH parts was also computed. They were then compared with the probe-tone judgments by using multiple correlation. The multiple correlation, $R_{2,2361} = 0.56$, $p < 0.0001$, showed that consonance correlated with the probe-tone judgments, such that the degree of judged fit increased with the consonance between the probe and the music. The consonance of the probe tone both with LH and with RH parts contributed significantly ($p < 0.0001$), with closely equal weights for the LH and RH parts.

7 Discussion

Our objective was to explore how the experience of music in real-time might be measured and modeled. The primary data came from a method intended to assess how the sense of key develops and changes over time. The concurrent probe-tone method was devised, in which the probe tones were sounded throughout the duration of the music. Musically trained listeners judged, at each point in time, how well the probe tone fit with the ongoing music. The organ *Duetto* BWV 805 by J S Bach was selected because it contains modulations, or shifts, in tonality, and also because there are points when no tonal center is strongly indicated.

To represent these dynamic changes in perceived tonality, we used the continuous spatial medium of a self-organizing map (SOM, Kohonen 1997). The map was trained with the 24 K-K vectors for major and minor keys (Krumhansl and Kessler 1982). Previously a map of musical keys in the shape of a torus had been derived from these data by multidimensional scaling. This multidimensional-scaling map of musical keys was interpretable in terms of the circle of fifths and relative and parallel major–minor relationships. However, it made a number of measurement assumptions and, more importantly for representing ambiguous or mixed perceptions of tonality, points intermediate between those for the 24 keys were not well specified. Moreover, the method did not afford representing the strength of key.

The SOM addressed each of these problems. Although its measurement assumptions were completely different from that of multidimensional scaling, the SOM configuration was virtually identical to the multidimensional-scaling solution. That is, the units that best matched the K-K vectors were arranged as in the earlier map. In a previous paper based on a self-organizing approach, Tillmann et al (2000) derived a map of musical keys., However, their analysis considered only major keys and recovered only the circle of fifths. It did not include minor keys and thus did not recover the additional structure that specifies the relative and parallel relationships between major and minor keys.

Unlike the earlier multidimensional-scaling map of musical keys, the units between keys in the SOM were also identified, each being associated with its own vector of weights. This meant that the match between any vector (such as those coming from the concurrent probe-tone experiment) and any unit could be determined. If, for example, the relative fitness judgments of the probe tones at some point of time strongly resembled the K-K vector of A minor, the activation of the A minor unit in the SOM would be strong, with activation grading off with distance from A minor. However, if multiple keys were weakly activated, then this would produce a diffuse pattern of activation over many units in the map.

This method made it possible to visually represent the way the sense of key develops and changes over time. Focused centers of strong activation showed that listeners readily found the A minor key of the key signature (within the first two measures), and traced modulations to near and more distant keys as they occurred in the music. In addition, there were passages during which no units were strongly activated; rather, many units were weakly activated. Representative samples were shown as gray-scale images imposed on the SOM, but the temporal nature of this process is best seen in the dynamic representation of the music (see, for example, http://www.cc.jyu.fi/~ptoiviai/bwv805/index.html).

It must be noted that the SOM, as used in the present study, was not intended to simulate any kind of perceptual learning that would occur in listeners through, for instance, the extraction of regularities present in Western music. Rather, it was used as a tool for visualizing the multidimensional data obtained from the continuous probe-tone experiment on one hand and from the key-finding models on the other. Previous research has shown, however, that exposing the SOM to excerpts of Western music will result in a map structure that is almost identical to the one used in the present study (see, eg, Leman and Carreras 1996, 1997). Therefore, we believe that the representation of the present SOM bears a resemblance to the mental representations of tonality.

The second objective of this project was to develop a dynamic model of tonality induction. The music was analyzed in two steps. The first, called the pitch vector, describes how the strength of each pitch changes with time. It assumes that the strength of a tone is zero when it is not sounded, jumps to a maximum of one at onset, decays exponentially over its duration, then is reset to zero at tone offset. The second, called the pitch memory vector, integrates the pitch vector over time, and again assumes an exponential decay. The decay parameters in both steps were set to match psychological results on durational accent and the duration of auditory sensory memory. Although the model began with a symbolic (MIDI) code of tone onset and offset, the analysis could instead be based on a pitch-extraction algorithm (such as used in Krumhansl et al 1999, 2000).

After the music was analyzed in this way, the results were used as input to each of two different models. The first model collapsed the pitch vectors over octaves (thus, corresponding to pitch classes). The resulting pitch class vector was then integrated to give the pitch memory vector. These pitch memory vectors were then correlated with the K-K vectors to get a measure of the strength of each of the 24 major and minor keys at each point in time (half-measure units of time were used). In addition, the pitch memory vectors were projected onto the SOM, as had been done with the concurrent probe-tone ratings.

The second model was developed to explore the possibility that tone transitions provide additional information about key. It created a transition-strength matrix giving the instantaneous strengths of transitions between all pairs of pitches. The transition strengths were scaled so that transitions between tones were stronger if they had longer durations and were more proximal in pitch and in time, according to principles of auditory scene analysis (Bregman 1990). These were then collapsed over octaves and integrated over time (on the assumption of an exponential decay) to give the dynamic tone transition matrix. To measure the strength of each of the 24 major and minor keys at any point in time, the matrix was then correlated with judgments of the similarity of tone pairs in tonal contexts (Krumhansl 1990).

Both models produced results very similar to those of the concurrent probe-tone data, and the two models matched the data equally well. They appeared to account for slightly different aspects of the listeners' responses, however. That the second model, using tone transitions, did not outperform the first, which used only tone distributions, was unexpected. Previous results (eg Krumhansl 1990; Krumhansl et al 1999, 2000) have shown that listeners have implicit knowledge of the frequency of tone transitions in familiar musical styles. It may be that using similarity judgments of tone pairs to determine key strength was not optimal. Using statistical summaries of tone transitions in Western tonal-harmonic music might produce a better result. However, the equal performance of the two models can be partially understood because of the strong dependence of perceived tone similarity on perceived tonal hierarchies as measured in probe-tone judgments (Krumhansl 1990, page 131 ff).

The final part of the project was to obtain real-time measures of tension. Musicians and nonmusicians both performed these tasks; the tension ratings of these subject groups were found to be similar to each other. This finding is in accordance with earlier studies (Krumhansl 1996). Of special theoretical interest in light of Lerdahl's tonal pitch space model (1988, 1996, 2001) is how tension relates to key distance, as this is one factor predicted to produce tension. In the present study, tension increased with distance from the initial (and final) key of A minor—but only in one direction. Specifically, high tension ratings were associated with perceived keys based on E and B (with more sharps), that is with keys based on the dominant and the supertonic of the principal key of the composition. Low tension ratings, on the other hand, were associated with perceived keys based on F, that is on the submediant of the principal key. Thus, key distance had an effect on tension, but its effect was directional, a factor not considered in the tonal pitch space model. Other studies have also found directional asymmetries in perceived key distance (Thompson and Cuddy 1989, 1992; Cuddy and Thompson 1992), suggesting that asymmetries such as these warrant further empirical measurement and modeling.

Another factor of interest, both in connection with Lerdahl's model and more generally, is consonance. The construction of the Bach *Duetto* as two separate LH and RH single-voice melodies allowed us to consider two different kinds of consonance: the consonance of the melodic intervals in each of the separate lines (melodic consonance), and the consonance of the intervals formed by simultaneously sounded tones (harmonic consonance). Both kinds of consonance had a relationship with tension, such that tension decreased with consonance. The effects, though, were relatively weak (similar to results of Krumhansl 1996; Lerdahl and Krumhansl, in preparation). One finding of potential interest was that musicians were more influenced by harmonic consonance, whereas nonmusicians were more influenced by melodic consonance. The consonance of the probe tone with the tones in the music was also considered. This had a relatively strong association with the concurrent probe-tone judgments, which is not surprising given previous analyses showing that probe-tone judgments correlate with measures of tonal consonance (Krumhansl 1990, page 55 ff).

To sum, on an empirical level this paper continued to explore the potential of real-time tasks to elucidate the experience of music in time. The concurrent probe-tone task produced results showing that the process of tonality induction is dynamic, with constant changes in perceived key strengths. These results were found to be associated with musical tension. On a modeling level, the article demonstrates that a SOM can be trained by psychological data and subsequently used as a continuous spatial medium in which to represent perceived tonality in a dynamic graphic form. In addition, two dynamic models of tonality induction were developed. One of these considered the information that tone transitions might provide to the tonality induction process. It defined a dynamic tone transition matrix based on principles of auditory scene analysis.

Its results matched the listeners' data well, but no better than a simpler model taking into account only tone distributions. At a more general level, we hope that the present convergence between psychological data and computational modeling encourages further explorations at the interface of these two approaches to the study of music.

Acknowledgments. The authors would like to thank John Sloboda and an anonymous reviewer for helpful comments. This work was supported by the Academy of Finland and the Fulbright Foundation.

References

Bharucha J J, 1999 "Neural nets, temporal composites, and tonality", in *The Psychology of Music* 2nd edition, Ed. D Deutsch (San Diego, CA: Academic Press)

Bregman A S, 1990 *Auditory Scene Analysis* (Cambridge, MA: MIT Press)

Castellano M A, Bharucha J J, Krumhansl C L, 1984 "Tonal hierarchies in the music of North India" *Journal of Experimental Psychology: General* **113** 394 – 412

Cohen A J, 1991 "Tonality and perception: Musical scales primed by experts from *The Well-Tempered Clavier* of J. S. Bach" *Psychological Research* **53** 305 – 314

Cuddy L L, Cohen A J, Mewhort D J K, 1981 "Perception of structure in short melodic sequences" *Journal of Experimental Psychology: Human Perception and Performance* **7** 869 – 883

Cuddy L L, Lunney C A, 1995 "Expectancies generated by melodic intervals: Perceptual judgments in melodic continuity" *Perception & Psychophysics* **57** 451 – 462

Cuddy L L, Thompson W F, 1992 "Asymmetry of perceived key movement in chorale sequences: Converging evidence from a probe-tone analysis" *Psychological Research* **54** 51 – 59

Darwin C J, Turvey M T, Crowder R G, 1972 "An auditory analogue of the Sperling partial report procedure: Evidence for brief auditory storage" *Cognitive Psychology* **3** 255 – 267

Deutsch D, 1972 "Effect of repetition of standard and comparison tones on recognition memory for pitch" *Journal of Experimental Psychology* **93** 156 – 162

Deutsch D, 1975 "Disinhibition in pitch memory" *Perception & Psychophysics* **17** 320 – 324

Deutsch D, 1978 "Delayed pitch comparisons and the principle of proximity" *Perception & Psychophysics* **23** 227 – 230

Fraisse P, 1982 "Rhythm and tempo", in *Psychology of Music* Ed. D Deutsch (New York: Academic Press)

Fredrickson W, 1995 "A comparison of perceived musical tension and aesthetic response" *Psychology of Music* **23** 81 – 87

Fucks W, 1962 "Mathematical analysis of the formal structure of music" *IEEE Transactions of Information Theory* **8** 225 – 228

Gjerdingen R O, 1994 "Apparent motion in music?" *Music Perception* **11** 335 – 370

Goldin D Q, Kanellakis P C, 1995 "On similarity queries for time-series data: Constraint specification and implementation", in *Proceedings of the First International Conference on Principles and Practice of Constraint Programming. Lecture Notes in Computer Science No. 976* Eds U Montanari, F Rossi (Berlin: Springer)

Grey J M, 1977 "Multidimensional perceptual scaling of musical timbres" *Journal of Acoustical Society of America* **61** 1270 – 1277

Hutchinson W, Knopoff L, 1978 "The acoustic component of Western consonance" *Interface* **7** 1 – 29

Iverson P, Krumhansl C L, 1993 "Isolating the dynamic attributes of musical timbre" *Journal of the Acoustical Society of America* **94** 2595 – 2603

Kessler E J, Hansen C, Shepard R N, 1984 "Tonal schemata in the perception of music in Bali and the West" *Music Perception* **2** 131 – 165

Kohonen T, 1997 "Self-organizing Maps (Berlin: Springer)

Krumhansl C L, 1979 "The psychological representation of musical pitch in a tonal context" *Cognitive Psychology* **11** 346 – 374

Krumhansl C L, 1990 *Cognitive Foundations of Musical Pitch* (New York: Oxford University Press)

Krumhansl C L, 1995a "Music psychology and music theory: Problems and prospects" *Music Theory Spectrum* **17** 53 – 80

Krumhansl C L, 1995b "Effects of musical context on similarity and expectancy" *Systematische Musikwissenschaft* **3** 211 – 250

Krumhansl C L, 1996 "A perceptual analysis of Mozart's piano sonata K. 282: Segmentation, tension, and musical ideas" *Music Perception* **13** 401 – 432

Krumhansl C L, 1997 "An exploratory study of musical emotions and psychophysiology" *Canadian Journal of Experimental Psychology* **51** 336 – 352

Krumhansl C L, 1998 "Topic in music: An empirical study of memorability, openness, and emotion in Mozart's string quintet in C major and Beethoven's string quartet in A minor" *Music Perception* **16** 119 – 132

Krumhansl C L, 2000 "Music and affect: Empirical and theoretical contributions from experimental psychology", in *Musicology and Sister Disciplines: Past, Present, Future* Ed. D Greer (Oxford: Oxford University Press) pp 88 – 99

Krumhansl C L, Cuddy L L (in press) "Musikalische Bezugssysteme: Strukturelle Beziehungen bei der Wahrnehmung von Tonhöhen", in *Allgemeine Musikpsychologie: Allgemeinpsychologische Grundlagen musikalischen Handels* Eds T H Stoff, R Oerter *Enzyklopädie der Psychologie* vol. E/VII/1

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

Krumhansl C L, Louhivuori J, Toiviainen P, Järvinen T, Eerola T, 1999 "Melodic expectation in Finnish spiritual folk hymns: Convergence of statistical, behavioral, and computational approaches" *Music Perception* **17** 151 – 195

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

Krumhansl C L, Schenck D L, 1997 "Can dance reflect the structural and expressive qualities of music: A perceptual experiment on Balanchine's choreography of Mozart's Divertimento No 15" *Musicæ Scientiæ* **1** 63 – 85

Krumhansl C L, Schmuckler M A, 1986 "The *Petroushka* chord" *Music Perception* **4** 153 – 184

Krumhansl C L, Shepard R N, 1979 "Quantification of the hierarchy of tonal functions within a diatonic context" *Journal of Experimental Psychology: Human Perception and Performance* **5** 579 – 594

Krumhansl C L, Toivanen P, Eerola T, Toiviainen P, Järvinen T, Louhivuori J, 2000 "Cross-cultural music cognition: Cognitive methodology applied to North Sami yoiks" *Cognition* **75** 1 – 46

Leman M, 1995 "A model of retroactive tone-center perception" *Music Perception* **12** 439 – 471

Leman M, Carreras F, 1996 "The self-organization of stable perceptual maps in a realistic musical environment", in *Proceedings of the Journées d'Informatique Musicale Caen* Ed. G Assayah (Caen: University of Caen IRCAM) pp 156 – 169

Leman M, Carreras F, 1997 "Schema and Gestalt: Testing the hypothesis of psychoneural isomorphism by computer simulation", in *Music, Gestalt, and Computing—Studies in Cognitive and Systematic Musicology* Ed. M Leman (Berlin: Springer) pp 144 – 168

Lerdahl F, 1988 "Tonal pitch space" *Music Perception* **5** 315 – 350

Lerdahl F, 1996 "Calculating tonal tension" *Music Perception* **13** 319 – 363

Lerdahl L, 2001 *Tonal Pitch Space* (New York: Oxford University Press)

Lerdahl F, Krumhansl C L, 2001a *Modeling Tonal Tension and Attraction in Chromatic Contexts: Chopin* (Kingston, Ontario: Society for Music Perception and Cognition)

Lerdahl F, Krumhansl C L, 2001b *Modeling Tonal Tension and Attraction in Chromatic Contexts: Wagner and Messiaen. Symposium on Neo-Riemannian Theory* (Buffalo, NY: University of Buffalo Press)

Lerdahl F, Krumhansl C L, Hannon E, Fineberg J, 2000 "Modeling tension and attraction", in *Art Meets Science: Collaboration Between Music Theorists and Music Psychologists* (Organizers R Parncutt, S Larson) Toronto 2000: Musical intersections

Lerdahl F, Krumhansl C L, in preparation "Musical tension in diatonic, chromatic, hexatonic, and octatonic music"

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

Madson C K, Fredrickson W E, 1993 "The experience of musical tension: A replication of Nielsen's research using the continuous response digital interface" *Journal of Music Therapy* **30** 46 – 63

Monahan C B, Carterette E C, 1985 "Pitch and duration as determinants of musical space" *Music Perception* **3** 1 – 32

Nielsen F V, 1983 *Oplevelse of Musikalsk Spænding* (Copenhagen: Akademisk Forlag)

Parncutt R, 1994 "A perceptual model of pulse salience and metrical accent in musical rhythms" *Music Perception* **11** 409 – 464

Ramsay J O, Silverman B W, 1997 *Functional Data Analysis* (New York: Springer)

Schmuckler M A, 1989 "Expectation in music: Investigations of melodic and harmonic processes" *Music Perception* **7** 109 – 150

Schmuckler M A, 1999 "Testing models of melodic similarity" *Music Perception* **16** 295 – 326

Schubert E, 1999 *The Measurement and Time Series Analysis of Emotion in Music* PhD dissertation, University of New South Wales, Sydney, Australia

Schubert E, 2001 "Continuous measurement of self-report emotional responses to music", in *Music and Emotion* Eds. P N Juslin, J A Sloboda (Oxford: Oxford University Press) pp 393 – 414
Schubert E, 2001 – 2002 "Correlation analysis of continuous response to music: correcting for the effects of serial correlation" *Musicæ Scientiæ* **2001 – 2002** Special Issue, 213 – 236
Shmulevich I, Yli-Harja O, 2000 "Localized key finding: algorithms and applications" *Music Perception* **17** 531 – 544
Temperley D, 1999 "What's key for key? The Krumhansl – Schmuckler key-finding algorithm reconsidered" *Music Perception* **17** 65 – 100
Thompson W F, Cuddy L L, 1989 "Sensitivity to key change in chorale sequences: A comparison of single voices and four-voice harmony" *Music Perception* **7** 151 – 168
Thompson W F, Cuddy L L, 1992 "Perceived key movements in four-voice harmony and single voices" *Music Perception* **9** 427 – 438
Tillmann B, Bharucha J J, Bigand E, 2000 "Implicit learning of tonality: A self-organizing approach" *Psychological Review* **107** 885 – 913
Toiviainen P, Kaipainen M, Louhivuori J, 1995 "Musical timbre: similarity ratings correlate with computational feature space distances" *Journal of New Music Research* **24** 282 – 298
Toiviainen P, Snyder J S (in press) "Tapping to Bach: Resonance-based modeling of pulse" *Music Perception*
Toiviainen P, Tervaniemi M, Louhivuori J, Saher M, Huotilainen M, Näätänen R, 1998 "Timbre similarity: Convergence of neural, behavioral, and computational approaches" *Music Perception* **16** 223 – 241
Treisman A M, 1964 "Verbal cues, language, and meaning in selective attention" *American Journal of Psychology* **77** 206 – 219
Vos P G, Leman M (Eds), 2000 "Tonality induction" *Music Perception* **17** 401 – 548
Williams P, 1980 *The Organ Music of J S Bach* volume 1 (Cambridge: Cambridge University Press)
Youngblood J E, 1958 "Style as information" *Journal of Music Theory* **2** 24 – 35