3 Navigating Tonal Space

Petr Janata
Center for Mind and Brain
Department of Psychology
267 Cousteau Place
University of California
Davis, CA 95618
tel: (530) 297-4471
pjanata@ucdavis.edu

Abstract

The question of how tonal structures in music are perceived and represented by the human mind has been approached by multiple disciplines, primarily music theory and cognitive psychology, and more recently, neuroscience. A parsimonious model of tonal space as the surface of a torus has emerged from various types of theoretical considerations and empirical data. Here I provide a brief overview of different variants of a very data-driven approach to modeling tonal space based on self-organizing maps (SOMs), focusing primarily on an ecologically inspired model (Leman and Carreras 1997) that allows one to project any desired auditory stimulus to the toroidal surface. I illustrate this with examples of the tonal trajectories charted by short chord progressions.

Tonal Theory for the Digital Age (Computing in Musicology 15, 2007), 39–50.

3.1 Recent Neuroscientific Research Related to Tonality

The past fifteen years have seen a rapid increase in the number of published studies examining various facets of music perception and production from a neuroscience perspective (Peretz and Zatorre 2003; Peretz and Zatorre 2005). What might this rush to uncover the neural underpinnings of music reveal about the way in which music is perceived?

Considerable effort has been devoted to the basic idea that the brains of musicians and non-musicians alike support musical knowledge, especially that of tonal structures and contexts. Notes and chords that are considered unlikely to occur in a particular harmonic or melodic context are readily perceived as deviant events. They elicit brain responses that are domain-general markers of expectancy violations (Besson and Macar 1987; Janata 1995; Patel 1998; Koelsch, Gunter et al. 2000).

In such studies, basic music theory has constrained the rather simple stimuli that have been used. However, as the neuroscience community moves increasingly toward using more natural musical stimuli, the need arises for means of modeling important underlying dimensions of musical pieces, such as tonality, without having to resort to painstaking and perhaps even controversial analyses of the pieces' harmonic structure. At the same time, such efforts, if they take into account known neural, perceptual, and cognitive constraints of the human brain, might prove useful to music theorists and musicologists (Huron 2006).

The topic of tonality in western tonal music is one that has been approached from theoretical, cognitive, computational, and neuroscientific perspectives. Accordingly, it provides one with an opportunity to illustrate how these different approaches can inform each other and perhaps combine to create an understanding of how tonal spaces play themselves out *in the minds of listeners*. I emphasize the minds of the listeners because the average person who finds pleasure in listening to any one of the many genres of Western tonal music is unencumbered by theories of how either music or the brain works.

The affective responses of a listener are shaped by the mental facilities that the listener uses to parse and organize the music they hear. These mainly include attention and various forms of memory, mental abilities that place limits on the bits of information that can be held in mind and associated with one another.

3.2 Tonal Space in Time

When thinking about the way in which we perceive and react to music, it is important to consider the timescales over which we do so, and the degree to which such limits are imposed by restricted attentional and mnemonic capacities of our brains. For example, estimates of the amount of time that auditory information is held in sensory memory duration range from 2–6 seconds (Treisman 1964; Lu, Williamson et al. 1992). This limit parallels the limits observed on grouping of events into perceptually coherent rhythmic patterns (Krumhansl 2000), and it coincides with observations that moment-to-moment tension ratings of a piece of music are shaped by local rather than global harmonic considerations (Bigand and Parncutt 1999).

Emotional appraisals of and physiological responses to excerpts of music unfold over similarly short timescales (Schubert 2004; Bigand, Vieillard et al. 2005; Koelsch, Fritz et al. 2006; Korhonen, Clausi et al. 2006; Steinbeis, Koelsch et al. 2006). While music-theoretical models of how tonal structures manifest themselves in human minds need not be concerned with details of how information is maintained over time by the brain, physiologically inspired models of music processing do have to take such information into account.

3.3 Geometries of Tonal Space

One property of tonal space that independent methods converge on is its toroidal shape (Krumhansl 1990; Leman and Carreras 1997; Lerdahl 2001; Burgoyne and Saul 2005). Superficially, the space takes form when the major and minor keys are arranged on the surface of the torus in a way that preserves their distance relationships to each other.

Important harmonic relationships such as the Circle of Fifths are evident within this organization, providing the toroidal model with face validity. However, a question arises when one considers the notion of "distance" between locations of keys on the torus: what are appropriate distance metrics? It is on this point that different disciplines diverge.

Music-theoretic and mathematic models may define distance relationships along various constructs such as the circle of fifths, circle of thirds, chroma circle, and other tonal/chordal relationships that seem salient (Krumhansl 1990; Lerdahl 2001). Distance relationships can also be defined in terms of psychological variables such as ratings of relatedness between a tonal context and probe events (Krumhansl, 1990). Finally, distance relationships can be based on the statistics of pitch-class distributions in key-defining musical material. Such comparisons underlie the self-organizing-map (SOM) approaches that I summarize below. These different approaches to defining distances in tonal space are not mutually exclusive. They all give rise to a toroidal model, and they have informed each other in various ways.

3.4 Self-Organizing Maps (SOMs)

Perhaps the main benefit of the SOM approach is that it allows one to uncover structure in a source of data without making assumptions about the relationships among elements in the source data. The method is agnostic with regard to music theory. Instead, the strong theoretical premise is that nervous systems learn to identify recurring patterns of sensory input, i.e. that they are sensitive to the statistics of their environments. Thus, because segments of music that are considered to be in G major will have very similar pitch-class distributions, elements in a neural network can be trained to recognize when that pitch-class distribution occurs on its inputs. The characteristic mapping of keys comes about because the learning algorithm adjusts the strengths (weights) of the connections between the input elements and output-layer elements within a neighborhood region, causing similar distributions across elements on the input to activate similar regions on the output surface. Overall, the SOM approach treats the listener as a statistical probability extractor.

3.4.1 Probe-Tone SOMs

Three types of SOM models of tonality have been developed. One starts with probe-tone profiles. A probe-tone profile for a key reflects average subjective ratings of how well each of the twelve pitch-classes is perceived to fit into that key (Krumhansl and Kessler 1982; Krumhansl 1990). Probe-tone profiles are formed for each major and minor key by shifting the canonical major and minor profiles across the twelve pitch-classes.

The full complement of profiles then serves as input to a SOM in which the output surface has a toroidal topology (Toiviainen and Krumhansl 2003). During training, the SOMs input patterns to locations on the output surface based on the similarities of the input vectors, such that similar input patterns result in spatially similar output patterns.

Not surprisingly, the resulting spatial arrangement of keys is very similar to the multi-dimensional scaling solutions of the probe-tone rating data that gave rise to the key profiles in the first place. One advantage of this SOM approach is that arbitrary probe-tone profiles can be presented to the trained network. Thus, a probe-tone profile that is obtained during a modulating segment in a piece of music can be projected to the toroidal surface to ask whether multiple key regions are activated.

3.4.2 Pitch-Class SOMs

Instead of using subjective ratings as input to an SOM, it is also possible to use distributions of pitch-class information that are accumulated via different means. In one method that also used a 12-element pitch-class vector as input, different chords were presented to a hierarchical SOM in which the input layer projected to an intermediate chord layer which projected, in turn, to a key layer (Tillmann, Bharucha et al. 2000). The map self-organized into the circle of fifths. When confronted with stimulus materials from several studies of perceived tonality, the model successfully predicted many of the results.

3.4.3 Acoustic Waveform SOMs

Finally, SOMs converge on the toroidal structure in models that start with the acoustic waveform as input (Leman and Carreras 1997; Janata, Birk et al. 2002). The modeling approach adopted by Leman and Carreras (1997), available as the IPEM toolbox for Matlab (*http://www.ipem.ugent.be/Toolbox*), uses models of known physiological mechanisms for defining transformations of the auditory input and subsequent representations.

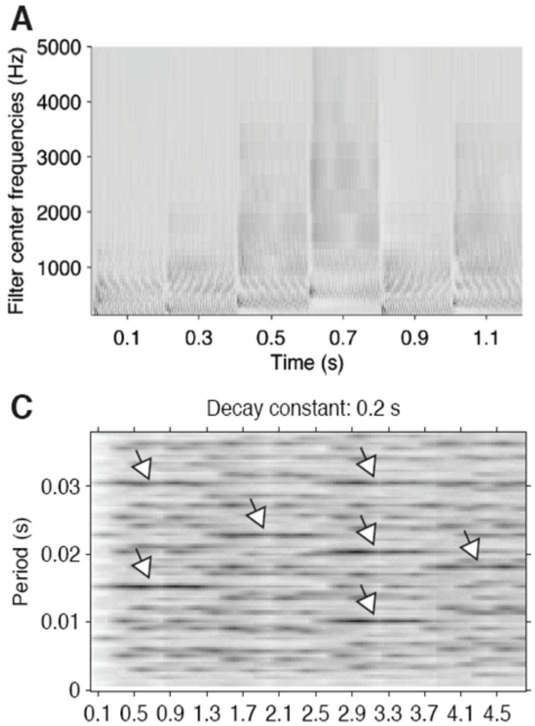


Figure 3.1. *Examples of the representations of auditory signals at different stages of processing as modeled using the IPEM Toolbox.* ***(a)*** *An auditory nerve image shows the firing pattern of different auditory nerve fibers/channels (y-axis) across time that arises at the output of a cochlear model. Note the differences in temporal fine structure across the different channels. Here, the responses to the six 200 ms notes of the first arpeggiated chord in a melody rendered with a synthetic clarinet timbre are shown.* ***(b)*** *Periodicity pitch images represent the time-varying pitch content of the signal. Each of the 24 distinct vertical columns shows the steady state portion of a single 200 ms note. The vertical spacing between the darker bars corresponds to the period of the fundamental frequency. Thus, the initial four notes are ascending in pitch.* ***(c)*** *An example of short time-scale pitch distribution estimates obtained by performing leaky integration (0.2 s decay to half maximum amplitude) over the periodicity pitch image shown in b. While local information about individual notes is still visible, the representation of repeating periods is enhanced. In this case, the four time segments that can be discerned as changes in the pattern of prominent bands (white arrows) correspond to the first four arpeggiated chords in the harmonic sequence of the melody. Each arpeggiated chord consisted of 6 notes.* ***(d)*** *In this example, leaky integration of the periodicity pitch image with a longer time constant (2 s) gives rise to a representation in which the pitch content spanning multiple harmonic transitions is captured at any given moment in time. This representation is taken to reflect the time-varying tonal context of the auditory input. The vector of periodicities at each time point serves as a single training vector to a SOM. Thus, single output units in the SOM are associated with preferred periodicity vectors.*

The initial processing of the sound files mimics the transformation from an acoustic to neural signal in the cochlea (Vanimmerseel and Martens 1992). The result is an "auditory nerve image" that contains the time-varying firing patterns that might be observed across auditory nerve fibers corresponding to different critical bands (Figure 3.1a).

Periodicities in the firing patterns are estimated using autocorrelation and pooled across channels in the auditory nerve image resulting in "periodicity pitch images" (Figure 3.1b) (Langner 1992; Cariani 1999).

Note that in these images a single pitch is represented as a series of evenly spaced peaks at different time lags along the y-axis, rather than as a single peak at a single period value. The periodicity pitch images are then smoothed using leaky integration and a time-constant of choice to provide running periodicity pitch distribution estimates (Figure 3.1c and 3.1d).

Momentary snapshots (single time-windows) of these integrated periodicity patterns then serve as inputs to a SOM algorithm (*http://www.cis.hut.fi/projects/somtoolbox/*), just as probe-tone profiles or pitch-class vector representations of chords served as input in the examples mentioned above.

3.4 Perception of Tonal Regions in a Modulating Melody

Regions of the SOM can be associated with specific keys by presenting the SOM with input vectors that are considered to be key-defining and finding the region of maximal activation in the SOM. The unfolded labeled tori in Figure 3.2 illustrate the arrangement of key regions in an SOM that was trained with a short (*c*. 8-minute) melody that modulated through all of the 24 major and minor keys (Janata, Birk et al. 2003).

As mentioned above, the consequence of the SOM training algorithm is that similar input patterns project to neighboring regions on the output surface. Thus, changing pitch distributions on the input leads to reshaping and/or displacement of the activation peak on the toroidal surface. Figure 3.2 illustrates how several patterns of harmonic motion are manifested on the surface. In this case, a series of stimuli in B major were projected onto the torus that was trained with the modulating melody mentioned above. In other words, the melody served as the basis for defining key regions and was then probed using a set of novel stimuli in B major that included the tonic triad, ascending diatonic scale, and four different chord progressions. Both short-term period distributions corresponding to local tonal chord contexts (Fig. 1c) and longer-term period distributions that correspond to more extended tonal contexts (Fig. 1d) were projected onto the SOM to give a sense of the activation dynamics at these different timescales. For example, a short time-constant of 0.2 s causes the individual notes of the ascending B major scale to activate separate spots on the torus, whereas a longer time constant causes the center of mass of the activation to

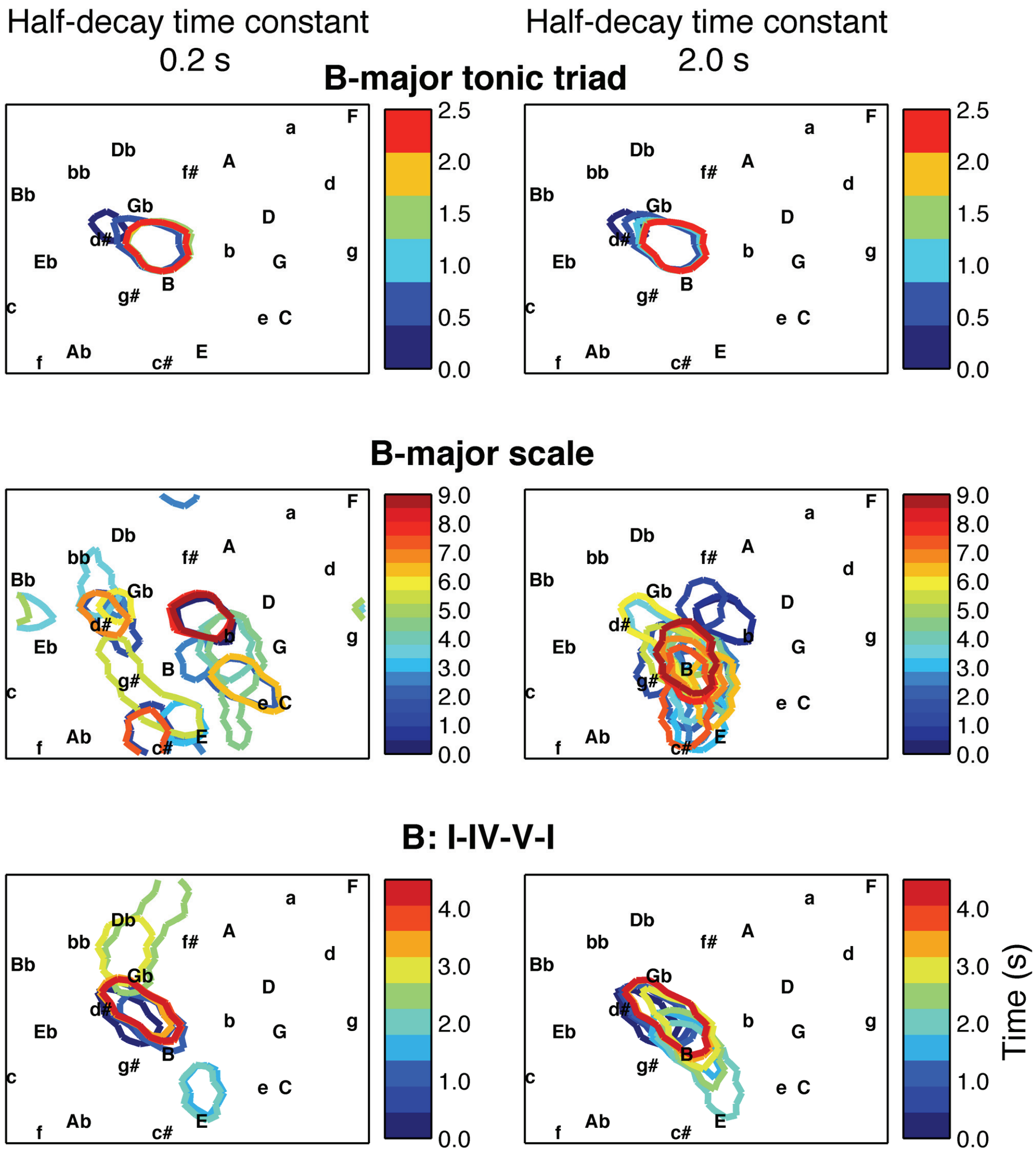
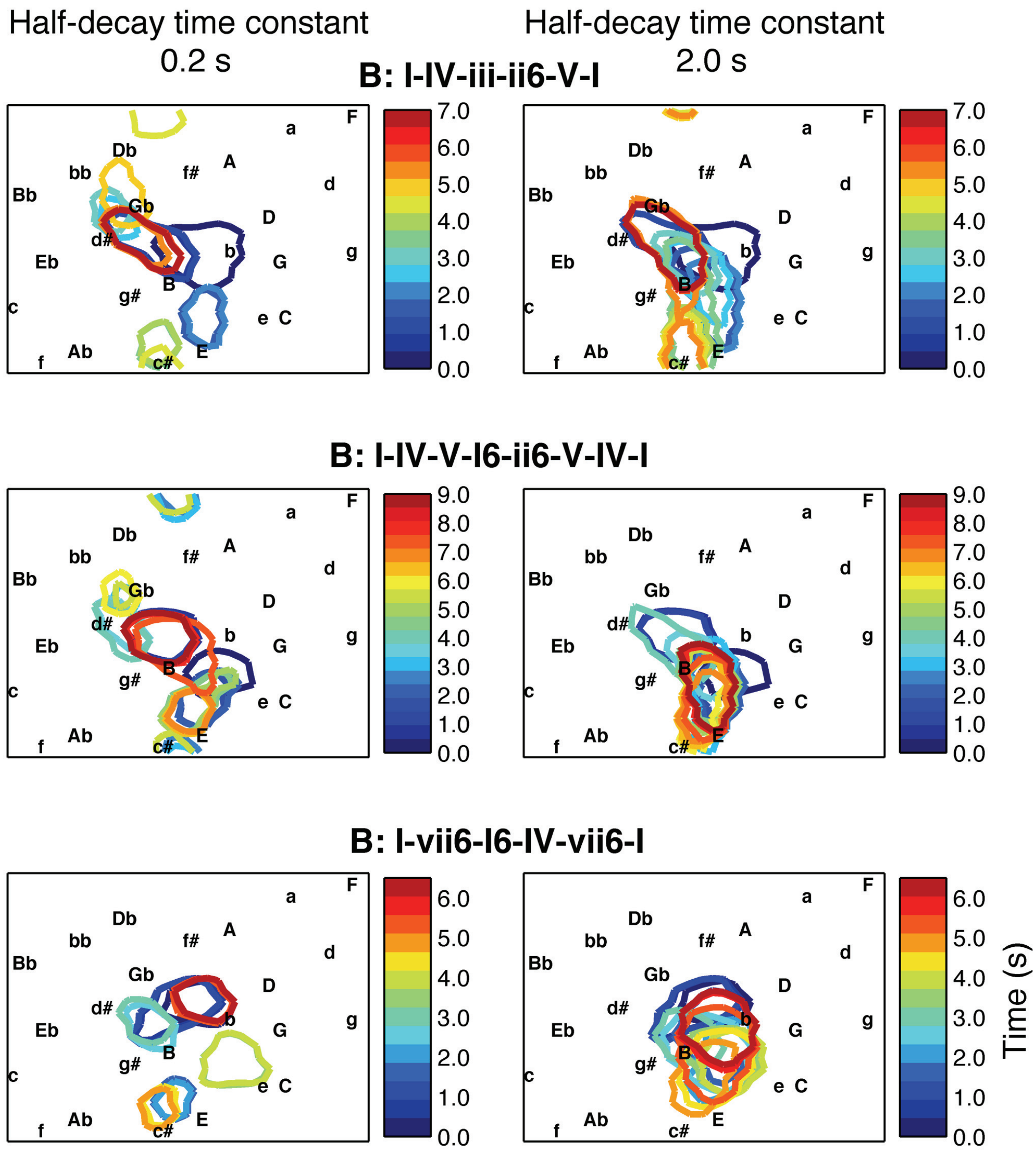


Figure 3.2. *Activation trajectories on the toroidal surface. Each panel depicts the surface of a SOM on a flattened torus. The left and right edges are contiguous as are the bottom and top edges. The SOM was trained using a melody that systematically modulates through all 24 major and minor keys (Janata, 2003). Key region labels were assigned to locations on the surface that showed the strongest activation when presented with segments of the melody that corresponded to each key (excluding modulation regions). Six musical stimuli in B Major (the tonic triad, the ascending and descending diatonic scale, and four chord progressions shown in Figure 3.3) were rendered using the Grand Piano patch on the Apple Software Synth,*



processed through the auditory model, and projected onto the SOM. The time-varying activation patterns are shown using a single contour line approximately every 500 ms. The contour line circumscribes the extent of activation that is ≥95% of the maximum activation at that time point. The grayscale value is used to represent elapsed time from the onset of the stimulus. For each stimulus, the panel on the left shows the projection of the tonal context images that arose from leaky integration with a short time constant, and the panel on the right corresponds to the projection of tonal context images calculated using a longer time constant. Please see the online version for a color version of this figure.

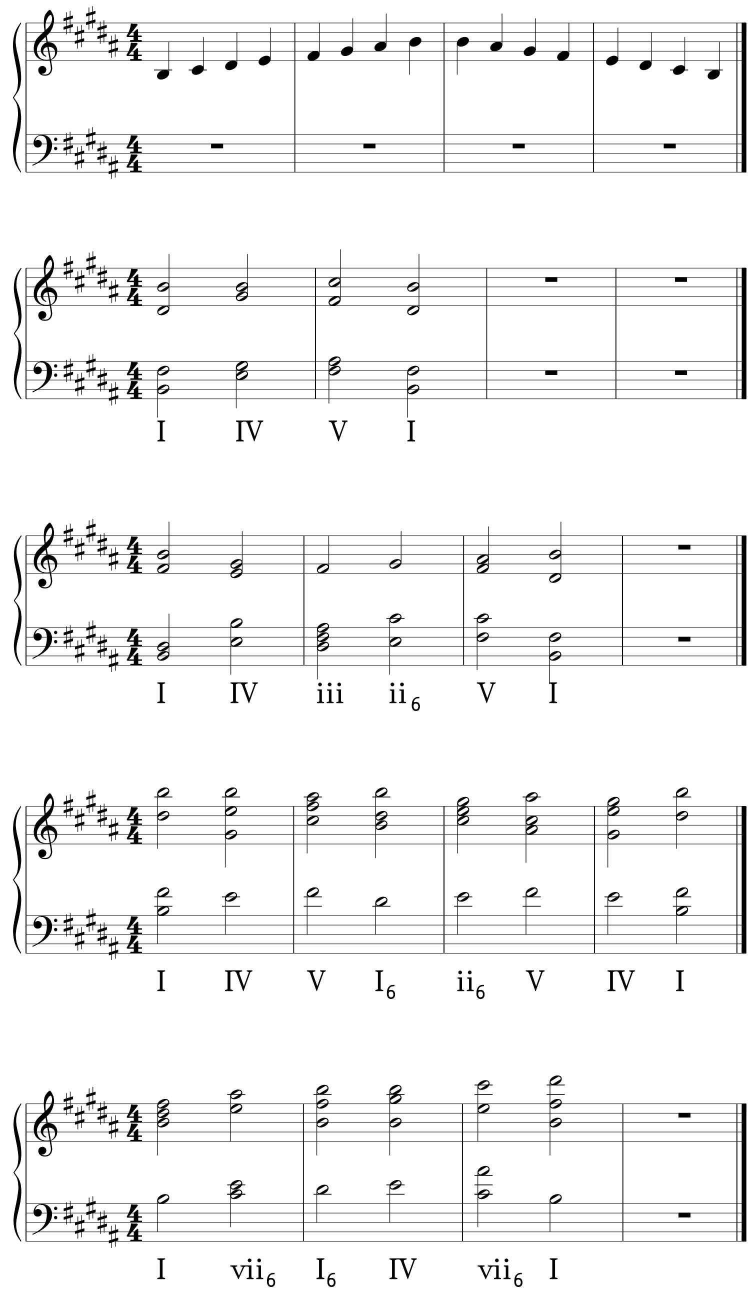


Figure 3.3. *Musical stimuli (ascending and descending B Major scale, four sets of chord progressions) used to generate Figure 3.2 (b)-(f). A quarter note = 120.*

dwell in the B-major region. The chord progression examples in Figure 2 further illustrate that although a region of the torus in the vicinity of the B-major label is activated in all cases, the activation is biased toward different key regions, and the biasing depends on the harmonic structure of the preceding sequence.

3.5 Discussion

The observations reported here echo earlier results of a behavioral experiment that mapped the developing and changing sense of key on the surface of a torus (Krumhansl and Kessler 1982). Specifically, each chord in a modulating or non-modulating sequence resulted in a displacement of the perceived momentary location on the surface of the torus, supporting the notion that perceived locations in tonal space are strongly influenced by local information and are dynamic. The simulations shown in Figure 2 indicate that the motility of the activation peak on the torus is governed by the time-constant that is used to shape the time-varying pitch distributions, with shorter time-constants resulting in larger displacements from one event to the next. Therefore, it is conceivable that, by matching the perceived trajectories against multiple simulated trajectories in which the time constants have been varied, one might estimate the listener's own time constant for integrating tonal information.

There is a considerable amount of converging behavioral evidence that listeners' journeys through tonal space are influenced strongly by local tonal information (Krumhansl and Kessler 1982; Tillmann, Bigand et al. 1998; Bigand and Parncutt 1999; Leman 2000), and such local information appears to be well represented in the time-varying activation patterns of tonal space represented on a toroidal surface. Although unconstrained by music theory, different SOM approaches, constrained by considerations of psychological and neural processes, have converged on similar representations of tonal space onto which subjects' percepts (Toiviainen and Krumhansl 2003) or arbitrary musical stimuli can be projected with relative ease (Leman and Carreras 1997; Janata, Birk et al. 2002; Toiviainen 2005). While the degree to which the current toroidal models will be successful in capturing tonal relationships that span longer timeframes (Lerdahl and Jackendoff 1983; Lerdahl 2001) remains unclear, they are likely to enjoy some utility in identifying brain processes that respond to the movement of music on the timescale captured by this type of model of tonal space (Janata, Birk et al. 2002).

Acknowledgments

I thank Rawi Nanakul for preparing the musical material.

References

Besson, M. and F. Macar (1987). "An Event-Related Potential Analysis of Incongruity in Music and Other Non-Linguistic Contexts," *Psychophysiology* 24/1: 14–25.

Bigand, E. and R. Parncutt (1999). "Perceiving Musical Tension in Long Chord Sequences," *Psychological Research* 62/4: 237–54.

Bigand, E., S. Vieillard et al. (2005). "Multidimensional Scaling of Emotional Responses to Music: The Effect of Musical Expertise and of the Duration of the Excerpts," *Cognition & Emotion* 19/8: 1113–39.

Burgoyne, J. A. and L. K. Saul (2005). *Visualization of Low-Dimensional Structure in Tonal Pitch Space*. International Computer Music Conference 2005, Barcelona, Spain.

Cariani, P. (1999). "Temporal Coding of Periodicity Pitch in the Auditory System: An Overview," *Neural Plast* 6/4: 147–72.

Huron, David (2006). *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge, MA: MIT Press.

Janata, P. (1995). "ERP Measures Assay the Degree of Expectancy Violation of Harmonic Contexts in Music," *Journal of Cognitive Neuroscience* **7**: 153–64.

Janata, P., J. L. Birk et al. (2003). "Online Detection of Tonal Pop-Out in Modulating Contexts," *Music Perception* 20/3: 283–305.

Janata, P., J. L. Birk et al. (2002). "The Cortical Topography of Tonal Structures underlying Western Music," *Science* 298/5601: 2167–70.

Koelsch, S., T. Fritz, et al. (2006). "Investigating Emotion with Music: An FMRI Study," *Human Brain Mapping* 27: 239–50.

Koelsch, S., T. Gunter, et al. (2000). "Brain Indices of Music Processing: 'Nonmusicians' are Musical," *Journal of Cognitive Neuroscience* 12/3: 520–41.

Korhonen, M. D., D. A. Clausi, et al. (2006). "Modeling Emotional Content of Music using System Identification," *IEEE Trans Sys Man Cybern B Cybern* 36/3: 588–99.

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

Krumhansl, Carol. L. (2000). "Rhythm and Pitch in Music Cognition," *Psychology Bulletin* 126/1: 159–79.

Krumhansl, Carol. L. and E. J. Kessler (1982). "Tracing the Dynamic Changes in Perceived Tonal Organization in a Spatial Representation of Musical Keys," *Psychological Review* 89/4: 334–68.

Langner, G. (1992). "Periodicity Coding in the Auditory System," *Hearing Research* 60/2: 115–42.

Leman, M. (2000). "An Auditory Model of the Role of Short-Term Memory in Probe-Tone Ratings," *Music Perception* 17/4: 481–509.

Leman, M. and F. Carreras (1997). "Schema and Gestalt: Testing the hypothesis of Psychoneural Isomorphism by Computer Simulation," *Music, Gestalt, and Computing—Studies in Cognitive and Systematic Musicology*. Berlin: Springer Verlag, 144–68.

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

Lerdahl, Fred and Ray Jackendoff (1983). *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.

Lu, Z. L., S. J. Williamson et al. (1992). ”Behavioral Lifetime of Human Auditory Sensory Memory Predicted by Physiological Measures,” *Science* 258 (5088): 1668–70.

Patel, A. D. (1998). ”Syntactic Processing in Language and Music: Different Cognitive Operations, Similar Neural Resources?” *Music Perception* 16 (1): 27–42.

Peretz, I. and R. J. Zatorre, eds. (2003). *The Cognitive Neuroscience of Music*. New York: Oxford University Press.

Peretz, I., and R. J. Zatorre (2005). ”Brain Organization for Music Processing,” *Annual Review of Psychology* 56: 89–114.

Schubert, E. (2004). ”Modeling Perceived Emotion with Continuous Musical Features,” *Music Perception* 21/4: 561–85.

Steinbeis, N., S. Koelsch et al. (2006). ”The Role of Harmonic Expectancy Violations in Musical Emotions: Evidence from Subjective, Physiological, and Neural Responses,” *Journal of Cognitive Neuroscience* 18/8: 1380–93.

Tillmann, B., J. J. Bharucha et al. (2000). ”Implicit Learning of Tonality: A Self-Organizing Approach,” *Psychology Review* 107/4: 885–913.

Tillmann, B., E. Bigand et al. (1998). ”Effects of Global and Local Contexts on Harmonic Expectancy,” *Music Perception* 16/1: 99–117.

Toiviainen, P. (2005). ”Visualization of Tonal Content with Self-Organizing Maps and Self-Similarity Matrices,” *ACM Computers in Entertainment* 3/4: Article 3C.

Toiviainen, P., and Carol. L. Krumhansl (2003). ”Measuring and Modeling Real-Time Responses to Music: The Dynamics of Tonality Induction,” *Perception* 32/6: 741–66.

Treisman, A. M. (1964). ”Verbal Cues, Language, and Meaning in Selective Attention,” *American Journal of Psychology* 77: 106–219.

Vanimmerseel, L. M., and J. P. Martens (1992). ”Pitch and Voiced-Unvoiced Determination with an Auditory Model,” *Journal of the Acoustical Society of America* 91/6: 3511–26.

Submitted: 21 December 2006. Final revisions: 14 September 2007.