

UC Office of the President

Recent Work

Title

A combined model of sensory and cognitive representations underlying tonal expectations in music: from audio signals to behavior.

Permalink

<https://escholarship.org/uc/item/53b2r818>

Journal

Psychological review, 121(1)

ISSN

0033-295X

Authors

Collins, Tom
Tillmann, Barbara
Barrett, Frederick S
[et al.](#)

Publication Date

2014

DOI

10.1037/a0034695

Supplemental Material

<https://escholarship.org/uc/item/53b2r818#supplemental>

Peer reviewed

TITLE: A combined model of sensory and cognitive representations underlying tonal expectations in music: from audio signals to behavior

AUTHORS: Tom Collins^{*‡}, Barbara Tillmann[#], Frederick S. Barrett^{*†}, Charles Delbé[#], and Petr Janata^{*†}

AFFILIATIONS: ^{*}Center for Mind and Brain, University of California, Davis, USA

[‡]Department of Computational Perception, Johannes Kepler University, Linz, Austria

[#]Lyon Neuroscience Research Center, Lyon, France

[†]Department of Psychology, University of California, Davis, USA

RUNNING HEAD: sensory and cognitive contributions to tonal expectations

KEYWORDS: tonality, torus, priming, music information retrieval, pitch

CODE AND DATA: <http://atonal.ucdavis.edu/resources/software/jlmt/tonmodcomp/>

CORRESPONDENCE:

Petr Janata

Center for Mind and Brain

267 Cousteau Place

Davis, CA 95618

phone: 530-297-4471

fax: 530-297-4400

e-mail: pjanata@ucdavis.edu

ABSTRACT

Listeners' expectations for melodies and harmonies in tonal music are perhaps the most studied aspect of music cognition. Long debated has been whether faster response times (RTs) to more strongly primed events (in a music theoretic sense) are driven by *sensory* or *cognitive* mechanisms, such as repetition of sensory information or activation of cognitive schemata that reflect learned tonal knowledge, respectively. We analyzed over 300 stimuli from seven priming experiments comprising a broad range of musical material, using a model that transforms raw audio signals through a series of plausible physiological and psychological representations spanning a sensory–cognitive continuum. We show that RTs are modeled, in part, by information in periodicity pitch distributions, chroma vectors, and activations of tonal space – a representation on a toroidal surface of the major/minor key relationships in Western tonal music. We show that in tonal space, melodies are grouped by their tonal rather than timbral properties, whereas the reverse is true for the periodicity pitch representation. While tonal space variables explained more of the variation in RTs than did periodicity pitch variables, suggesting a greater contribution of cognitive influences to tonal expectation, a stepwise selection model contained variables from both representations and successfully explained the pattern of RTs across stimulus categories in four of the seven experiments. The addition of closure – a cognitive representation of a specific syntactic relationship – succeeded in explaining results from all seven experiments. We conclude that multiple representational stages along a sensory–cognitive continuum combine to shape tonal expectations in music.

Music permeates human cultures around the world and is believed to have an evolutionary history that parallels that of language (Brown, 2000). In the broadest sense, music can be thought of, at a basic featural level, as the structured organization of auditory objects (sounds) into temporally extended sequences. What differentiates the many musical systems and musical genres encountered across the world's human cultures from each other, as well as music from other structured sound patterns such as language, birdsong, and a variety of environmental sounds that have periodic structure, are the principles by which auditory objects are organized relative to each other. Developing an understanding of how organizing principles in music come to be represented and operated on by the human mind and brain is likely to yield insights into the general biological and psychological mechanisms responsible for perceiving, associating, remembering, acting upon, and responding affectively to sequences of environmental signals.

In this article we focus on Western tonal music, as is common in research on music perception and cognition. Using computational modeling and data from a range of behavioral experiments, we address an issue that is frequently debated in the psychological and neuroscientific literature pertaining to the perception of tonal structure (Butler, 1989, 1990; Krumhansl, 1990a, 1990b; Leman, 2000; Parncutt & Bregman, 2000): do *sensory* or *cognitive* sources of information generate expectations for tonal events? Processes associated with forming and evaluating musical expectancies underlie our ability to detect “wrong notes,” and have broad implications for how we understand music and its emotional impact on us (Huron, 2006). Prior to describing the basis for the sensory–cognitive debate in the section titled *Priming of tonal knowledge*, we clarify our

use of the term *tonal structure* because the concept can refer to related yet different things.

At the level of single auditory objects, tonal structure pertains to the relationships of the frequencies a tone comprises. If the tone comprises multiple frequencies and those frequencies (often called overtones) are integer multiples (harmonics) of a fundamental frequency, the complex tone is associated with the percept of pitch. The perceived pitch of the complex tone is equivalent to the perceived pitch of the fundamental frequency played in isolation (Oxenham, 2012). Manipulating the presence and relative amplitudes of the different frequency components in a complex tone will shape the tone's timbral qualities, and may also influence the perceived pitch (Krumhansl & Iverson, 1992). These observations highlight that low-level *sensory* properties of musical sounds, i.e. the relationships among constituent frequencies, can shape the perceptual consequences based on the simple notion that evidence biases perception in the direction of the focus of the evidence. An example of this is the phenomenon of the *missing fundamental*: a pitch with fundamental frequency x Hz is perceived when overtones ordinarily associated with this fundamental frequency are present, but there is in fact no x Hz component in the sound (Schnupp, Nelken, & King, 2011). While the aspect of tonal structure that we consider in this article pertains to the relationships among multiple auditory objects, more specifically the circumscribed set of pitches that make up major and minor scales and keys in Western tonal music, the idea that sensory evidence accumulates and biases perception nonetheless carries forward from individual musical tones to collections of musical tones.

Henceforth, our use of the term *tonal structure* refers to the relationships among individual pitches (and by extension, musical notes), rather than the relationships among the frequencies that make up a single note. Specific relationships among multiple pitches create, among other things, a sense of key or tonal center, e.g. the key of E major. The notion of key, elaborated in more detail in the next section, is critical for the perception of Western tonal music: notes either belong to a key, or by contrast fall outside a key and are thereby more likely to be perceived as contextually deviant “wrong” notes. The fact that we experience notes to vary in the degree to which they fit into the ongoing musical context as we listen to a piece of music belies the existence of one or more mental representations that generate expectations of what notes we will probably hear in a given context (Krumhansl, 1990a).

As we will review below, an extensive body of cognitive psychology research on tonal structure is consistent with a *cognitive* view, which holds that mental representations of key structure are learned and stored in long-term memory. However, numerous studies have provided evidence for a *sensory* view in which tonal membership judgments can be explained by models of tonal information accumulated in short-term memory (Bigand, Tillmann, & Poulin-Charronnat, 2006). According to the sensory view, perceptual judgments will be faster and more accurate when there is a high degree of overlap in the pitch/frequency content of the to-be-judged target event and the preceding prime/context. In the cognitive view, a perceptual judgment can be facilitated even when there is no sensory overlap of the target and prime, as long as the target event is strongly related to the content of the prime (via the stored cognitive representation of pitch relationships determined by the key), relative to other possible target events.

While the sensory–cognitive debate has served to dichotomize these theoretical viewpoints on tonal perception, the actual mechanisms and representations that support tonal perception may not adhere to such a strict dichotomy. Here, we consider tonal perception on a continuum along which acoustic signals are transformed into higher-order musical knowledge across a series of processing stages. While the representations at each of the processing stages may be regarded as “sensory” or “cognitive” to differing degrees, they are all taken as available sources of information on which perceptual judgments related to tonal structure can be based. Before describing our model in relation to the sensory–cognitive debate and other models of musical expectation, we briefly review some requisite music terminology and concepts.¹

TONAL STRUCTURE IN WESTERN TONAL MUSIC

The basic unit for our discussion of tonal structure is that of *pitch class* or, synonymously, *pitch chroma*. Consider the span of an octave – a doubling in frequency. In equal-tempered Western tonal music, this span is divided into twelve pitches/notes that are separated from each other equally on a logarithmic scale. The interval between two adjacent pitches is referred to as a *semitone*. Each of the notes is given a letter name, e.g. C, D, E, etc., and possibly a symbol to designate it as a note that is one semitone lower (♭, flat) or higher (♯, sharp) than the note bearing only the letter as its name. Across octaves, notes with the same name are judged as highly similar to each other, a perceptual phenomenon known as octave equivalence (Deutsch, 1999). Thus they are ascribed to the same pitch class and play the same functional role in establishing tonal structure. Rules specify which sets of pitch classes belong to major and minor *keys*. For each of the twelve major and minor keys (one major and one minor for each pitch class), seven pitch

classes belong to the key and five fall outside the key. Two pitches played simultaneously form a harmonic interval, and chords are built from interval combinations. For instance, the *major triad* (e.g., C, E, G) consists of two superimposed thirds (C, E; E, G), as does the *minor triad* (C, E \flat , G), where the middle pitch class has been flattened from E to E \flat . Intervals and chords differ in their consonance/dissonance and tonal implications, depending on the specific combination of pitches (Hutchinson & Knopoff, 1978; Plomp & Levelt, 1965).

The relationship between the concepts of pitch class and both musical and psychological representations of key is illustrated most easily as a *key profile* (Figure 1A). Perceptual key profiles are typically obtained using the probe-tone method in which a sense of key is instantiated with a context, whether using a single chord, a sequence of chords (Figure 1B), or a melody, and is followed by a single note (probe; Figure 1B) about which a judgment must be made. Although the exact shape of key profiles varies as a function of listener expertise, age, measurement, and context (Krumhansl & Cuddy, 2010), a basic observation remains consistent across studies: notes that strongly define a key – the first scale degree (tonic) and fifth scale degree (dominant) – are psychologically privileged, in that they are judged to fit better with the preceding context than are other scale degrees, and judgments about them are made more quickly and accurately. Usually, pitch classes that belong to the key (called diatonic notes) are also judged to fit better (or judged more quickly/accurately) than those that do not belong (called chromatic or non-diatonic notes) (Janata & Reisberg, 1988). Key profiles are a summary of behavioral results that point to the existence of *tonal hierarchies*. A tonal hierarchy is a more general concept that spans the music theoretic and psychological

literature. In essence, tonal hierarchies specify the relative music theoretic, psychological, or statistical prominence of certain notes and chords over others within a given key, and they form a basis for analyzing the sequential and hierarchical structure of musical passages and associated psychological responses to those passages.

The idea of tonal hierarchies, expressed in the pattern of relative weightings for 12 pitch classes (as reflected in the key profiles), is prevalent not only within music cognition but also in the younger field of music information retrieval (MIR). In MIR there is a heavy reliance on chroma vectors, which also consist of relative weightings for the 12 pitch classes. Here, we can consider key profiles and chroma vectors as synonymous – reflecting the relative presence or probability of a pitch class in a given context – since perceptual key profiles have been shown to closely match the distributional statistics of pitch class information in corpora of Western tonal music (Krumhansl, 1990a). This observation simultaneously suggests that knowledge of tonal structure might be obtained via statistical learning mechanisms and fuels the debate about whether perceptual judgments of tonal material are supported by sensory priming or cognitive priming, that is, priming based on information stored in long-term memory (Krumhansl & Cuddy, 2010). The context in Figure 1B for instance contains three occurrences of pitch-class C, and no occurrences of F \sharp , so a sensory argument based on pitch content in short-term memory suffices for explaining the stronger goodness-of-fit rating for C than for F \sharp in Figure 1A. A sensory argument seems less convincing for explaining the difference in ratings between C and G in Figure 1A, however, as both have three occurrences in Figure 1B. Cognitive processes may contribute in the absence of

sensorially related primes, with the prime instantiating a strong enough cognitive schema to which the probe can be compared.

Although a tonal hierarchy for single pitch classes is shown in Figure 1A, the concept extends to a hierarchy of chords built on the different scale degrees (for more details see Rosen, 1972, pp. 23–29). Thus a chord constructed from the notes of the tonic triad (pitch classes C, E, and G in the key of C major) can be used to prime the key of C major. A chord built on the second most salient note within the key profile – the dominant (V, 5th scale degree) – is also a very stable chord within an instantiated key, and commonly regarded as the second most stable chord within a key. The dominant serves as the tonic chord of the key that is adjacent, in a clockwise direction, on the circle of fifths (Figure 1C) to the instantiated key. The tonic chord of the key that is one step in the counter-clockwise direction on the circle of fifths from the instantiated key is built on the 4th scale degree, and is therefore the sub-dominant (IV) of the instantiated key. The sub-dominant is commonly regarded as the third most stable chord within a key. Together, the I, IV, and V chords form the basis for harmonic progressions in many forms of Western classical and popular music. Chords built on the other scale degrees further flesh out the harmonic hierarchy.

The circle of fifths represents the tonal proximity of the different keys in spatial distance (the more related, the closer they are) and typically a piece of tonal music will establish a key center and not stray too far from this key on the circle. By using chordal (also called harmonic) contexts, Krumhansl and Kessler (1982) were able to demonstrate a higher-order organization of key relationships that captured the relative psychological distances between the different major and minor keys. The most parsimonious

arrangement of the keys is on the surface of a torus (Figure 1D). Importantly, the map of keys based on psychological distances respects critical music theoretic relationships. Two helical orbits describe the circle of fifths for major and minor keys (Figure 1C), and the interlacing of those orbits captures the relatedness between each major key and both of its related minor keys, the relative and parallel minors.² For example, F \sharp minor is the relative minor of A major (“f \sharp ” appears to the left of “A” in Figure 1D), and A minor is the parallel minor of A major (“a” appears above “A” in Figure 1D).

Throughout this article we refer to the representation of keys on a toroidal surface as *tonal space*. It is important to note that each location on the torus represents a key profile with a different weighting. The labeled locations are those that correlate maximally with the perceptual profile observed for that key. Thus the toroidal surface can be thought of as a distribution of key profiles. While the present research pursues the toroidal arrangement of key profiles advanced by Krumhansl and Kessler (1982) (due to its empirical basis), it would be remiss not to mention that very similar lattice arrangements of pitch classes have for centuries been the speculation of scientists and music theorists, beginning with the *Tonnetz* of Euler (1739, p. 147), the *Chart of regions* of Schoenberg (1954/1983) and more recently Longuet-Higgins (1979, pp.316–317), Balzano (1980, p. 73), Shepard (1982, p. 327), and Lerdahl (2001, p. 66). This long-standing fascination with describing the structure of musical pitch relations reflects the importance of tonality in cognition, and more generally the human propensity to impose order on naturally occurring phenomena.

PRIMING OF TONAL KNOWLEDGE

As alluded to above, a sensory–cognitive debate emerged surrounding the explanation of tonal hierarchies (Butler, 1989, 1990; Krumhansl, 1990a, 1990b; Leman, 2000; Parncutt & Bregman, 2000) and the results of other experiments that have investigated tonal structure processing (Bharucha & Stoeckig, 1987; Bigand, Poulin, Tillmann, Madurell, & D'Adamo, 2003; Delbé, 2009; Marmel, Tillmann, & Delbé, 2010; Tekman & Bharucha, 1998). Sensory accounts of tonal expectancy argue that observed tonal hierarchies arise solely on the basis of shared sensory features of a target event and the preceding prime context that can be maintained in a short-term memory store. For instance, it has been suggested that a chord whose tones and/or overtones coincide with those of a preceding chord could be more anticipated than would be a continuation chord containing no coincident tones and/or overtones (Bharucha & Stoeckig, 1987; Bigand et al., 2003; Schmuckler, 1989). A cognitive account of tonal expectancy, on the other hand, draws on listeners' schematic knowledge of tonal space that is stored in long-term memory and can be activated by a contextual stimulus. This activation leads to stronger anticipation of certain notes or harmonies than others, even in the absence of those notes or harmonies in the preceding context. Interpreted within the context of the toroidal model of tonal space, activation of a particular key region serves to prime the tonal hierarchy associated with that location of tonal space, even though a particular pitch class may not have been present in the context that primed the key region.

Cast more broadly, the sensory–cognitive debate in music is about the relative dominance of perceptual and repetition priming versus priming of learned representations, in this case short-term sensory memory for present pitch distributions and long-term

memory of tonal hierarchy structure, respectively. Research in the psychology of language has long distinguished between various forms of priming in the course of spoken and written word recognition, including, repetition priming, rime priming, semantic priming, cross-modal priming (e.g. Hecht, Reiner, & Karni, 2009; Hillinger, 1980; Radeau, Besson, Fonteneau, & Castro, 1998). Similarly, the distinction between repetition priming and semantic priming in the perception of environmental sounds has been explored (De Lucia et al., 2010). Following on observations from other domains that various types of primed representations may interact to varying degrees, as in the case of morphology in word recognition (Gonnerman, Seidenberg, & Andersen, 2007), our primary goal in this article is to adjudicate on the relative contributions of different priming mechanisms in music cognition.

Paralleling other domains in psychology, the priming paradigm emerged in the music cognition field as a means of probing listeners' implicit knowledge of tonal structure (Bharucha & Stoeckig, 1986, 1987).³ In tonal priming experiments, participants are typically asked to make speeded judgments in response to a musical event at the end of a short excerpt of music, such as to indicate by pressing a button whether the final event is in-tune or mistuned, or played by one timbre or another. The nature of the discrimination task is not of utmost importance, so long as it does not involve an explicit judgment of tonal relatedness or expectancy, as do goodness-of-fit ratings. Rather, by manipulating the tonal content of the short excerpts, it is possible to investigate how the tonal function of a final chord or note, relative to the preceding context, implicitly affects participants' RTs (Bharucha & Stoeckig, 1986, 1987). Thus just as RTs in a lexical decision task are used to make inferences about the implicitly processed semantic

relatedness of prime and target words, intonation or timbral judgments are used to make inferences about the implicitly processed tonal relatedness between target notes or chords and the preceding tonal context. While participants' accuracy on a given task is important, RTs are the focus here. Research has shown that the greater the listener's expectancy for the final event, the shorter the RT (see Bharucha & Stoeckig, 1986; Bharucha & Stoeckig, 1987 for music; see Neely, 1991 for language). Explicit measures requiring judgments of completion, expectation or tension are still commonly used, however (e.g., Pearce, Ruiz, Kapasi, Wiggins, & Bhattacharya, 2010).

Experimental dissociation of sensory and cognitive influences on the perception of tonal hierarchies

In this article, we focus on modeling RTs for terminal events in 303 chord sequences and melodies drawn from seven studies in the tonal priming literature discussed briefly below (Table 1; examples of the musical stimuli are provided in Supplemental Figures S1–S4). The studies investigated various aspects of tonal structure processing and the development of musical expectations in time in non-musician listeners for both harmonic and melodic materials. The stimuli – like the majority of experimental approaches in the field – are mainly restricted to tonal aspects of 18th century European music, but we acknowledge a broader aim that encompasses tonal systems from any time or place, even custom-built systems such as the Bohlen-Pierce scale used by Loui, Wu, Wessel, and Knight (2009). The restriction is legitimate, however, because most of the experimental evidence regarding mental representations of tonal structure has been obtained using the Western tonal music system.

The studies described below are of particular interest because taken together they address the sensory–cognitive debate while simultaneously addressing issues regarding the perception of the hierarchical organization of tonal structure, i.e. the relative importance of different tonal functions. In some of the experiments, sensory and cognitive influences were confounded, whereas in others they were purposefully dissociated. Moreover, taken as a group, the RTs from these studies cannot be modeled using other models of tonal knowledge because various limitations of those models preclude them from representing at least some of the stimulus material.⁴ We now describe the studies we focus on and our reasons for doing so, followed by a description of relevant computational models of tonal expectation.

The strongest contrast between positions in the tonal hierarchy appears in Tillmann, Janata, and Bharucha (2003), where half of the chord sequences ended on the tonic chord (called related), and the other half ended on a major chord with the root one semitone away from the established tonic (an out-of-key event called unrelated, Supplemental Figure S1). RTs were faster for related than unrelated targets. Our reasons for including this dataset are (a) that it presents a clear test case in which any model of tonal processing must succeed, (b) the coarseness of the tonal distinction (two out of key notes) is representative of stimuli used in the majority of neuroimaging studies on tonality (Koelsch, Gunter, Friederici, & Schröger, 2000; Koelsch et al., 2001; Koelsch et al., 2002; Koelsch, Gunter, Wittfoth, & Sammler, 2005; Leino, Brattico, Tervaniemi, & Vuust, 2007; Maess, Koelsch, Gunter, & Friederici, 2001), and (c) it represents a clear case where sensory and cognitive approaches to priming make the same predictions for perceived differences. The confounding of sensory and cognitive predictions has

motivated the use of more controlled material to provide evidence for cognitive processing of musical structure, as in the remaining experiments that we considered.

Stimuli from (Bigand et al., 2003) were selected to represent a more subtle manipulation of harmonic material, in which chord sequences with tonic (I) and subdominant (IV) endings were contrasted (Supplemental Figure S2).⁵ In this material, all occurrences of the target chord were removed from the context, so that it would be awkward to offer a sensory interpretation of any resulting differences. Thus these stimuli control for sensory influences and explore cognitive influences somewhat further. Despite the reduction in direct sensory priming, the RTs were faster for the related tonic ending than for the less-related subdominant ending, thus supporting the cognitive interpretation.

To complement stimulus material using chord (harmonic) sequences, we also selected data from experiments that employed melodic stimuli. Marmel, Tillmann, and Dowling (2008) used melodies that ended on either the tonic (I) or the subdominant (IV; Supplemental Figure S3A). Conforming to tonal hierarchy predictions, RTs were faster for tonic endings than for subdominant endings. Priming studies of tonal structure that investigated more subtle distinctions in the tonal hierarchy of within-key targets have primarily compared RTs for the tonic (I), which is functionally situated at the top of the tonal hierarchy, with the subdominant (IV), which is less stable but nonetheless serves a prominent tonal function. To explore the processing of tonal functions lower down in the tonal hierarchy, we used data from Marmel and Tillmann (2009). Half of the stimuli in this experiment ended on the third scale degree, the mediant (III), and the other half ended on the seventh scale degree, leading tone (VII; Supplemental Figure S3B). The

mediant and leading tone are both members of the diatonic scale that defines a key, though with differing key profile values (Figure 1A). As a member of the tonic triad, the mediant has the third highest value in the canonical key profile, whereas the leading tone is regarded as “the least stable” tone in a key (with the lowest value of the in-key tones) because of its tendency to resolve to the tonic, i.e. the note that most commonly follows the leading tone is the tonic. RTs were indeed faster for the mediant than for the leading tone. The third set of RT data for melodic contexts was from Marmel et al. (2010). The motivation for this study was to determine whether expectation for the target event is driven solely by sensory properties (the aforementioned overtone structure) of the context, or if cognitive processes are involved. To this end, a 2 x 2 factorial design was used with one factor manipulating whether the melody ended on the tonic or the subdominant (Supplemental Figure S3C) and the second factor manipulating the acoustic complexity of the stimulus (piano or pure-tone timbre, the latter having no overtone structure). Within each timbral category RTs were shorter for tonic endings than for subdominant endings. Between the two timbral categories, RTs were shorter for the piano timbre than for the pure-tone timbre. The within-timbre finding suggests that cognitive processes influence expectation for the target event.

All of the studies listed above investigated *relative facilitation*. That is, they demonstrated facilitated processing of the more strongly expected tone/chord (most often the tonic chord) in relation to a less-expected tone/chord. However, these studies did not provide any information about facilitation and inhibition, that is, whether the processing of a tonal center and the development of tonal expectations represent a cost or a benefit. This question was addressed by a pair of studies (Tillmann, Janata, Birk, & Bharucha,

2003; Tillmann, Janata, Birk, & Bharucha, 2008) that adapted the approach of psycholinguistic priming studies by introducing comparisons with neutral baseline contexts. For musical priming, facilitation and inhibition were measured relative to so-called *baseline* chord sequences that did not establish a tonal center (by virtue of jumping about the circle of fifths, Figure 1C, from unrelated chord to unrelated chord). The hypothesis was that processing of tonic endings would be facilitated when a tonal center was established, but it was not known whether processing of subdominant endings would be: (1) facilitated, but to a lesser degree; (2) inhibited compared with sequences that do not establish a tonal center; (3) or no different compared to sequences without a tonal center. As such, stimuli from Tillmann, Janata, Birk, et al. (2003) are chord sequences that end on either the tonic or the subdominant, and are paired with baseline sequences (Supplemental Figures S4A and S4B). The stimulus set in Tillmann et al. (2008) was broadened to include chord sequences with dominant endings to investigate non-musicians' perception of the most fine-grained differences in tonal function (Supplemental Figure S4C). Both studies found that compared with RTs to baseline sequences, responses to tonic (I) endings were facilitated, and responses to subdominant (IV) endings were inhibited. Responses to dominant (V) endings were neither facilitated nor inhibited (Tillmann et al., 2008). In other words, RTs for the standard chord sequences were shortest for tonic endings, followed by dominant endings, and the longest RTs are for subdominant endings. This I, V, IV hierarchy mirrors prior cognitive psychology studies of harmonic function that used explicit relatedness judgments (Bharucha & Krumhansl, 1983; Krumhansl, Bharucha, & Kessler, 1982) and music-theoretic accounts (Rosen, 1972).

COMPUTATIONAL MODELS OF TONAL EXPECTATION

There are relatively few computational models of musical expectancy (and only a handful that are replicable or for which an implementation is available), which is remarkable given that: (1) there is a large body of experimental research on chordal (e.g., Bigand et al., 2003; Krumhansl & Kessler, 1982), melodic (e.g., Krumhansl & Shepard, 1979; Marmel et al., 2010), and temporal (e.g., Boltz, 1989) aspects of expectancy; (2) the topic of musical expectancy has also received considerable attention from music theorists, dating back at least as far as Meyer (1956), especially in relation to the emotional responses that music engenders (for recent reviews see Huron, 2006; Rohrmeier & Koelsch, 2012; Trainor & Zatorre, 2009). The benefit and contribution of computational modeling is to make theories of musical expectancy concrete and testable, to point out potential for ambiguity in textual descriptions, to compare the strengths and weaknesses of different theories, and so to shed light on the underlying neural mechanisms of expectancy.

Computational models in music vary along two primary, but non-orthogonal dimensions (Table 2). One of those dimensions is aligned with the sensory–cognitive continuum notion. That is, a model may seek to extract algorithmically from an input source one or more numeric representations that are considered counterparts to psychological representations that support perceptual and/or cognitive judgments – for example periodicity pitch, chroma vector, and tonal space representations described below. Models vary in the number of representational levels they encompass (*Model description* column in Table 2). Moreover, the algorithms used to extract those representations or transform between them strive to present facsimiles of psychological

or neural mechanisms that accomplish those transformations. The second primary dimension along which models differ is the source of input (*Input* column in Table 2): from musical scores (symbolic) or from the actual audio signals to which experiment participants listen (acoustic).

Both symbolic and acoustic models have been used to model ostensibly equivalent representations, whether sensory or cognitive, but each type of model comes with limitations or challenges for representing what may reasonably be considered relevant information for the listener at different stages of the sensory–cognitive continuum. Specifically, beginning with a symbolic representation, while computationally easier, makes multiple and often problematic assumptions about how an excerpt or piece of music might be perceived. For instance, a note in a chord may be played more softly than the others so that it is barely audible to the listener. In a symbolic representation the relative loudness of notes of a chord typically is not notated and therefore appears the same to a model that operates on symbolic input, whereas models that operate on the audio input are more likely to be influenced by amplitude differences between notes. This difference may prove important for accurately modeling the sensory representation stages, in particular interactions of pitch and timbre. Acoustic models may thus be considered more general, and may avoid the problem of *scriptism* in music perception Cook (1994). However, they come with greater computational and algorithmic demands, thus placing the psychological or neural verisimilitude of each transformational step under greater scrutiny. Scrutiny of the *Input* column in Table 2 indicates that while acoustic and symbolic models of tonal expectancy coexisted between 1987 and 2002, the last decade has been dominated by models that handle symbolic input

only. Here we focus on three models in detail (Bharucha, 1987; Janata et al., 2002; Leman, 2000) with reference to many of the models summarized in Table 2.

The periodicity pitch (PP) model and other pitch models

Leman's (2000) PP model, which is part of the IPEM Toolbox (Leman, Lesaffre, & Tanghe, 2001), takes into account the information emerging from peripheral auditory processing of incoming sound signals⁶. Leman's (2000) aim was to show that a short-term memory process consisting of an autocorrelation mechanism for pitch estimation (Cariani & Delgutte, 1996a, 1996b) coupled with a standard leaky integration model of neural information accumulation could explain tonal probe-tone judgments (Krumhansl & Kessler, 1982), thus presenting a challenge to Krumhansl's (1990a) cognitive account of tonal expectation.

In the first stage (see numbered stages in Figure 2), an auditory nerve image (ANI) is produced from the auditory stimulus. This simulates the cochlea's transduction of an acoustic signal to a neural signal (Van Immerseel & Martens, 1992). Columns of the ANI matrix are samples in time, and rows represent firing patterns of auditory nerve fibers across the different critical bands (see also Janata, 2007). In the second stage, a periodicity pitch image is produced, representing estimated periodicities in the firing patterns. Depicted in Figure 2, the matrix for the periodicity pitch image results from autocorrelation within channels of the auditory nerve image and pooling across channels thus mimicking neural correlates of pitch perception (Cariani & Delgutte, 1996a, 1996b). For melodic stimuli, the smaller the vertical spacing between bands, the higher the fundamental frequency.

In the third stage, a context image (CI) is produced by leaky integration of the periodicity pitch image over time. That is, values in a given matrix column of the CI depend not only on the corresponding matrix column of the periodicity pitch image, but also on a number of preceding columns, controlled by a chosen time constant. A CI calculated using a *global* 4 s time constant will appear more smeared in time than a CI calculated using a *local* 0.1 s time constant. In psychological terms, leaky integration with the *global* time constant represents a simple model of echoic memory.

In the fourth and final stage of the PP model (not shown in Figure 2), Leman (2000) proposed calculating Pearson's r (hereafter correlation coefficient) between corresponding columns of the local and global context images, at a time sample just after the onset of the target event, to measure how well the momentary sensory information matches the echoic memory buffer. Leman (2000) compared these correlation values with probe-tone profiles and found a significant match, thus supporting a sensory account of tonal expectancy that does not rely on the notion of tonal templates stored in long-term memory, and so challenging Krumhansl's (1990a) cognitive account (see Bigand, Delbé, Poulin-Charronnat, Leman, & Tillmann, submitted, for converging data). We consider the periodicity pitch (PP) context images as the representational stage that models the effect of sensory priming on tonal expectations. Other models (Huron & Parncutt, 1993; Parncutt & Bregman, 2000) have been proposed that involve converting notated pitches into a semi-acoustic representation by approximating the overtone structure of each note, but fall short of using the acoustic signal directly.

MUSACT and other cognitive models

The MUSACT model (Bharucha, 1987) was conceived as a cognitive account of tonal priming. Given that it has been used extensively for modeling behavioral data (e.g. Bigand, Madurell, Tillmann, & Pineau, 1999; Bigand et al., 2003; Tillmann, Janata, Birk, et al., 2003; Tillmann et al., 2008), and because its basic architecture was derived using a self-organizing map (SOM) approach in a paper that is widely cited in association with the claim that knowledge of tonal structure is stored in long-term memory in non-musician listeners via implicit learning mechanisms (Tillmann, Bharucha, & Bigand, 2000), we consider the MUSACT model in some detail here. It is a connectionist model with twelve units representing tones, 24 units representing major and minor chords, and twelve representing major keys. Tone units are connected to chords to which they belong, and similarly, chord units are connected to keys to which they belong. A note or chord presented to MUSACT will cause activation of corresponding tone units, which spreads towards chord and key units (bottom-up activation), weighted by the strengths of the connections between the layers (Bharucha, 1987). Activation spreads from key units back towards chord and tone units (top-down activation), and after several reverberation cycles reaches equilibrium (i.e. change in activation is below a given threshold).

When an entire stimulus is presented to MUSACT, “activations due to each chord are accumulated, and their pattern of activation decays over time (according to recency). The global pattern of activation in the units at the end of the sequence represents the influence of the overall context. The activation levels are interpreted as levels of expectation for subsequent events ... and [as potential] predict[ors of] harmonic priming” (Bigand et al., 2003). While explanations of MUSACT are detailed enough to be

replicable, and an implementation exists (Bigand et al., 1999), it is restricted to symbolic input, it does not represent some chords (e.g., dominant seventh or diminished chords), and it does not represent minor keys. For these reasons, we do not re-implement MUSACT for our simulations of RT data, but do refer to its successes and failures in previously reported simulations.

Both MUSACT (Bharucha, 1987) and the tonal space (TS) model (Janata et al., 2002), explained in the next section, capture some probabilistic aspects of music implicitly. For instance, contrast the chord progression $G \flat, A \flat, D \flat$, which would have a relatively high probability of occurrence in a piece in $D \flat$ major, with a low probability sequence, such as $G \flat, A, G \flat$. One can imagine the location of maximum intensity on the toroidal surface shifting from $G \flat$ to $A \flat$ to $D \flat$, and then measuring the mean distance between successive maxima. The mean distance would be small for a high probability sequence such as $G \flat, A \flat, D \flat$, because the labels are close on the torus, compared with a large mean distance for a low probability sequence, such as $G \flat, A, G \flat$, where the labels are further apart (Figure 1D and also Janata, 2007). Bharucha (1987) claims (but did not test) that models such as MUSACT capture probabilistic aspects “by incrementally altering the connection strengths so as to bring the expectations generated by the network in line with the transition probabilities of the music” (p. 26).

More explicitly probabilistic than either the TS model or MUSACT are Temperley’s (2004, 2007) Bayesian model for the likelihood of a pitch-class set in a piece (see also Kim, Kim, & Chung, 2011), Conklin and Witten’s (1995) n -gram model for predicting the next note of a melody, which has been revised and assessed in experimental settings (Pearce et al., 2010; Pearce & Wiggins, 2006), and Toivianen and

Krumhansl's (2003) tone transition model, which unlike Pearce et al.'s (2010) model can handle polyphonic input. Outputs from each of these probabilistic models could be construed as predictors of RTs for tonal priming stimuli, but at present all are restricted to symbolic input and few implementations have been made available. It is worth noting that Toivainen and Krumhansl (2003) compared a model based on pitch class distribution (zeroth order transition probabilities) with the tone transition model (first order), and found the models performed equally well as explanatory variables for listeners' judgments of fit. A third model containing both variables had the most explanatory power, suggesting that the variables accounted for judgments in different ways, and that it was worthwhile to test a linear combination.

Huron (2006), von Hippel and Huron (2000), Aarden (2003), Schellenberg (1997), and Narmour (1990) have addressed more abstract probabilistic matters, such as the likelihood of a descending melodic interval following an ascending interval. However, it is not immediately obvious how their findings could be applied to an entire melody, passage of polyphony, or changes related to tonal function of tones or chords. The same restricting observation applies to models of tonal music that focus on calculating the distance between chord pairs (Callender, Quinn, & Tymoczko, 2008; Milne, Sethares, Laney, & Sharp, 2011; Woolhouse, 2009). Research on tonal tension (Farbood, 2012; Lerdahl & Krumhansl, 2007) could also be considered relevant to modeling tonal priming data. Implementations for this research are not yet available, however, and Lerdahl and Krumhansl's (2007) method is not fully algorithmic.

The tonal space (TS) model

Janata et al.'s (2002) TS model is an extension to Leman's (2000) PP model: it projects the output from the final stage of Leman's model implemented in the IPEM Toolbox to the surface of a torus, and is therefore closely related to a model that served as a precursor to the IPEM Toolbox (Leman, 1995; Leman & Carreras, 1997).⁷ The repeated exposure to, and memory for, pitch distribution information that a human listener experiences was simulated with a well-established proxy for implicit learning training: a self-organizing map (SOM) algorithm (Kohonen, 1995). Given the emphasis on long-term memory for tonal hierarchy relationships, the TS model can be considered a more *cognitive* representation along the sensory–cognitive continuum than the PP model.

In Janata et al.'s (2002) TS model, the map was trained using CI vectors (PP images integrated with a 2 s time constant) extracted from a melody purposefully composed to modulate through all 24 major and minor keys (Janata, Birk, Tillmann, & Bharucha, 2003; Janata et al., 2002) over the course of approximately 8 minutes. At the start of training, the SOM consists of units distributed across the surface of a torus (as in Figure 1D), each with a corresponding and random-valued weight vector associating it with the vector of periodicities in the CI. On each iteration of training, the Euclidean distance between a randomly selected (without replacement) CI vector and each weight vector is computed. The unit whose weight vector is closest to the CI vector is called the best matching unit (BMU). The weight vectors are modified using a weight update rule to more closely match the CI vector, with the magnitude of change increasing with proximity to the BMU. After training, BMUs for input vectors that represent each major and minor key appear as two helical orbits of the torus, corresponding to separate but

interlaced circles of fifths for major and minor keys, as indicated by labels on the unfolded tori in Figure 2.⁸

Toiviainen and Krumhansl (2003) also investigated tonality induction with an SOM, but unlike Janata et al. (2002), Toiviainen and Krumhansl's (2003) map is limited to symbolic input (i.e., MIDI information) rather than real audio information, and is trained with shifted versions of a canonical probe-tone profile rather than actual music. Generally, it is preferable to train a model with actual music (e.g. Leman & Carreras, 1997), even though using idealized and observed data may lead to differences, such as slight displacements of the locations of key labels from their theoretical locations, e.g. the closer apposition of the C major label to E minor on the toroidal surface in Figure 2 than would be expected by theory.

The output of the fourth stage of Janata et al.'s (2002) TS model is an $m \times n \times p$ array, depicted in Figure 2, where p is the number of time samples, and at each time sample an $m \times n$ matrix represents the instantaneous activation on the torus (tonal space). As TS activations are calculated from CIs, there remains a choice of time constant for the length of echoic memory. We calculate the correlation coefficient pointwise between local ($t = 0.1$ s) and global ($t = 4$ s) TS matrices. This is analogous to the calculations performed in the PP model, but applied to the TS activations.

As mentioned above, Krumhansl and Kessler (1982) posited the existence of a torus-shaped schema for probe-tone profiles and distance relationships among major and minor keys, based on behavioral data and multidimensional scaling. To the extent that the toroidal topography is learned, it represents a more cognitive stage of abstraction. Janata et al. (Janata, 2005; Janata et al., 2002) presented fMRI evidence that a number of brain

areas tracked the time-varying activation of tonal space while participants listened to a modulating melody and performed timbre and tonality deviance detection tasks.

Specifically, the analyses identified brain areas in which the time-varying BOLD signal was correlated with the pattern of movement on the toroidal surface governed by the implied harmonic structure of the melody, a phenomenon referred to as *tonality tracking*.

The researchers found that discrete locations within the rostral medial prefrontal cortex responded most strongly to different key regions in tonal space, suggesting the presence of a topography of keys in this region of the brain (though the organization in the brain was not obviously that of a torus). Within listeners, the exact topography changed from session to session suggesting the presence of a more complex interaction between a representation of tonal space and the momentary experiential state of the listener.

Applying the same analysis to fMRI data obtained when participants listened to original audio excerpts of popular memory-evoking music, Janata (2009) found tonality tracking in a distributed set of brain areas, including medial prefrontal and ventrolateral prefrontal regions. Together with the results of Krumhansl and Kessler (1982), the fMRI results suggest that the tonal space model provides a level of psychological representation that can link basic music perception to other cognitive and affective processes in the mind of the listener.

The chroma vector (CV) representation

Above, we characterized the PP and TS models as sensory and (at least partially) cognitive representations, respectively. There is a third representation that is very relevant for models and discussions of the sensory/cognitive transition: the chroma vector (CV), a vector consisting of a weighting for each of the 12 pitch classes $\{C, C\#, D, \dots\}$.

Thus a chroma vector is a compact representation of the distribution of pitch class information across a time segment of music. Given evidence that anterior regions of the auditory cortex respond more strongly to changes in pitch chroma than pitch height (Warren, Uppenkamp, Patterson, & Griffiths, 2003), it is reasonable to postulate the accumulation of pitch chroma information in short term memory to provide chroma vector estimates.

Chroma vectors are also referred to as pitch class profiles, and as discussed above, they are analogous to key profiles. This representation has been at the core of discussions of tonal perception and music information retrieval approaches to describing tonal structure. Tonal hierarchies (key profiles) are chroma vectors that represent the psychological goodness-of-fit of each pitch class into an established key (Krumhansl, 1990a). It has been argued that, “relations between keys and between individual chords and keys are all mediated through the internalized hierarchy of tonal functions at the level of individual tones” (Krumhansl & Kessler, 1982, pg. 366). Not surprisingly, chroma vectors are the most common basis for key-finding and tonality algorithms (Chai & Vercoe, 2005; Gomez, 2006; Krumhansl, 1990a; Sapp, 2005; Serrà, Gomez, Herrera, & Serra, 2008; Temperley, 2007; Temperley & Marvin, 2008; Toiviainen & Krumhansl, 2003), and arriving at CVs is the objective of various methods for analyzing musical audio (Bello & Pickens, 2005; Chai & Vercoe, 2005; Gomez, 2006; Lartillot, Toiviainen, & Eerola, 2008; Serrà et al., 2008).

Because of the ubiquity of the CV representation, both in the psychology of music and MIR, we believe it important to examine how well this representation is able to model the RT data from tonal priming studies. Thus we incorporated it into our auditory

model, as a representation on an indirect route from PP to TS (stages 5-7 in Figure 2, and Appendix A). In terms of a sensory–cognitive continuum, we consider CVs to be situated somewhere between PP and TS variables because, on the one hand, they can represent accumulated pitch probability distributions across some integration window, i.e. they can be maintained in echoic or short term memory and need not be learned, whereas on the other hand, they can represent the key profiles that have been used to characterize listeners' long-term tonal knowledge (Krumhansl, 1990a).

Comparison of periodicity pitch, chroma vector, and tonal space representations

Prior to modeling the full collection of RT data from the seven tonal priming experiments described above, we provide an illustrative example to demonstrate that the PP, CV, and TS representations we will use are indicative of different stages of processing along a sensory–cognitive continuum. Figure 3 shows the outputs of the three representational stages in response to four versions of one melody (Marmel et al., 2010). Sensory (timbral) and cognitive (tonal function) factors were crossed to create the four versions. *Timbre* is something of a catchall term in music theory and music psychology, referring to the qualities of a sound (and properties of its overtone structure) that are not captured by the concepts of pitch, loudness, duration, or spatial position. Two instruments such as the flute and oboe can play the same pitch, but sound distinctive due to differing overtone structures, and so are said to have different timbres. For the timbral manipulation, two versions used a dull piano timbre (solid lines) and two used pure tones (dashed lines). For the tonal manipulation, two versions ended on the tonic tone (dark gray lines) and, by alteration of a single pitch class in the context, the key of the remainder of the melody was changed such that the terminal tone was the subdominant

(light gray lines). Each of the four traces indicates the time-varying correlation between the local ($t = 0.1$ s) and global ($t = 4$ s) images for that representation type.

Figure 3A shows output from Leman's (2000) PP model. It is evident that the primary grouping is by the sensory factor, timbre, seen as greater overlap of line style (timbre) than line color (tonal function). Figure 3B shows the output of the leaky-integrated chroma vector representation. As in the periodicity pitch representation, there is little differentiation of the different tonal contexts for the pure tone stimuli, and somewhat greater differentiation for the melodies rendered with piano tones. Often this differentiation occurs after the appearance of the altered pitch class (indicated by the arrows). For example, following the second altered pitch class in Figure 3B, there are two subsequent negative-going deflections that differentiate between Piano IV and Piano I — following labels (V, I) and (VI, II). The differences can be explained in terms of the altered pitch class persisting in the global ($t = 4$ s) context image. However, this differentiation does not persist clearly until the end of the melody. By contrast, in Figure 3C which shows the output from the TS model, it is clear that the grouping is primarily by tonal function (though small differences between pure and piano tone melodies remain). This example illustrates that the projection from the PP representation to the TS representation marks an important transition from a *sensory* to *cognitive* representation.

SIMULATION OF RT DATA

In the preceding sections, we presented the theoretical basis for considering the perception of tonal structure along a sensory–cognitive continuum and summarized previous work describing how an audio signal is transformed into representations within each of these theoretical spaces. In order to determine the relative contributions of each

of these spaces to perceptual judgments about target musical events, it is necessary to derive measurements of the activation of each of these spaces that can serve as explanatory variables in a statistical model. Selection of the appropriate variables is driven by two principles. First, a measurement/variable should, when possible, have a basis in theory. Second, when multiple theoretically plausible variables or combinations thereof are available, a model-fitting approach is taken to identify the best set of variables. Here, we used standard multiple regression techniques in our model-fitting effort.⁹

Model variables

For our simulation of RT data, explanatory variables were derived for each of the PP, CV, and TS representational spaces. Details regarding the different dimensions on which a total of 24 explanatory variables varied and how the variables were calculated are provided in the Supplemental Material and summarized in Supplemental Table S1. Here, we provide an overview of the theoretic motivation for the variables and illustrate the calculation of the variable type that was most important for modeling the RT data.

One class of explanatory variables was derived by correlating the *local* and *global* activations of a space, where local and global refer to the activations obtained by integrating the activity with short (0.1 s) and long (4 s) time constants, respectively. These time constants were based on previous work (Delbé, 2009; Janata, 2007; Leman, 2000). This measure estimates how well the activation caused by the current event fits with the activation integrated over the preceding context, and is therefore a single-valued estimate of moment-to-moment contextual congruency (Delbé, 2009; Leman, 2000). We contrasted this measure with one that considers the activation at only a single integration timescale, e.g. 0.1 s, such as the *maximum value* of the activation within the space.

A second class of variables was based on the average *absolute* value of a measure, such as the local/global correlation, within some time window following the onset of the target event (either an early window from 0 – 200 ms, or a late window from 201 – 600 ms). We note that Leman (2000) and Delbé's (2009) variables were all *absolute*, making use of post-target values only.

A third class of variables captured the *relative* value of the measure, obtained by subtracting its pre-target value from its post-target value. Several theoretical considerations underlie our use of a measure of relative change. Within our model, the change from pre- to post- values in the correlation of local/global contexts reflects the degree to which the present event diverges from or converges toward the global activation pattern. In the context of music, such local-change estimates may contribute to implication (divergence) and realization (convergence) processes in melody perception (Narmour, 1990) or the buildup and release of tonal tension. More generally, the evidence from studies of reward processing indicates that the relative, rather than absolute, value of a reward drives activity within the brain's reward processing circuitry comprising striatal and orbitofrontal areas (Elliott, Agnew, & Deakin, 2008). Given the putative role that these brain regions play in supporting musical expectation processes, such as in the anticipation of reward ("chill" moments) when listening to musical stimuli (Salimpoor, Benovoy, Larcher, Dagher, & Zatorre, 2011) or the moment-to-moment affective evaluation of musical stimuli about which a purchasing decision must be made (Salimpoor et al., 2013), the pattern of moment-to-moment context violating/reinforcing changes may be as important, if not more important, than the absolute degree of how well local events fit with the global context.

The variable classes described above allowed us to assess the degree to which relative values, absolute values, and patterns of change within the PP, CV, and TS spaces underlie listeners' performance in tonal priming tasks. Here we provide an example, in relation to Figure 4, of the calculation of one PP variable that we will label, x_{PP} .

Analogous variables x_{CV} and x_{TS} were calculated from chroma vector (CV) and tonal space (TS) matrices, but the latter in particular, being an $m \times n \times p$ array, is more difficult to depict. The PP variable we discuss is inspired by prior work (Delbé, 2009; Leman, 2000), though it is the first use of comparable variables from the CV and TS spaces.

Figure 4A shows an example stimulus from Experiment 1 (Tillmann, Janata, & Bharucha, 2003). It was processed using Leman's (2000) periodicity pitch (PP) model (explained above), where column vectors of the output matrix comprise periodicity pitch distribution estimates at successive time points (Figure 4B). The amount of context/information retained from previous vectors at the current time point is dependent on the time constant. The difference between local (0.1 s) and global (4 s) time constants is evident upon comparing Figures 4B and 4C, which contain integrated PP images (or context images) for the 0.1 and 4 s time constants, respectively. The local context image exhibits relatively crisp transitions from one chord to the next in comparison with the global context image, where there is more leakage of values from one time point to the next, simulating a memory buffer.

The x_{PP} variable is derived from the correlation of local and global context images within specific time windows (Figure 4D). We hypothesize that in the region of the target event (its onset is indicated by the vertical solid line at 3.5 s), the correlation time course is a useful predictor of RTs. For example in Figure 4A, there is an abrupt change from a

C major tonality to a B major chord at the target onset. Such a transition is extremely unexpected because a B major chord does not occur naturally within the key of C major. This transition is associated with a sudden decrease in correlation between short (local) and long (global) context images (Figure 4D). In a tonal priming context, the B major target chord engenders a longer RT, compared to a stimulus that remains in C major at the target onset (dashed line in Figure 4D). For this stimulus, $x_{PP} = -.259$, which is the difference between the mean correlation .469 in the 0-200 ms post-target window, and the mean correlation .728 in the 100 ms pre-target window. The x_{PP} variable is referred to as *relative* because it is derived from the difference between pre- and post-target values.

The explanatory variables described above (and in the Supplemental Material) were derived from the audio files for 303 stimuli from the seven tonal priming experiments listed in Table 1. We constructed linear regression models to explain the variance in RTs across this set of stimuli. The mean RT across participants for each stimulus, minus the overall within-experiment mean RT, constituted the dependent variable.¹⁰

We present our multiple regression analysis results in three stages. First, we examine the ability of a simple model, containing only the x_{PP} , x_{CV} , and x_{TS} variables described above, to fit the data. Second, we compare the relative explanatory power of the PP, CV, and TS spaces utilizing a full set of variables obtained for those spaces. Third, we identify an optimal model, given our variables, using stepwise regression, and we compare the ability of this model to explain differences among stimulus categories with the differences among stimulus categories that were established with behavioral data.

Regression using a simple model

Regression of RT on x_{PP} was significant ($F(1, 301) = 81.87, p < .001, s = 97.91, R^2 = .21$). Regression of RT on x_{TS} was also significant ($F(1, 301) = 85.12, p < .001$), with a smaller error standard deviation ($s = 97.50$) and larger value for proportion of variance explained ($R^2 = .22$) than for x_{PP} . Notably, a model including both x_{TS} and x_{PP} explained significantly more of the variance in RT ($F(2, 300) = 53.64, p < .001, s = 94.93, R^2 = .26$) than did models containing either variable in isolation. This finding suggests that both sensory (x_{PP}) and cognitive (x_{TS}) variables are necessary to account for RT patterns in tonal priming experiments. The x_{CV} variable was significant in isolation ($F(1, 301) = 43.31, p < .001, s = 103.25, R^2 = .13$), but explained considerably less variance than either x_{TS} or x_{PP} . As such, it did not merit inclusion in a model that already contained x_{TS} and x_{PP} variables, whereas the x_{PP} variable did merit inclusion in a model that already contained x_{TS} and x_{CV} ($F(3, 299) = 35.65, p < .001, s = 95.09, R^2 = .26$).

Our initial findings can be summarized as follows: (1) the tonal space variable x_{TS} is a stronger predictor for RT than the periodicity pitch variable x_{PP} ; (2) considered together, TS and PP variables lead to a significantly stronger model for RT than either variable considered in isolation; (3) the chroma vector variable x_{CV} is relatively less useful than either x_{PP} or x_{TS} for predicting RTs in tonal priming experiments.

Relative explanatory power of the PP, CV, and TS spaces

We conducted an ANOVA using a subset of 17 of the 24 explanatory variables.¹¹ The full model explained 33% of the total variance, $F(17, 285) = 8.14, p < .001, s = 93.10, R^2 = .33$. We grouped the ANOVA according to representational space (PP, CV, and TS) to determine whether the group of explanatory variables based on TS activations

explained a greater proportion of variation in the RT data than the group based on PP or CV. Table 3 shows the ANOVA results, arranged in descending order of mean square. As a group, the TS variables explained more of the variance ($p < .001$, Cohen's $f^2 = .10$) than did the PP variables ($p < .001$, $f^2 = .08$), which in turn explained more than the CV variables ($p < .05$, $f^2 = .04$).

Stepwise regression model

To obtain a final model we used stepwise selection, a standard variable selection technique that begins by comparing the predictive power of each of the 24 explanatory variables in isolation (Supplemental Table S2). The strongest predictor is included in the stepwise model, and entry/elimination of subsequent variables into/from the stepwise model is merited only if they account for significantly more of the variance than was accounted for previously. Appendix B provides a summary of the entries into and eliminations from the model at each step.

The model that resulted from stepwise selection is

$$RT = 110.77 - 97.94 \cdot x_{TS} - 245.20 \cdot x_{PP} + 0.13 \cdot y_{PP} - 4.09 \cdot y_{CV}, \quad (1)$$

with test statistic $F(4, 298) = 32.00$, $p < .001$, $s = 92.83$, and $R^2 = .30$. Thus the overall proportion of variance explained by this model was greater than that achieved considering only the x_{PP} , x_{CV} , and x_{TS} variables, and approached the R^2 of the 17-variable model. The coefficients for the first two variables x_{TS} and x_{PP} in (1) were negative as expected: as correlations between the local and global images decreased, RTs increased. The third variable, y_{PP} , was a *maximum value* variable and had a positive coefficient. A

post-hoc examination to determine whether certain variables fit certain datasets better than others, revealed an interpretation of y_{PP} with respect to Experiment 5 (Marmel et al., 2010), in which stimuli with a piano timbre were observed to have shorter RTs on average than stimuli with a pure-tone timbre. In the case of a piano timbre, energy in the PP vector decays quickly after the onset and will therefore have a smaller maximum value than for pure tones, where the energy is sustained and more concentrated after the onset. Peaks and troughs in PP vectors of pure-tone stimuli are more extreme than in PP vectors of piano stimuli, even after normalizing for overall energy. Behavior of the y_{PP} variable is in the same direction as the RT trend, hence the positive coefficient.

To summarize, the stepwise model (1) substantiates our first finding concerning the strength of the TS representation, because the TS variable x_{TS} is the strongest predictor overall, and so the first to enter the model. It substantiates our second finding about needing both TS and PP variables, as the PP variable x_{PP} is second to enter the model. Finally, it substantiates our third finding concerning the weaker predictive power of CV variables, as these are not as prominent in the stepwise model as TS or PP variables (first three variables). Together, the ANOVA and stepwise regression results demonstrate that our initial findings, based on only a single variable from each of the representational spaces, are robust to different analytic approaches and variable pools (see Supplemental Material for more details).

Using the stepwise model to predict RT

To test whether the stepwise model could adequately explain RT differences between stimulus categories across the different experiments, we first fitted an RT for each experimental stimulus using the coefficients obtained in the stepwise model (1), and

then performed *t*-tests to assess differences between categories. Figure 5 shows observed and fitted RTs calculated from the stepwise model in (1) across different conditions within each experiment. The stepwise model accounts well for RT trends in Experiments 1 (Tillmann, Janata, & Bharucha, 2003), 3 (Marmel et al., 2008), 4 (Marmel & Tillmann, 2009), and 6 (Tillmann, Janata, Birk, et al., 2003), but less well for Experiments 2 (Bigand et al., 2003), 5 (Marmel et al., 2010), and 7 (Tillmann et al., 2008).

To quantify the success of our simulations, we assessed the result of simulated RT *t*-tests for those between-category comparisons where the original observed difference in RTs was statistically significant. If the corresponding simulated difference was also significant in the same direction as the observed difference, we counted this as a modeling success, and as a failure otherwise. For example, in Experiment 1 we compared the observed RTs for category Rel B (standing for a stimulus in the key of B major that ends on a B major tonic chord, i.e., is related to the established key) with those of Un B (standing for a stimulus in the key of B major that ends on C major chord, i.e., is unrelated to the established key) using a paired *t*-test and observed a significant difference ($p < .001$). The difference between the fitted RTs for stimuli from the Rel B and Un B categories was also statistically significant in the same direction ($p < .001$), and therefore counted as a modeling success. For Experiments 6 and 7 (Tillmann, Janata, Birk, et al., 2003; Tillmann et al., 2008), in which there was a greater number of stimuli in baseline categories, a two-sample *t*-test was used instead of a paired *t*-test. Overall, there were 20 observed-simulated comparisons to be made over the seven experiments, 80% (16/20) of which were modeling successes ($p < .05$).¹² The results are summarized in Table 4, with a check mark in the sixth column indicating complete success of the model

in simulating the results of the study, i.e. successfully simulating *all* relevant comparisons in the study.

Discussion

Multiple representational spaces on a sensory/cognitive continuum

Given the historical debate regarding sensory versus cognitive determinants of tonal expectancies, our primary goal was to determine whether variables calculated from the tonal space (TS) representation – further along the sensory/cognitive continuum – contributed more to explaining RTs than did variables from the more sensory periodicity pitch (PP) representation. We found that this was the case, but TS variables alone did not model the RT data as well as a linear combination of PP and TS variables, suggesting that listener judgments are influenced by a combination of sensory and cognitive properties. Evidence for this was that a stepwise model (1) containing PP and TS variables explained a greater proportion of variation in RTs ($R^2 = .30$) than any of the proposed TS variables did individually (in Supplemental Table S2, the maximum value of $R^2 = .22$). As an additional investigation, stepwise selection was also run with variable pools consisting of: (1) PP variables only; (2) CV variables only; (3) TS variables only; (4) PP variables only for the first step, and then all variables (PP, CV, TS) for subsequent steps. None of these selection processes resulted in a higher value of R^2 than that for the stepwise model in equation (1). Together, these observations underscore that tonal expectations and judgments about musical events are shaped by multiple stimulus features, requiring the combination of multiple sensory and cognitive representational stages within a unified model.

We also extended the work of Leman (2000) and Delbé (2009) by assessing the contributions variables based on *relative* and *absolute* activation values. We found that variables based on correlations of local and global activations outperformed variables based on maximum activation values, affirming the importance of modeling contextual congruency of the present moment. Nevertheless, absolute activation magnitude in the PP representation did capture important timbral differences that served to explain a systematic difference in RTs to pure tone and piano timbre stimuli. We also found evidence for the importance of another relational attribute, namely the change in contextual congruency caused by a target event. Both of the leading terms in the stepwise regression equation were measures of the change in local/global correlation caused by the target event rather than the overall degree of local/global correlation.

Chroma vectors and hierarchical tonal structures

As discussed in the section *Tonal Structure in Western Tonal Music*, a representation based on chroma vectors or pitch class profiles is attractive given its compactness and ubiquity in the psychological and MIR literature. We found, however, that measures of chroma vector activations were not particularly successful in entering into either the stepwise models or in accounting for more of the behavioral variance than did the TS measures. We conclude that CVs, while functioning as an adequate computational intermediate, are not an optimal representational stage for modeling tonal expectations, at least when measures based on correlations of local and global CV profiles are used (see Appendix A for further discussion of the CV representation in relation to the PP and TS representations).

Our conclusion regarding the relative inefficacy of the CV representation may seem controversial in light of a hierarchical framework of tonal structure – in which individual tones lay a foundation atop of which chord functions and keys are successively constructed (Bharucha & Krumhansl, 1983; Bharucha, 1987; Krumhansl et al., 1982; Krumhansl & Kessler, 1982) – because it shifts the relevant representational level for explaining behavior from that of the tonal hierarchy embodied in a key profile to the representational space of relationships among keys. Additionally, in contrast to the aforementioned studies, our model does not explicitly include a level of chord information as an intermediate step to establishing a sense of key, although chord functions associated with a key may be distributed within key regions on the toroidal surface (Janata, 2007; Lerdahl, 2001). The parsimony of a direct projection from either the PP or CV space directly to the TS space, in contrast to a projection via a chordal intermediate, arises from the ability to represent arbitrary and complex harmonies without the need to attach a symbol, i.e., create a representational category, for every collection of tones that might co-occur. Moreover, the model accommodates melodic material also.

Simulation failures

Fitted RTs of tonic categories were involved in all four of the simulation failures. In Figures 5B and 5G, for instance, the observed RT for the tonic conditions is significantly faster than the subdominant and dominant conditions, but this was not the case for the fitted RT. This pattern of failure indicates that there was an element of listeners' expectations that is inadequately modeled by simply considering the fit of a harmonic event with the established tonal context, and that there exists at least one additional process that shapes listeners' expectations for a terminal event. The most

parsimonious explanation for this result is that listeners are influenced, in part, by the degree of closure that the terminal event imparts (Aarden, 2003). Because the sequences in these experiments were of fixed length and because judgments were made for terminal events, it was easy to guide expectations in time toward the closing event. More importantly, a phrase ending with an authentic cadence (V – I, Figure 1B) imparts a stronger sense of closure than phrases ending with a half cadence (ii – V) or ending I – IV. Even though the latter two transitions can be regarded as high probability transitions (Huron, 2006; Piston, 1948), it is nonetheless uncommon for those transitions to comprise the penultimate and final events in a phrase and/or piece. In the next section we elaborate our model to incorporate the concept of closure.

SYNTAX CONSIDERED

As noted above, using the stepwise model (1) we could not account for faster observed RTs in response to the tonic relative to other tonal functions for some of the experiments (Figures 5B and 5G). We propose that the most parsimonious explanation for this discrepancy is that listeners are influenced, in part, by the degree of closure that the terminal event imparts. Although the closure variable that we are about to define can be regarded as a syntactic variable, the present purpose is not to develop a full model of tonal syntax. A full model might be used to calculate expectation values at arbitrary transitions within a musical phrase. Here we focus on a restricted but ubiquitous case in which a listener expects musical phrases and pieces to end with a resolution to the tonic. Nonetheless, the addition of a closure variable marks an additional representational stage within the domain of possible *cognitive* representations that might guide listener expectations (Bigand & Parncutt, 1999; Lerdahl & Krumhansl, 2007; Tillmann et al.,

2008). The need to do so is suggested by experimental results, indicating that syntactic structure providing closure on a local level, e.g. a cadence, is a stronger determinant of perceived tension than is the global harmonic structure (Bigand & Parncutt, 1999).

Neuroimaging evidence for processing of tonal violations

A large number of electrophysiological and functional neuroimaging studies have also investigated the processing of tonal structure, primarily in relation to the processing of structure of language. One can and should question whether syntax in language and music are at all comparable. Indubitably there are analogies to be drawn between words forming sentences and notes forming melodies (or chords forming harmonic sequences), but as Patel (2012) observes in a recent review, “the syntactic architecture of these sequences [in music] differs from linguistic syntax in a number of ways” (p. 207). For example, there is no clear linguistic analogy for polyphony or counterpoint in music, which are terms for describing simultaneous horizontal melodic lines that combine vertically to imply various harmonies. For more examples of the similarities and differences between language and music, see Jackendoff and Lerdahl (2006).

While delving into the similarities and differences of syntax in music and language is beyond the scope of this article, we nonetheless offer a brief review of the neuroimaging literature in which parallels between music and language have been examined, as these studies provide additional context for understanding why tonal expectations are commonly viewed in terms of cognitive schemata rather than as a product of short-term memory processes. Patel, Gibson, Ratner, Besson, and Holcomb (1998) manipulated the identity of a target chord occurring partway through a short sequence, so that it was in key, from a nearby key, or from a distant key. The authors

observed two ERP markers of violations caused by targets from distant (unrelated) keys: a large positivity distributed across posterior sites on the scalp (P600) and a right-lateralized anterior negativity (RATN). The P600 component elicited by linguistic and musical violations had the same properties, whereas the RATN was specific to music.

When investigating the effect of placing a deviant Neapolitan sixth chord at different locations in an otherwise standard chord progression, Koelsch et al. (2000) also found a right-lateralized anterior negativity, which they termed an early right-anterior negativity (ERAN). A Neapolitan sixth chord is a triad based on the flattened second degree of the major scale. In C major, it consists of the pitches D \flat , F, A \flat , with F most often in the bass (first inversion). It contains two chromatic tones (D \flat and A \flat), and so is one of the most coarse tonal distinctions that can be constructed with triads. They interpreted the ERAN as a musical counterpart to the early left-anterior negativity (ELAN) that is commonly observed in response to syntactic violations in language. Koelsch (2009) and Koelsch et al. (2001) differentiated between the ERAN and a more basic sensory auditory mismatch response called the mismatch negativity (MMN), which is an anterior negativity with peak latency 150-180 ms observed in connection with a physically deviant event, such as an out of tune note or unusual timbre. The ERAN, often bilaterally distributed, has become one of the most studied phenomena in the cognitive neuroscience of music, and has been used to make the case for expectancy formation via cognitive mechanisms rather than short-term sensory templates (Koelsch, 2009).

Studies using functional neuroimaging methods, including functional magnetic resonance imaging (fMRI) and magnetoencephalography (MEG), have suggested that the ventrolateral prefrontal cortex (specifically Broca's Area and its right hemisphere

homologue) is an important site in the processing of violations of tonal expectations. These include fMRI blood-oxygen-level-dependent (BOLD) responses and source-modeling of MEG responses to Neapolitan sixth chords (Koelsch et al., 2005; Maess et al., 2001), responses to tone clusters and sudden modulations¹³ at the end of brief chord sequences (Koelsch et al., 2002), and BOLD responses to contextually unrelated target chords in a tonal priming paradigm involving speeded consonance/dissonance judgments about target chords (Tillmann, Janata, & Bharucha, 2003).

Taken together, the neuroscientific findings reviewed above raise the possibility of a shared neural resource for processing syntactic relationships in music and language (Patel, 2003). Harmonic and linguistic violations mutually influence processing in the other modality when musical and linguistic materials are presented concurrently, using either manipulations of syntactic (Hoch, Poulin-Charronnat, & Tillmann, 2011; Koelsch, Jentschke, Sammler, & Mietchen, 2007; Perruchet & Poulin-Charronnat, 2013; Slevc, Rosenberg, & Patel, 2009) or semantic structures of language material (Hoch et al., 2011; Perruchet & Poulin-Charronnat, 2013; Poulin-Charronnat, Bigand, Madurell, & Peerean, 2005; Steinbeis & Koelsch, 2008). The influencing of cognitive operations in the language domain by the concurrent processing of tonal structure has also been interpreted in support of a view that tonal expectations arise from activation of tonal schema stored in long-term memory, and would thereby appear to support cognitive accounts over sensory accounts of tonal priming.

It is important to note that while most of the cognitive neuroscientific research on the processing of contextually anomalous harmonies has been discussed in terms of “musical syntax” (Koelsch et al., 2000; Patel, 2003), there has been relatively little

formal dissociation between (a) the expectedness of a harmonic event given the overall key (tonal center) primed by the preceding context and (b) the expectedness of the event given the specific sequence of preceding chords. While the magnitudes of ERP responses have been illustrated to depend on the tonal function or sequential position of a target (Janata, 1995; Koelsch et al., 2000; Leino et al., 2007; Poulin-Charronnat, Bigand, & Koelsch, 2006), the varying magnitude of the responses has been interpreted in terms of expectedness relative to the tonal center or build-up of the tonal context, rather than the stricter definition of syntax that would pertain to the expectedness of sequence items defined in terms of their tonal functions and the probability of encountering a specific sequence of tonal functions. Both types of tonal expectation can be regarded as cognitive processes. In this article we have focused on modeling the contribution of expectations generated by a primed tonal context, rather than the specific syntactic structure of tonal sequences (Rohrmeier, 2011), as historically, primed tonal contexts have been the dominant approach in the field.

The case for modeling closure as a specific syntactic process

The principle of resolving to the tonic is a cornerstone of Western tonal music, and influential music analytic methods are framed in terms of the musical prolongations (elaborations) that intervene between the opening tonal center of a piece and the return to that tonal center (Lerdahl, 2001; Schenker, 1973). Results are mixed regarding the extent to which listeners are aware of closure of large-scale tonal structure. Cook (1987) found that ratings of “coherence” and “sense of completeness” in Classical and Romantic period pieces did not depend on whether or not large terminal sections of the pieces had been transposed to different keys, except for the shortest duration tested (30 sec), showing a

limited sensitivity to the large-scale tonal structure. Marvin and Brinkman (1999), on the other hand, found that listeners performed better than chance when asked if the closing and opening keys were the same. It is possible that differing procedures – indirect questions about coherence versus direct questions about opening/closing keys, and the specific implementations of the modulations between keys – are responsible for the apparently contradictory results. There is also evidence that the prolongational structure of a piece underlies the tension/relaxation patterns that shape a listener's experience (Lerdahl, 2001; Lerdahl & Krumhansl, 2007) and that cadential structures (typical chord sequences that serve to establish a sense of closure) play a powerful role in shaping expectation processes on a local level (Bigand & Parncutt, 1999).

As in language, hierarchical structures are important to music (Lerdahl & Jackendoff, 1983), and phrase endings are salient structural markers as shown by both behavioral (Bigand & Parncutt, 1999; Stoffer, 1985) and electrophysiological measures (Hantz, Kreilick, Kananen, & Swartz, 1997; Knosche et al., 2005). Although multiple musical properties can contribute to a sense of closure, a harmonic progression from the subdominant (IV) to dominant (V) to tonic (I) is regarded as a major contributor (Lerdahl, 2001). Thus, it suffices for our purposes to consider the concept of closure at a phrasal level, specifically, the terminal phrase of a movement or composition.

Quantifying closure

We considered the following to be important aspects for a sense of closure: (1) the final event should fit well into the global context, i.e., the correlation of local ($t = 0.1$ s) and global ($t = 4$ s) TS activations should be high, and (2) the local/global correlation of the penultimate event should be lower than, or possibly as high as, that for the final event.

Accordingly, we calculated local/global correlations for the final event (r_n) and penultimate event (r_{n-1}) using TS activations at the 0.1 and 4 s time constants, averaged over a 100-300 ms post-event window. When an excerpt ends on the tonic, the mean correlation for the final event r_n should be close to 1, as there will be a high level of agreement between the local ($t = 0.1$ s) and global ($t = 4$ s) tonal space activations. The value of r_{n-1} may well be lower, for example if the penultimate chord is the dominant. However, when an excerpt ends on the subdominant or dominant (less common in tonal music, but present as stimulus categories in the experiments that we model), r_n should be lower than for tonic endings, but r_{n-1} may well be higher, especially if the chord sequence ends I–IV or I–V. Excerpts with tonally less-related endings than these should have still lower values of r_{n-1} and r_n .

To capture the space of possible transitions, we took a probabilistic approach. The forecasts for (r_{n-1}, r_n) can be expressed as a two-dimensional probability density function on $[-1, 1] \times [-1, 1]$, such as the theoretical distribution in Figure 6A. This distribution can be used to calculate the *closure probability* for a given event correlation pair (r_{n-1}, r_n) . For instance, if an excerpt's penultimate and final event correlations map to a point within the black strip in Figure 6A, then this excerpt will have higher closure probability than an excerpt whose terminal event correlation pair maps to the gray or white segments.

The theoretical distribution in Figure 6A was motivated by an empirical probability density function that we calculated for a collection of 2315 popular and classical music tracks. For each track, an onset detector (Tomic & Janata, 2008) was used to locate the penultimate and final events, and then the mean event correlations (r_{n-1}, r_n)

were calculated and assigned to a $[-1, 1] \times [-1, 1]$ grid using a bin spacing of 0.025. The resulting empirical closure distribution is shown in Figure 6B. As the empirical distribution is somewhat noisy, it is useful to apply smoothing and thresholding filters. Figure 6C shows the empirical distribution from Figure 6B smoothed with a rotationally symmetric Gaussian lowpass filter of size 7 bins and standard deviation 10, and thresholded such that probability values less than 0.0025 were set to 0. The connection between the smoothed empirical distribution (Figure 6C) and hypothesized closure distribution (Figure 6A) is evident. The strip of high probability in the former is not as pronounced in the latter, but can be explained as follows. Many of the tracks in the collection end by repeating the tonic, $X - I - I$, where X may be chord IV, chord V, or some other chord. Our method does not capture the $X - I$ transition, but the $I - I$ transition. This causes the high strip of probability in the smoothed empirical distribution (Figure 6C) to be less pronounced than in the hypothesized distribution (Figure 6A).

Closure incorporated into the stepwise regression model

For the three distributions shown in Figure 6, closure probabilities were calculated for each of the 303 priming stimuli considered in this paper, and the process of stepwise selection was rerun with the three closure variables (hypothesized, empirical, smoothed empirical) included in the variable pool. The resulting model was,

$$\begin{aligned} RT = & -57.37 - 111.80 \cdot x_{TS} - 1.4210^4 \cdot p_{\text{clos}} - 357.01 \cdot x_{PP} \\ & + 225.83 \cdot z_{PP} + 0.14 \cdot y_{PP} - 3.44 \cdot y_{CV}, \quad (2) \end{aligned}$$

with test statistic $F(6, 296) = 25.67, p < .001, s = 90.31$, and $R^2 = .34$. The main difference between the stepwise models in (1) and (2) is that p_{clos} , the closure probability calculated from the hypothesized distribution (Figure 6A), enters the model in (2), and R^2 increases. The signs of the coefficients are unchanged, and the coefficient for closure probability is negative as expected (high probability of closure leads to a faster RT).

The model in (2) was better at simulating the RT data. The number of observed comparisons that were successfully fit increased from 16 out of 20 (80%) to 18 out of 20 (90%) at the $p < .05$ level. Figure 7 shows a comparison of observed and fitted RTs calculated from the stepwise model in (2), and it is an evident improvement over Figure 5, which was calculated from the model in (1). With the modeling of closure, we arrived at a model of tonal priming that accounts for all of the trends in Experiments 1–7 (Bigand et al., 2003; Marmel & Tillmann, 2009; Marmel et al., 2010; Marmel et al., 2008; Tillmann, Janata, & Bharucha, 2003; Tillmann, Janata, Birk, et al., 2003; Tillmann et al., 2008), except for the tonic-subdominant trend in Experiment 2 (Bigand et al., 2003), which is on the border of significance ($p = .053$), and the subdominant-baseline trend in Experiment 7 (Tillmann et al., 2008), which similarly borders significance ($p = .077$).

GENERAL DISCUSSION

How perception of structured information in the environment is shaped by “bottom-up” sensory features and “top-down” schemata based on memory for structured relationships is of general interest within psychology and neuroscience. Within the music cognition field, this issue has played itself out in a long-standing debate focusing on the degree to which tonal expectations are driven by sensory information or cognitive schemata of tonal structures. Our contribution to addressing this issue was to determine

the degree to which RT data from tonal priming experiments can be explained using variables derived from different representational levels of a model that begins with the actual audio signals of the experimental stimuli and transforms them across a series of stages that approximate known physiological mechanisms and psychological constructs. We focused on three levels, those of periodicity pitch, chroma vectors or key profiles, and tonal space. These three levels span the transition from “sensory” representations that do not rely on learned knowledge, to “cognitive” representations, as defined by the requirement of prior learning about the space of multiple possible stimulus relationships.

Comparison of model outcomes

Table 4 summarizes the simulation successes and failures of variants of our model and the two previously published models that are closely related and could serve as benchmarks: the PP model (Bigand et al., submitted; Delbé, 2009; Leman, 2000) and the MUSACT model (Bharucha, 1987) in columns two and three, respectively. With regards to the previously reported simulations summarized in columns 2 and 3, Delbé (2009) ran the PP model (column 2) on stimuli from Tillmann, Janata, and Bharucha (2003), and Bigand et al. (2003a). The model simulated the RT trend in the former experiment but not in the latter. Marmel et al. (2010) reported that the PP model could simulate the RT trend within their piano timbral category, but not within the pure-tone timbral category. Prior to the present work, no simulations with the PP model had been conducted for stimuli from the remaining four experiments listed in Table 4. According to Delbé (2009), MUSACT (column 3) like the PP model is capable of simulating the RT trend from Tillmann, Janata, and Bharucha (2003), but not from Bigand et al. (2003a). MUSACT can also simulate the RT trend observed by Marmel and Tillmann (2009). Being restricted to symbolic input,

however, MUSACT cannot be tested for the piano-pure timbre distinction present in Marmel et al. (2010). According to Tillmann et al. (2008), MUSACT had mixed success for predicting the observed trends in Tillmann, Janata, Birk, et al. (2003) and Tillmann et al. (2008), failing in the latter experiment to place the dominant category correctly in the observed increasing RT order of tonic, dominant ~ baseline, subdominant. These simulations indicated the need to combine sensory and cognitive features in the same model, as accomplished in our present work.

Columns four to six were constructed by applying the criteria for success and failure discussed above to simulated RTs from the most significant univariate PP, CV, and TS models. Columns seven and eight summarize the outcomes of the stepwise models (1–2) described above, where stepwise model 1 allowed for a combination of PP, CV, and TS variables, and stepwise model 2 added the concept of closure. When determining the overall success of simulating the results of any given experiment, we adopted the criterion that all relevant comparisons had to be simulated successfully.

Did our simulations achieve our main aim, which was an improved model of tonal expectation? The stepwise model in (1), which took into account the degree to which local events fit into a globally established tonal center, succeeded in simulating observed RT trends for more experiments than any existing model of tonal priming (4 of 7). Nonetheless it still could not account for some trends in Experiments 2 (Bigand et al., 2003 p. 168), 5 (Marmel et al., 2010) and 7 (Tillmann et al., 2008), in particular, the faster RTs for the tonic target events. We did not expect this partial failure at the outset, given that a significant proportion of the behavioral and neuroimaging literature on tonal expectation, in particular harmonic expectation, is implicitly framed in terms of how well

events fit into zeroth-order probability distributions, that is, without regard for sequential information (c.f., Tillmann et al., 2008).

To be able to account for the simulation failures, we had to consider further components of music cognition, the broadest one of which is that of harmonic syntax, i.e., a specification of structural relationships among tonal functions. Though the importance of harmonic syntax for describing the tonal trajectory of a piece of music and associated expectations is undisputed by music theorists (for various interpretations, see Huron, 2006; Lerdahl, 2001; Piston, 1948; Schenker, 1973; Swain, 2002), explicit consideration of the perception of chord transitions and the locations of those transitions within phrasal structure has yet to gain momentum in the field of music psychology (c.f. Bigand et al., 2006; Rohrmeier & Koelsch, 2012; Tillmann & Marmel, 2012). This is despite extensive examination of pairwise relatedness judgments between chords within an instantiated tonal context, and order effects (Contextual Asymmetry) for those chord pairs, in early work on cognitive representations of tonal structure (Bharucha & Krumhansl, 1983; Krumhansl et al., 1982; Krumhansl & Kessler, 1982). As a discrete step in the direction of explicitly modeling syntax, we augmented our model with a quantification of a specific syntactic relationship, namely the transition between the final two events of a musical phrase.

Upon adding a closure variable to the stepwise selection model (2), we were able to explain RT trends in all seven experiments, with borderline-significant results in Experiments 2 (Bigand et al., 2003) and 7 (Tillmann et al., 2008). Comparison of Figures 7F and 5F, and Figures 7G and 5G reveals that the closure variable increases modeling power by reducing RTs for tonic categories. Thus the model in (2) achieved our main aim

of an improved model of tonal expectation, and illustrated that multiple sensory and cognitive representations are a necessary part of a tonal expectation model.

Is it a cause for concern that we were able to account for only 34% of the variance in the RT data using the final model (2)? Inspection of Figures 5 and 7 reveals that, between experiments, there are a number of scaling differences among stimulus categories as well as differences in the offsets of fitted RT means relative to mean RT for the experiment. Because we fit the data from all of the experiments simultaneously, thus allowing these differences to persist, a large proportion of the total variance remains unexplained. Given that the RT data we simulated were collected in three different laboratories across a span of several years, these differences could be due to heterogeneity in the participant samples and/or perceptual and cognitive characteristics of the testing situations. As we could not identify, control for, or measure such differences, we could not try to account for them in the final model. Nevertheless, as described in the following section, the final model we obtained is sufficiently robust to account for RT differences measured for a novel set of stimuli, thus further establishing it as a new benchmark in the field.

Timbral invariance in tonal space

Experiment 5 (Marmel et al., 2010) is particularly illustrative of the improvement afforded by a model that takes both sensory and cognitive representations into account. The original paper reported that the sensory PP model could simulate RT differences within the piano timbral category, but not within the pure-tone timbral category. MUSACT, meanwhile, could simulate the RT difference for tonic and subdominant categories, but it could not be used to compare piano and pure-tone timbres. Our stepwise

model (2) successfully simulated these behavioral data (Figure 7E). As discussed in the section, *Comparison of periodicity pitch and tonal space representations*, the PP representations group stimuli primarily by timbre whereas the TS representations group primarily by tonal function (Figure 3). Remarkably, tonal relatedness of the terminal note to the preceding melodic context was varied by modifying only two tones toward the beginning of the melodies. Even though the remaining notes of the melody were identical, the TS model adequately captured the shift in tonal center imparted by this slight modification.

While Figure 3C suggests that some timbral sensitivity remains in the TS representation, the TS representation nonetheless captures the concept of timbral invariance, that is, the property that a melody played by different instruments sounds “the same” and establishes the same tonal center. As a test of the ability of the combined model to simulate tonal priming data in the face of extreme variation in timbre, we used the stepwise model in (2) to predict RTs for stimuli used in Experiment 2b from Tillmann et al. (2006). The chord sequences from Tillmann et al. (2006) were the same as those used by Bigand et al. (2003) and used in the simulations above. They consisted of twelve six-chord sequences with two different endings (chords seven and eight). For the “related” condition, chords seven and eight functioned as dominant (V) followed by tonic (I), and in the “less related” condition, chords seven and eight functioned as tonic (I) followed by subdominant (IV). Participants performed a timbre discrimination task for the final target chord. Crucially, each of the first seven chords in a sequence was rendered with different timbres chosen from a set (also with different timbral options for attack and sustain portions of the sound). The behavioral results indicated that facilitated processing of the

tonic persisted, despite the potential for interference from the changing timbres. Our stepwise model (2) predicted a mean RT of 638.24 ms for “related” stimuli, and a mean of 673.75 ms for “less related” stimuli. A paired t -test revealed a significant difference for fitted RTs ($t(11) = -3.27, p < .01$), thus matching the difference in the observed RTs.

A number of behavioral studies have found interactions between pitch and timbre (Krumhansl & Iverson, 1992; Melara & Marks, 1990). In particular, the study by Warrier and Zatorre (2002) found that intonation judgments about a final note were better, in spite of a timbre change on the final note, when the preceding melody established a tonal center than when pairs of tones were presented in isolation. We may speculate that the tonal context effect in this experiment is linked to the cognitive TS representation, which displays greater timbral invariance than the PP representation, and could underlie the improved intonation judgments observed for pitches with strong tonal expectations (Janata & Paroo, 2006; Marmel et al., 2008; Navarro Cebrian & Janata, 2010).

Relational versus absolute processing

Although our discussion of tonal expectation is framed in terms of discrete representational spaces, it is important to emphasize that the raw activations of these spaces are not themselves the variables that enter into the statistical models of RTs. Additional operations were specified on those spaces that distill psychologically relevant parameters for use in the statistical models. Here we considered relational and absolute measures, both alone and in combination.

One form of relational information was the average correlation between distributions within specified time-windows relative to event onset: the mean correlation (MC) variable. Based on leaky integration time constants of 0.1 and 4 s, this measure

indicates how well the current (local) event fits with the longer-term (global) event distribution, and effectively serves as a measure of expectation fulfillment or violation.

The second form of relational information compared the post-event value of a measure with the pre-event value of the measure, and therefore represented a temporally local change-detection mechanism. When combined with the MC variable, a measure was created that indicated momentary change in local/global fit, irrespective of the overall degree of local/global fit. In other words, a slight decrement in MC in a passage of overall lower MC would be seen as the same as a slight decrement in MC in a passage of overall higher MC. This measure, derived both in the PP and TS representations, proved most effective in explaining RTs. The fact that mean correlation (MC) variables have been successful, both here and in prior work that considered only the PP representation (Leman, 2000; Delbe, 2009), at modeling behavioral responses in tonal expectation paradigms suggests that listeners utilize a goodness-of-fit measure when listening to music (see also, Krumhansl & Kessler, 1982).

Based on our results, we suggest that it is the moment-to-moment change in goodness-of-fit, rather than overall goodness-of-fit that is of greatest relevance to listeners. We speculate that the local patterns of expectancy violation and fulfillment associated with event transitions have attached to them reward values that make it possible, if not pleasing, to listen to pieces of music in which it is difficult to determine a stable sense of tonal center (low absolute levels of local/global correlation) but certain event transitions nonetheless affirm what sense of tonal center there is (a positive change in the local/global correlation from before the event to after the event). Of interest will be to determine whether activity in brain regions associated with tonal expectation

processing, such as the inferior frontal gyrus, anterior insula, and striatum, reflects absolute or relative local/global correlation estimates.

Limitations and future work

Syntactic knowledge

The study of harmonic expectations in Western tonal music has often been cast in terms of syntax, particularly in relation to language (Koelsch et al., 2005; Koelsch et al., 2007; Patel, 2003). However, the use of the term syntax in studies of music has largely confounded the likelihood of an event given a probability distribution of events (zeroth-order probability) with the likelihood of an event given the specific order of one or more preceding events (first and higher-order transition probabilities) or its position within a hierarchical branching structure. Even though experiments have mainly tested musical expectations based on zeroth-order probabilities, we believe the latter two conceptions of syntax are more appropriate because they acknowledge the importance of specific transitions and the positioning of transitions within a larger structure. One example is that of phrase-ending transitions that impart a sense of closure. We found it necessary to implement this concept in order to improve our ability to model the priming data.

The space of syntactic relationships within music is vast, as it relates to hierarchical tonal relationships found within extended passages of music (Lerdahl, 2001; Rohrmeier & Koelsch, 2012). Describing syntactic relationships of harmonic information within an information theoretic framework is a logical extension of the toroidal model of TS because TS inherently reflects a distribution of probability distributions. An information theoretic approach based on *n-grams* has been used successfully to model

behavioral and neural expectation processes operating within melodies (Pearce et al., 2010), but for a symbolic representation not the complete audio signal.

We foresee an extension of the series of *cognitive* processing stages in our model through the coupling of the tonality model with a temporal model that utilizes the same input processing stages (Tomic & Janata, 2008). The combined model should enable the extraction of tonal symbols, i.e. categories that primarily reflecting complex collections of pitches (chords of arbitrary complexity), as a series of events. Such series of tonal symbols could then be used in syntactic models that jointly consider both temporal and tonal probabilities (Hazan et al., 2009; Schmuckler, 1989).

Another potential approach for embedding chord-level categorical and syntactic information directly into the toroidal model of tonal space is to use an SOM as a model of how these relationships are learned. This is of interest given that the TS representation used in this paper is also based on an SOM, as are related models of tonality (Leman & Carreras, 1997; Tillmann et al., 2000; Toiviainen & Krumhansl, 2003). More generally, SOMs are widely applicable as models learning based on feature similarity, both in visual and auditory domains (e.g. Dufau et al., 2010; Janata, 2001). Delbé (2008) used the temporal or recurrent SOM of Voegtlin (2002) to build a sensory model of tonal priming, and it appears to encode relative transition probabilities between periodicity pitch images. The map was trained using PP images extracted at each beat from the synthesized audio of ten chorales by J. S. Bach. Delbé (2008) demonstrated that adjacent chords on the circle of fifths activate adjacent collections of neurons on the map, but stopped short of a systematic examination of the relationship between transition probability and map proximity for a broader range of periodicity pitch image pairs. Using activity of the

winner neuron as a correlate of RT, Delbé (2008) also showed that the map could simulate RT trends from (Bigand et al., 2003). This is one of the experiments (Experiment 2 above) that prior models had difficulty simulating, and as such the temporal SOM approach merits further investigation.

A considerable challenge in studying tonal syntax is the structuring of stimulus materials and tasks such that closure effects are avoided. Both priming and subjective measures require that the to-be-judged event is clearly marked, something that is most easily achieved by making that event the terminal event, though a recent attempt at marking events with visual cues has been successful (Tillmann & Marmel, 2012).

Other musical systems

The toroidal representation of tonal space arises by virtue of the pitch distributional statistics present in Western tonal music. Consequently, one is left to wonder how well this model would capture listeners' expectations when listening to music with different pitch-distributional statistics, whether those arise from different cultural norms (Demorest et al., 2010; Krumhansl, 1990a), manipulations of the standard Western tonal pitch class set (Oram & Cuddy, 1995; Saffran, Johnson, Aslin, & Newport, 1999), or use of nonstandard scale or structure systems (Loui et al., 2009; Tillmann & Poulin-Charronnat, 2010). While answering this question is beyond the scope of the present paper, we offer some suggestions.

In the case of listeners raised on predominantly Western tonal music, the TS model could be thought of as a filter through which the listener processes the material of unknown systems (Curtis & Bharucha, 2009). The idea that listeners listen through an invariant culturally determined filter is challenged, however, by observations that

listeners adapt their expectations to the style-specific characteristics of the musical stimuli with which they are presented (Castellano, Bharucha, & Krumhansl, 1984; Kessler, Hansen, & Shepard, 1984; Krumhansl, 1990a; Oram & Cuddy, 1995), pointing out the role rapid learning and short term memory systems for guiding perception. Thus the main questions become whether the pitch distributional statistics of other musical systems are readily represented on a toroidal surface or whether some other manifold is more appropriate, and whether the derived measures of how well the local distribution matches the global distribution, e.g. mean correlation, are adequate for capturing variance in behavioral responses. Our finding that the CV representation did not fare as well as PP or TS representations in the statistical models suggests that representational spaces do vary in their ability to explain the behavioral data.

Despite the many experimental results that remain to be modeled and the need to extend the model to fully account for tonal syntax, we believe the utility and theoretical import of a modeling framework that derives multiple representational stages along a sensory-to-cognitive continuum, starting with the musical audio signals used in experiments that probe the psychological and neural response to those signals, is high. We encourage other researchers in the field to make use of our freely available modeling tools for examining predicted tonal relationships in their stimulus materials.

ACKNOWLEDGMENTS

We thank Malcolm Moyer and Oliver Janata for help with stimulus preparation, Frédéric Marmel for providing stimulus material and experimental data, and three anonymous reviewers for helpful comments on earlier versions of the paper. This work was supported by NSF Grant #1025310 to Petr Janata. Charles Delbé was supported by a

travel grant from the LabEx CeLyA (“Centre Lyonnais d'Acoustique”, ANR-10-LABX-60).

References

- Aarden, B. J. (2003). *Dynamic melodic expectancy*. Ohio State University. Retrieved from http://rave.ohiolink.edu/etdc/view?acc_num=osu1060969388
- Balzano, G. J. (1980). The Group-Theoretic Description of 12-Fold and Microtonal Pitch Systems. *Computer Music Journal*, 4(4), 66-84. doi: 10.2307/3679467
- Bello, J. P., & Pickens, J. (2005). *A robust mid-level representation for harmonic content in music signals*. Paper presented at the Proceedings of the International Symposium on Music Information Retrieval, London, UK.
- Bharucha, J., & Krumhansl, C. L. (1983). The representation of harmonic structure in music - hierarchies of stability as a function of context. *Cognition*, 13(1), 63-102. doi: 10.1016/0010-0277(83)90003-3
- Bharucha, J. J. (1987). Music cognition and perceptual facilitation: a connectionist framework. *Music Perception*, 5, 1-30.
- Bharucha, J. J., & Stoeckig, K. (1986). Reaction time and musical expectancy: priming of chords. *J Exp Psychol Hum Percept Perform*, 12(4), 403-410.
- Bharucha, J. J., & Stoeckig, K. (1987). Priming of chords: spreading activation or overlapping frequency spectra? *Percept Psychophys*, 41(6), 519-524.
- Bigand, E., Delbé, C., Poulin-Charronnat, B., Leman, M., & Tillmann, B. (submitted). Does musical syntax processing tell us more about auditory memory than about syntactic-like computations:
the unanswered question.
- Bigand, E., Madurell, F., Tillmann, B., & Pineau, M. (1999). Effect of global structure and temporal organization on chord processing. *Journal of Experimental Psychology: Human Perception and Performance*, 25(1), 184-197.
- Bigand, E., & Parncutt, R. (1999). Perceiving musical tension in long chord sequences. *Psychol Res*, 62(4), 237-254.
- Bigand, E., Poulin, B., Tillmann, B., Madurell, F., & D'Adamo, D. A. (2003). Sensory versus cognitive components in harmonic priming. *Journal of Experimental Psychology-Human Perception and Performance*, 29(1), 159-171. doi: 10.1037/0096-1523.29.1.159
- Bigand, E., Tillmann, B., & Poulin-Charronnat, B. (2006). A module for syntactic processing in music? *Trends in Cognitive Sciences*, 10(5), 195-196. doi: 10.1016/j.tics.2006.03.008

- Boltz, M. G. (1989). Rhythm and "good endings": Effects of temporal structure on tonality judgments. *Perception and Psychophysics*, *46*, 9–17.
- Brown, S. (2000). The "musilanguage" model of music evolution. In N. L. Wallin, B. Merker & S. Brown (Eds.), *The Origins of Music*. Cambridge, MA: MIT Press.
- Butler, D. (1989). Describing the perception of tonality in music: A critique of the tonal hierarchy theory and a proposal for a theory of intervallic rivalry. *Music Perception*, *6*, 219-241.
- Butler, D. (1990). Response to Carol Krumhansl. *Music Perception*, *7*, 325-338.
- Callender, C., Quinn, I., & Tymoczko, D. (2008). Generalized voice-leading spaces. *Science*, *320*(5874), 346-348.
- Cariani, P. A., & Delgutte, B. (1996a). Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *J Neurophysiol*, *76*(3), 1698-1716.
- Cariani, P. A., & Delgutte, B. (1996b). Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase invariance, pitch circularity, rate pitch, and the dominance region for pitch. *J Neurophysiol*, *76*(3), 1717-1734.
- Castellano, M. A., Bharucha, J. J., & Krumhansl, C. L. (1984). Tonal hierarchies in the music of North India. *Journal of Experimental Psychology: General*, *113*(3), 394-412.
- Chai, W., & Vercoe, B. (2005, September). *Detection of key change in classical piano music*. Paper presented at the Proceedings of the International Symposium on Music Information Retrieval, London, UK.
- Chew, E. (2002). The spiral array: An algorithm for determining key boundaries. In C. Anagnostopoulou, M. Ferrand & A. Smaill (Eds.), *Music and artificial intelligence: Proceedings of the international conference on music and artificial intelligence* (pp. 18-31). Berlin: Springer.
- Conklin, D., & Witten, I. H. (1995). Multiple viewpoint systems for music prediction. *Journal of New Music Research*, *24*, 51-73.
- Cook, N. (1987). The perception of large-scale tonal closure. *Music Perception*, *5*(2), 197-205.
- Cook, N. (1994). Perception: a perspective from music theory. In R. Aiello & J. A. Sloboda (Eds.), *Musical perceptions* (pp. 64–95). Oxford: Oxford University Press.
- Curtis, M. E., & Bharucha, J. J. (2009). Memory and musical expectation for tones in cultural context. *Music Perception*, *26*(4), 365-375.

- De Lucia, M., Cocchi, L., Martuzzi, R., Meuli, R. A., Clarke, S., & Murray, M. M. (2010). Perceptual and Semantic Contributions to Repetition Priming of Environmental Sounds. *Cerebral Cortex*, *20*(7), 1676-1684. doi: 10.1093/cercor/bhp230
- Delbé, C. (2008). Recurrent self-organization of sensory signals in the auditory domain. In R. M. French & E. Thomas (Eds.), *From association to rules: Connectionist models of behavior and cognition* (pp. 178-187). Singapore: World Scientific Publishing.
- Delbé, C. (2009). *Music, psychoacoustics and implicit learning: Toward an integrated model of music cognition*. (PhD), University of Burgundy, France. Retrieved from <http://leadserv.u-bourgogne.fr/%7Edelbe/Thesis.pdf>
- Demorest, S. M., Morrison, S. J., Stambaugh, L. A., Beken, M., Richards, T. L., & Johnson, C. (2010). An fMRI investigation of the cultural specificity of music memory. *Social Cognitive and Affective Neuroscience*, *5*(2-3), 282-291.
- Deutsch, D. (1999). The processing of pitch combinations. In D. Deutsch (Ed.), *The Psychology of Music* (2nd ed., pp. 349-411). San Diego: Academic Press.
- Dufau, S., Lété, B., Touzet, C., Glotin, H., Ziegler, J. C., & Grainger, J. (2010). A developmental perspective on visual word recognition: New evidence and a self-organizing model. *European Journal of Cognitive Psychology*, *22*(5), 669-694.
- Elliott, R., Agnew, Z., & Deakin, J. F. W. (2008). Medial orbitofrontal cortex codes relative rather than absolute value of financial rewards in humans. *European Journal of Neuroscience*, *27*(9), 2213-2218. doi: 10.1111/j.1460-9568.2008.06202.x
- Farbood, M. M. (2012). A parametric, temporal model of musical tension. *Music Perception*, *29*(4), 387-428.
- Gomez, E. (2006). Tonal description of polyphonic audio for music content processing. *Inform Journal on Computing*, *18*(3), 294-304. doi: 10.1287/ijoc.1040.0126
- Gonnerman, L. M., Seidenberg, M. S., & Andersen, E. S. (2007). Graded semantic and phonological similarity effects in priming: Evidence for a distributed connectionist approach to morphology. *Journal of Experimental Psychology: General*, *136*(2), 323-345. doi: 10.1037/0096-3445.136.2.323
- Hantz, E. C., Kreilick, K. G., Kananen, W., & Swartz, K. P. (1997). Neural responses to melodic and harmonic closure: An event-related-potential study. *Music Perception*, *15*(1), 69-98.
- Hazan, A., Marxer, R., Brossier, P., Purwins, H., Herrera, P., & Serra, X. (2009). What/when causal expectation modelling applied to audio signals. *Connection Science*, *21*(2), 119 - 143.

- Hecht, D., Reiner, M., & Karni, A. (2009). Repetition priming for multisensory stimuli: Task-irrelevant and task-relevant stimuli are associated if semantically related but with no advantage over uni-sensory stimuli. *Brain Research*, *1251*(0), 236-244. doi: <http://dx.doi.org/10.1016/j.brainres.2008.10.062>
- Hillinger, M. (1980). Priming effects with phonemically similar words. *Memory & Cognition*, *8*(2), 115-123. doi: 10.3758/bf03213414
- Hoch, L., Poulin-Charronnat, B., & Tillmann, B. (2011). The influence of task-irrelevant music on language processing: Syntactic and semantic structures. [Original Research]. *Frontiers in Psychology*, *2*. doi: 10.3389/fpsyg.2011.00112
- Huron, D. (2006). *Sweet Anticipation: Music and the Psychology of Expectation*: MIT Press.
- Huron, D., & Parncutt, R. (1993). An improved model of tonality perception incorporating pitch salience and echoic memory. *Psychomusicology*, *12*, 154-171.
- Hutchinson, W., & Knopoff, L. (1978). The acoustic component of western consonance. *Interface*, *7*, 1-29.
- Jackendoff, R., & Lerdahl, F. (2006). The capacity for music: What is it, and what's special about it? *Cognition*, *100*(1), 33-72. doi: <http://dx.doi.org/10.1016/j.cognition.2005.11.005>
- Janata, P. (1995). ERP measures assay the degree of expectancy violation of harmonic contexts in music. *Journal of Cognitive Neuroscience*, *7*(2), 153-164.
- Janata, P. (2001). Quantitative assessment of vocal development in the zebra finch using self-organizing neural networks. *Journal of the Acoustical Society of America*, *110*(5), 2593-2603.
- Janata, P. (2005). Brain networks that track musical structure. *Annals of the New York Academy of Sciences*, *1060*(1), 111-124.
- Janata, P. (2007). Navigating tonal space. In W. B. Hewlett, E. Selfridge-Field & E. Correia (Eds.), *Computing in musicology: Tonal theory for the digital age* (Vol. 15, pp. 39-50). Stanford: Center for Computer Assisted Research in the Humanities.
- Janata, P. (2009). The neural architecture of music-evoked autobiographical memories. *Cerebral Cortex*, *19*, 2579-2594. doi: [10.1093/cercor/bhp008](http://dx.doi.org/10.1093/cercor/bhp008)
- Janata, P., Birk, J. L., Tillmann, B., & Bharucha, J. J. (2003). Online detection of tonal pop-out in modulating contexts. *Music Perception*, *20*(3), 283-305.

- Janata, P., Birk, J. L., Van Horn, J. D., Leman, M., Tillmann, B., & Bharucha, J. J. (2002). The cortical topography of tonal structures underlying Western music. *Science*, 298(5601), 2167-2170.
- Janata, P., & Paroo, K. (2006). Acuity of auditory images in pitch and time. *Perception and Psychophysics*, 68(5), 829-844.
- Janata, P., & Reisberg, D. (1988). Response-time measures as a means of exploring tonal hierarchies. *Music Perception*, 6(2), 161-172.
- Kessler, E. J., Hansen, C., & Shepard, R. N. (1984). Tonal Schemata in the Perception of Music in Bali and in the West. *Music Perception*, 2(2), 131-165.
- Kim, S.-G., Kim, J. S., & Chung, C. K. (2011). The effect of conditional probability of chord progression on brain response: An MEG study. *PLoS ONE*, 6(2), 1-9.
- Knosche, T. R., Neuhaus, C., Haueisen, J., Alter, K., Maess, B., Witte, O. W., & Friederici, A. D. (2005). Perception of phrase structure in music. *Human Brain Mapping*, 24(4). doi: 10.1002/hbm.20088
- Koelsch, S., Gunter, T. C., Friederici, A. D., & Schröger, E. (2000). Brain indices of music processing: "Nonmusicians" are musical. *Journal of Cognitive Neuroscience*, 12(3), 520-541.
- Koelsch, S., Gunter, T. C., Schroger, E., Tervaniemi, M., Sammler, D., & Friederici, A. D. (2001). Differentiating ERAN and MMN: An ERP study. *Neuroreport*, 12(7), 1385-1389.
- Koelsch, S., Gunter, T. C., v Cramon, D. Y., Zysset, S., Lohmann, G., & Friederici, A. D. (2002). Bach speaks: a cortical "language-network" serves the processing of music. *Neuroimage*, 17(2), 956-966.
- Koelsch, S., Gunter, T. C., Wittfoth, M., & Sammler, D. (2005). Interaction between syntax processing in language and in music: An ERP study. *Journal of Cognitive Neuroscience*, 17(10), 1565-1577.
- Koelsch, S., Jentschke, S., Sammler, D., & Mietchen, D. (2007). Untangling syntactic and sensory processing: An ERP study of music perception. *Psychophysiology*, 44(3), 476-490. doi: 10.1111/j.1469-8986.2007.00517.x
- Kohonen, T. (1995). *Self-organizing maps*. Berlin: Springer.
- Krumhansl, C. L. (1990a). *Cognitive foundations of musical pitch*. New York: Oxford University Press.
- Krumhansl, C. L. (1990b). Tonal hierarchies and rare intervals in music cognition. *Music Perception*, 7(3), 309-324.

- Krumhansl, C. L., Bharucha, J. J., & Kessler, E. J. (1982). Perceived harmonic structure of chords in three related musical keys. *J Exp Psychol Hum Percept Perform*, 8(1), 24-36.
- Krumhansl, C. L., & Cuddy, L. L. (2010). A Theory of Tonal Hierarchies in Music. In M. Riess Jones, R. R. Fay & A. N. Popper (Eds.), *Music Perception* (Vol. 36, pp. 51-87): Springer New York.
- Krumhansl, C. L., & Iverson, P. (1992). Perceptual interactions between musical pitch and timbre. *J Exp Psychol Hum Percept Perform*, 18(3), 739-751.
- 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(4), 334-368.
- 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(4), 579-594.
- Lartillot, O., Toiviainen, P., & Eerola, T. (2008). A Matlab toolbox for music information retrieval. In C. Preisach, H. Burkhardt, L. Schmidt-Thieme & R. Decker (Eds.), *Data analysis, machine learning and applications* (pp. 261-268). Berlin: Springer.
- Leino, S., Brattico, E., Tervaniemi, M., & Vuust, P. (2007). Representation of harmony rules in the human brain: Further evidence from event-related potentials. *Brain Research*, 1142, 169-177. doi: 10.1016/j.brainres.2007.01.049
- Leman, M. (1995). A Model of Retroactive Tone-Center Perception. *Music Perception*, 12(4), 439-471.
- 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., & Carreras, F. (1997). Schema and Gestalt: Testing the hypothesis of psychoneural isomorphism by computer simulation. In M. Leman (Ed.), *Music, Gestalt, and Computing – Studies in Cognitive and Systematic Musicology* (pp. 144–168). Berlin: Springer.
- Leman, M., Lesaffre, M., & Tanghe, K. (2001). *Introduction to the IPEM toolbox for perception-based music analysis*. Paper presented at the Proceedings of the FWO Research Society on Foundations of Music, Ghent, Belgium.
- Lerdahl, F. (2001). *Tonal Pitch Space*. New York: Oxford University Press.
- Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA.: MIT Press.

- Lerdahl, F., & Krumhansl, C. L. (2007). Modeling Tonal Tension. *Music Perception*, 24(4), 329-366. doi: doi:10.1525/mp.2007.24.4.329 %U <http://caliber.ucpress.net/doi/abs/10.1525/mp.2007.24.4.329>
- Longuet-Higgins, H. C. (1979). Review Lecture: The Perception of Music. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 205(1160), 307-322. doi: 10.1098/rspb.1979.0067
- Loui, P., Wu, E. H., Wessel, D. L., & Knight, R. T. (2009). A Generalized Mechanism for Perception of Pitch Patterns. *Journal of Neuroscience*, 29(2), 454-459. doi: 10.1523/jneurosci.4503-08.2009
- Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A. D. (2001). Musical syntax is processed in Broca's area: an MEG study. *Nature Neuroscience*, 4(5), 540-545.
- Margulis, E. H. (2005). A model of musical expectancy. *Music Perception*, 22(5), 663-714.
- Marmel, F., & Tillmann, B. (2009). Tonal priming beyond tonics. *Music Perception*, 26(3), 211-221.
- Marmel, F., Tillmann, B., & Delbé, C. (2010). Priming in melody perception: tracking down the strength of cognitive expectations. *Journal of Experimental Psychology: Human Perception and Performance*, 36, 1016-1028.
- Marmel, F., Tillmann, B., & Dowling, W. J. (2008). Tonal expectations influence pitch perception. *Perception and Psychophysics*, 70(5), 841-852. doi: 10.3758/pp.70.5.841
- Marvin, E. W., & Brinkman, A. (1999). The effect of modulation and formal manipulation on perception of tonic closure by expert listeners. *Music Perception*, 16(4), 389-407.
- Melara, R. D., & Marks, L. E. (1990). Interaction among auditory dimensions: Timbre, pitch, and loudness. *Perception and Psychophysics*, 48(2), 169-178.
- Meyer, L. B. (1956). *Emotion and meaning in music*. Chicago, IL: University of Chicago Press.
- Milne, A. J., Sethares, W. A., Laney, R., & Sharp, D. B. (2011). Modelling the similarity of pitch collections with expectation tensors. *Journal of Mathematics and Music*, 5(1), 1-20.
- Narmour, E. (1990). *The analysis and cognition of basic melodic structures: the implication-realization model*. Chicago, IL: University of Chicago Press.

- Navarro Cebrian, A., & Janata, P. (2010). Electrophysiological correlates of accurate mental image formation in auditory perception and imagery tasks. *Brain Research*, *1342*, 39-54.
- Neely, J. H. (1976). Semantic priming and retrieval from lexical memory: Evidence for facilitatory and inhibitory processes. *Memory and Cognition*, *4*(5), 648-654.
- Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264-336). Hillsdale, NJ: Erlbaum.
- Oram, N., & Cuddy, L. L. (1995). Responsiveness of Western adults to pitch-distributional information in melodic sequences. *Psychological Research*, *57*(2), 103-118.
- Oxenham, A. J. (2012). Pitch Perception. *The Journal of Neuroscience*, *32*(39), 13335-13338. doi: 10.1523/jneurosci.3815-12.2012
- Parncutt, R., & Bregman, A. S. (2000). Tone profiles following short chord progressions: Top-down or bottom-up? *Music Perception*, *18*(1), 25-57.
- Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, *6*(7), 674-681.
- Patel, A. D. (2012). Language, music, and the brain: a resource-sharing framework. In P. Rebuschat, M. Rohrmeier, J. A. Hawkins & I. Cross (Eds.), *Language and music as cognitive systems* (pp. 204-223). Oxford: Oxford University Press.
- Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience*, *10*(6), 717-733.
- Pearce, M. T., Ruiz, M. H., Kapasi, S., Wiggins, G. A., & Bhattacharya, J. (2010). Unsupervised statistical learning underpins computational, behavioural, and neural manifestations of musical expectation. *Neuroimage*, *50*(1), 302-313.
- Pearce, M. T., & Wiggins, G. A. (2006). Expectation in melody: The influence of context and learning. *Music Perception*, *23*(5), 377-405.
- Perruchet, P., & Poulin-Charronnat, B. (2013). Challenging prior evidence for a shared syntactic processor for language and music. *Psychonomic Bulletin & Review*, *20*(2), 310-317. doi: 10.3758/s13423-012-0344-5
- Piston, W. (1948). *Harmony*. New York: W. W. Norton.
- Plomp, R., & Levelt, W. J. M. (1965). Tonal consonance and critical bandwidth. *Journal of the Acoustical Society of America*, *38*(4), 548-560.

- Poulin-Charronnat, B., Bigand, E., & Koelsch, S. (2006). Processing of musical syntax tonic versus subdominant: An event-related potential study. *Journal of Cognitive Neuroscience, 18*(9), 1545-1554.
- Poulin-Charronnat, B., Bigand, E., Madurell, F., & Peereman, R. (2005). Musical structure modulates semantic priming in vocal music. *Cognition, 94*(3), B67-B78. doi: 10.1016/j.cognition.2004.05.003
- Radeau, M., Besson, M., Fonteneau, E., & Castro, S. L. (1998). Semantic, repetition and rime priming between spoken words: behavioral and electrophysiological evidence. *Biological Psychology, 48*(2), 183-204.
- Rohrmeier, M. A. (2011). Towards a generative syntax of tonal harmony. *Journal of Mathematics and Music, 5*(1), 35-53.
- Rohrmeier, M. A., & Koelsch, S. (2012). Predictive information processing in music cognition. A critical review. *International Journal of Psychophysiology, 83*(2), 164-175.
- Rosen, C. (1972). *The classical style: Haydn, Mozart, Beethoven*. New York: W. W. Norton & Company.
- Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*(1), 27-52.
- Salimpoor, V. N., Benovoy, M., Larcher, K., Dagher, A., & Zatorre, R. J. (2011). Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nature Neuroscience, 14*(2), 257-U355. doi: 10.1038/nn.2726
- Salimpoor, V. N., van den Bosch, I., Kovacevic, N., McIntosh, A. R., Dagher, A., & Zatorre, R. J. (2013). Interactions Between the Nucleus Accumbens and Auditory Cortices Predict Music Reward Value. *Science, 340*(6129), 216-219. doi: 10.1126/science.1231059
- Sapp, C. S. (2005). Visual hierarchical key analysis. *Computers in Entertainment, 3*(4), 1-19.
- Schellenberg, G. E. (1997). Simplifying the implication-realization model of melodic expectancy. *Music Perception, 14*(3), 295-318.
- Schenker, H. (1973). *Harmony* (E. Mann Borgese, Trans.). Cambridge, MA: MIT Press.
- Schmuckler, M. A. (1989). Expectation in music: investigation of melodic and harmonic processes. *Music Perception, 7*(2), 109-150.
- Schnupp, J., Nelken, I., & King, A. (2011). *Auditory neuroscience: Making sense of sound*. Cambridge, MA: MIT Press.

- Schoenberg, A. (1954/1983). *Structural functions of harmony*. London, UK: Faber and Faber.
- Serrà, J., Gomez, E., Herrera, P., & Serra, X. (2008). Statistical Analysis of Chroma Features in Western Music Predicts Human Judgments of Tonality. *Journal of New Music Research*, 37(4), 299-309. doi: 10.1080/09298210902894085
- Shepard, R. N. (1982). Geometrical approximations to the structure of musical pitch. *Psychol Rev*, 89(4), 305-333.
- Slevc, L. R., Rosenberg, J., & Patel, A. (2009). Making psycholinguistics musical: Self-paced reading time evidence for shared processing of linguistic and musical syntax. *Psychonomic Bulletin & Review*, 16(2), 374-381. doi: 10.3758/16.2.374
- Steinbeis, N., & Koelsch, S. (2008). Shared neural resources between music and language indicate semantic processing of musical tension-resolution patterns. *Cerebral Cortex*, 18(5), 1169-1178. doi: 10.1093/cercor/bhm149
- Stoffer, T. H. (1985). Representation of phrase structure in the perception of music. *Music Perception*, 3(2), 191-220.
- Swain, J. P. (2002). *Harmonic Rhythm: Analysis and Interpretation*. Oxford: Oxford University Press.
- Tekman, H. G., & Bharucha, J. J. (1998). Implicit knowledge versus psychoacoustic similarity in priming of chords. *Journal of Experimental Psychology-Human Perception and Performance*, 24(1), 252-260.
- Temperley, D. (2004). Bayesian models of musical structure and cognition. *Musicae Scientiae*, 8(2), 175-205.
- Temperley, D. (2007). *Music and Probability*. Cambridge: The MIT Press.
- Temperley, D., & Marvin, E. W. (2008). Pitch-class distribution and the identification of key. *Music Perception*, 25(3), 193-212. doi: 10.1525/mp.2008.25.3.193
- Tillmann, B., Bharucha, J. J., & Bigand, E. (2000). Implicit learning of tonality: a self-organizing approach. *Psychol Rev*, 107(4), 885-913.
- Tillmann, B., Bigand, E., Escoffier, N., & Lalitte, P. (2006). The influence of musical relatedness on timbre discrimination. *European Journal of Cognitive Psychology*, 18(3), 343-358.
- Tillmann, B., Janata, P., & Bharucha, J. J. (2003). Activation of the inferior frontal cortex in musical priming. *Cognitive Brain Research*, 16, 145-161.

- Tillmann, B., Janata, P., Birk, J. L., & Bharucha, J. J. (2003). The costs and benefits of tonal centers for chord processing. *Journal of Experimental Psychology: Human Perception and Performance*, 29(2), 470-482.
- Tillmann, B., Janata, P., Birk, J. L., & Bharucha, J. J. (2008). Tonal centers and expectancy: facilitation or inhibition of chords at the top of the harmonic hierarchy? *Journal of Experimental Psychology: Human Perception and Performance*, 34(4), 1031-1043.
- Tillmann, B., & Marmel, F. (2012). Musical expectations within chord sequences: Facilitation due to tonal stability without closure effects. *Psychomusicology*, in press.
- Tillmann, B., & Poulin-Charronnat, B. (2010). Auditory expectations for newly acquired structures. *Quarterly Journal of Experimental Psychology*, 63, 1646-1664.
- Toiviainen, P., & Krumhansl, C. L. (2003). Measuring and modeling real-time responses to music: The dynamics of tonality induction. *Perception*, 32(6), 741-766.
- Tomic, S., & Janata, P. (2008). Beyond the beat: Modeling metric structure in music and performance. *Journal of the Acoustical Society of America*, 124(6), 4024-4041.
- Trainor, L. J., & Zatorre, R. J. (2009). The neurological basis of musical expectations. In S. Hallam, I. Cross, M. Thaut & A. D. Patel (Eds.), *The Oxford handbook of music psychology*. Oxford, UK: Oxford University Press.
- Van Immerseel, L. M., & Martens, J. P. (1992). Pitch and voiced/unvoiced determination with an auditory model. *Journal of the Acoustical Society of America*, 91(6), 3511-3526.
- Voegtlin, T. (2002). Recursive self-organizing maps. *Neural Networks*, 15(8-9), 979-991.
- von Hippel, P., & Huron, D. (2000). Why do skips precede reversals? The effect of tessitura on melodic structure. *Music Perception*, 18(1), 59-85.
- Warren, J. D., Uppenkamp, S., Patterson, R. D., & Griffiths, T. D. (2003). Separating pitch chroma and pitch height in the human brain. *Proc Natl Acad Sci U S A*, 100(17), 10038-10042.
- Warrier, C. M., & Zatorre, R. J. (2002). Influence of tonal context and timbral variation on perception of pitch. *Perception and Psychophysics*, 64(2), 198-207.
- Woolhouse, M. (2009). Modelling tonal attraction between adjacent musical elements. *Journal of New Music Research*, 38(4), 357-379.

Footnotes

¹ *Oxford music online*, including *The new Grove dictionary of music and musicians*, provides a convenient resource with definitions and discussion of the musical terms we use in this article. Available from <http://www.oxfordmusiconline.com>

² The relative minor scale shares the same notes but has a different tonic while the parallel minor scale shares the same tonic but has different notes.

³ The tonal priming paradigm was adapted from psycholinguistics (Neely, 1976), where prime-target word pairs such as *lion-tiger* and *table-tiger* exemplify differing degrees of semantic relatedness, with more or less related primes leading to faster or slower word processing, respectively. Experimentally, the speed of word processing can be measured by investigating response times (RTs) for participants' discriminating between an existing word of their language and a non-word foil (i.e., the lexical decision task).

⁴ We recognize that the studies we focus on are only a subset of the many empirical studies that have addressed the issue of tonal expectations, and that the priming methodology that they have in common is but one of several methods by which tonal expectations can be assessed. Nonetheless we believe this focus is justified on three grounds: 1) the breadth of manipulations pertaining to the sensory/cognitive debate across a methodologically homogeneous set of studies (which is unmatched elsewhere in the relevant literature), 2) unavailability of suitable datasets containing item-level behavioral data and audio material, and 3) length constraints on the present paper.

⁵ For this experiment we used RTs collected for one of two conditions by Tillmann, Bigand, Escoffier, and Lalitte (2006), on a timbre discrimination task. The stimuli were

from Bigand et al. (2003), so we will continue to cite this paper in reference to Experiment 2.

⁶ Available from <http://www.ipem.ugent.be/?q=node/27>

⁷ A MATLAB implementation of the TS model is available from <http://atonal.ucdavis.edu/resources/software/jlmt/>

⁸ The implementation makes use of the SOM Toolbox, available from <http://www.cis.hut.fi/projects/somtoolbox>

⁹ While it is true that the underlying theoretical/computational model is currently implemented as a sequential feedforward model, the purpose of the statistical model is different. The purpose of the statistical model is to determine the relative influence of different stages of the model on listener responses. We assume that the information from each processing stage remains available to the listener's decision-making processes. For this reason, the order of entry of the variables into a multiple regression model need not be constrained by the underlying sequence of transformations.

¹⁰ Zeroing the means is justified, as RT data were gathered for different tasks (with varying difficulty levels), by different researchers and in different papers. We are not interested in identifying the source of between-experiment differences in RT, which may themselves have their own independent causes that are not being modeled here.

¹¹ A subset was necessary in order to avoid rank deficiency of the correlation matrix due to pairs of variables that were highly correlated. These were typically variable pairs derived from the early and late post-target windows, in which case we retained the early window variable.

¹² Three stimuli from Experiment 1 (one from each of categories Rel B, Rel C, and Un C) were excluded from the paired *t*-tests, because audio for the corresponding stimulus from the Un B (unrelated stimulus in B major) category was missing.

¹³ A tone cluster is a chord consisting of three or more adjacent notes, and a modulation is a strongly established change of key.

Table 1

Summary of tonal priming studies used for the RT modeling

Study	Experimental manipulation	Sensory priming interpretation?	Discrimination task	Number of stimuli (303 total)
Tillmann, Janata, and Bharucha (2003)	Chord sequences ending on tonic or out-of-key chord	Yes	Diss-onance	63
Bigand et al. (2003, experiment 1)	Chord sequences ending on tonic (I) or subdominant (IV)	No	Diss-onance	24
Marmel et al. (2008, experiment 2)	Melodies ending on tonic (I) or subdominant (IV)	No	Intonation	24
Marmel and Tillmann (2009, experiment 2)	Melodies ending on mediant (III) or leading tone (VII)	No	Intonation	24
Marmel et al. (2010, experiments 1 and 2)	Pure tone or piano timbre melodies ending on tonic (I) or subdominant (IV)	No for pure tones; possibly for piano tones	Timbre	48
Tillmann, Janata, Birk, and Bharucha (2003, experiment 2)	Key-inducing and non-inducing chord sequences ending, when key-inducing, on tonic (I) or subdominant (IV)	No overall	Diss-onance	48
Tillmann et al. (2008, experiment 3)	As immediately above, but with inclusion of additional dominant (V) condition	No overall	Diss-onance	72

Table 2

Computational models of tonal expectation in reverse chronological order

Name	Model description	Input	Polyphonic*
Farbood (2012)	Combines tonal, metric, and dynamic measures of tension, using a sliding time window	Symbolic	Yes
Milne et al. (2011)	Calculate the distance between pairs of chords (see also voice leading approaches of Callender et al., 2008; Woolhouse, 2009)	Symbolic	Yes
Rohrmeier (2011)	28 rules that enable a tree structure to be placed over a passage of music	Symbolic	Yes
IDyOM (Pearce et al., 2010)	Variable-length n -gram models applied to multiple viewpoints of the musical surface	Symbolic	No
Lerdahl and Krumhansl (2007)	Tonal tension plotted against time, based on the perceptual distance between two chords and reductions (not fully implemented)	Symbolic	Yes
Temperley (2004, 2007)	Bayesian model for the likelihood of a pitch-class set in a piece	Symbolic	Yes
Aarden (2003)	Mainly experimental evidence contrasting mid- and end-of-phrase probe-tone profiles	Symbolic	No
Toiviainen and Krumhansl (2003)	Tone-transition model, using a pitch memory vector and calculation of the transition strengths between all pitch pairs	Symbolic	Yes
Tonal space model (Janata et al., 2002)	Projection of periodicity pitch images to a toroidal surface representing major/minor key relationships in Western tonal music	Acoustic	Yes

Spiral array model (Chew, 2002)	Helical representation of circle of fifths, with pitch centroids forming chord helices, and chord centroids forming key helices	Symbolic	Yes
PP model (Leman, 2000)	Post-target correlation between local and global integrated periodicity pitch images	Acoustic	Yes
Parncutt and Bregman (2000)	Similar to Huron and Parncutt (1993), but for chroma vectors	Semi-acoustic	Yes
Schellenberg (1997)	Simplified Narmour's (1990) implication-realization theory. See also Margulis (2005)	Symbolic	No
Conklin & Witten (1995)	Variable-length n -gram models applied to multiple viewpoints of the musical surface	Symbolic	No
Huron and Parncutt (1993)	Weighting pitches that could be heard in an acoustic spectrum, weights decay over time	Semi-acoustic	Yes
MUSACT (Bharucha, 1987)	Connectionist model with units for tones, chords, and keys, with connections in both directions between these three layers	Symbolic	Yes

* Polyphonic refers to the ability of the model to represent the simultaneous occurrences of multiple notes

Table 3

Analyses of variance for variables grouped by representational space

Source	df	<i>F</i>	Cohen's f^2	<i>p</i>
Tonal space	6	4.85***	.10	< .001
Periodicity pitch	5	4.71***	.08	< .001
Chroma vector	6	2.12	.04	.051
error	285	(8668)		

Note. This table is based on Type 2 sums of squares – the reduction in residual sum of squares obtained by adding that group of variables to a model consisting of all other terms apart from the variables group in question. Value enclosed in parentheses represents mean square error. * $p < .05$.

*** $p < .001$.

Table 4

Summary of simulation success and failure for reported experiments and models

Experiment	PP model (Leman, 2000)	MUSACT (Bharucha, 1987)	PP variable	CV variable	TS variable	Stepwise model in (1)	Stepwise model in (2)
1. Tillmann, Janata, and Bharucha (2003)	✓	✓	✓	✓	✓	✓	✓
2. Bigand et al. (2003)	✗	✗	✗	✗	✗	✗	✓*
3. Marmel et al. (2008)	?	?	✓	✓	✓	✓	✓
4. Marmel and Tillmann (2009)	?	✓	✓	✗	✓	✓	✓
5. Marmel et al. (2010)	✗	✗	✗	✗	✗	✗	✓
6. Tillmann, Janata, Birk, et al. (2003)	?	✓	✗	✗	✗	✓	✓
7. Tillmann et al. (2008)	?	✗	✗	✗	✗	✗	✓*

Note. The best-performing variables in each space (PP, CV, and TS) can be read from Supplemental Table S2. The stepwise models are specified in (1-2). Key: ✓ = success; ✗ = failure; ? = not tested. *There were borderline-significant p -values of .053 in experiment 2 (Bigand et al., 2003) and .077 for the subdominant-baseline comparison in experiment 7 (Tillmann et al., 2008).

Figure captions

Figure 1. (A) The canonical major-key profile obtained by Krumhansl and Kessler (Krumhansl & Kessler, 1982) in a probe-tone experiment in which participants rated the goodness-of-fit of probe tones that followed a harmonic context such as the authentic cadence (a sequence of chords built on scale degrees IV, V, and I shown in B). The profile shown here applies to the key of C major, and is labeled along the ordinate with the names of the 12 pitch classes. The seven pitch classes that belong to the C major diatonic scale are indicated by the Roman numerals corresponding to the seven scale degrees. (B) Musical notation showing a sequence of chords corresponding to the subdominant (IV), dominant (V), and tonic (I) functions in the key of C major, followed by a silence (rest) and a probe tone that does not belong to the C major diatonic scale. (C) The *circle of fifths* for the major keys (outside set of uppercase labels) and minor keys (inner set of lowercase labels). The proximity of the major and minor keys reflects the *relative minor* relationship in which the major and minor keys share the diatonic pitch class set but not the tonic. The existence of natural, melodic, and harmonic minor variants, complicates the exact relationship between major and minor keys. (D) A toroidal representation of the musical and psychological distance relationships among musical keys. The key labels are positioned based on the training of a self-organizing map using a modulating melody (Janata et al., 2002). The dark gray line has been added to illustrate how the circle of fifths for the major keys wraps around the torus. The gray line represents the circle of fifths for the minor keys. The interlacing of the major and minor circles of fifths allows for multiple major/minor relationships to be respected.

Figure 2. Schematic summary of Leman's (2000) periodicity pitch model (steps 1-3), Janata et al.'s (2002) tonal space model (steps 1-4), and a chroma vector model (steps 1, 2, 5, 6). Red indicates values of high relative intensity and blue indicates low values. The gray background (stages 5–7), indicates an “indirect route” from the periodicity pitch model representation (step 3) to tonal space.

Figure 3. (A) Plot of the local/global context image correlation across time for four melodies from Experiment 5 (Marmel et al., 2010). Instrumental timbre was varied (piano tones = solid line; pure tones = dashed line), as was tonal relatedness (dark gray lines for “related” melodies ending on I; light gray lines for “less related” melodies ending on IV); (B) Plot of the corresponding chroma vector correlation across time for the same four melodies; (C) Plot of the corresponding tonal space correlation across time. Roman numerals on the *x*-axis indicate the scale degrees of the constituent melody notes. Arrows above these plots indicate the positions of the two notes for which the pitch class differed between melodies, thus establishing different keys for the two melodies.

Figure 4. Illustration of the calculation of explanatory variables used in the statistical models of RT data from tonal priming experiments. (A) Staff notation for a stimulus from Tillmann, Janata, and Bharucha (2003); (B) Context image (integrated periodicity pitch images at successive time points) using the 0.1 s time constant for the corresponding stimulus. The grayscale indicates relative intensity from low (white) to high (black); (C) Context image for the same stimulus, using the 4 s time constant; (D) Plot against time of the correlation coefficients calculated for corresponding column vectors from the context

images shown in B and C. The dashed line indicates the correlation coefficient for the stimulus ending on a related C major chord instead of the unrelated B major chord shown in A. Horizontal bars are the windows across which mean correlations are calculated, as explained in the text. The markers at the ends of the bars are for the sake of clarity. The vertical bar in each panel indicates the position of the target (eighth) chord.

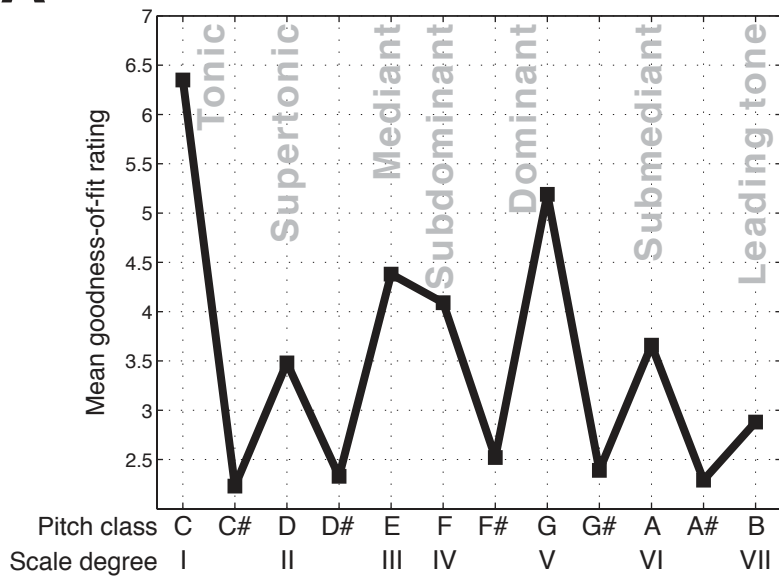
Figure 5. Plots of observed RTs and fitted RTs from the stepwise model (1) across the different categories within each experiment. For each experiment, the RTs are expressed relative to the mean observed RT in that experiment. The error bars indicate 95% confidence intervals. (A) In Experiment 1 (Tillmann, Janata, & Bharucha, 2003), “Rel B” = stimulus in B major that ends on B major (related), “Rel C” = stimulus in C major that ends on C major, “Un B” = stimulus in B major that ends on C major (unrelated), “Un C” = stimulus in C major that ends on B major; (B) In Experiment 2 (Bigand et al., 2003), “I NTC” = stimulus that ends on the tonic chord with No Target chord (tonic) in the Context, “IV NTC” = stimulus that ends on the subdominant chord with No Target chord (subdominant) in the Context; (C) In Experiment 3 (Marmel et al., 2008), “I” = stimulus that ends on the tonic scale degree, “IV” = stimulus that ends on the subdominant scale degree; (D) In Experiment 4 (Marmel & Tillmann, 2009), “III” = stimulus that ends on the mediant scale degree, “VII” = stimulus that ends on the leading tone; (E) In Experiment 5 (Marmel et al., 2010), “Pno I” = stimulus with a piano timbre that ends on the tonic scale degree, “Pno IV” = stimulus with a piano timbre that ends on the subdominant scale degree, “Pure I” = stimulus with a pure-tone timbre that ends on the tonic scale degree, “Pure IV” = stimulus with a pure-tone timbre that ends on the

subdominant scale degree; (F) In Experiment 6 (Tillmann, Janata, Birk, et al., 2003), “I” = standard chord sequence that ends on the tonic, “IV” = standard chord sequence that ends on the subdominant, “BL” = baseline chord sequence that establishes no tonal center; (G) In Experiment 7 (Tillmann et al., 2008) the categories are as in F, with the addition of a dominant category (V).

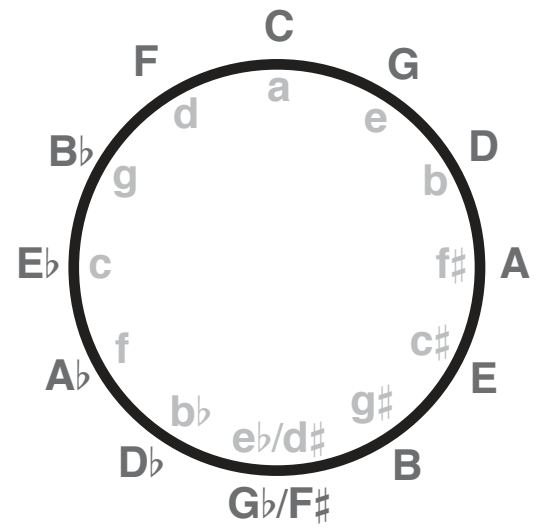
Figure 6. Transition probability matrices used for modeling closure. Each matrix plots the probabilities of observing pairs of mean post-event correlations of the local/global context in tonal space for the final two events in a piece of tonal music. The x -axis corresponds to correlations for the final event (r_n), and the y -axis corresponds to correlations for the penultimate event (r_{n-1}). Probabilities range from zero (white) to a maximal value (black) corresponding to the largest number of observations for a correlation combination normalized by the total number of observations. (A) The hypothesized closure distribution; (B) The empirical closure distribution obtained by analyzing the audio of over 2000 tracks of tonal music; (C) Smoothed and thresholded version of the empirical closure distribution.

Figure 7. Plots of observed and fitted RTs from the stepwise model (2) across different categories within each experiment. The error bars indicate 95% confidence intervals. Category labels are maintained from Figure 5.

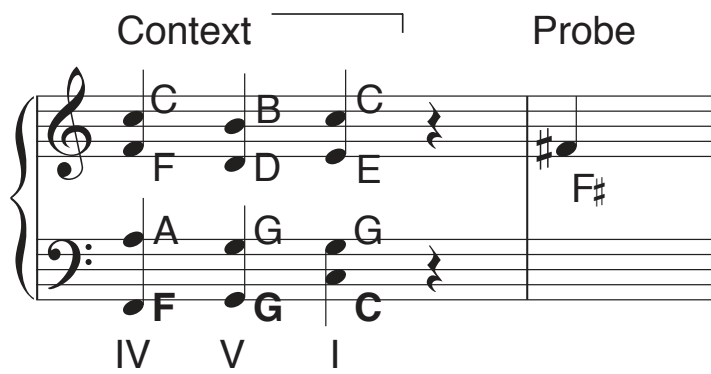
A



C



B



D

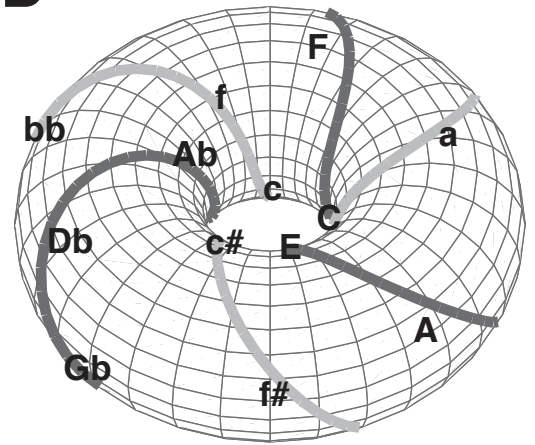


Figure 1

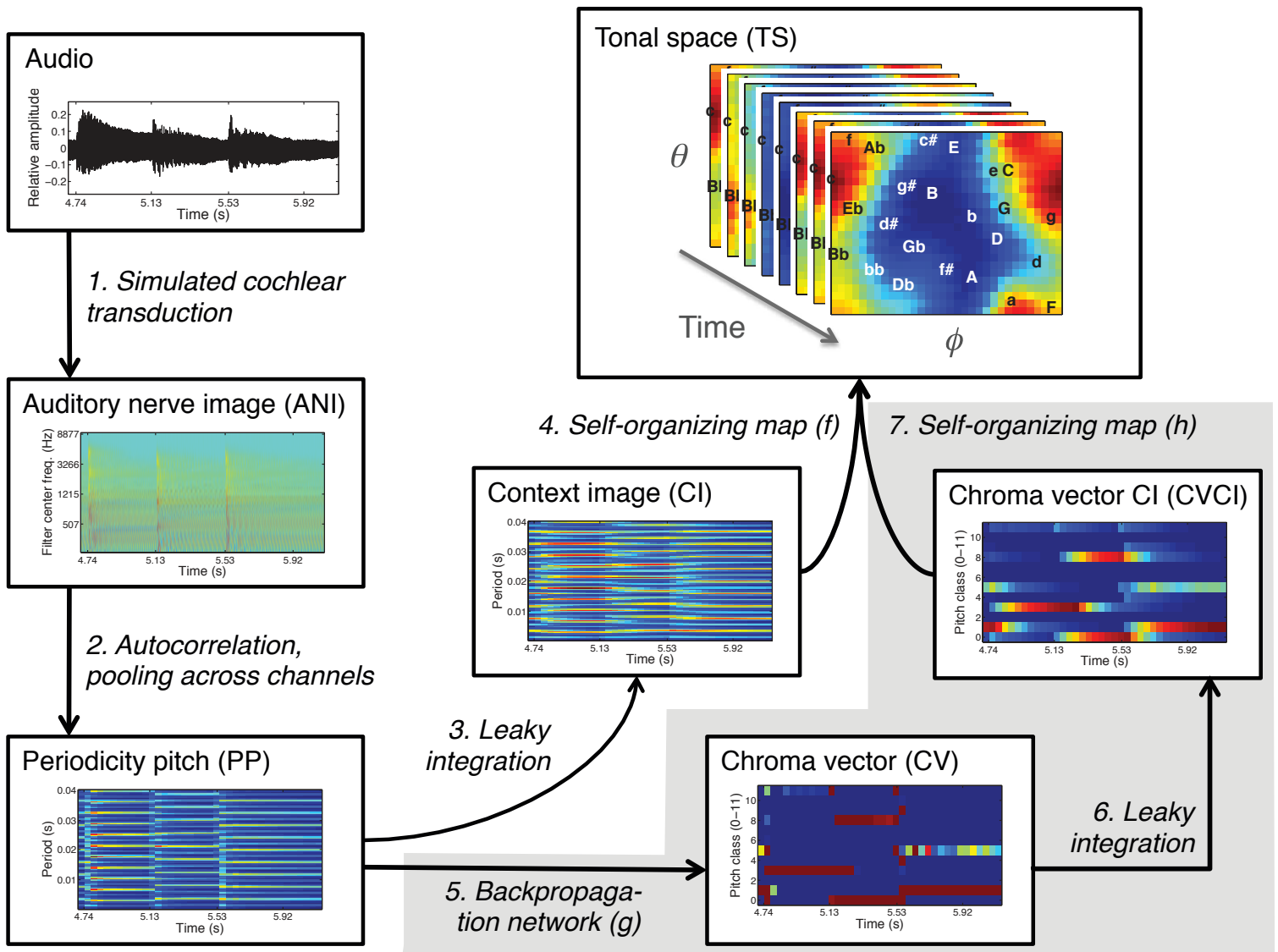


Figure 2

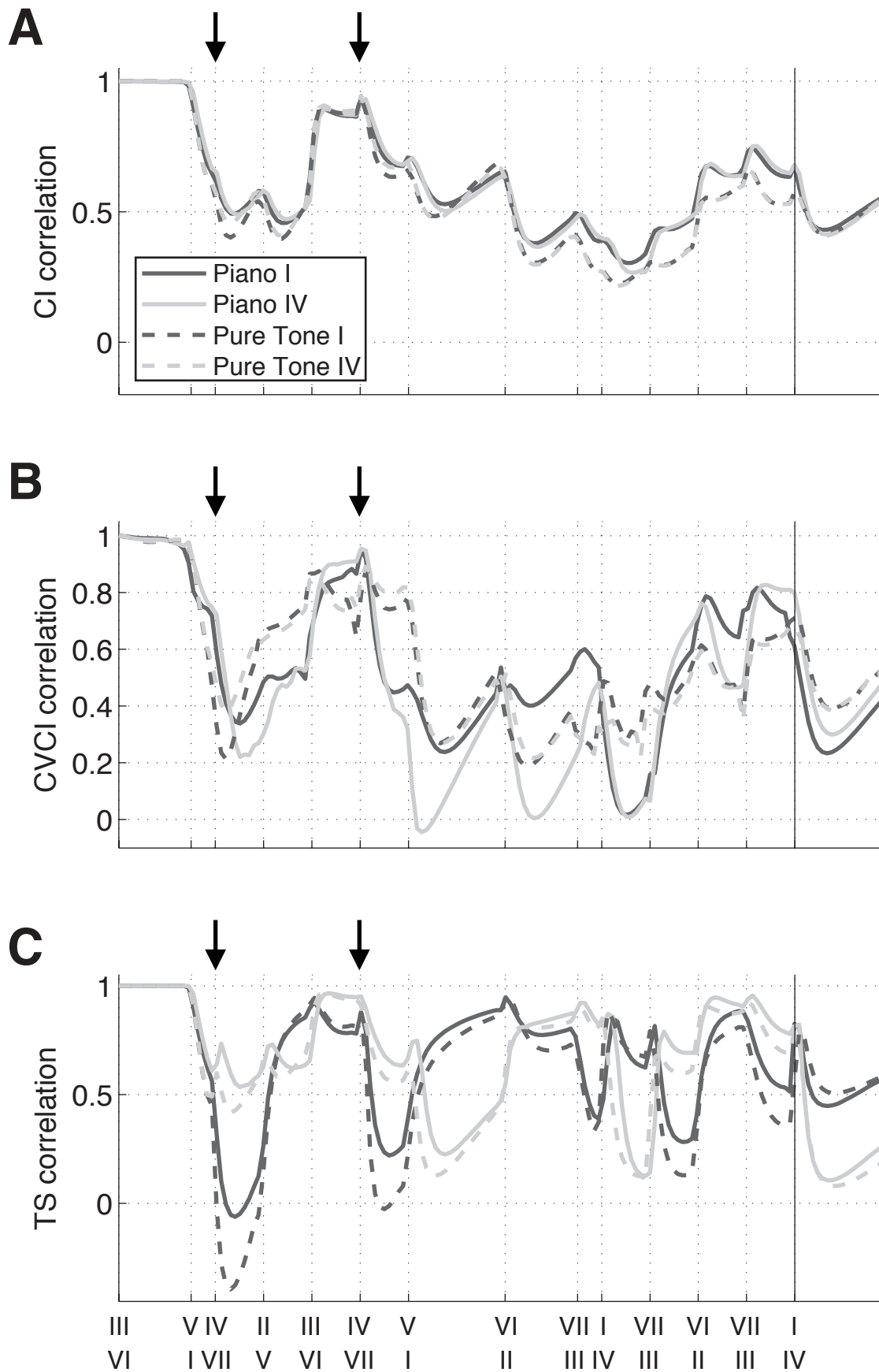


Figure 3

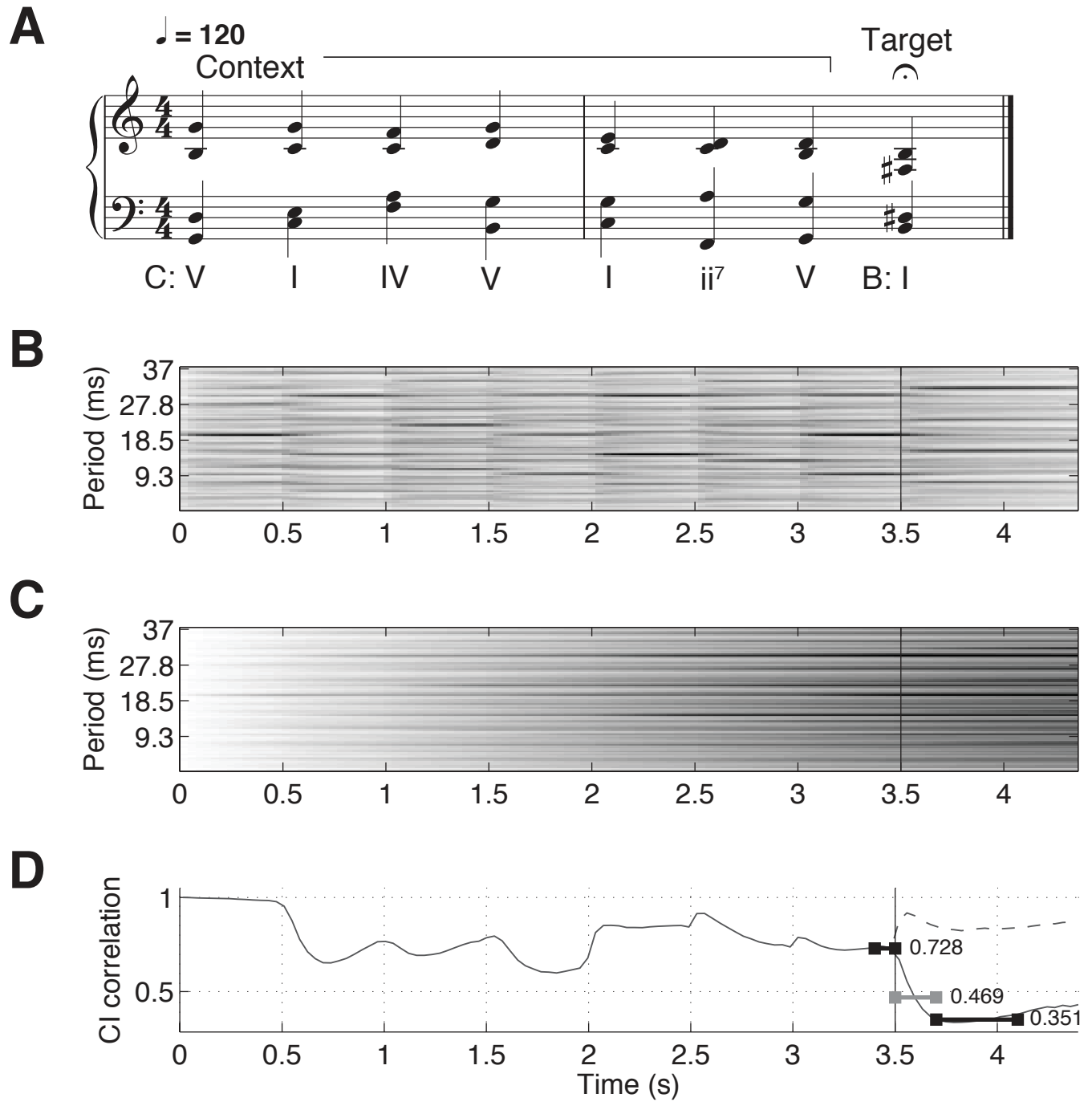


Figure 4

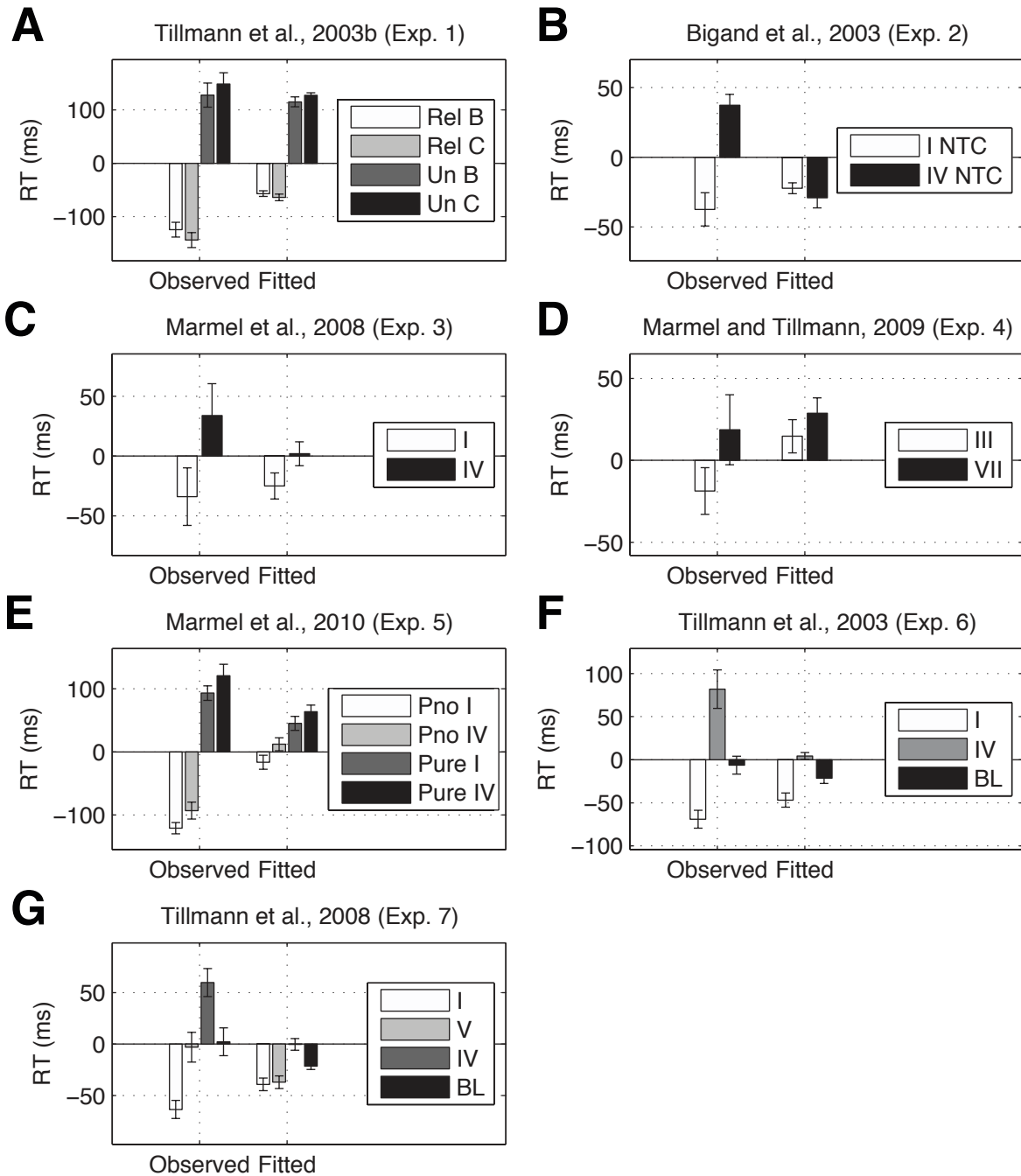
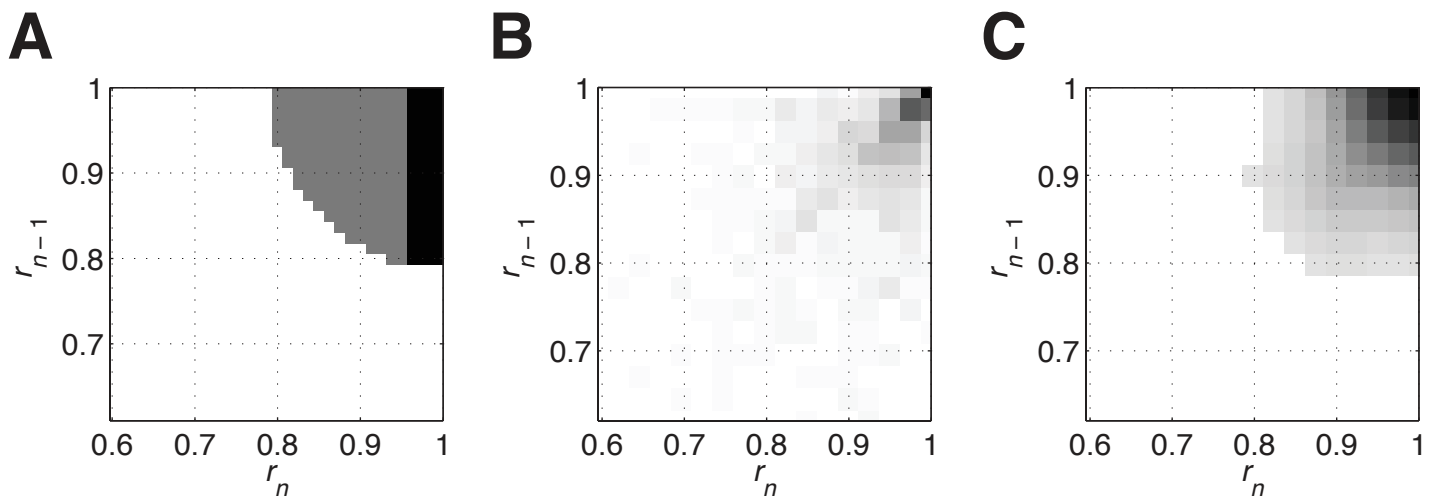


Figure 5

**Figure 6**

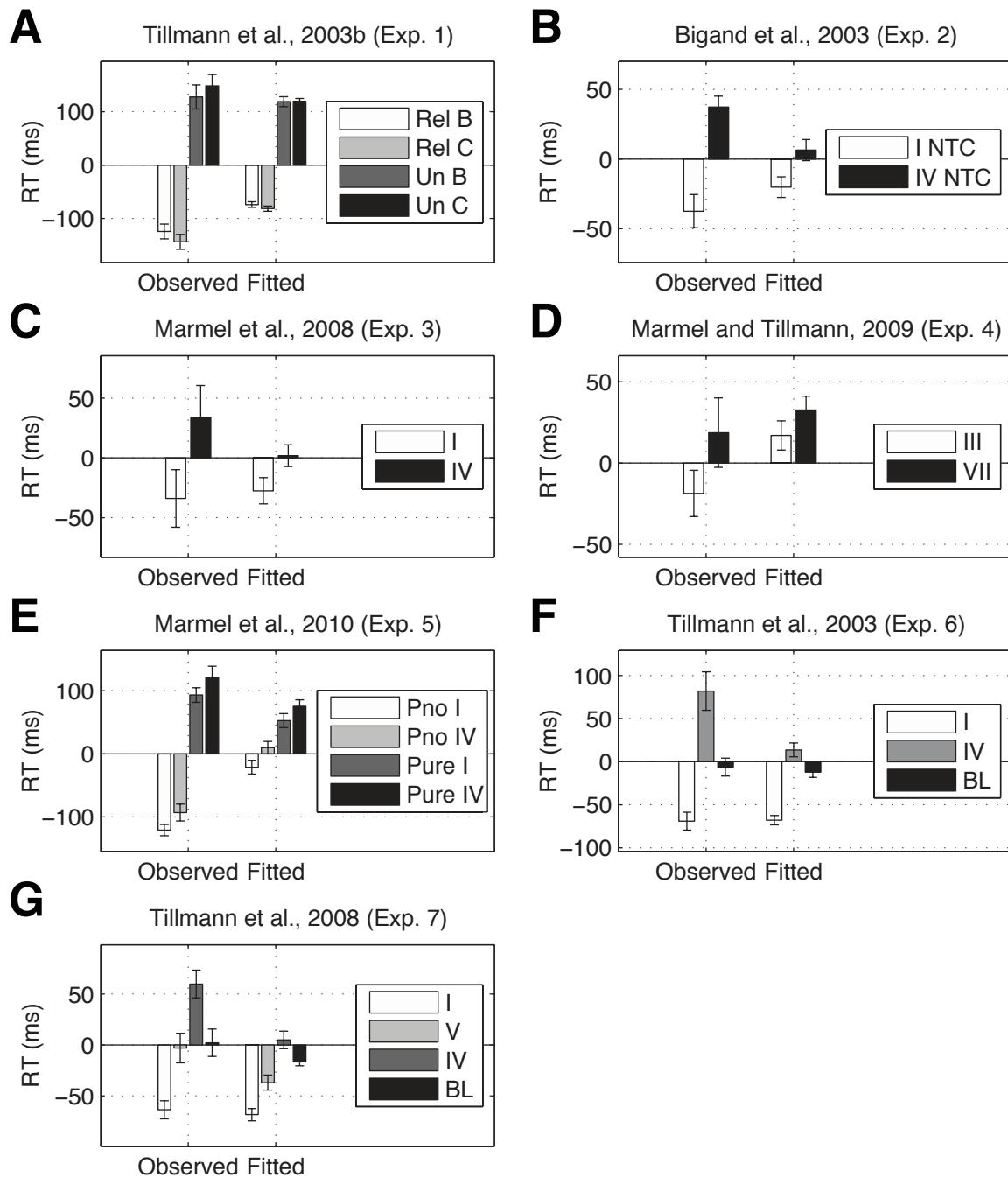


Figure 7

APPENDIX A

THE CHROMA VECTOR (CV) REPRESENTATION

EQUIVALENCE OF TONAL SPACE (TS) REPRESENTATIONS ACHIEVED VIA DIFFERENT PROJECTION ROUTES

To the extent that CVs serve as a basis for projecting to a tonal space representation on the toroidal surface (Toivainen & Krumhansl, 2003), it is of interest to determine whether TS representations achieved via the different projection routes shown in Figure 2 are equivalent. We refer to the route comprising stages 1–4 in Figure 2 as *direct*, and the route comprising stages 1, 2 and 5–7 as *indirect*. Because the PP stage is common to both routes, three maps of interest arise from that stage:

1. Mapping a periodicity pitch vector, \mathbf{u} , to a tonal space matrix, \mathbf{W} , (PP \rightarrow TS; direct route);
2. Mapping a periodicity pitch vector, \mathbf{u} , to a chroma vector, \mathbf{v} , (PP \rightarrow CV; indirect route, step 1);
3. Mapping a chroma vector, \mathbf{v} , to a tonal space matrix, \mathbf{W}' , (CV \rightarrow TS; indirect route, step 2).

Thus the equivalence question can be stated as follows: given a periodicity pitch vector \mathbf{u} , is the tonal space matrix \mathbf{W} obtained via map 1 equivalent to the tonal space matrix \mathbf{W}' , obtained via map 2 to the chroma vector \mathbf{v} , followed by map 3 to tonal space? Formally, we defined a map $f: U \rightarrow W$ from the space U of PP vectors to the space W of TS matrices, a map $g: U \rightarrow V$ from the PP vectors to the space V of CVs, and a map $h: V \rightarrow W$ from the CVs to the TS matrices. For notational simplicity, maps f and h subsume the leaky integration step. Labels f and h in Figure 2 suggest otherwise, but are

intended to help connect the diagram with the text. Map f (the direct route) is embodied in the model described above in *The tonal space model* section. To achieve the indirect route, we had to implement maps g and h .

Projection from a periodicity pitch vector to a chroma vector to tonal space

Map g was accomplished by training a neural net using a supervised learning algorithm (Levenburg-Maquardt backpropagation algorithm in MATLAB's Neural Network Toolbox). The training data consisted of a three-octave chromatic scale and chord types in twelve semitone transpositions (including major, minor, diminished, augmented, major seventh, dominant seventh, minor seventh, half-diminished seventh, and each of the four seventh chords with a natural ninth above it), each rendered in fourteen timbres. This gave a total of 2534 ($= 14 \text{ timbres} \times [37 \text{ scale notes} + 12 \text{ transpositions} \times 12 \text{ chord types}]$) training stimuli. The resulting trained weight matrix that maps a given PP vector to a CV is shown in Figure A1. Banding patterns in the weight matrix reflect harmonics in each of the chroma.

We validated the effectiveness of the CV projection g , by assessing pitch-class detection performance using twenty (ten major, ten minor) four-part chorale harmonizations by Johann Sebastian Bach (1685-1750), and the corresponding soprano melodies as a ground truth.¹ Table A1 summarizes the pitch-class detection performance using evaluation metrics from the Music Information Retrieval Evaluation eXchange (MIREX) Multiple F0 Estimation task (Downie, 2008), which is part of the annually organized comparative evaluation of MIR algorithms on predefined tasks.

¹ The chorale harmonizations, obtained from <http://www.jsbchorales.net>, are listed as BWV2.6, 3.6, 4.8, 5.7, 6.6, 7.7, 9.7, 10.7, 11.6, 13.6, 14.5, 16.6, 17.7, 18.5, 19.7, 20.7, 25.6, 26.6, 29.8, and 30.6.

We implemented map h using the same process as used for map f , that is, presenting CVs, to which leaky integration had been applied, to the self-organizing map algorithm used for map f .

Comparison of tonal space representations attained via direct and indirect routes

We sought to assess the equivalence of mappings f and $g \circ h$. For an arbitrary periodicity pitch vector $\mathbf{u} \in U$, the equivalence of the tonal space matrices $\mathbf{W} = f(\mathbf{u})$ and $\mathbf{W}' = h(g(\mathbf{u}))$ was determined using the pointwise correlation coefficient between matrices \mathbf{W} and \mathbf{W}' , denoted $r(\mathbf{W}, \mathbf{W}')$. In order for $r(\mathbf{W}, \mathbf{W}')$ to be meaningful, we had to ensure that \mathbf{W} and \mathbf{W}' were rotated equivalently, i.e. that the positions of tonal center labels were matched. Different orientations were expected given the different dimensions of the weight matrices associated with f and h and random initial weight assignments in the weight matrices. The matrix \mathbf{W}' was first rotated by using the plane parameterization shown in Figure A1 panels C–F, where ϕ is the angle that a point on the surface makes with the axis passing through the center of the torus and the center of the tube, and θ is the angle that a point makes with the perpendicular axis passing through the center of the tube and the surface. Rotation about these axes in 3D is equivalent to translation in the x - y plane (modulo 2π). The translation by (ϕ', θ') was applied in each case, where ϕ' and θ' were estimated by averaging over angles that produced the maximum correlation between tori, for 24 isolated triads (twelve major and twelve minor) with roots from G3 up to F#4.

Once the appropriate rotation of map h was determined, each of the 303 audio stimuli studied in this paper was processed to give periodicity pitch vectors $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_L$, where L depends on the length of the stimulus. Letting $\mathbf{W}_i = f(\mathbf{u}_i)$ and $\mathbf{W}'_i = h(g(\mathbf{u}_i))$,

where $i = 1, 2, \dots, L$, the mean and standard deviation of the pointwise correlation coefficients $r(\mathbf{W}_1, \mathbf{W}_1')$, $r(\mathbf{W}_2, \mathbf{W}_2')$, \dots , $r(\mathbf{W}_L, \mathbf{W}_L')$ were calculated for each stimulus.

We divided the 303 excerpts into two categories depending on whether they were wholly diatonic or contained chromaticism. For the purposes of this article, an excerpt of music is diatonic if the union of its pitch classes is a subset of at least one major scale, and it is chromatic otherwise. In general the situation is a little more nuanced regarding minor scales, where there are two variants. Minor mode stimuli were not used in the studies considered here, however.

The chromatic stimuli either jumped about the circle of fifths (the "baseline" sequences from Tillmann, Janata, Birk, & Bharucha, 2003; Tillmann, Janata, Birk, & Bharucha, 2008) or began as diatonic progressions but ended out of key (Tillmann, Janata, & Bharucha, 2003). The mean correlation for diatonic stimuli was .86 ($SD = 0.11$), and the mean correlation for chromatic stimuli was 0.74 ($SD = 0.14$). Applying Fisher's transformation and a two-sample t -test, this is a highly significant difference ($t(301) = 9.82, p < .001$).

The correlations for chromatic stimuli were lower than for diatonic stimuli, and inspection of the correlation time courses revealed that there were correlation troughs after locally chromatic chord progressions. Figure A1B shows the correlation time courses between direct and indirect projection routes for a diatonic stimulus (solid line) and its paired chromatic stimulus (dashed line) from Tillmann et al. (2008). Tonal space matrices were extracted for the two projection routes and the two stimuli, at the timepoint indicated by the gray vertical line, and are plotted in Figure A1C–F. The TS activations are very similar for the diatonic stimulus (Figure A1C, E). Although the details of the

topography are disparate for the chromatic stimulus (Figure A1D, F), giving rise to the low correlation, they are similar in that the activation values across the image occupy a narrow sub-range of the overall range observed for the diatonic stimulus. From the perspective of modeling tonality, low correlations between projection routes for a highly chromatic stimulus are not surprising: chromaticism is to the establishing of a tonal center what noise is to establishing a clear signal. It is interesting to note that these simulations provide a way of illustrating that the composition of the baseline sequences was effective in preventing the build-up of a clear tonal center.

To check whether one projection route held more predictive power than the other when it came to modeling RTs to chromatic stimuli, we performed two sets of linear regressions: one regressing mean RT against explanatory variables calculated from the direct route to tonal space; the other regressing the same mean RTs against explanatory variables from the indirect route to tonal space. The explanatory variables from the two projection routes are positively correlated (mean $r = .85$, $SD = 0.11$). That is, even though the absolute morphology of the tonal space activation is different, the values of the explanatory variables, which are primarily based on relative measures, are highly similar. Whether the chromatic nonhomogeneity can serve any purpose (e.g., as a cognitive marker of chromaticism) remains to be seen, but it is an interesting empirical finding. Absent the superiority of either route in explaining RTs to chromatic stimuli, we used direct route projections to tonal space for our modeling work, given our use of the direct projection route in prior work (Janata, 2005, 2007, 2009; Janata et al., 2002).

DISCUSSION OF CV MODELING OUTCOMES

The relative inability of the CV variables to explain variance in RTs with the efficacy shown by their PP and TS counterparts is all the more remarkable given our ability to show: (1) the backpropagation network indicated in Figure 2 and Figure A1A provides a robust PP-CV mapping, with 83% accuracy for detecting the pitch class of melodic audio, and 76% accuracy for four-part chorale harmonizations; (2) functional equivalence, at the TS level, of periodicity pitch information projected either directly to TS or indirectly to TS via CVs when the chord sequences established a clear tonal center.

An important difference between the CV and TS representations is their compactness. A CV is a compact space consisting of 12 discrete elements, whereas TS is a surface with the number of elements determined by the sampling density, in our case 768 elements distributed in a 24×32 grid. In this regard, the TS representation is less parsimonious. However, the TS effectively represents a distribution of CVs. When applying a self-organizing algorithm to a large set of CVs comprising the pitch distributions in a corpus of music that spans all major and minor keys, each location on the torus comes to represent a slightly different CV probability distribution than its neighbors. Distributing CV probability distributions across a broader space allowed the variables obtained by correlating temporally local with temporally global activation distributions to be better able to model RTs when those variables were calculated in TS rather than for CVs.

To obtain a different perspective on the structure of the CV and TS representations that might explain why correlation measures applied in each of the spaces behave differently, we calculated pairwise correlations between activation time points

(using a leaky integration time constant of 4 s) for the PP, CV, and TS spaces drawn from a modulating melody originally used to train the projection from PP space to TS (Janata, Birk, Tillmann, & Bharucha, 2003; Janata et al., 2002). Figure A2 shows that the effect of projecting from PP space to either CV space or TS is a broadening of the distributions of correlations such that TS contains a much greater incidence of activation patterns that are strongly positively correlated or strongly negatively correlated with each other. The latter activation patterns correspond to keys that are at opposing sides of the circle of fifths, e.g., C major and F \sharp major. Figures A2B and A2C show the joint distributions of correlations between PP and TS and between CV and TS calculated for the same time points, thus providing an illustration of how the correlation between a pair of time points might be transformed when it is projected into TS. Both the transforms from PP space to TS and from CV space to TS are sigmoidal. Thus if two periodicity pitch or chroma vectors are positively correlated, then the correlation between the corresponding tonal space matrices is likely to be more positive still (similarly, more negative for weaker positive correlations in the case of PP space or for negative correlations in CV space). For explaining response times, the greater predictive strength of correlation-based variables derived from TS (compared to either PP or CV) appears to be associated with the polarization of correlation values in TS. Although a full exploration of the reasons for the differences in distributions is beyond the scope of this paper, we suspect it is a consequence of the different neighborhood relationships present in the different geometries of the three representational spaces.

Table A1. *Pitch-class detection accuracy, precision, and recall for chorale harmonizations*

Test set	Accuracy	Precision	Recall
Melodies	.83	.83	1
Four parts	.76	.91	.82
Both	.80	.87	.91

Note. Using the abbreviations for true positive (TP), false positive (FP), true negative (TN), and false negative (FN) for whether ground truth elements were identified successfully or not, accuracy = $(TP + TN)/(TP + TN + FP + FN)$, precision = $TP/(TP + FP)$, and recall = $TP/(TP + FN)$.

FIGURE CAPTIONS

Figure A1. (A) Trained weight matrix for map g from the space of periodicity pitch vectors to the space of chroma (pitch class) vectors. The color map varies from blue (low intensity) to red (high intensity); (B) Correlation time courses between tonal space matrices for a diatonic stimulus (solid line) from Tillmann et al. (2008), and the paired chromatic stimulus (dashed line). The gray vertical line indicates the time at which tonal space matrices were extracted and plotted in C–F; (C) Tonal space matrix for the diatonic stimulus, obtained via the direct route (steps 1–4 in Figure 2); (D) Tonal space matrix for the chromatic stimulus, obtained via the direct route; (E) Tonal space matrix for the diatonic stimulus, obtained via the indirect route (steps 1, 2, 5–7 in Figure 2); (F) Tonal space matrix for the chromatic stimulus, obtained via the indirect route. The color map indicates relative intensity ranging from low (blue) to high (red).

Figure A2. Distributions of correlations of activation patterns within leaky-integrated ($t = 4$ s) periodicity pitch (PP), chroma vector (CV), and corresponding tonal space (PP_TS; CV_TS) representations. Correlations were computed between pairs of time points in a melody that modulates through all major and minor keys (Janata et al., 2003). Twenty percent of the melody was sampled at random. (A) Distributions of correlations within each representational space. The ordinate indicates the proportion of the total number of measured correlations. Thus, the area under each of the curves sums to 1. (B) Joint distribution of corresponding correlations in periodicity pitch and tonal spaces. (C) Joint distribution of corresponding correlations in chroma vector and tonal spaces. The gray colormap in B and C indicates the proportion of observed correlations, increasing from white to black.

References cited in Appendix A

Downie, J. S. (2008). The music information retrieval evaluation exchange (2005–2007):

A window into music information retrieval research. *Acoustical Science and Technology*, *29*(4), 247-255.

Janata, P. (2005). Brain networks that track musical structure. *Annals of the New York Academy of Sciences*, *1060*(1), 111-124.

Janata, P. (2007). Navigating tonal space. In W. B. Hewlett, E. Selfridge-Field & E. Correia (Eds.), *Computing in musicology: Tonal theory for the digital age* (Vol. 15, pp. 39–50). Stanford: Center for Computer Assisted Research in the Humanities.

Janata, P. (2009). The neural architecture of music-evoked autobiographical memories. *Cerebral Cortex*, *19*, 2579-2594. doi: [10.1093/cercor/bhp008](https://doi.org/10.1093/cercor/bhp008)

Janata, P., Birk, J. L., Tillmann, B., & Bharucha, J. J. (2003). Online detection of tonal pop-out in modulating contexts. *Music Perception*, *20*(3), 283-305.

Janata, P., Birk, J. L., Van Horn, J. D., Leman, M., Tillmann, B., & Bharucha, J. J. (2002). The cortical topography of tonal structures underlying Western music. *Science*, *298*(5601), 2167-2170.

Tillmann, B., Janata, P., & Bharucha, J. J. (2003). Activation of the inferior frontal cortex in musical priming. *Cognitive Brain Research*, *16*, 145-161.

- Tillmann, B., Janata, P., Birk, J. L., & Bharucha, J. J. (2003). The costs and benefits of tonal centers for chord processing. *Journal of Experimental Psychology: Human Perception and Performance*, 29(2), 470-482.
- Tillmann, B., Janata, P., Birk, J. L., & Bharucha, J. J. (2008). Tonal centers and expectancy: facilitation or inhibition of chords at the top of the harmonic hierarchy? *Journal of Experimental Psychology: Human Perception and Performance*, 34(4), 1031-1043.
- Toiviainen, P., & Krumhansl, C. L. (2003). Measuring and modeling real-time responses to music: The dynamics of tonality induction. *Perception*, 32(6), 741-766.

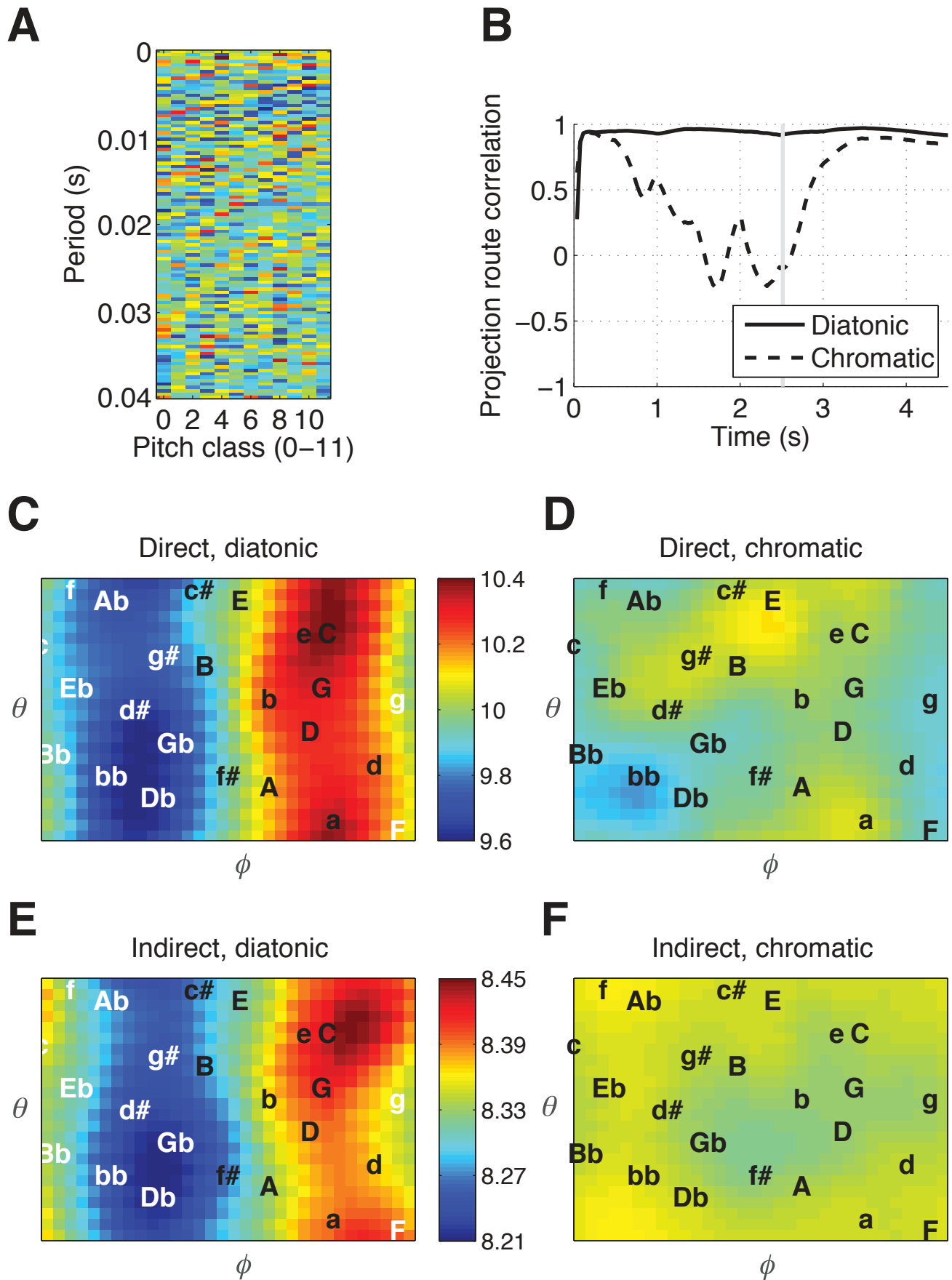


Figure A1

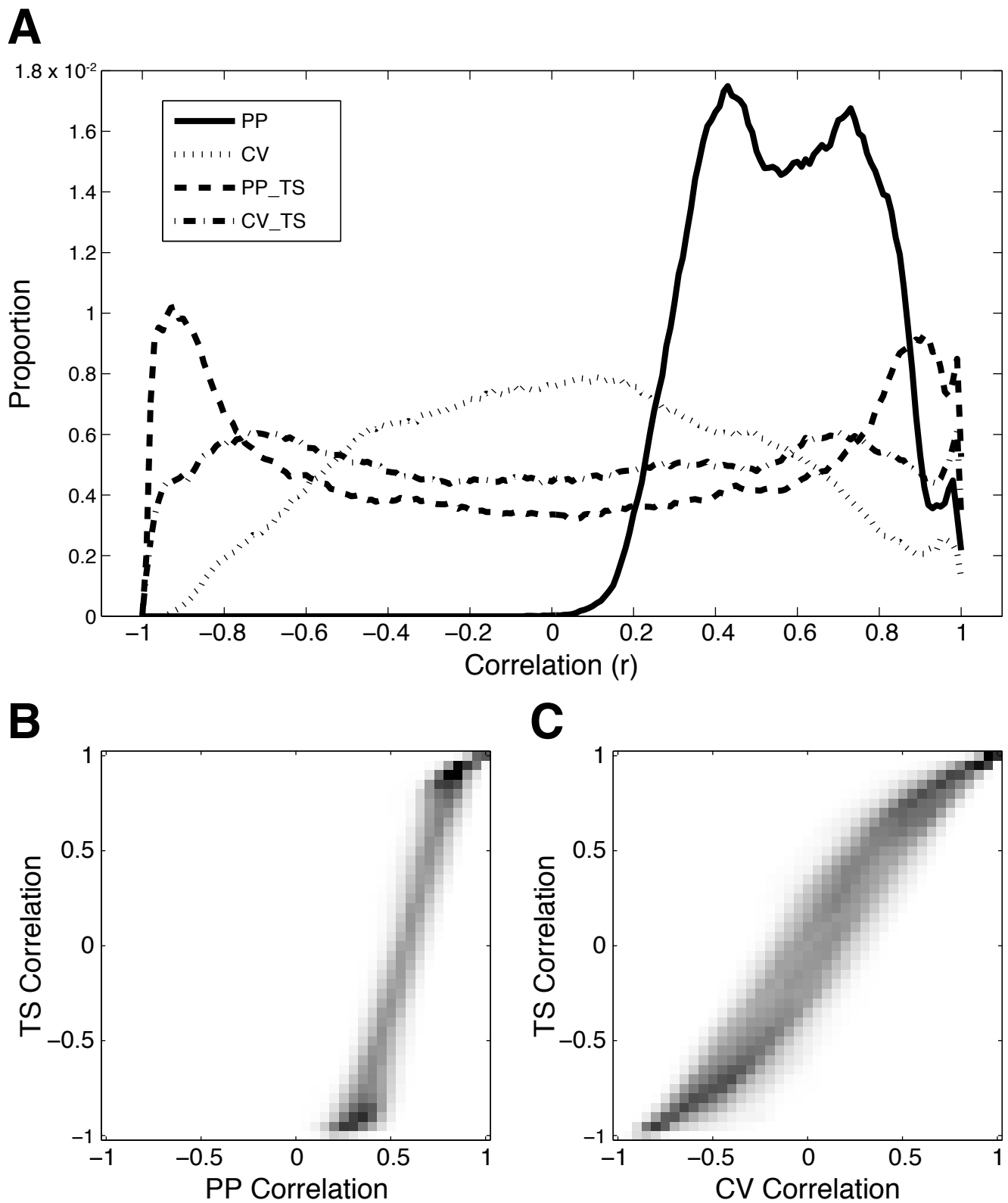


Figure A2

APPENDIX B

SUMMARY OF THE STEPWISE SELECTION PROCESS UNDERLYING

EQUATION 1

Variable	<i>B</i>	<i>SE B</i>	<i>b</i>	<i>R</i> ²
Step 1 (addition), 3 strongest out-of-model predictors were:				
$x_{TS} = \{\text{TS, MC, rel, early}\}$	-142.66	15.46	-51.76***	.22
$\{\text{PP, MC, abs, early}\}$	-392.18	43.34	-50.98***	.21
$\{\text{TS, MC, abs, late}\}$	-106.16	12.35	-48.94***	.20
<i>Outcome: add x_{TS} to model.</i>				
Step 2 (elimination), weakest in-model predictor was:				
$x_{TS} = \{\text{TS, MC, rel, early}\}$	-142.66	15.46	-51.76***	.00
<i>Outcome: keep x_{TS} in model as it is significant.</i>				
Step 3 (addition), 3 strongest out-of-model predictors were:				
$x_{PP} = \{\text{PP, MC, rel, early}\}$	-231.07	55.24	-30.04***	.26
$\{\text{PP, MC, abs, early}\}$	-117.49	39.39	-18.33***	.24
$\{\text{PP, MC, rel, late}\}$	-117.54	42.03	-20.55**	.24
<i>Outcome: add x_{PP} to model.</i>				
Step 4 (elimination), 2 weakest in-model predictors were:				
$x_{PP} = \{\text{PP, MC, rel, early}\}$	-231.07	55.24	-30.04***	.22
$x_{TS} = \{\text{TS, MC, rel, early}\}$	-88.93	19.79	-32.27***	.21
<i>Outcome: keep x_{PP} in model as it is significant.</i>				
Step 5 (addition), 3 strongest out-of-model predictors				
$\{\text{PP, MC, rel, late}\}$	219.83	97.74	38.43*	.28
$y_{PP} = \{\text{PP, MV, abs, late}\}$	0.05	0.02	11.68*	.27
$\{\text{CV, MC, rel, late}\}$	49.42	26.51	15.56	.27
<i>Outcome: add $\{\text{PP, MC, rel, late}\}$ to model.</i>				
Step 6 (elimination), 3 weakest in-model predictors were:				
$\{\text{PP, MC, rel, late}\}$	219.83	97.74	38.43*	.26
$x_{PP} = \{\text{PP, MC, rel, early}\}$	-501.74	132.26	-65.22***	.24
$x_{TS} = \{\text{TS, MC, rel, early}\}$	-94.16	19.79	-34.16***	.22
<i>Outcome: keep $\{\text{PP, MC, rel, late}\}$ in model as it is significant.</i>				
Step 7 (addition), 3 strongest out-of-model predictors				
$y_{PP} = \{\text{PP, MV, abs, late}\}$	0.06	0.03	12.68*	.29
$\{\text{PP, MV, abs, early}\}$	0.05	0.03	11.66*	.29
$\{\text{CV, MV, rel, early}\}$	3.48	2.33	10.01	.28
<i>Outcome: add $y_{PP} = \{\text{PP, MV, abs, late}\}$ to model.</i>				

Table continued overleaf

Table B1 continued				
Variable	<i>B</i>	<i>SE B</i>	<i>b</i>	<i>R</i> ²
Step 8 (elimination), 3 weakest in-model predictors were:				
$y_{PP} = \{PP, MV, abs, late\}$	0.06	0.03	12.68*	.28
$\{PP, MC, rel, late\}$	236.84	97.30	41.41**	.27
$x_{PP} = \{PP, MC, rel, early\}$	-508.	131.33	-66.09 **	.25
<i>Outcome: keep $y_{PP} = \{PP, MV, abs, late\}$ in model as it is significant.</i>				
Step 9 (addition), 3 strongest out-of-model predictors were:				
$y_{CV} = \{CV, MV, abs, late\}$	-3.50	1.29	-21.22**	.31
$\{CV, MV, rel, late\}$	5.40	2.40	15.53*	.30
$\{TS, MV, abs, late\}$	$-1.53 \cdot 10^5$	$7.64 \cdot 10^4$	-19.16*	.30
<i>Outcome: add $y_{CV} = \{CV, MV, abs, early\}$ to model.</i>				
Step 10 (elimination), 3 weakest in-model predictors were:				
$\{PP, MC, rel, late\}$	154.15	100.96	26.95	.30
$y_{CV} = \{CV, MV, abs, early\}$	-3.50	1.29	-21.22**	.29
$x_{PP} = \{PP, MC, rel, early\}$	-430.	133.08	-55.94**	.28
<i>Outcome: eliminate $\{PP, MC, rel, late\}$ from model as it is not significant.</i>				
Step 11 (addition), 3 strongest out-of-model predictors were:				
$\{PP, MC, rel, late\}$	154.15	100.96	26.95	.31
$\{TS, MC, rel, late\}$	47.56	33.65	25.43	.31
$\{CV, MV, abs, late\}$	5.24	3.74	29.10	.31
<i>Outcome: predictors not significant, end stepwise selection.</i>				

Note. Please see Supplemental Table S1 for an explanation of the groupings of variable names in curly brackets. For the addition steps, R^2 is the amount to which proportion of variance explained would rise if the variable is added. For the elimination steps, R^2 is the amount to which proportion of variance explained would fall if the variable is eliminated. * $p < .05$. ** $p < .01$. *** $p < .001$. Abbreviations: *PP* – periodicity pitch; *CV* – chroma vector; *TS* – tonal space; *MC* – mean correlation; *MV* – maximum value; *abs* – absolute; *rel* – relative

SUPPLEMENTAL MATERIAL ACCOMPANYING, “A COMBINED
MODEL OF SENSORY AND COGNITIVE REPRESENTATIONS
UNDERLYING TONAL EXPECTATIONS IN MUSIC: FROM AUDIO
SIGNALS TO BEHAVIOR,” BY COLLINS, TILLMANN, BARRETT,
DELBÉ, AND JANATA.

CALCULATING EXPLANATORY VARIABLES FROM PERIODICITY PITCH,
CHROMA VECTOR, AND TONAL SPACE REPRESENTATIONS

For our statistical analyses of the RT data, explanatory variables were derived from each of the representational spaces (PP, CV, and TS). Each explanatory variable refers to a combination of attributes (variable classes) shown in Supplementary Table S1. These attribute values give rise to the sets of abbreviations used to label the explanatory variables.

Table S1. *Four attributes involved in calculating variables from processed audio*

Attribute	Options	Labels	x_{PP} , x_{CV} , and x_{TS}	y_{PP}	y_{CV}	z_{PP}
1. Representational space	Periodicity pitch, chroma vector, or tonal space	PP, CV, TS	PP, CV, TS, respectively	PP	CV	PP
2. Calculation type	Mean correlation (MC) or maximum value (MV)	MC, MV	MC	MV	MV	MC
3. Window comparison	Post (absolute) or pre-post (relative)	abs, rel	rel	abs	abs	abs
4. Post-target window (ms)	[0, 200] (early) or [201, 600] (late)	early, late	early	late	early	early

Note. The last four columns contain examples of attribute combinations that give rise to variables described in the *Regression results*. For instance x_{PP} arises from the combination {PP, MC, rel, early}.

Representational space

Representational space refers to whether a periodicity pitch (PP), chroma vector (CV) or tonal space (TS) representation is used as the basis for the variable.

Calculation type

Calculation type indicates whether a variable is based on the *mean correlation (MC)* of local (0.1 s time constant) and global (4 s time constant) context images within a specific time window, or instead on the *maximum value (MV)* of the mean image in a specific time window (using only the 4 s time constant constant). The rationale for using the maximum value in a mean image is that it provides a coarse estimate of the degree to which the energy distribution is concentrated within a specific region of the representational space. This information might be related to the clarity of a tonal center, in which a high MV would indicate a clearly established tonal center, while a low MV would indicate that the map is more uniformly activated at a lower amplitude. Note that one can attain a high MC independently of the shape of the distribution.

Window comparison

We use the term “relative” (*rel*) to refer to the difference between pre- and post-target values, and “absolute” (*abs*) to refer to the post-target values only. The pre-target window was 100 ms in duration, and is indicated by the black horizontal bar to the left of the vertical line in Figure 4D (value .728). In Figure 4D we show an early post-target correlation of .469 (gray horizontal bar) and a late post-target correlation of .351 (black horizontal bar to the right of the gray bar). The relative window comparison presumes that responses to target events are driven not as much by the absolute level of activation (either .469 or .351 in this case), but rather by the amount of change that the activation state of the target event represents from the activation state immediately preceding the onset of the target, that is, by a local differencing operation. In Figure 4D, the change is $-.259 \approx .469 - .728$ for the early post-target window, and $-.377 \approx .351 - .728$ for the late window. By contrast, the absolute readout implies that the absolute state of the model at each moment is the relevant parameter that listeners are tracking.

Post-target window

We used two post-target windows. An *early* window spanned 0 – 200 ms following target onset, while a *late* window spanned 201 – 600 ms. Initially, the late window was split into two sections (201 – 400 and 401 – 600 ms), but collapsed subsequently into a single window due to high correlations.

In total we considered a pool of 24 (= 3 representational spaces \times 2 calculation types \times 2 window comparisons \times 2 post-target windows) variables. As an example of the abbreviation system used in Table S2, the last variable discussed in relation to Figure 4D, taking the value -0.377 , is labeled {PP, MC, rel, late}. It comprised the mean correlation between two ($t = 0.1$ s and $t = 4$ s) leaky-integrated PP images (hence PP, MC), using values from either side of the target onset (relative or rel), and a window of 201-600 ms following the target onset (late).

The correlation structure of the matrix containing the 24 explanatory variables for all of the 303 stimuli was rank deficient, indicating the presence of highly correlated variables. These were general pairs of early/late variables. For multiple regression analyses that simultaneously estimated the variance associated with all variables in the model, we removed the *late* variable of a pair. A reduced set of 17 explanatory variables was thus created. To estimate the maximum value of R^2 that can be achieved for this RT data, a model was fitted consisting of the entire reduced set of 17 explanatory variables. For this model, $R^2 = .33$, $s = 93.10$. This model established an upper boundary for all analyses.

STEPWISE SELECTION

To address which weighted sum of variables, potentially drawn from different representational stages, was best able to model the observed RTs, we performed stepwise selection that began with the original set of 24 explanatory variables. Stepwise selection begins by comparing univariate models. A univariate model contains one explanatory variable. The univariate model that most reduces the residual sum of squares (RSS) becomes the stepwise model (if its explanatory variable is significant at the .05 level). The results of the first stage of stepwise selection are shown in Supplementary Table S2, and afford the interested reader the opportunity to see the explanatory power of each variable used in isolation. In the next stage, the least contributing variable in the model is

removed, unless it is significant. Next, models are formed by adding each remaining variable, in turn, to the stepwise model. The stepwise model is updated to include the remaining variable that most reduces the RSS (if significant at the .05 level). The process of adding and removing variables is repeated until no further changes can be made according to these rules. Tables similar in format to Supplementary Table S2 are constructed for each stage of stepwise selection, and the final table showing the addition and elimination of variables that resulted in Equation 1 is provided in Appendix B.

Table S2. *Individual fittings for the first stage of stepwise selection, in descending order of R^2*

Variable	B	$SE B$	b	R^2
$x_{TS} = \{TS, MC, rel, early\}^{\dagger\dagger}$	-142.66	15.46	-51.76***	.22
$\{PP, MC, abs, early\}^{\dagger}$	-392.18	43.34	-50.98***	.21
$\{TS, MC, abs, late\}^{\dagger}$	-106.16	12.35	-48.94***	.20
$\{TS, MC, rel, late\}$	-86.93	10.78	-46.48***	.18
$\{PP, MC, rel, late\}$	-259.39	33.13	-45.35***	.17
$\{TS, MV, rel, early\}^{\dagger}$	$-2.87 \cdot 10^6$	$3.73 \cdot 10^5$	-44.70***	.16
$\{TS, MC, abs, early\}^{\dagger}$	-113.03	15.31	-43.17***	.15
$\{TS, MV, rel, late\}^{\dagger}$	$-9.83 \cdot 10^5$	$1.33 \cdot 10^5$	-43.13***	.15
$z_{PP} = \{PP, MC, abs, late\}$	-248.28	34.66	-42.07***	.15
$x_{CV} = \{CV, MC, rel, early\}^{\dagger\dagger}$	-161.73	24.57	-39.10***	.13
$x_{PP} = \{PP, MC, rel, early\}^{\dagger\dagger}$	-239.28	38.31	-37.34***	.11
$\{CV, MC, rel, late\}$	-105.57	19.24	-33.24***	.09
$\{CV, MC, abs, late\}^{\dagger}$	-118.50	21.98	-32.71***	.09
$\{CV, MC, abs, early\}^{\dagger}$	-113.28	22.98	-30.13***	.07
$\{PP, MV, rel, early\}^{\dagger}$	-0.91	0.22	-24.73***	.05
$\{PP, MV, rel, late\}^{\dagger}$	-0.31	0.08	-24.48***	.05
$\{CV, MV, rel, early\}^{\dagger}$	-28.24	7.34	-23.90***	.05
$\{CV, MV, rel, late\}^{\dagger}$	-7.73	2.17	-22.22***	.04
$\{PP, MV, abs, early\}^{\dagger\dagger}$	0.06	0.03	14.17*	.02
<hr style="border-top: 1px dashed black;"/>				
$y_{PP} = \{PP, MV, abs, late\}$	0.04	0.03	9.69	.01
$\{TS, MV, abs, early\}^{\dagger}$	$6.37 \cdot 10^4$	$4.87 \cdot 10^4$	8.30	.01
$y_{CV} = \{CV, MV, abs, early\}^{\dagger}$	0.59	1.05	3.57	.00
$\{CV, MV, abs, late\}$	-0.51	1.14	-2.83	.00
$\{TS, MV, abs, late\}$	$-9.77 \cdot 10^3$	$5.08 \cdot 10^4$	-1.22	.00

Note. * $p < .05$. *** $p < .001$. The dotted indicates a cut-off point above which variables are significant at the .05 level. One dagger \dagger indicates a member of the reduced set of seventeen variables; two daggers indicate significance in this model ($p < .05$).

EXAMPLES OF STIMULUS MATERIALS

Three hundred and three stimuli were obtained from seven tonal priming experiments. Several stimulus examples, highlighting manipulations of interest, are illustrated in musical notation in Figures S1–S4.







<p>Condition ending on I, Related to context in C (Rel C)</p>	<p>Rel C</p> 
<p>$\text{♩} = 120$</p> 	<p>Un C</p>  <p>VII</p>
<p>Condition ending on VII, Unrelated to context in C (Un C)</p>	<p>Rel B</p> 
<p>Condition ending on I, Related to context in B (Rel B)</p>	<p>Un B</p>  <p>♭II</p>
	
<p>Condition ending on ♭II, Unrelated to context in B (Un B)</p>	

Figure S1. Stimuli from Tillmann, Janata, and Bharucha (2003). “Rel B” = stimulus in B major that ends on B major (related), “Rel C” = stimulus in C major that ends on C major, “Un B” = stimulus in B major that ends on C major (unrelated), “Un C” = stimulus in C major that ends on B major.

Condition ending on I,
No Target in Context (NTC)

$\text{♩} = 100$

Condition ending on IV,
No Target in Context (NTC)

Figure S2. Stimuli from Bigand, Poulin, Tillmann, Madurell, and D'Adamo (2003a), demonstrating chord sequences with tonic and subdominant endings. Exclusion of target chords from the context is one control for sensory influences. “I NTC” = stimulus that ends on the tonic chord with No Target chord (tonic) in the Context, “IV NTC” = stimulus that ends on the subdominant chord with No Target chord (subdominant) in the Context.

A $\text{♩} = 76$

B $\text{♩} = 76$

C $\text{♩} = 76$

Figure S3 displays three sets of musical stimuli (A, B, and C) in 4/4 time, with a tempo of 76 beats per minute. Each set consists of two staves of music. Stimuli A and C are in a key with three flats (E-flat major/C minor). Stimuli B are in a key with one flat (F major/D minor). The stimuli are labeled with Roman numerals indicating their ending scale degrees: I (tonic), IV (subdominant), III (mediant), and VII (leading tone). In stimulus A, the bottom staff has some notes marked with a circled 'q'. In stimulus B, the bottom staff has some notes marked with a circled 'b'. In stimulus C, the bottom staff has some notes marked with a circled 'b'.

Figure S3. (A) Stimuli from Marmel, Tillmann, and Dowling (2008), demonstrating melodies with tonic and subdominant endings. “I” = stimulus that ends on the tonic scale degree, “IV” = stimulus that ends on the subdominant scale degree. (Adapted from Marmel et al. (2008), Figure 1A) (B) Stimuli from Marmel and Tillmann (2009), demonstrating melodies with mediant and leading tone endings. “III” = stimulus that ends on the median scale degree, “VII” = stimulus that ends on the leading tone; (C) Stimuli from Marmel et al. (2010), demonstrating melodies with tonic (I) and subdominant (IV) endings. The extra condition here, compared with Marmel et al. (2008), was to investigate the effect of a piano timbre versus a pure-tone timbre. Thus, the melodies shown were rendered in each of these timbres.

A $\text{♩} = 120$

B

C $(\text{♩} = 120)$

Figure S4. (A) Stimuli from Tillmann et al. (2008) demonstrating a chord sequence with a tonic ending “I”, and a baseline sequence that crisscrosses the circle of fifths “I BL”; (B) From the same experiment, a chord sequence with a subdominant ending “IV”, and a paired baseline sequence “IV BL”; (C) A chord sequence with dominant ending “V” and its paired baseline sequence “V BL” (adapted from Tillmann, Janata, Birk, and Bharucha (2008), Figure 1B). Tillmann, Janata, Birk, and Bharucha (2003) used similar chord sequences, but restricted to tonic and subdominant endings and paired baseline sequences.