A spectral pitch class model of the probe tone data and scalic tonality

Journal Item

A Spectral Pitch Class Model of the Probe Tone Data and Scalic Tonality

Andrew J. Milne
MARCS Institute, University of Western Sydney, NSW, Australia

Robin Laney & David B. Sharp
The Open University, Milton Keynes, UK

In this paper, we introduce a small family of novel bottom-up (sensory) models of the Krumhansl and Kessler (1982) probe tone data. The models are based on the spectral pitch class similarities between all twelve pitch classes and the tonic degree and tonic triad. Cross-validation tests of a wide selection of models show ours to have amongst the highest fits to the data. We then extend one of our models to predict the tonics of a variety of different scales such as the harmonic minor, melodic minor, and harmonic major. The model produces sensible predictions for these scales. Furthermore, we also predict the tonics of a small selection of microtonal scales—scales that do not form part of any musical culture. These latter predictions may be tested when suitable empirical data have been collected.

Received: January 30, 2013, accepted June 17, 2014.

Key words: tonal hierarchies, probe tone data, spectral pitch class similarity, tonality, microtonality

The Krumhansl and Kessler (1982) probe tone data comprise the perceived "fits" of twelve chromatically pitched *probe tones* to a previously established major or minor tonal context. Ten participants gave ratings on a seven-point scale, where "1" designated *fits poorly* and "7" designated *fits well*. These well-known results are illustrated in Figure 1.

The major or minor tonal context was established by playing one of four musical *elements*: just the tonic triad I, the cadence IV–V–I, the cadence II–V–I, the cadence VI–V–I. For example, to establish the key of C major, the chord progressions Cmaj, Fmaj–Gmaj–Cmaj, Dmin–Gmaj–Cmaj, and Amin–Gmaj–Cmaj were used; to establish the key of C minor, the chord progressions Cmin, Fmin–Gmaj–Cmin, Ddim–Gmaj–Cmin, and A♭maj–Gmaj–Cmin were used. A *cadence* is defined by Krumhansl and Kessler (1982) as "a strong key-defining sequence of chords that most frequently contains the V and I chords of the new key" (p. 352); the above three cadences are amongst the most common in Western music. Each element, and its twelve probes, was listened to four times by each participant. As shown in Table 1, for each context, the ratings of fit were highly correlated over its four different elements—mean correlations for the different elements were $r(10) = .90$ in major and $r(10) = .91$ in minor—so the ratings were averaged to produce the results shown in Figure 1. This implies that there were a total of $10 \times 4 \times 4 = 160$ observations per probe tone and mode, hence a total

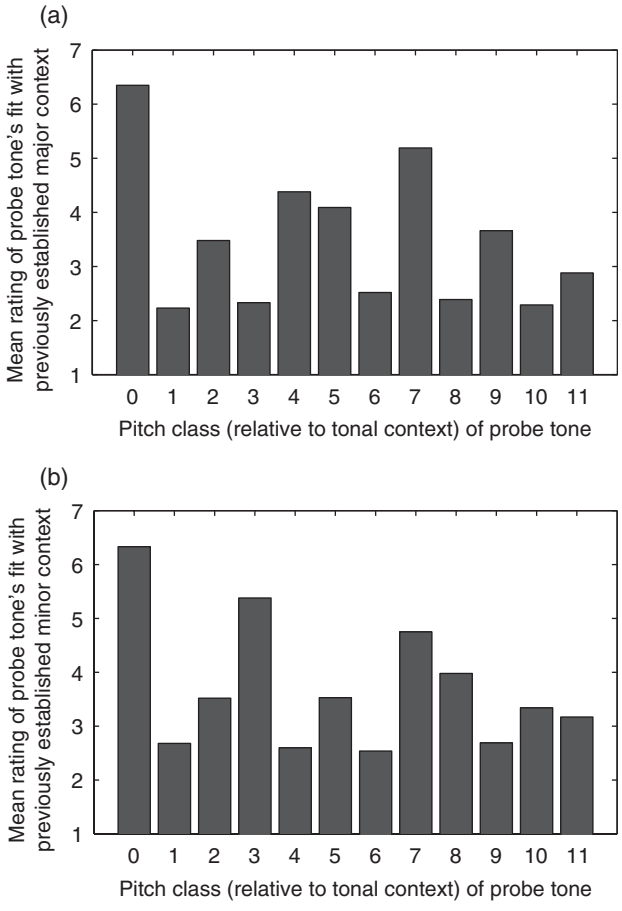

FIGURE 1. Krumhansl and Kessler's major and minor tonal hierarchies.

Music Perception, volume 32, issue 4, pp. 364–393, issn 0730-7829, electronic issn 1533-8312. DOI: 10.1525/mp.2015.32.4.364

TABLE 1. *Intercorrelations (df = 10) of the Fit Data for Each of the Context-Setting Elements.*

	Major				Minor			
	I	IV–V–I	II–V–I	VI–V–I	I	IV–V–I	II–V–I	VI–V–I
I	1.00	.97	.93	.85	1.00	.95	.89	.96
IV–V–I	.97	1.00	.86	.80	.95	1.00	.85	.97
II–V–I	.93	.86	1.00	.96	.89	.85	1.00	.84
VI–V–I	.85	.80	.96	1.00	.96	.97	.84	1.00

of $160 \times 24 = 3840$ observations in total. All listeners had a minimum of five years' formal instruction on an instrument or voice, but did not have extensive training in music theory.

All context elements and probes were played with *octave complex tones* (also known as OCTs or Shepard tones). Such tones contain partials that are separated only by octaves (i.e., they contain only 2^{n-1}th harmonics, where $n \in \mathbb{N}$), and the centrally pitched partials have a greater amplitude than the lower and higher pitched partials; precise specifications are given in Krumhansl and Kessler (1982). Octave complex tones have a clear pitch chroma but an unclear pitch height; in other words, although they have an obvious pitch, it is not clear in which octave this pitch lies. The stated purpose of using OCTs was to "minimize the effect of pitch height differences between the context and probe tones, which strongly affected the responses of the least musically oriented listeners in [an] earlier study" (Krumhansl, 1990, p. 26). However, OCTs are unnatural acoustical events—no conventional musical instrument produces such spectra; they have to be artificially synthesized. Musical instruments typically produce *harmonic complex tones* (HCTs) in which most harmonics are present and such timbres contain a greater multiplicity of interval sizes between the harmonics (e.g., frequency ratios such as 3/2, 4/3, 5/3, and 5/4, in addition to the 2/1 octaves found in OCTs). Krumhansl and Kessler (1982, p. 341) describe the OCT timbre as "an organlike sound, without any clearly defined lowest or highest component frequencies." The use of OCTs, rather than HCTs, may affect the resulting ratings of fit; that is, if HCTs had been used instead, it is possible the results may have been—to some extent—different, even after taking account of pitch height effects. For example, Parncutt (2011, p. 1339) points out that the experimental data obtained by Budrys and Ambrazevičius (2008) indicates HCTs may reverse the fits of the minor third and perfect fifth—pitch classes 3 and 7—in the minor context.

Issues related to many of the design choices in the probe tone experiment, including the use of Shepard tones, the use of a small number of musical experts as participants, and the length of experiment are discussed at length in Auhagen and Vos (2000). However, it is clear that any specific experiment has to make trade-offs between possibly incompatible goals.

The probe tone data are considered to be one of the most important sets of empirical data related to the perception of tonality. For example, the results can be generalized to predict aspects of music that were not explicitly tested in the experiment. Notably, the degree of fit can be used to model the stability or "tonicness" of the pitches and chords found in major-minor tonality—as originally suggested by Krumhansl (1990, pp. 16 & 19) and reiterated by Parncutt (2011, p. 333). Also, the data have been used to model perceived inter-key distances (Krumhansl & Kessler, 1982), and to predict the key—dynamically—of music as it plays (Krumhansl, 1990; Toiviainen & Krumhansl, 2003). However, Temperley (1999) has noted that key-finding performance is improved if the probe tone profile is adjusted so as to increase the weights of the fourth and seventh scale degrees. Furthermore, there is no obvious way to use these data to account for some other important aspects of tonality: Why is the primary major scale the diatonic, while the primary minor scale is the nondiatonic harmonic minor scale?[1] Why does the seventh degree (leading tone) of the major scale lose much of its activity when it is the fifth of the iii (mediant) chord? Why are certain root progressions favored over others (e.g., descending fifths are more common than ascending—particularly the cadential V–I)?

Causal Explanations

An important question raised by the probe tone data set is what is its origin—what causes the tonal hierarchy to take the form it does? There are two broad approaches to this question. *Top-down* models attempt to explain

[1] We use the term diatonic to refer exclusively to the scale with two steps sizes—L for large, and s for small—arranged in the pattern (L L s L L L s), or some rotation (mode) thereof. The harmonic minor and ascending melodic minor are, therefore, non-diatonic.

the data as a function of long-term memory—the fit of a scale degree to a tonic is a function of the implicitly learned prevalence of that scale degree (i.e., its familiarity). Conversely, *bottom-up* approaches attempt to explain the data without recourse to statistical knowledge of this kind. Typically, a bottom-up model will transform the context-setting elements and the probe according to a short-term memory model where salience decreases over time (Leman, 2000; Parncutt, 1994) and may make transformations that reflect plausible neurological, psychoacoustical, or other cognitive processes. Examples of neurological processes include the neural oscillations modeled by Large (2011); examples of psychoacoustic processes include virtual pitch perception (Leman, 2000; Parncutt, 1989, 1994, 2011), examples of other cognitive process include Gestalt grouping principles or the impact of structural properties of scales like interval cycles (Woolhouse & Cross, 2010).

The importance of bottom-up models is that they provide a causal explanation for the shape of the probe tone data (and the corresponding scale degree prevalences in Western music) that is further back in the causal chain and, hence, has greater *explanatory power*.[2] It is plausible there is a causal loop (across time) whereby, in one direction, prevalence increases fit (through familiarity) while, in the other direction, increased fit increases prevalence (due to composers and performers privileging high-fit pitches). But, if there is a sensory or other bottom-up reason for favoring certain pitches regardless of their familiarity, this both causally precedes and continually feeds into this causal loop from the outside, thereby stabilizing the system around values consistent with the bottom-up processes. With no bottom-up component, a pure top-down model can make no prediction about which specific forms the probe-tone data could plausibly take because any initial random choice of scale degree prevalences would stabilize into a corresponding tonal hierarchy.

Taken to the extreme, a bottom-up explanation means long-term implicit learning is completely unnecessary to explain perceived fit and stability. We might hypothesize that, given a collection of pitches in short-term memory, we are able to mentally "calculate" or "feel" the sensory fit of any current pitch or chord each time it occurs. However, even if bottom-up processes play an important role, it would be implausible to dismiss the impact of long-term memory (the importance of long-term memory has been established in numerous music perception experiments such as Francès, 1988; Lynch, Eilers, Oller, & Urbano, 1990; Schellenberg & Trehub, 1999; Trehub, Schellenberg, & Kamenetsky, 1999). For instance, it is likely that certain scales (e.g., the diatonic and harmonic minor) are so commonly used that we learn where the best fitting chords are without having to mentally assess their sensory fit each time. Furthermore, if composers favor pitches and chords with high sensory fit, their increased prevalence will further amplify their perceived fit. It is also likely we become familiar with specific sequences (ordered sets) of pitch classes and chords that exemplify musically useful patterns of fit such as those used in cadences, which induce tension and then resolution. For example, as we discuss in later sections, movements from chords containing pitch classes with low fit to those with high fit may provide particularly effective resolutions that strongly define a tonic. These examples suggest that long-term memory fit templates may be quite diverse in form, consisting of a variety of pitch and chord-based fragments rather than just the two overarching major and minor hierarchies described by the probe tone data.

Often it may be difficult to make a clean distinction between bottom-up and top-down models. For example, a model may be formulated and presented by its author as bottom-up but it may also be possible to interpret it as actually being top-down (e.g., see our discussion of Butler's, 1989, model in the following section). This means any assertion as to how a given model affects the dependent variable must be examined to see if there may be an alternative explanation. Furthermore, a model may comprise both types of process. However, it is often possible to characterize a model as being *essentially* bottom-up or essentially top-down according to the relative importance of its components. The key distinction is that bottom-up models may include top-down components that support the bottom-up processes—they enhance their effect but don't essentially change them (as in the causal loop described in the previous paragraph). Such models would still be reasonably classified as bottom-up. Other models may have bottom-up components that are subsumed by top-down effects. Such models would be reasonably characterized as essentially top-down. Other models may be down to a complex interaction of bottom-up and top-down processes in which both play an essential role, and these would be most reasonably characterized as both bottom-up and top-down.

Throughout this paper, we have attempted to categorize each of the models we discuss (including our own)

[2] See Deutsch (1997) and Lewandowski and Farrell (2011) for comprehensive discussions of explanation versus prediction.

into top-down or bottom-up categories and, perhaps more importantly, we explore each model's ability to actually explain why the probe tone data take the specific form they do.

Summary of the Spectral Pitch Class Similarity Models

In this paper we will present a small family of spectral pitch class similarity models that provide a bottom-up explanation for the probe tone data. We then extend one of the models (using parameter values as optimized to the probe tone data) to predict the tonicness of pitch classes and chords in a variety of scales, including microtonal. We will give the full mathematical specification of these models in the following section. But, before proceeding, we feel it will be helpful to provide an overview of how they work and the music perception assumptions upon which they rest.

To model the pitch perception of any musical sound, we use a *spectral pitch class vector*. Each of the 1,200 elements of this vector represents a different log-frequency in cents (modulo the octave), while the value of that element is a model of the expected number of partials (frequency components) perceived at that log-frequency. Figure 2 illustrates a spectral pitch class vector model of a major triad (bottom) and a harmonic complex tone a perfect fourth higher (top). We model the fit of any two such tones or chords by calculating the cosine similarity of their respective spectral pitch class vectors (the resulting similarity value lies between 0 and 1).

This model rests upon a number of assumptions, which are now detailed. First, we model pitch as proportional to log-frequency and model each pitch as having a *salience* value, which is its probability of being perceived. Second, we model each spectral component (partial) of a tone or chord as a pitch class (i.e., its log-frequency is represented modulo the octave). This is a model of octave equivalence in that any two pitches an octave apart are the same (they are in the same pitch class). Third, we smear each spectral component in the log-frequency domain to model perceptual inaccuracy—for example, we might expect that two spectral components separated by one cent are likely to be perceived as having identical pitch. The width of this smearing—called *smoothing width* (σ)—is a nonlinear parameter in our models. Fourth, we treat the harmonics of each tone as reducing in salience smoothly as a function of their harmonic number. This is to ensure the spectra used by the model are broadly representative of those produced by musical instruments as well as modeling the increased resolvability of lower versus higher partials (e.g., Moore, 2005). The steepness at which they reduce is another nonlinear parameter called *roll-off* (ρ). Figure 2 uses roll-off and smoothing width values as optimized to the probe tone data—note how the partials are smeared into a Gaussian shape across log-frequencies, and that the peaks reduce for higher-numbered harmonics (in the top figure, harmonics 1, 2, 4, and 8 are centered at pitch class 5.00, harmonics 3, 6, and 12 are centered at pitch class 0.02, harmonics 5, and 10 are centered at pitch class 8.86, and so forth).

As discussed in more detail in the next section, different researchers have modeled the probe tone experiment's context elements in a number of different ways: First, each of the eight different context-setting elements (four major and four minor) may be separately modeled and fitted (e.g., Parncutt, 1994). Second, the

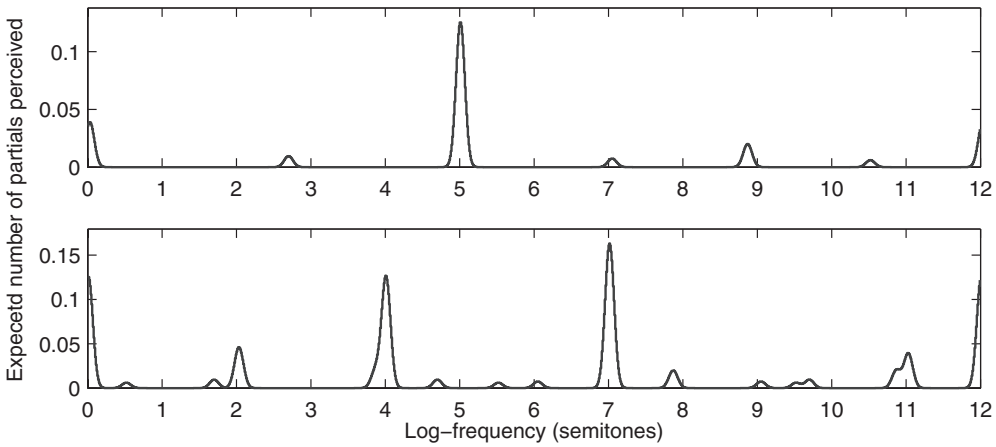

FIGURE 2. The spectral pitch class vectors for a major triad (bottom) and a pitch class five semitones higher than the former's root (top). The parameters are as optimized to the probe tone data ($\rho = 0.67$ and $\sigma = 5.95$).

four major context elements may be aggregated into a single major context, and the four minor context elements into a single minor context (e.g., Butler, 1989; Parncutt, 1989). Third, all eight elements may be represented by a single tonic pitch class (as in Krumhansl's consonance model, 1990). We will refer to this as the *tonic-as-pitch-class* concept. Fourth, all elements may be modeled by their respective tonic triad (so all the major context elements are modeled with the tonic major triad, the minor context elements with the tonic minor triad). This latter method is Parncutt's (2011) *tonic-as-triad* concept.

In one of our models, we follow the tonic-as-triad concept and model the context-setting elements with a single tonic triad. In the other two models we allow the root of the tonic triad (i.e., the tonic pitch class) to have a greater weight than the other tonic triad tones. This is to reflect the greater salience of the root in a major or minor triad (Parncutt, 1988) and to allow the model to be situated anywhere on the continuum between tonic-as-triad and tonic-as-pitch-class. Ignoring the precise form of the context elements by representing them as a tonic triad or tonic pitch class (or somewhere between) is sensible when the elements serve a cadential function (Krumhansl and Kessler, 1982, chose these specific contexts precisely because they are common cadences). This is because, by definition, the purpose of a cadence is to strongly induce a feeling of tonicness for the final chord and it seems reasonable to assume this tonic will be our predominant perception just prior to the probe tone. The tonic-as-triad concept also seems to mirror the probe tone data in that the four profiles for the differing major context elements are very highly correlated (and the same for the minor context elements)—as shown in Table 1.

Furthermore, we model the context chord tones and probe tones as full harmonic complex tones (HCTs) not as Shepard tones, which have only octave spaced partials. This implies our model assumes the auditory system adds a full harmonic spectrum to a Shepard tone (through nonlinear processes such as those observed in Lee, Skoe, Kraus, & Ashley, 2009, and modeled by Large & Almonte, 2012), or that the probe tones in the experiment act as a trigger (through long-term memory) for the responses that would have occurred with HCTs of identical pitch classes. It is important to point out that even in the latter case, the spectral origin of the model holds—although the model may now comprise a long-term component it is still founded upon an important bottom-up component that provides its explanatory power.

Extending the Probe Tone Models

As we demonstrate in the later section *A Model of Scalic Tonality*, an interesting feature of probe tone models is that, if we equate tonicness with fit, they can be used to model the tonicness of pitch classes or chords given a scale. For example, we can treat the harmonic minor scale as an abstract entity that represents a set of possible pitches, but impose no additional structure by giving all its pitches equal weight. This enables us to talk of a *scalic tonality* whereby any unique collection of pitch classes (a scale) has unique tonal implications—even in the absence of a pre-existing corpus using that scale. In that section, we use the same spectral pitch class similarity model—as optimized to the probe tone data—to model the affinity of triads to a selection of Western scales (Guidonian hexachord, diatonic, harmonic minor, melodic minor, and harmonic major) and a selection of microtonal scales.

To be more concrete, we model the spectral pitch classes induced by all HCTs in a scale (as if it is a big chord) by placing them into a spectral pitch class vector as described above. Each scale tone is equally weighted, but the salience of each partial (as a function of its harmonic number) and the width of the smearing is identical to the optimal values used to fit the probe tone data. The cosine distance between this vector and the spectral pitch class vector of any given chord (produced in the same as way as for the scale) is used to model the fit—and hence tonicness—of that chord given the scale.

In the subsection *Fit Profiles for 12-TET Scales*, we additionally suggest some related mechanisms that may help to answer the three questions posed earlier (at the end of *The Probe Tone Experiment* subsection). These are that resolutions are strengthened when a worst-fitting pitch class moves to the root of a best-fitting triad, and that we also need to consider the fit of each pitch class within the chord it is part of. At the moment, however, these mechanisms are not instantiated in a formal mathematical model and, until they are, they should be thought of as preliminary findings or suggestions. We hope to formally embody these latter principles and test them against novel empirical data in future work.

Models of the Probe Tone Data

To provide the context for our model of the probe tone data, in this section we survey a variety of other existing models of these data. Most of these are also usefully summarized in Parncutt (2011) so we will keep our account brief, but we will also highlight a few areas where we take a different stance to Parncutt. In order

to fairly compare the predictive power of the models (ours is nonlinear), we use cross-validation statistics in addition to conventional correlation. When exploring the predictive power of the models, the main focus is on their fit to the aggregated probe tone data (all of which are very highly correlated); however, for some of the models, we additionally discuss what happens when they are applied to the data arising from each context-setting element. We also explore the extent to which each model contributes a plausible and generalizable bottom-up explanation.

Before discussing each of the models in turn, Table 2 summarizes their relevant statistical properties with respect to the probe tone data (we also provide a table of intercorrelations in Appendix A).[3] When comparing models we feel it is important to consider correlation values over all 24 data points because the same underlying process should apply to the major and minor contexts—separately correlating them is equivalent to calculating the *r*-values of two linear regressions with different intercept and slope parameters. Because there is no a priori reason to expect the two sets of parameters to be different, this procedure is not ideal—precisely the same model should be used for both major and minor. For this reason, in the cross-validation statistics, we apply a single set of parameter values to both major and minor. However, it is still useful to see how well each model performs with respect to the major and minor contexts, so we also supply more conventional correlations for each context. An important reason for using cross-validation correlation is to allow our nonlinear models to be fairly compared with the mostly linear models that have been proposed so far. Utilizing uncross-validated statistics would be inappropriate, because the additional flexibility of a model with additional nonlinear parameters may allow it to fit the noise rather than the process underlying the data, thereby giving it an unwarranted advantage. Cross-validation statistics provide a way for models with differing levels of flexibility (complexity) to be fairly compared, and ensure they are not overfitting the data.

The models are ordered by their cross-validated correlations over all 24 data points and, when these are not available, by the mean of their (not cross-validated) correlations for the major and minor contexts. This provides an indication of their ranking in terms of predictive power. However, it is useful to bear in mind that if we consider these 24 data points to be a sample from a population of participants, replications, contexts, and so forth, the correlation confidence intervals are wide; for example, for a correlation of $r(22) = .95$, the 95% confidence interval is from .89 to .98. Indeed, even if we consider the probe tone data to perfectly represent the expected population values, the best performing models are still very close. For example, the Bayesian Information Criterion (BIC) of Milne 14*c*, Lerdahl 88, and Parncutt 89 are −45.42, −43.60, and −46.33, respectively (lower is better); typically, differences in BIC values are only considered meaningful when greater than 2.

We used 20 runs of 12-fold cross validation, which means the data set of 24 probe tone fit values is split into a *training set* of 22 probe fit values and a *validation set* of 2 probe fit values. The parameters of each model are optimized to the training set (for the linear models these parameters are the intercept and slope; for our models there are additional nonlinear parameters). The modeled values for the two validation data points are then calculated. This procedure is done 12 times, in each case a different training and validation set is used, such that each validation set never contains a data point used in a previous validation set. This ensures we end up with 24 modeled values corresponding to all 24 data points. The cross-validation statistic of interest is then calculated for these values (e.g., cross-validation correlation). Cross-validation statistics have an unknown variance, but this variance can be reduced by repeating the process multiple times with different validation sets and taking the mean value of the statistic. As mentioned above, we performed 20 runs of the 12-fold cross-validation. We give a more technical explanation of the cross-validation statistics in Appendix B.

It is worth pointing out that the modeled data do not need to replicate much of the experimental data's fine structure in order to achieve what appears to be a reasonably good correlation value. For example, let us define a *basic triad model* as one that gives the tonic chords' pitches a value of 1, and all other pitch classes a value of 0; the resulting statistics are surprisingly impressive looking: $r_{CV}(22) = .82$ and major and minor correlations of $r(10) = .83$ and $r(10) = .89$, respectively. We suggest that any model with similarly valued statistics is probably struggling to describe much of the fine structure of the data; we place this basic triad model into the table to serve as a benchmark.

KRUMHANSL 90B: CORPUS PREVALENCE MODEL

Krumhansl (1990) suggested a model for the probe tone data, $r_{CV}(22) = .83$, which is that they are correlated with the distribution (prevalences) of scale degrees in existing music. This is a purely top-down model of

[3] The interval cycle theory of Woolhouse and Cross (2010) is not included in Table 2 because it does not produce a single model of the probe tone data. This theory is discussed later in this section.

TABLE 2. *Cross-validation Correlations of Each Model's Predictions with the Major and Minor Profiles Combined (df = 22).*

	$r_{cv}(22)$	$r_{maj}(10)$	$r_{min}(10)$	Type	Parameters
Milne 14*c*	.96	.98	.97	bottom-up	nonlinear
Lerdahl 88	.95	.98	.95	top-down	linear
Parncutt 89	.95	.99	.94	top-down	linear
Parncutt 94	—	.96	.95	bottom-up	nonlinear
Parncutt 11*a*	.93	.94	.95	bottom-up	linear
Milne 14*b*	.92	.98	.94	bottom-up	nonlinear
Milne 14*a*	.91	.96	.93	bottom-up	nonlinear
Parncutt 11*b*	.90	.93	.92	bottom-up	linear
Large 11	—	.97*	.88*	bottom-up	nonlinear
Smith 97	.87	.91	.88	bottom-up	linear
Butler 89	.84	.90	.86	top-down	linear
Krumhansl 90*b*	.83	.89	.86	top-down	linear
Basic triad	.82	.83	.89	—	linear
Leman 00	—	.87	.84	bottom-up	nonlinear
Krumhansl 90*a*	.57	.76	.53	bottom-up	linear
Null	−.68	.00	.00	—	linear

Note: The cross-validation correlations are the means of these statistics taken over twenty runs of 12-fold cross-validation. We also show the correlations (not cross-validated) for the major and minor contexts separately. The null model is an intercept-only model—i.e., all probe fit values are modeled by their mean. The remaining models are described in the main text. The models are ordered by their cross-validation statistics or, where these are missing, by the mean of their major and minor context correlations. The correlation statistics for the Large model are starred to indicate different nonlinear parameter values were used for the major and minor contexts—with unified parameter values these correlations will be lower. The models are categorized according to whether they are essentially bottom-up or top-down; these labels should be taken with some caution because there is always some ambiguity about precisely which underlying processes a model instantiates.

music perception, in that the perceived fits of the probe tones are hypothesized to be down to nothing more than learning: if we frequently hear the fourth scale degree, we will tend to feel that scale degree has a good fit; if we rarely hear altered scale degree $\flat\hat{2}/\sharp\hat{1}$, we will tend to feel that scale degree has a poor fit.

This model provides a straightforward explanation for our perception of scale degree fit, but the scope of this explanation is limited because it cannot explain why the probe tone data/scale degree prevalences take the specific profile they do. Indeed, an implicit assumption of this model is that this profile is down to nothing more than chance—for some unknown reason, composers favored certain scale degrees and hence listeners came to feel these scale degrees fitted better. Composers (who are also listeners) continued to write music that utilized these learned patterns of fit (because such music made sense to them and their listeners), and so listeners (some of whom are composers) continued to have their learning of these patterns reinforced. And so forth, in a circular pattern of causal effects: music perception is the way it is because music is the way it is, and music is the way it is because music perception is the way it is, ad infinitum. Presumably, this theory predicts that on a "parallel Earth"—identical in all respects to ours except for random fluctuations—a completely different profile of pitch class fits might have developed. Of course, this may be true. But it is quite plausible that there are innate perceptual, cognitive, or core knowledge (Spelke & Kinzler, 2007) principles that might contribute to making one, or a small number, of actual fit profiles possible or more likely.

LERDAHL 88: PITCH SPACE MODEL

Lerdahl's (1988) basic pitch space has five levels: (1) tonic, (2) tonic and fifth, (3) major tonic triad, (4) diatonic major scale, (5) chromatic scale. He points out that the five levels in this basic pitch space correlate well with the major context's probe tone data (p. 338). He does not, however, suggest a formal model for the minor context. To address this, it is necessary to create a conceptually related "minor pitch space" for the minor context. Lerdahl's model (and its extension to the minor context) is predictively very effective, $r_{cv}(22) = .95$. However, it is deficient in terms of explanatory power because important aspects of the basic pitch space itself are derived from (or require) top-down explanations.

Lerdahl provides a bottom-up explanation for the first three levels, which is that the height of a level should correlate with "the degree of sensory consonance of adjacent intervals" within it (Lerdahl, 2001, p. 272; he defines sensory consonance psychoacoustically as a function of both roughness and clarity of the root, p. 321)). The perfect fifth in the second level is the most consonant interval, and the major triad on the third level is the most consonant triad (although the minor triad is similarly consonant and seems a reasonable alternative). The fourth level—which is critical for

TABLE 3. *Correlations (df = 10) of the Lerdahl 88 Model and the Fit Data for Each of the Context-setting Elements.*

Major				Minor			
I	IV–V–I	II–V–I	VI–V–I	I	IV–V–I	II–V–I	VI–V–I
.94	.88	.98	.95	.92	.95	.86	.92

Note: The mean correlation is .93.

producing high correlations with the data—is the diatonic major scale. Although Lerdahl gives a number of bottom-up explanations for privileging the diatonic scale,[4] he gives only a top-down explanation for choosing its Ionian (i.e., major) mode, rather than the Mixolydian or Lydian—he privileges the former due to its prevalence (2001, p. 41). The predictive power of the basic pitch space, therefore, relies on a long-term memory explanation, so we class this model as top-down.

To extend Lerdahl's model to account for the minor context, Parncutt (2011) created a "minor pitch space." This builds up the levels in the same way, but has a minor triad (rather than a major triad) on the third level, and has the harmonic minor scale (rather than the diatonic major scale) on the fourth level. The resulting major (basic) and minor pitch spaces are highly correlated with their respective probe tone data, $r_{CV}(22) = .94$.

However, in one respect, this minor pitch space is not in keeping with Lerdahl's conceptualization of the basic pitch space because it uses a nondiatonic scale (the harmonic minor), which does not have the property of coherence, for the fourth level. It is more in keeping with Lerdahl's theory to use the coherent Aeolian (natural minor scale), Dorian, or Phrygian mode—rather than the harmonic minor—for the fourth level. The Aeolian is probably the most prevalent (hence familiar) of these three modes, and using it in this model gives a higher correlation with the minor context's data than Parncutt's harmonic minor version. It is this Aeolian version of Lerdahl's model that we include in Table 2.

This latter model is predictively extremely effective and provides amongst the highest cross-validated correlations, $r_{CV}(22) = .95$. As discussed in the introduction, the probe tone data in each major or minor context are highly correlated across the four different elements (I, IV–V–I, II–V–I, and VI–V–I). Because this model has a good fit with the aggregated data, and it produces the same predictions across the four elements of each context, it also has good fits with profiles resulting from each context-setting element (as shown in Table 3). However, as an essentially top-down model, it has limited explanatory power.

[4] Balzano's principles of uniqueness, coherence, and simplicity, and Clough and Douthett's maximal evenness (Lerdahl, 2001, pp. 50–51 & p. 269).

BUTLER 89: AGGREGATE CONTEXT PITCH MULTIPLICITY MODEL

Butler (1989) presents his model as utilizing nothing more than short-term memory, in which case, it is an explanatory bottom-up model. However, as we shall see, it is actually more likely that this is a top-down model of a possible long-term memory process.

He models the probe tone ratings simply by the number of times their pitches occur in each context's elements (i.e., the chord progressions I, IV–V–I, II–V–I, and VI–V–I). These four elements were aggregated into a chord collection containing IV, II, VI, three Vs, and four Is. The model counts the number of occurrences of each scale degree in this collection: there are six $\hat{1}$s (in the four Is, the IV, and the VI); there are zero $\sharp\hat{1}/\flat\hat{2}$s; there are four $\hat{2}$s (in the three Vs and the II); and so on. The resulting counts for the major and minor contexts' elements fit the data well, $r_{CV}(22) = .84$. As a short-term memory model, it is bottom up and provides a meaningful explanation for why, given an immediate context element, certain pitches (probes) fit better than others: currently heard pitches that are also salient in short-term memory are perceived to fit better than pitches that are not also salient in short-term memory—we are "comfortable" with, or "less surprised" by, repetition. It also implies that there is not necessarily a stable tonal hierarchy that serves as a fixed template against which currently heard pitches are compared.

However, it is questionable whether this model can be considered to be a short-term memory model. As Krumhansl (1990, p. 62) points out, the different context elements were presented to listeners in separate blocks, not intermixed within the same block and, for this reason, it is implausible that short-term memory—which typically completely decays within 20 seconds (Peterson & Peterson, 1959)—could be responsible for aggregating the four elements (this point is also amplified by Woolhouse & Cross, 2010). If Butler's model is applied to each context element separately and then averaged over them, the fit with the probe tone data is

substantially poorer, averaged $r_{CV}(22) = .74$. So, when corrected to more accurately reflect short-term memory processes, the model becomes predictively weak.[5] Furthermore, Krumhansl and Kessler (1982, p. 343) found the ratings produced by the differing context elements to be "very similar," whereas the modeled data produced by the differing context elements are not.

As pointed out by Parncutt (2011, p. 341), a mechanism that could account for the aggregation of the four context elements being correlated with the data would be that the aggregated chord context is a good summary of the prevalences of chords in Western music. However, this transforms the model into a purely top-down model, where the fit of probe tones is solely down to their prevalence. In other words, viewed from this perspective, Butler's model is the same as Krumhansl's prevalence model; the difference being that Krumhansl statistically analyses a corpus, while Butler statistically analyses a set of common cadences—and both have similar scale degree prevalences. For this reason, we class this model as top-down.

PARNCUTT 89: AGGREGATED CONTEXT PITCH CLASS SALIENCE MODEL

Parncutt (1989) adapted Butler's model in two ways. First, he used a different aggregation of the contexts' elements: IV, II, VI, three Vs, and six Is. The difference is that the tonic triad element is counted six rather than four times, this is because Parncutt counts the tonic triad three times for the context element that comprises only the I chord. Despite Krumhansl's criticism (1990, p. 62) that this does not reproduce the stimuli used in the experiment, it is actually quite reasonable because the ratings produced by the four context elements were averaged to produce the final sets of probe tone data (so, counting the I element three times, gives it equivalent weight to each of the other three elements; Parncutt, 1989, p. 159). Second, he included not just the notated pitches in the context elements, but also their pitch class (or chroma) salience profiles. The precise mechanism by which the pitch class saliences are generated for a harmonic complex tone is detailed in Parncutt (1989, Sec. 4.4.2). In summary, the salience of any given pitch class is calculated from a combination of the weights of harmonics and subharmonics with corresponding pitch classes—these subharmonics and harmonics extending from each notated pitch. The subharmonics are, overall, weighted significantly higher than the harmonic pitches, so this is primarily a virtual (subharmonic) pitch model.

When applied to the aggregated elements in each context, the model produces one of the best fits to the data, $r_{CV}(22) = .95$. But when applied to each context element separately—as shown in as shown in Table 4—the model performs less well; the mean correlation is $r(10) = .87$. This means it suffers from the same problems as Butler's: it cannot really be interpreted as a model of short-term memory processes; rather, it is a model of a possible long-term memory process, where the aggregated cadences serve as proxies for prevalent chords in Western music. So the model has limited explanatory scope—although it may explain the data given the prevalence of a small set of chords, it does not explain why those chords, in particular, are prevalent.

LEMAN 00: SHORT-TERM MEMORY MODEL

Leman (2000) utilizes a short-term memory model whose inputs are derived from a model of the auditory system. The latter comprises 40 bandpass filters, half-wave rectification and simulations of neural firings induced by the filters, and periodicity detection (autocorrelation) applied to those firings. Autocorrelation automatically detects frequencies that are subharmonics of the input frequencies. In this respect it is, therefore, similar to Parncutt's chroma salience model. The resulting signals, produced in response to the context element, are stored in a short-term (echoic) memory model that decays over time and, at the time at which the probe is presented, this represents the "global image" of the context element. The length of the decay (the half-life of the signal) is a free parameter. This global image is correlated with a "local image" produced by each of the 12 probe tones (for each of the four context elements in both major and minor). The twelve correlation values (for the twelve probes) are averaged over the four major and four minor context elements (in the same way as Krumhansl's data), and these are used to model the probe tone data.

The model produces correlations towards the lower end of those discussed here, $r(10) = .85$ for major and $r(10) = .83$ for minor. However, Leman chooses a decay parameter of 1.5 seconds, when his Table 3 shows that the maximum decay value tested (5 seconds) would have fit the probe tone data better (he chooses the lower time value because fitting the probe tone data is not his only criterion). With the 5 second decay time, the correlations improve, but only slightly, $r(10) = .87$ for major and $r(10) = .84$ for minor.

Because of the nonlinear decay time parameter, and without easy access to the original model, we have not

[5] The only practicable way to perform the cross-validations was to allow for the parameters, within each training fold, to vary across the different context elements. There is, however, no a priori reason why they should be different over different context elements. If they had have been kept the same, the resulting statistics would have been even lower.

TABLE 4. *Correlations (df = 10) of the Parncutt 89 Model and the Fit Data for Each of the Context-setting Elements.*

Major				Minor			
I	IV–V–I	II–V–I	VI–V–I	I	IV–V–I	II–V–I	VI–V–I
.88	.92	.86	.98	.94	.90	.54	.92

Note: The mean correlation is .87.

calculated its cross-validation correlations. However, since the $r(22)$ statistics will be lower than .87—which is the highest $r(10)$ statistic gained by the 5 second decay time model of the major context's data—it is safe to conclude that, in terms of prediction, this is one of the worst performing models and probably no better than the "basic triad" benchmark model.

KRUMHANSL 90A: CONSONANCE MODEL

Krumhansl's (1990) other model is bottom-up and attempts to provide a more substantive explanation than the prevalence model. It also predicts rather poorly, $r_{CV}(22) = .57$. This model hypothesizes that the probe tone fits are due to the consonance of the corresponding pitch class and the tonic pitch class (the first scale degree). Clearly, this model will struggle to obtain high correlations with the empirical data because it produces identical predictions for the major and minor contexts (they both have the same tonic pitch class).

Krumhansl uses consonance values that are the averages of a variety of bottom-up models of consonance (Helmholtz, 1877/1954, Hutchinson & Knopoff, 1978; Kameoka & Kuriyagawa, 1969; Malmberg, 1918), and one set of empirically derived consonance ratings (Malmberg, 1918). This means the model, as a whole, is essentially bottom-up and has wide explanatory scope—it provides an explanation for the probe tone ratings based on innate perceptual processes. However, it is also worth noting that—as Krumhansl points out (1990, p. 55)—there is something of a mismatch between the model's explanation and the experimental procedure used to get the empirical data: the probe tones were played after the context-setting chords, not simultaneously, so harmonic consonance/dissonance does not play a direct role in the experimental stimuli. For this model to make sense, it must be additionally assumed that the listeners were mentally simulating harmonic intervals comprising the tonic and the probe, and then determining their consonance/dissonance values either directly or from long-term memory. This is plausible, given the musical experience of the participants, but it is an indirect explanation.

SMITH 97: CUMULATIVE CONSONANCE MODEL

Like Krumhansl, Smith (1997) also uses consonance—but in a different way—to explain the data from the bottom up. He takes a tonic pitch and finds a second pitch with the greatest consonance. To these two pitches, he then finds the third pitch that makes the most consonant three-tone chord (in all cases, consonance is calculated as the *aggregate dyadic consonance*, which is the sum of the consonances of all interval classes in the chord; Huron, 1994). To this three tone chord, he finds the pitch of the fourth tone that creates the most consonant four-tone chord. And so forth, until all 12 pitch classes are utilized.

If the first pitch is C, the second pitch is G, and the third pitch is either E or E♭ (the major and minor triads have equal aggregate consonance because they contain the same three interval classes, 3, 4, and 5). Because there are two possible three-tone chords, the resulting cumulatively constructed scales bifurcate at this juncture. For the major triad C–E–G, the fourth tone is A; for the minor triad C–E♭–G, the fourth pitch is B♭ Continuing this process, leads to the following two sequences of pitch classes: C–G–E–A–D–F/B–A♭–G♭/B♭–D♭/E♭, and C–G–E♭–B♭–F–D/A♭–B–D♭/A–E/G♭ (where X/Y denotes that X and Y have the same ranking). When each pitch class is assigned a value according to its ranking (e.g., in the first sequence, C = 1, G = 2, E = 3, A = 4, D = 5, F = 6.5, B = 6.5, A♭ = 8, etc.), they provide a predictively effective model of their respective major and minor probe tone ratings, $r_{CV}(22) = .87$.

This model has reasonable predictive power (though its predictive performance is towards the lower end of the models discussed here) and, like Krumhansl's 90*a* consonance model, has potential for good explanatory power if the consonance values it uses are derived from a psychoacoustic or other bottom-up model. Smith actually uses interval class consonance values derived by Huron (1994) from empirical data collected by Kameoka and Kuriyagawa (1969), Hutchinson and Knopoff (1978), and Malmberg (1918), not from modeled data. Using empirical data means that the

consonance values are likely to be correct and do not have to rely upon possibly inaccurate models (Huron, 1994). However, this weakens the explanatory scope of Smith's model—ideally, a bottom-up consonance model would be substituted at some stage. Like Krumhansl's consonance model, this model also suffers from the indirect relationship between harmonic consonance (the model's variables) and melodic fit (what the experiment actually measures).

LARGE 11

Ed Large's (2011) model is appealing because it is founded on the neural oscillations caused by interaction of hypothesized banks of excitatory and inhibitory neurons. It is, in this respect, a principally bottom-up model that attempts a purely physical explanation. It additionally allows for aspects of top-down learning to be incorporated through the mechanism of Hebbian learning (as described below). To be more precise, Large models a neural oscillator as resulting from interacting populations of excitatory and inhibitory neurons. Each oscillator has a natural frequency (eigenfrequency). Multiple such neural oscillators are arranged in banks in order of their oscillation frequency (a gradient frequency oscillator network) and every oscillator can be connected (coupled) to every other oscillator in the same bank. Furthermore, more than one bank can be used and there can be connections between oscillators in different banks. The coupling strengths of the connections between pairs of oscillators can be varied to model Hebbian learning, which neatly allows the model to incorporate top-down learning as well. Another parameter controls the nonlinearity of the connections.

Given an auditory stimulus comprising frequencies f_1 and f_2, this mechanism results in additional oscillations (distortion products) not present in the stimulus. These additional frequencies occur at harmonics (nf_1 and nf_2), subharmonics (f_1/n and f_2/n), differences ($f_2 - f_1$), summations ($f_1 + f_2$), and integer ratios (mf_1/n and mf_2/n), where m and n are natural numbers.

To model the probe tone data, Large uses a gradient frequency network with oscillators spaced at 10 cent intervals (the overall log-frequency range spanned is not provided in the paper). Each oscillator is coupled to other oscillators at low integer frequency ratios close to 12-TET (16/15, 9/8, 6/5, 5/4, 4/3, 17/12, 3/2, 8/5, 5/3, 16/9, 15/8, and 2/1) because low integer ratios are stable resonances in such oscillator networks and 12-TET is presumed to have been learned through the Hebbian process. The nonlinearity of the couplings is a free parameter denoted ε. The network was stimulated so as to give stable oscillations at all pitches in the tonal context (Large is not specific about whether the four contexts were aggregated or run separately and then aggregated). The stabilities of the oscillations resulting from this stimulus were used to model the probe tone profiles and result in correlations of $r(10) = .97$ for major and and $r(10) = .88$ for minor. The major profile values are amongst the best of the models considered here, but the minor values are worse than the benchmark "basic triad" model shown in Table 2. It is also important to point out ε was separately optimized for the major and minor profiles ($\varepsilon = 0.78$ in major and 0.85 in minor). As we noted earlier, parameters' values should ideally be invariant across major and minor (as they are in the Leman and Milne models); for example, considering Large is modeling a physical system, why would the nonlinearities of the neuronal connections be different for major and minor contexts? With a unified parameter value, the fit of the model will be less than the above figures—though without access to the original, it is impossible to ascertain what a single correlation value over all 24 stimuli would be.

A possible concern about this model is that there are a large number of parameters whose values can be arbitrarily chosen prior to formal optimization. For example, there are the choices of how many banks, which pairs of oscillators should be connected and how different banks should be connected. Each bank of n oscillators, indexed by i and j, has parameters including: the bifurcation α, nonlinear saturation $\beta_1, \beta_2, \ldots, \beta_n$ (typically these are constrained to take the same value), frequency detuning $\delta_1, \delta_2, \ldots, \delta_n$ (typically these are constrained to take the same value), and connection strengths c_{ij} (also these are typically constrained). Although explanation is given for some of the parameter values, it is not clear from the published paper which values were chosen for the probe tone model, or why, and to what extent different choices would have affected the model's predictions.

WOOLHOUSE 10

Woolhouse and Cross' (2010) model calculates the sum total of interval cycles between any arbitrary pitch class set and the diatonic scale (for the major context) and the harmonic minor scale (for the minor context). The *interval cycle* between any two pitch classes is the number of times that interval can be stacked until it reaches the same pitch class (assuming 12-tone equal temperament). For example, a major third has an interval cycle of three because it takes three stackings to return to the same pitch class (e.g., C–E–G♯–C). The resulting sum is taken to be a model of the "tonal attraction" of the two pitch class sets.

We will not discuss this model in depth here because this theory does not actually produce a single model of the probe tone data. There are 2,044 different pitch class sets, so this results in 2,044 different models for each fit profile. There is no principled method to choose any one of these models over any other—other than choosing the best fitting, which would result in a model that is too flexible to have any value (essentially the choice of pitch class set becomes a free parameter). For this reason, Woolhouse seeks to show there is a statistical link between interval cycles and the probe tone data by calculating two distributions. The first is the distribution of correlations between the probe tone data and 127 interval cycle models generated by pitch class sets comprising pitch classes only from the respective context. The second is the distribution of correlations between the probe tone data and 4,094 interval cycle models generated by all possible pitch class sets. He then shows these two distributions are different (using a Kolmogorov-Smirnoff test) and that the expected correlation value of the former is higher than the latter. As such there is, therefore, no single interval cycle model of the probe tone data. It would have been informative to see how well the average of all 127 interval cycle models containing only context pitch classes correlate with the probe tone data, but that information is not supplied in the paper.

PARNCUTT 11 & 94: VIRTUAL PITCH CLASS MODELS

Parncutt's 11 model (Parncutt, 2011) is a predictively effective bottom-up model, $r_{CV}(22) = .93$. It builds on Parncutt's (1988) model of virtual pitch classes, and the concept of "tonic as triad," which is explored in Parncutt (2011). (The model described here was first presented in 2011, though aspects of it date back to 1988.) This concept treats the tonic as a triad—a major or minor chord built upon the tonic pitch class—and it can be seen as a break from a more traditional concept of "tonic-as-pitch-class."[6] For example, the tonic of the key C major is not the pitch class C, but the triad Cmaj; the tonic of the key B♭ minor is not the pitch class B♭, but the triad B♭min.

The tonic-as-triad concept implies that the context-setting elements—whose purpose is to induce a strongly defined key and all of which end in the tonic triad—can be effectively represented by the tonic triad. For instance, the cadence Fmaj–Gmaj–Cmaj is used to establish the chord Cmaj as a strong and stable tonic chord, so it is unsurprising if our attention is more clearly focused on the Cmaj chord than on the preceding chords. Indeed, even if the elements were, for example, Fmaj–Gmaj, or only G7, even though the Cmaj is not actually played it is still easy to imagine it as the most expected (and best fitting) continuation. The tonic triad, therefore, effectively summarizes our response to the context-setting elements used in the experiment; importantly, it also effectively summarizes our response to tonal context-setting devices (cadences) in general.

The probe tone ratings are modeled from the weights of the *virtual pitches* that are internally generated in response to the notated pitches in the tonic triad. (By *internally generated*, we mean that virtual pitches are produced by some aspect of the auditory or cognitive system—they are not physically present in the stimulus prior to entering the ear.) Virtual pitches are typically modeled to occur at subharmonics below the notated pitch (the first N subharmonics of a notated pitch with frequency f occur at frequencies f, $f/2$, $f/3, \ldots, f/N$). There is well-established evidence that virtual pitches are generated from physical frequencies—for example, if the fundamental is removed from a harmonic complex tone, its pitch still heard as corresponding to that missing fundamental, and combination tones produced by multiple sine waves are clearly audible phenomena. However, the extent to which HCTs (or OCTs) produce salient virtual pitches at pitch classes different to that of their fundamental is less obviously demonstrable.

In Parncutt's model, the pitch of each subharmonic is modeled in a categorical fashion; that is, it is categorized by the pitch class it is closest to. For example, the seventh subharmonic below C4 corresponds to a pitch 31 cents above D1, but is modeled by the pitch class (category) D. The model, therefore, hypothesizes that pitch discrepancies of the order of a third of a semitone have no impact on whether that pitch is mentally categorized as a specific chromatic pitch class.[7] For any given notated pitch, its virtual pitch classes are weighted: the virtual pitch class corresponding to the notated pitch class itself has weight 10; the virtual pitch class seven semitones (a perfect fifth) below has weight 5; the virtual pitch class four semitones (a major third) below has weight 3; the virtual pitch class ten semitones (a minor seventh) below has weight 2; the virtual pitch class two semitones (a major second) below has weight 1. These

[6] An early description of the tonic-as-triad concept is given in Wilding-White (1961).

[7] Parncutt (1988, p. 70) argues such pitch differences can be ignored because the seventh harmonic of an HCT can be mistuned by approximately half a semitone before it sticks out. Conversely, it could be argued that when musicians' pitches go off by more than about 20 cents, the notes are generally perceived as out-of-tune, and so do not comfortably belong to their intended (or any other) chromatic pitch class category.

TABLE 5. *Correlations (df = 10) of the Parncutt 11a Model and the Fit Data for Each of the Context-setting Elements.*

Major				Minor			
I	IV–V–I	II–V–I	VI–V–I	I	IV–V–I	II–V–I	VI–V–I
.93	.90	.89	.90	.93	.97	.80	.95

Note: The mean correlation is .91.

weights are justified on the grounds that they are numerically simple and are approximately proportional to the values achieved by taking a subharmonic series with amplitudes of $i^{-0.55}$, where i is the number of the subharmonic (a typical loudness spectrum for the harmonics produced by musical instruments), and summing the amplitudes for all subharmonics with the same pitch class (Parncutt, 1988, p. 74).

These virtual pitch classes, and their weights, are applied to the three notated pitches in the major or minor tonic triad; when virtual pitch classes from different notated pitches are the same, their weights are summed to model the overall virtual pitch class weights produced by a tonic triad. For example, in the chord Cmaj, the notated pitch C contributes a virtual pitch class C of weight 10, the notated pitch G contributes a virtual pitch class C of weight 5, the notated pitch E contributes a virtual pitch class C of weight 3; the three are combined to give a virtual pitch class C with a total weight of 18. The two sets of virtual pitch class weights for a major and minor triad closely fit their respective probe tone data, and do so with a plausible bottom-up (psychoacoustic) model.

The resulting model has a cross-validation correlation of $r_{CV}(22) = .93$. A natural explanation provided by this model would appear to be that the greater the commonality of the pitches evoked by the tonic triad (which represents the context) and those evoked by the probe, the greater the perceived fit. However, in this model (which is designated Parncutt 11*a* in Table 2), the probe tone itself is modeled with a single pitch, rather than as a collection of virtual pitch classes. It is not clear why the tonic triad should evoke virtual pitches, but the probe does not; the probe's missing virtual pitch classes seems like a conceptual inconsistency in this model. If the probe tone is given virtual pitch classes—in the same way as the tonic triad—the resulting predictions are still good, but slightly less accurate, $r_{CV}(22) = .90$. This is shown as Parncutt 11*b* in Table 2.

It is interesting to note that any tonic-as-triad model will produce the same values when applied to any of the four major contexts individually (similarly for the minor contexts). This is because the precise form of the contexts is ignored so long as they serve a cadential function. The intercorrelations of Parncutt 11*a* and each of the individual contexts' fit data are shown in Table 5.

Clearly, this model performs well for each of the contexts as well as to the aggregated data—something that does not occur with the Butler and Parncutt '89 models). Interestingly, in an earlier model, Parncutt (1994) utilized a similar virtual pitch class model that included all of the chords played in each context-setting element, but adjusted their weights to account for short-term memory decay (similar to that described for Leman 00). The memory half-life was a nonlinear parameter optimized to 0.25 seconds; this means the model incorporates the virtual pitch classes of the final tonic, and—to a much lesser degree—the virtual pitch classes of the preceding chords. This means the model produces different values for each of the contexts. As shown in Table 6, this model also performs well for each context-setting element and, when its predictions are averaged across the elements, it has a slightly better correlation than the Parncutt 11*a* model (as shown in Table 2, where it is designated Parncutt 94). We were unable to calculate the cross-validation statistics because we do not have access to the original model, but they are unlikely to be significantly better than Parncutt 11*a*. These results suggest that utilizing all the chords in a given context-setting element works slightly better than using just the tonic triad for predicting the response specific to that element, but using just the tonic triad for cadential contexts is sufficient for capturing the effects of harmonic tonality more generally; that is, averaged over a broader range of chord progressions.

MILNE 14: SPECTRAL PITCH CLASS SIMILARITY MODELS

For our models, we build upon Parncutt's central insight of the tonic as triad, but we use a different measure of the "distance" between the probe tones and this tonic—we use *spectral pitch class similarity* rather than virtual pitch class commonality. Spectral pitch class similarity uses plausible psychoacoustic assumptions

TABLE 6. *Correlations (df = 10) of the Parncutt 94 Model and the Fit Data for Each of the Context-setting Elements.*

Major				Minor			
I	IV–V–I	II–V–I	VI–V–I	I	IV–V–I	II–V–I	VI–V–I
.93	.91	.93	.93	.93	.98	.81	.95

The mean correlation is .92.

to give the similarity between the perceived pitch content of one tone (or chord) and those of another.

We will now provide a brief overview of the mathematical formalization of the model (a more complete description is provided in Appendix C, and the MATLAB routines can be downloaded from http://www.dynamictonality.com/probe_tone_files/). We model the pitch perception of each probe tone and tonic triad tone as taking the form of an HCT (harmonic complex tone). All such HCTs have 12 harmonics. The harmonics, indexed by $n = 1$ to 12, of each tone are weighted by the *roll-off* parameter ρ using $1/n^{\rho}$. This weighting is used as a simple model for their perceptual salience, which is conjectured to be lower for higher harmonics because they are typically acoustically quieter and less easy to perceptually resolve (because adjacent higher harmonics have smaller frequency ratios). The pitch of each harmonic is expressed in a cents (log-frequency) value relative to a reference frequency (e.g., middle C, which is 261.63 Hz) and then transformed modulo 1,200 (the octave in cents). More explicitly, the cents value of a frequency f is given by $1200 \log_2(f / f_{\text{ref}})$ mod 1200, where f_{ref} is the reference frequency. In other words, the pitch of each harmonic is represented as a finely grained pitch class.

Each such harmonic is embedded in a separate vector each with 1,200 elements indexed from zero to 1,199. For example, the first harmonic of an HCT with a nominal pitch of C4 would be represented by a value of 1 at the zeroth element of the first 1,200-element vector; the second harmonic would be represented by a value of $1/2^{\rho}$ at the zeroth element of a second vector, because 0 is the closest integer to $1200 \log_2(2)$ mod 1200; the third harmonic by a value of $1/3^{\rho}$ at the 702nd element of a third vector, because 702 is the closest integer to $1200 \log_2(3)$ mod 1200; and so on, until all twelve harmonics are embedded in twelve vectors. If the notated pitch had been G4, then all the above vectors would have the same elements but circularly transposed up by 700 cents (the 12-TET perfect fifth). Each of these twelve vectors is then circularly convolved by a discrete normal distribution with standard deviation σ, which is the *smoothing width* parameter. As illustrated in Figure 3, the convolution spreads (smears) the salience values across the log-frequency domain and models pitch perceptual uncertainty or noise in that, after convolution, there is a nonzero probability that two similar but nonidentical log-frequencies will be represented by the same (finely grained) pitch class.

The twelve convolved vectors are then summed to give a single 1,200-element spectral pitch class vector denoted **x**. If each of the weights given to the original harmonics is interpreted as a model of their probability of being perceived, the value of each element in the final pitch class vector models the expected number of partials perceived at that log-frequency pitch class.

Using the above-described procedures and parameters we embed the tonic triad in one vector and a probe tone in another. We model their fit with their cosine similarity, which takes a value between 0 and 1. Cosine similarity $s(\mathbf{x},\mathbf{y})$ is the cosine of the angle between the vectors **x** and **y** and it equals 1 when both vectors are parallel and 0 when they are orthogonal. More formally, $s(\mathbf{x}, \mathbf{y}) = \mathbf{x}\mathbf{y}' / \sqrt{\mathbf{x}\mathbf{x}'\mathbf{y}\mathbf{y}'}$, where **x** and **y** are row vectors and $'$ is the matrix transpose operator that converts a row vector into a column vector.

In two of our three models we allow for different weightings of the tonic triads' tones. In Model *a*, we give all their tones the same weights—that is, the saliences of the partials in its three pitch classes, as previously determined by ρ, are multiplied by 1 and so left unchanged. In Model *b*, two weightings are available—the tonic triads' roots have unity weight, while the remaining pitch classes have a weight of ω, which takes a value between 0 and 1; for example, if the tonic triads are Cmaj and Cmin, the saliences of the partials of the pitch class C are left unchanged, while the saliences of the partials of all the remaining pitch classes are multiplied by ω. In Model *c*, there are still two weightings, but this time the unity weight is applied to the roots of the major and minor tonics and also the third of the minor tonic, while the weighting of ω is applied to the remaining pitch classes; for example, if the tonics are Cmaj and Cmin, the weights of the partials of the pitch classes C and E♭ are unchanged, while the weights of the remaining pitch classes are multiplied by ω.

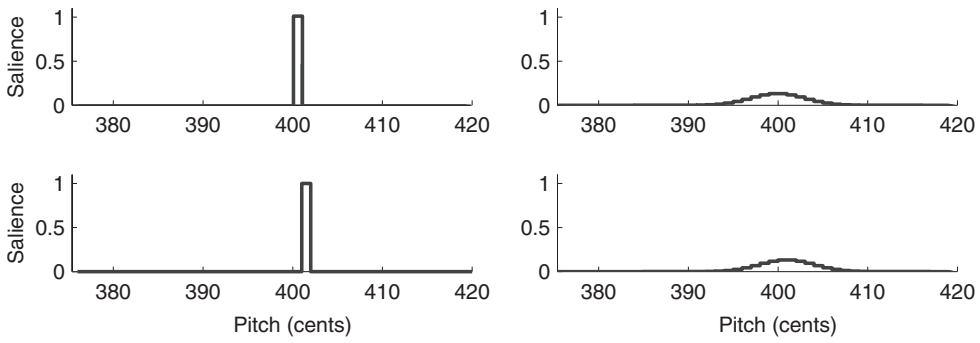

FIGURE 3. Discrete log-frequency embeddings of two partials—one at 400 cents, the other at 401 cents. On the left, no smoothing is applied, so their distance under any standard metric is maximal; on the right, Gaussian smoothing (standard deviation of 3 cents) is applied, so their distance under any standard metric is small.

Model *a* is a pure tonic-as-triad model (all its three pitch classes are equally weighted), but the separate weightings in *b* and *c* allow these models to be situated in continua between tonic-as-triad and tonic-as-pitch-class models. This is useful because it is plausible that, of the tonic triad's pitches, the tonic pitch is the most salient and tonic-like. Model *c* treats the third of the minor triad as an additional root and as a frequent substitute tonic. A bottom-up (sensory) justification for considering the root of a major triad and the root and third of a minor triad as having greater importance is because a typical sensory model will predict that these pitch classes closely correspond to those that are likely to be perceived as possible fundamentals (virtual pitches). For example, Parncutt's (1988) psychoacoustic model predicts the third of a minor triad to have a greater salience than the fifth (*salience*, in this context, is the extent to which it is heard as a fundamental pitch class after matching with a harmonic template). There are also top-down explanations for giving the third of a minor chord a higher weighting than the fifth—in Western music, the third of the minor chord is often treated as a stable root (minor chords in first inversion are not treated as dissonances) and, in minor keys, modulations to the relative major are very common (the tonic of the relative major is the third of the minor tonic's triad). We class models *b* and *c* as bottom-up because there are plausible bottom-up explanations, though we acknowledge that top-down aspects may be playing an important role here too and that the additional predictive abilities of *b* and *c* over *a* may be a result of top-down processes.

The above means that, in addition to the intercept and slope parameters (which are part of every model discussed so far due to the process of obtaining correlation values),[8] Model *a* has two nonlinear parameters (ρ and σ), while models *b* and *c* have three nonlinear parameters (ρ, σ, and ω). This nonlinearity means the parameter values cannot be optimized analytically, so we used MATLAB's fmincon routine to optimize them iteratively. We optimized each model so as to minimize the sum of squared errors between its predictions and the probe tone data—this is the same for all the models discussed in this paper, because obtaining correlation values automatically chooses intercept and slope values that minimize the sum of squared errors.

The optimized parameter values all seem quite plausible: for Model *a*, $\hat{\rho} = 0.52$ and $\hat{\sigma} = 5.71$; for Model *b*, $\hat{\rho} = 0.77, \hat{\sigma} = 6.99$, and $\hat{\omega} = 0.63$; for Model *c*, $\hat{\rho} = 0.67$, $\hat{\sigma} = 5.95$, and $\hat{\omega} = 0.50$.[9] The values of ρ are all similar to the loudnesses of partials produced by stringed instruments (a sawtooth wave, which is often used to synthesize string and brass instruments, has a pressure roll-off equivalent to a ρ of 1 and, using Steven's law, this approximates to a loudness roll-off equivalent to $\rho = 0.60$). Under experimental conditions, the frequency difference limen (just noticeable difference) corresponds to

[8] The correlation coefficient between a model's data and the empirical data is given by $\sqrt{(\hat{\mathbf{y}}-\mathbf{y})'(\hat{\mathbf{y}}-\mathbf{y})/(\mathbf{y}-\bar{\mathbf{y}})'(\mathbf{y}-\bar{\mathbf{y}})}$, where $'$ is the transpose operator which turns a column vector into a row vector, $\mathbf{y}$ is a column vector of the empirical data, $\bar{\mathbf{y}}$ is a column vector all of whose entries are the mean of the empirical data and, critically, $\hat{\mathbf{y}}$ is a column vector of the model's predictions *after* having been fitted by simple linear regression.

[9] With iterative optimization, there is always a danger that a local rather than global minimum of sum of squared errors is found; we tried a number of different start values for the parameters, and the optimization routine always converged to the same parameter values so we are confident they do represent the global optimum.

approximately 3 cents, which would be modeled by a smoothing width of 3 cents (Milne, Sethares, Laney, & Sharp, 2011, Online Supplementary: App. A). In a music experiment like the one being modeled, we would expect the smoothing to be somewhat wider, and the value of around 6 cents seems plausible. It is also worth noting that in an earlier experiment using a related model, our optimized values were $\hat{\rho} = 0.42$ and $\hat{\sigma} = 10.28$ (Milne, Laney, & Sharp, 2015; these values are similar to those found for this experiment, because using them instead has only a small negative impact on the resulting fit (reducing the correlation values by approximately 0.003). This also indicates that the model is robust over such changes to these parameters.

The optimized spectral pitch class similarity models are predictively effective—for models *a*, *b*, and *c*, respectively, the cross-validation statistics are $r_{\mathrm{CV}}(22) = .91$, $r_{\mathrm{CV}}(22) = .92$, and $r_{\mathrm{CV}}(22) = .96$. The predictions made by the three models are shown in Figure 4. They also have great explanatory power—like Parncutt's virtual pitch class model, we are using psychoacoustic principles to explain the specific shape taken by the probe tone data.

Like some of the other models discussed in this paper (e.g., Lerdahl 88 and Parncutt 11), each of ours produces the same outputs across the four contexts, and they also have high fits with the probe tone data for each of the individual contexts, as shown in Table 7.

However, there is one aspect of these models that does not bear a direct relationship with the experimental procedure. In the experiment, the stimuli were all OCTs, not HCTs. In our models, we use HCTs (if OCTs are used as variables, the models perform very poorly). (This is also the case in Krumhansl's and Smith's consonance models, because their consonance values are all derived from HCTs.) There are at least four possible explanations that can bridge the gap between the model's use of HCTs and the experiment's use of OCTs. First, nonlinearities in the auditory system—such as the distortion products measured in brainstem responses to simple chords Lee et al. (2009)—may add harmonics to the OCTs (e.g., a combination tone of any two adjacent OCT partials with frequencies *f* and 2*f*, has a frequency at 3*f* —a third harmonic). Second, when listeners were making their judgments of fit, the representations of the tonic triad and probe they retrieved from short-term memory may have been "contaminated" by long-term representations of HCTs with the same pitch (HCTs being much more familiar). Third, listeners may have recalled the levels of fit, stored in long-term memory, of equivalently sized HCT intervals. Fourth, listeners' judgments of the fit of the probe and the tonic triad are

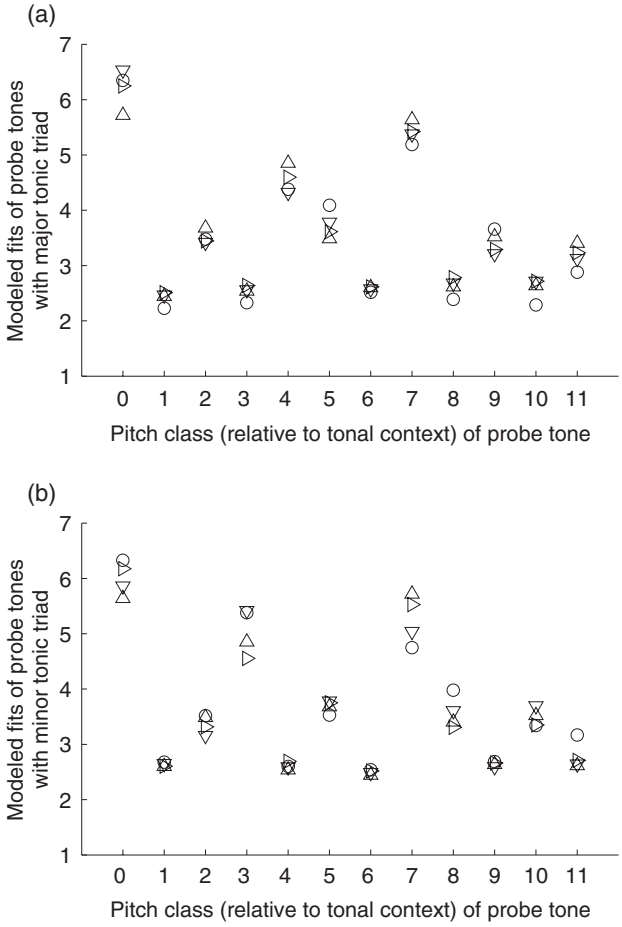

FIGURE 4. The circles show the probe tone data, the upwards pointing triangles show the data as modeled by Model *a*, the rightwards pointing triangles show the data as modeled by Model *b*, the downwards pointing triangles show the data as modeled by Model *c*.

due to musical prevalence, but these musical prevalences are themselves a function of the psychoacoustic process modeled here: specifically, composers usually work with HCTs (not OCTs) and build up a set of tonal prevalences based upon their desire to follow their innate and universal perceptual processes (and "consumers" support music that accords with their similar innate processes). In each of the latter explanations, top-down processes play a role of some kind. But at root, it is the sensory component of this model (spectral pitch class similarity) that actually dictates the final form of the probe tone data. In that sense, these are all essentially bottom-up models even if top-down processes may play an important role in supporting, and indeed strengthening, the patterns determined by spectral pitch class similarity.

TABLE 7. *Correlations (df = 10) of the Milne 14c Model and the Probe Tone Fit Data for Each of the Context-setting Elements.*

Major				Minor			
I	IV–V–I	II–V–I	VI–V–I	I	IV–V–I	II–V–I	VI–V–I
.97	.92	.97	.92	.94	.98	.86	.96

Note: The mean correlation is .94.

A Model of Scalic Tonality

In the previous section, we modeled the fit of pitch classes to a given tonic triad. The same model can also be used to model the tonicness of pitch classes or triads given a scale (when the scale is treated as a pitch class set). We call this a model of *scalic tonality*, because the tonicness of a chord is a function of the scale against which it is compared—even when the scale's pitches have equal weight.[10] To do this relies on an assumption that tonicness and fit are related—that is, that a pitch class or chord must have a high fit to be a tonic. Of course, there may be other factors that affect tonicness, but this is the focus of this model. We do, however, make some speculations about some possible processes that are related and may play an additional role.

In our model of scalic tonality, the spectral pitches of all of a given scale's pitch classes are embedded in one spectral pitch class vector, and the spectral pitches of each possible pitch class or triad are embedded into another, as described in the previous section (each spectral pitch class is given a salience value as determined by the roll-off parameter ρ and smeared according to the smoothing width parameter σ). In this way, the fit of the scale and the pitch class or chord—and hence the tonicness of the pitch class or chord—can be modeled by their spectral pitch class similarity. In all of the examples in this section, we used $\rho = 0.67$ and $\sigma = 5.95$, as optimized for Model *c* (we could have chosen the values as optimized for any of the three models, but Model *c*'s values fall between those of models *a* and *b*, so seemed a sensible choice; furthermore, the results are robust over the three sets of values). Also, the candidate tonic triads have equally weighted pitch classes, which means the model is effectively equivalent to Model *a* described in the previous section. In other words, the root-weighting parameter ω is not used in the scalic tonality model.

It should be noted that Parncutt uses a similar fit-based technique (using virtual rather than spectral pitch classes) to determine the pitch class tonics for the diatonic scale (Parncutt, 2011; Parncutt & Prem, 2008) in medieval music. However, his approach is inconsistent in the same way as in the Parncutt 11*a* model in that the scale pitch class set is modeled with virtual pitches, while the candidate tonic pitch classes are not. In the following examples, we additionally look for tonic triads as well as pitch classes, and we model the scale and candidate tonics consistently—their pitch classes have identical harmonic spectra and all pitch classes are equally weighted (with one noted exception).

For this scalic tonality model to make sense requires that we consider the scales as known entities (in either short-term or long-term memory). For a scale to be known, it must be perceived as a distinct selection of pitches or as a specific subset of a chromatic gamut of pitch classes. A composer or performer aids this by ensuring all scale pitch classes are played over a stretch of time short enough for them all to be maintained in short-term memory, and by utilizing scales that have relatively simple and regular structures (well-formed scales provide an excellent example of a scale type that is both simple and regular and, more generally, scales that are subsets of a relatively small gamut of "chromatic" pitches). Long-term memory is also likely to play an important role in that certain scales are learned through repetitive exposure.

Up to this point, we have used uppercase Roman numeral notation, so IV–V–I in a major key means all chords are major, while IV–V–I in a minor key means the first and last chords are minor. In the following sections we are dealing with specific scales, so we use upper case to denote major triads and lower case to denote minor. For example, the above minor tonality cadence is now denoted iv–V–i.

FIT PROFILES FOR 12-TET SCALES

In this section, we consider a variety of scales that can be thought of as subsets of the twelve pitch classes of twelve-tone equal temperament.

[10] It is worth noting that, for an abstract scale in which all pitch classes are equally weighted, a pure short-term memory model (such as Butler's, 1989) will give homogeneous fits for all in-scale pitch classes or chords. The additional structure resulting from the addition of harmonics, or subharmonics, makes the fits of different in-scale pitch classes and chords heterogeneous.

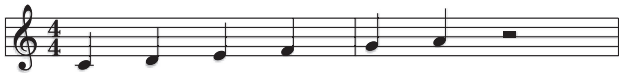

FIGURE 5. C (Guidonian) hexachord.

Major (Guidonian) hexachord. This six-tone scale formed the basis of much medieval music theory and pedagogy (Berger, 1987). It is equivalent to a diatonic scale with the fourth or seventh scale degree missing. For instance the C hexachord contains the pitches C, D, E, F, G, A. There is no B or B♭ to fill the gap between A and C. In modal music, the note used to fill the gap was either a *hard* B (a B♮) or a *soft* B (a B♭).[11] The choice of hard or soft was not notated but was made by performers to avoid simultaneous or melodic tritones—this practice is called *musica ficta* (Berger, 1987). This scale is illustrated in Figure 5.

In Figure 6, we will assume that pitch class 0 corresponds to C. Figure 6a shows that the pitch classes E and F (4 and 5), which are a semitone apart, are the least well-fitting of the hexachord tones. In Gregorian chant, the *finalis* (final pitch) was D, E, F, or G (corresponding to the modes *protus*, *deuterus*, *tritus*, and *tetrardus*). Of these modes, Figure 6a shows that the pitch classes with the highest fit are at D and G (2 and 7), which suggests these two modes have the most stable final pitches. This tallies with statistical surveys, referenced in Parncutt (2011), which indicate these two modes were the most prevalent. The relative fits of D and G are even higher when the hexachord has a Pythagorean tuning in which all its fifths have the frequency ratio 3/2—such tunings were prevalent prior to the fifteenth century (Lindley, 2013).

When we look at the modeled fit of each of the hexachord's major and minor triads with all the pitches in the hexachord, the results are quite different (Figure 6b). Here, every major or minor chord has identical fit with this scale. It is as if the Guidonian hexachord—when used for major/minor triad harmony—has no identifiable best-fitting tonic chord. As shown in the next example, all of this changes when that missing seventh degree is specified, thereby producing a specific diatonic scale.

Diatonic major scale. The diatonic scale—regardless of its mode—has numerous properties that make it perceptually and musically useful. A number of those properties follow from its well-formedness (Carey & Clampitt, 1989; Wilson, 1975) such as Myhill's property,

[11] The shape of the natural and flat symbols derive from two different ways of writing the letter "b."

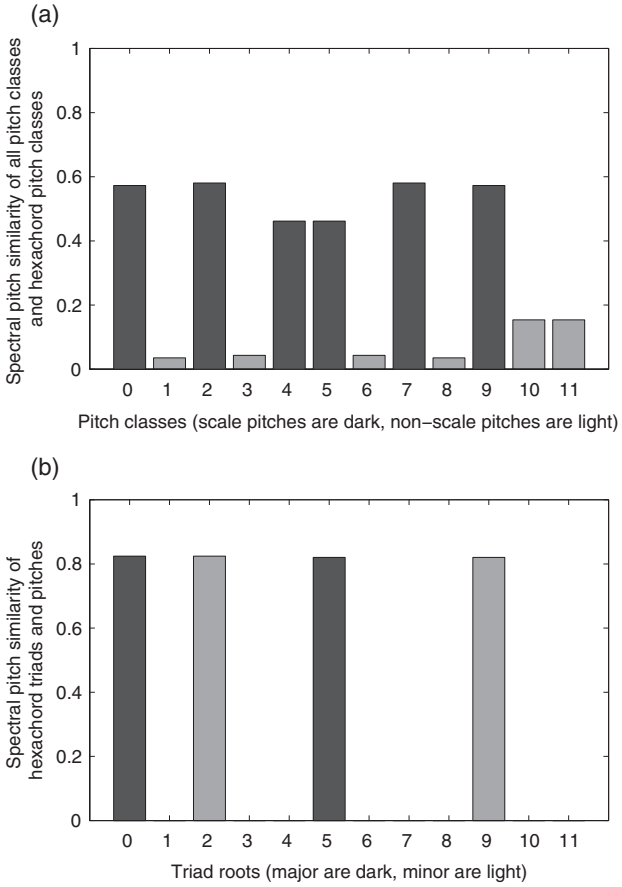

FIGURE 6. Modeled pitch class and chord fit with the Guidonian hexachord.

maximal evenness, uniqueness, coherence, and transpositional simplicity.[12] Furthermore, it contains numerous consonant intervals (approximations of low integer frequency ratios), and supports a major or minor triad on all but one of its scale degrees. For tonal-harmonic music, the major scale (e.g., C, D, E, F, G, A, B) is the

[12] *Myhill's property* is that every *generic interval* (e.g., second, third, fourth) comes in two *specific sizes* (as measured in a log-frequency unit like semitones or cents). *Maximal evenness* means an *N*-tone scale's large and small steps are arranged so as to most closely approximate an *N*-tone scale with equally sized steps. *Uniqueness* means each scale degree is surrounded by a unique set of specific intervals (this does not occur in equal-step scales or scales with patterns that repeat at sub-octave intervals like the diminished). *Coherence* means the interval size (in cents or semitones) spanned by any *n* consecutive scale notes is always larger than the interval size spanned by $n - 1$ consecutive scale notes; for instance, a diatonic scale in Pythagorean tuning is not coherent because the (augmented) fourth between F and B is larger than the (diminished) fifth from B to F. *Transpositional simplicity* means the scale can be transposed so as to produce a new scale that shares all but one pitch class with the untransposed scale.

most important and prevalent mode of the diatonic scale. The only other mode that comes close is the Aeolian (e.g., A, B, C, D, E, F, G, or C, D, E♭, F, G, A♭, B♭)—also known as the *natural minor scale*—which is one of the three scale forms associated with the minor scale (the other two are the harmonic minor, in which the Aeolian's seventh degree is sharpened, and the ascending melodic minor in which the sixth and seventh degrees are sharpened). The C major diatonic scale is illustrated in Figure 7.

The addition of a seventh tone to the hexachord—thereby making a diatonic scale—makes the fits of its triads more heterogeneous. Figure 8b illustrates this with the diatonic major scale—note how the Ionian and Aeolian tonic triads (the chords shown on pitch classes 0 and 9, respectively) are modeled as having greater fit than all the remaining triads. This, correctly, suggests they are the most appropriate tonics of the diatonic scale—the major scale's tonic and the natural minor scale's tonic, respectively. The tonicness of the diatonic vi chord is also reflected in its use as a substitute for the tonic (I) in deceptive cadences (Macpherson 1920, p. 106; Piston & Devoto, 1987, p. 191), and the frequent modulation of minor keys to their relative major (Piston & Devoto, 1987, p. 61). It is also interesting to observe that the fourth and seventh degrees of the major scale have lower fit than the remaining tones. This possibly explains why these two scale degrees function as leading tones in tonal-harmonic music—scale degree $\hat{7}$ resolving to $\hat{1}$, and $\hat{4}$ resolving to $\hat{3}$—for example, both these motions occur in the dominant seventh to tonic cadence (i.e., V^7–I). They function as leading tones because listeners anticipate that a poor-fitting, hence unstable, tone will move to a stable good-fitting tone.

There are five aspects of major-minor tonality not obviously explained by the above fit profiles: (a) in the diatonic scale, the Ionian tonic is privileged over the Aeolian tonic; (b) in the major scale, the seventh scale degree is typically heard as more active—more in need of resolution—than the fourth degree; (c) the importance of the V–I cadence; (d) the activity of the seventh degree of the major scale is significantly reduced when it is the fifth of the iii (mediant) chord in comparison to when it is the third of the V (dominant) chord. We propose two additional hypotheses that may account for these features.

A bottom-up hypothesis to explain the first two features is that the strongest sense of harmonic resolution is induced when a bad-fitting (low spectral pitch class similarity) tone moves by semitone to the *root* of a best-fitting (high spectral pitch class similarity) chord, where the spectral pitch class similarities are measured with respect to the scale. In the white-note diatonic scale,

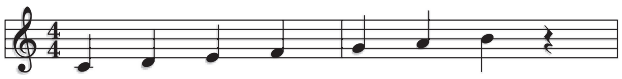

FIGURE 7. C major scale.

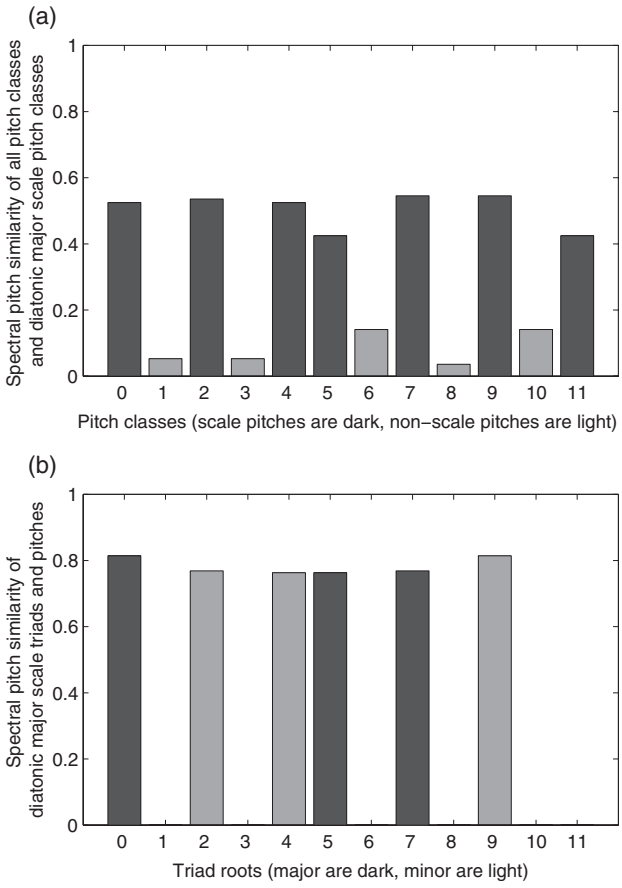

FIGURE 8. Modeled pitch class and chord fit with the major scale.

there are two best-fitting triads (Cmaj and Amin) and two worst-fitting pitch classes (B and F). This means that only Cmaj has a root (C) that can be approached by semitone from a worst-fitting pitch class (B); for Amin, the root (A) cannot be approached, by semitone, by either B or F. If we assume that this provides a built-in advantage to the Ionian mode, this introduces an interesting feedback effect. Let us now weight the pitch class C a little higher than the other tones to reflect its status as the root of a best-fitting triad that is approached, by semitone, by a worst-fitting pitch—the results of this are illustrated in Figure 9 where the weight of C is twice that of the other tones (possibly an extreme value, but it demonstrates the effect). Although the pitch class C is a member of both the C

major and A minor tonics, Figure 9b shows that increasing its weight disproportionately enhances the fit of the triad Cmaj over the triad Amin. It also decreases the fit of B (Figure 9a). It seems likely, therefore, that this results in a positive feedback loop: we hypothesize that the resolution of the poor-fitting B to the root of Cmaj increases the perceived fit of C; we model this by giving the C a greater weight, and this disproportionately increases the fit of Cmaj over Amin, and reduces the fit of B; this is likely to result in an even stronger resolution from B to the root of Cmaj (B is worse fitting than before, and Cmaj is better fitting) and this, in turn, will further enhance the fit of pitch class C and thereby enhance the fit of Cmaj over Amin, and so on in a positive feedback loop.

The third feature—the importance of the V–I cadence, which is typically described as the "strongest" or "most powerful" progression in tonal music (Piston & Devoto, 1987, p. 21; Pratt, 1996, p. 9)—also follows, in part, from the same hypothesis that resolution is enhanced by a low-fit pitch moving to the root of a high-fit triad. This favors the resolutions V–I or vii°–I (which contain the scale degrees $\hat{7}$–$\hat{1}$—a resolution to the tonic's root), over IV–I or ii–I (which contain the scale degrees $\hat{4}$–$\hat{3}$—a resolution to the tonic's third). It is also interesting to note that V^7–I and vii°–I, which provide the strongest tonal resolutions, contain both $\hat{7}$–$\hat{1}$ and $\hat{4}$–$\hat{3}$.

However, this suggests that iii–I would also provide an effective cadence because it too has the worst-fitting $\hat{7}$ resolving to the root of I. But such cadences are rare (Piston & Devoto, 1987, p. 21), and the activity of the seventh degree is typically felt to be much reduced when it is the fifth of the iii chord—a common use of the iii chord is to harmonize the seventh degree when it is descending to the sixth (Macpherson, 1920, p. 113). This may be explained by a second hypothesis, which is that we need to consider the fit of pitches not just in relation to their scalic context, but also in relation to their local harmonic (chordal) context. Against the context of a major or minor chord, the third is the worst-fitting pitch—see Figure 10 (all triad pitches are equally weighted), which shows that both chords' thirds (pitch class 4 for the major triad, and 3 for the minor) have lower fit than the root and fifth (pitch classes 0 and 7). This suggests that the higher fit of scale degree $\hat{7}$ in iii—due to it being the chord's fifth—makes it less active; while the lower fit of $\hat{7}$ in V—due to it being the chord's third—makes it more active. This hypothesis, therefore, explains the greater stability of the seventh degree in iii compared to V, and completes the explanation for the importance of the V–I, V^7–I, and vii°–I cadences.

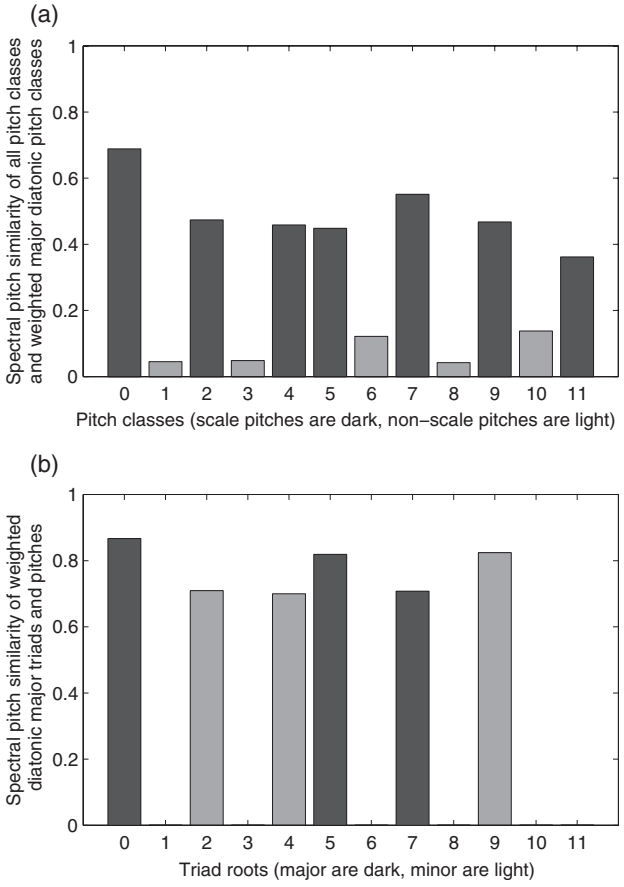

FIGURE 9. Modeled pitch class and chord fit with a major scale with a double-weighted tonic pitch class.

These additional hypotheses (the importance of semitone resolutions from poor-fit tones to roots of good-fit triads, and the decreased fit of pitches that are the thirds of chords) seem promising in that they may determine precisely which semitone movements will function as leading tone resolutions and which will not. In future work, we hope to precisely specify these effects, and use them to model responses to a variety of chord progressions and scalic contexts.

Harmonic minor scale. An important aspect of the minor tonality is that the harmonic minor scale is favored over the diatonic natural minor scale—particularly in common practice cadences where (the harmonic minor) V–i is nearly always used in preference to (natural minor) v–i (Piston & Devoto, 1987, p. 39). The harmonic minor scale is equivalent to the Aeolian mode with a sharpened seventh degree. This change has an important effect on the balance of chordal fits—and goes some way to explaining why this scale forms the basis of minor tonality in

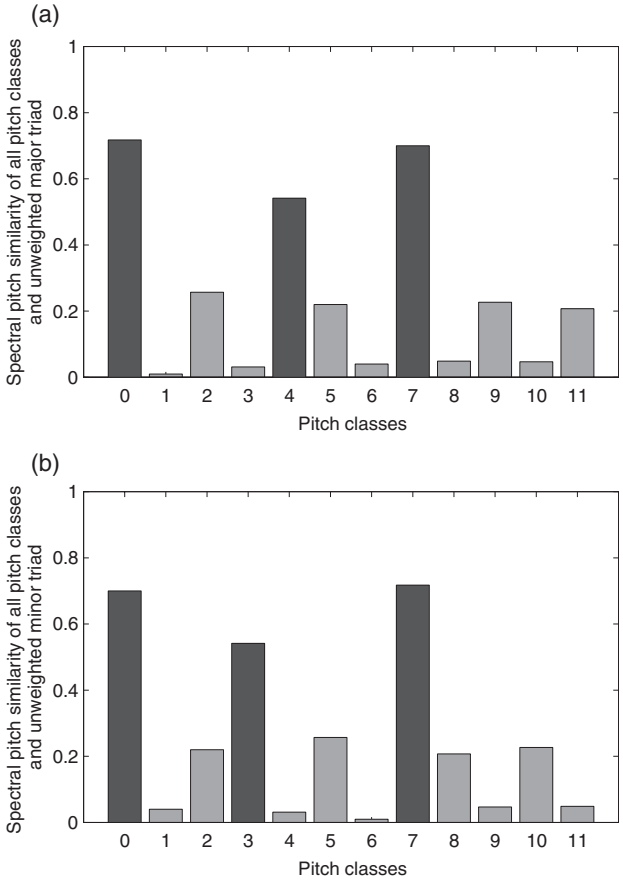

FIGURE 10. Modeled pitch class fits with unweighted major and minor triads.

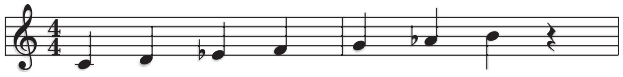

FIGURE 11. C harmonic minor scale.

Western music. The C harmonic minor scale is illustrated in Figure 11.

Figure 12a shows that $\hat{7}$ is clearly the worst-fitting scale degree; the next worst are $\flat\hat{6}$ and $\hat{2}$. Figure 12b shows that the best-fitting triad is i; furthermore, every pitch in this tonic i chord can be approached by the three most poorly fitting scale degrees which, therefore, act as effective leading tones: $\hat{7}$–$\hat{1}$, $\flat\hat{6}$–$\hat{5}$, and $\hat{2}$–$\flat\hat{3}$—as exemplified by a chord progression like Bdim7–Cmin, or G7♭9–Cmin. These properties appear to make this scale a context that provides unambiguous support of a minor triad tonic. Compare this to the diatonic mode, where there is an equally well-fitting major triad; for example, Macpherson (1920, p. 162)

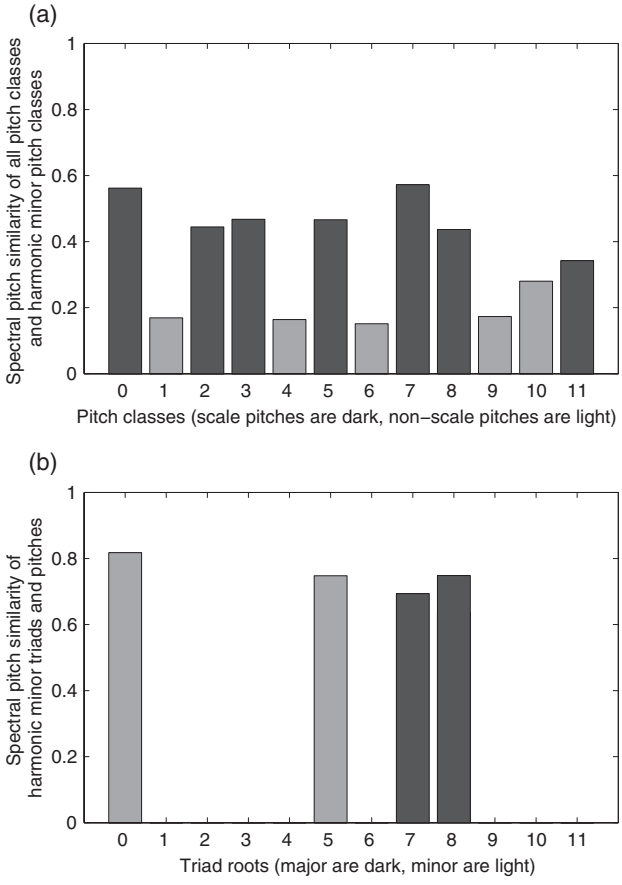

FIGURE 12. Modeled pitch class and chord fit with the harmonic minor scale.

writes that, "any chord containing the minor 7th usually requires to be followed as soon as possible by a chord containing the Leading-note . . . otherwise the tonality easily becomes vague and indeterminate, and the music may tend to hover somewhat aimlessly between the minor key and its so-called 'relative' major."

Ascending melodic minor scale. It is well-recognized in music theory that the harmonic minor scale provides effective harmonic support for a minor tonic, but that it is also melodically awkward due to the augmented second between its sixth and seventh degrees. When a melodic line is moving from the sixth to the seventh degree, this awkward interval is typically circumvented by sharpening the sixth degree—this produces the ascending melodic minor scale (the descending melodic minor scale is identical to the natural minor scale; Aeolian mode). The C ascending melodic minor scale is illustrated in Figure 13.

Figure 14b shows that, in terms of chord fits, this scale has returned to a similar situation as the Guidonian

hexachord: all chords have equal fit, hence there is no obvious tonic. This suggests that using this scale, for brief periods of time to improve the melodic line, will not disrupt a minor tonality previously established with the parallel harmonic minor scale. However, this scale cannot form the foundation of a minor tonality, because it has no specific tonal centre (when triads are used). Again, this seems to be in accord with conventional tonal music theory, which specifies that the primary function of this scale is to improve melodic connections rather than to provide the basis for harmony (the use of the raised sixth degree, like A♮ in C minor, is usually subject to strict melodic conventions—e.g., Schoenberg (1969, p. 18) advises that it should not move to the "natural" sixth, which is A♭ in C minor, or the "natural" seventh degree, which is B♭ in C minor).

Harmonic major scale. In the same way that sharpening the seventh degree of the Aeolian mode can make its tonic unambiguously the best-fitting, it is interesting to consider if there is a different alteration that can do the same for the Ionian mode. The alteration that seems to provide a similar benefit for the Ionian is to flatten its sixth degree, which forms the harmonic major scale. The harmonic major scale plays a notable role in Russian tonal music theory as exemplified by Rimsky-Korsakov (1885). The C harmonic major scale is illustrated in Figure 15.

In comparison to Figure 8b, Figure 16b shows how the I chord is now the uniquely best-fitting chord. This appears to indicate that flattening the sixth degree of the major scale strengthens the major tonality. This accords with Harrison's (1994, pp. 15-34) description of the chromatic iv in major as the tonic-strengthening dual of the "chromatic" V in minor. However, like the harmonic minor scale, this alteration creates an awkward sounding melodic interval—the augmented second between the sixth and seventh degrees—which maybe explains why this scale is not considered to be the primary major tonality scale.

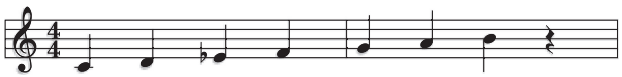

FIGURE 13. C ascending melodic minor scale.

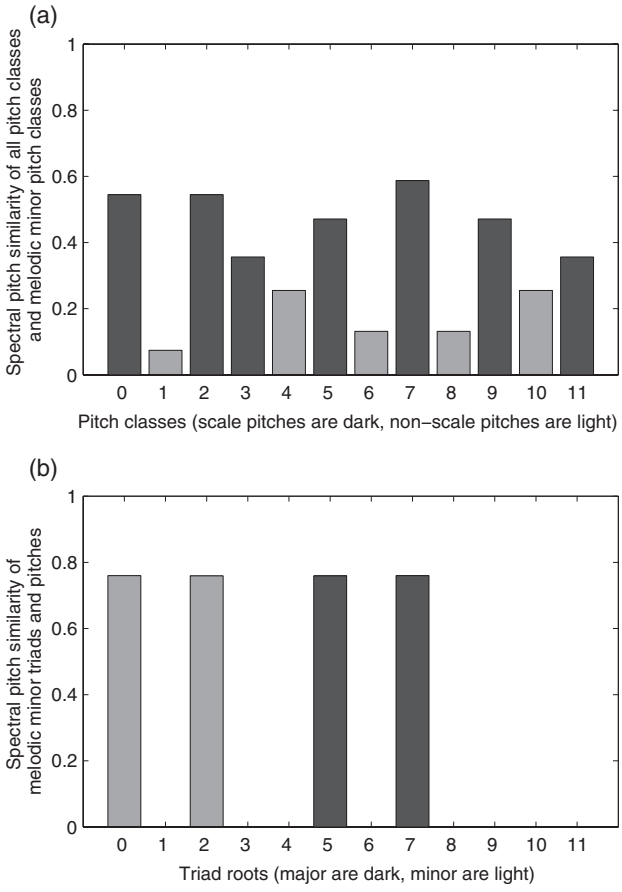

FIGURE 14. Modeled pitch class and chord fit with the ascending melodic minor scale.

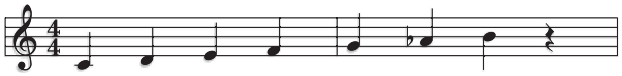

FIGURE 15. C harmonic major scale.

FIT PROFILES FOR MICROTONAL SCALES.

Unlike all of the previously discussed models, ours is generalizable to pitches with any tuning (e.g., microtonal chords and scales). It is interesting to explore some of the predictions of pitch class and chord fit made by the model given a variety of microtonal scales. All of the microtonal scales we analyze here are well-formed. We do this under the hypothesis that the simple and regular structure of such scales may make them easier to hold in short-term memory, or learn as part of long-term memory—all well-formed scales have a number of useful musical properties including the previously described Myhill's property, uniqueness, maximal evenness, transpositional simplicity.[13]

[13] Equal step scales are structurally simpler and more regular than well-formed scales, but they are actually too regular because their internal structure is completely uniform—every pitch class or chord bears the same relationship to all other scale pitches and chords. The structure of equal step scales cannot, therefore, support a different musical function on different scale degrees—such a musical function may be imposed by pitch repetition or a drone, but it is not inherent to the scale, merely to its usage.

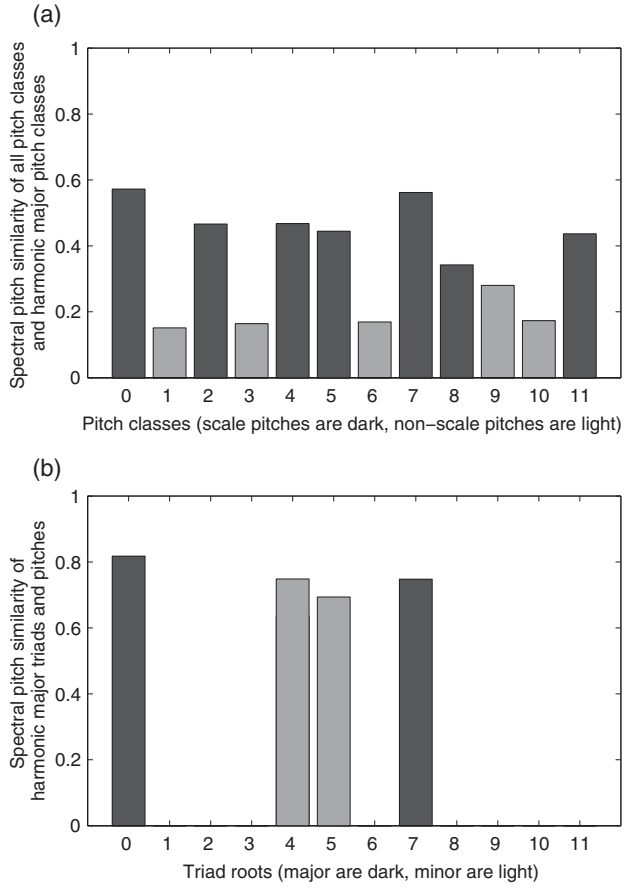

FIGURE 16. Modeled pitch class and chord fit with the harmonic major scale.

Quarter-comma meantone diatonic scale. This tuning was first described by Pietro Aaron in 1523 (cited in Barbour, 1951) who described a system of temperament where every perfect fifth is equally flattened slightly but all major thirds are perfectly tuned. This is around the time that modal music began its gradual transition into harmonic tonality, and may have been a prevalent tuning at that time. For that reason it is interesting to see what, if any, impact it has on the fit of the diatonic pitches and chords. One aspect that differentiates meantone tunings from 12-TET is that enharmonically equivalent pitches (e.g., C♯ and D♭) do not have identical tunings. For this reason, we use a gamut of 19 pitch classes (e.g., the chain-of-fifths from C♭ to E♯), which provides a sharp and a flat for every diatonic scale degree (e.g., C, D, E, F, G, A, B) except for the fourth (e.g., F) which has no flat, and the seventh (e.g., B) which has no sharp. Another difference is that its major and minor triads are, by any standard metric, closer to the low integer ratios of just intonation (4:5:6 and 10:12:15, respectively) than the 12-TET versions: the just intonation triads are, to the nearest cent, (0, 386, 702) and (0, 316, 702); the quarter-comma meantone triads, to the nearest cent, are (0, 386, 697) and (0, 310, 697); the 12-TET triads are (0, 400, 700) and (0, 300, 700).

For the diatonic scale degrees and chords, the overall pattern of fits is similar to that produced by 12-TET—as shown in Figure 17. The fourth and seventh scale degrees are still modeled as the worst fitting, and the Ionian and Aeolian tonic triads are still modeled as the best fitting. This suggests that this pattern and, hence, its tonal implications, are robust over such changes in the underlying tuning of the diatonic scale.

22-TET 1L, 6s porcupine scale. In the following three examples, we look at different well-formed scales that are subsets of 22-tone equal temperament. The names of these temperaments (*porcupine*, *srutal*, and *magic*) are commonly used in the microtonal community, and are explained in greater detail in Erlich (2006) and the website http://xenharmonic.wikispaces.com/. In all of these scales, the tunings—rounded to the nearest cent—of the major triads are (0, 382, 709), and the tunings of the minor triads are (0, 327, 709). These tunings are, by most standard metrics, closer to the just intonation major and minor triads than those in 12-TET. For each scale, the spectral pitch class similarities suggest one or more triads that will function as tonics. We do not, at this stage, present any empirical data to substantiate or contradict these claims; but we suggest that collecting such empirical data—tonal responses to microtonal scales—will be a useful method for testing bottom-up models of tonality. Audio examples of the scales, their chords, and some of the cadences described below, can be downloaded from http://www.dynamictonality.com/probe_tone_files/. The intervallic structure of these scales can also be gleaned from Figures 18a, 19a, and 20a, where the scale pitches are shown by dark bars against a light grey 22-TET "chromatic" background.

The porcupine scale has seven tones and is well-formed—it contains one large step of size 218 cents and six small steps of size 164 cents (hence its signature 1L, 6s), and the scale pitch classes are indicated with dark bars in Figure 18a. Figure 18b shows that the major triad on 18 and the minor triad on 9 are modeled as the best-fitting. This suggests that, within the constraints of this scale, they may function as tonics. The worst-fitting pitch classes are 6 and 12, which can both lead to the root of the minor triad on 9. Neither of these potential leading tones are thirds of any triads in this scale, which possibly reduces their effectiveness when using triadic harmony. However, the above suggests the

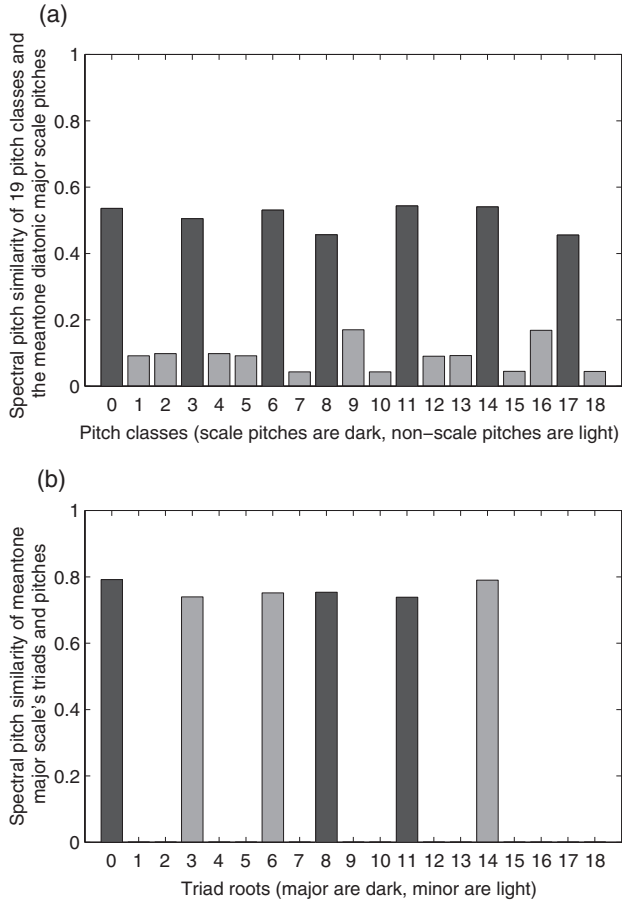

FIGURE 17. Modeled pitch class and chord fit with the 1/4-comma meantone diatonic major scale.

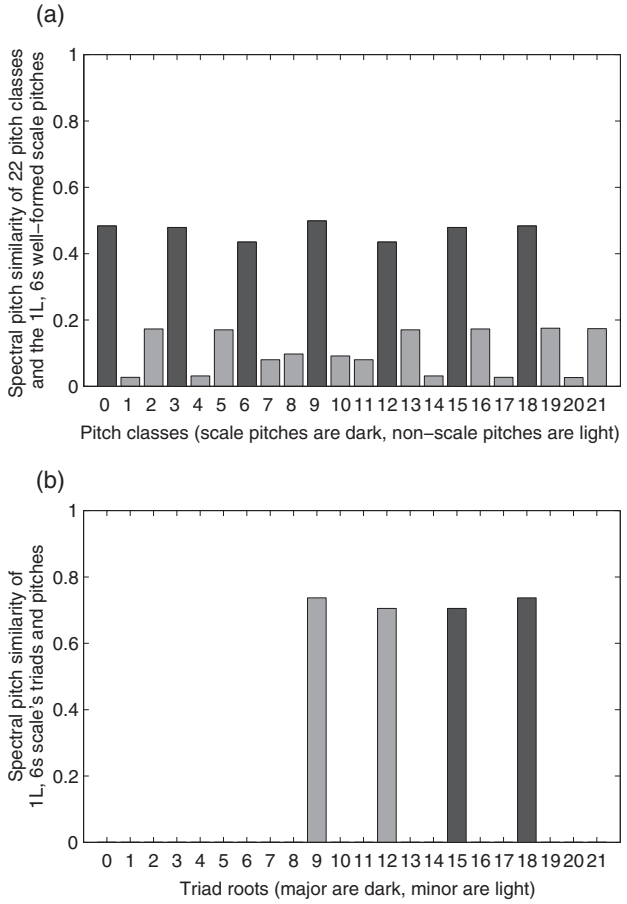

FIGURE 18. Modeled pitch class and chord fit with the porcupine 1L, 6s scale.

most effective cadences in this scale will be the minor chord on 12 leading to the minor chord on 9, the major chord on 15 (whose fifth is pitch class 6) leading to the minor chord on 9, or a variety of seventh chords containing both 6 and 12 like the dominant seventh built on 15 (whose third is 6 and seventh is 12) also leading to the minor chord on 9. Using Roman numerals, taken relative to the minor tonic on pitch class 9, these are ii–i, III–i, and III^7–i, respectively.

22-TET 2L, 8s srutal scale. This ten-tone microtonal scale—first suggested by Erlich (1998)—is unusual in that it repeats every half-octave (it is well-formed within this half-octave interval). This repetition accounts for why the fit levels—shown in Figure 19—also repeat at each half-octave. It contains two large steps of size 164 cents, and eight small steps of size 109 cents. The scale pitches are indicated with dark bars in Figure 19a. The modeled fits suggest there are two possible major triad tonics (on pitch classes 4 and 15) and two possible minor tonics (on pitch classes 2 and 13). The roots of both the minor chords can be approached by a poorer-fitting leading tone (pitch classes 0 and 11) than can the major (pitch classes 2, 6, 13, and 17). This suggests effective cadences can be formed with the major chord on 15 (whose third is pitch class 0) proceeding to the minor chord on 2 (or their analogous progressions a half-octave higher), or variety of seventh chords such as the dominant seventh on 4 (whose seventh is pitch class 0). Using Roman numerals relative to the minor tonic on 2 (or 13), these are VII–i and II^7–i, respectively. These cadences can be thought of as slightly different tunings of the familiar 12-TET progressions V–i and $\flat II^7$–i.

22-TET 3L, 7s magic scale. This microtonal scale also has ten tones, and is well-formed with respect to the octave (so no repetition at the half-octave)—it has three large steps of size 273 cents and seven small steps

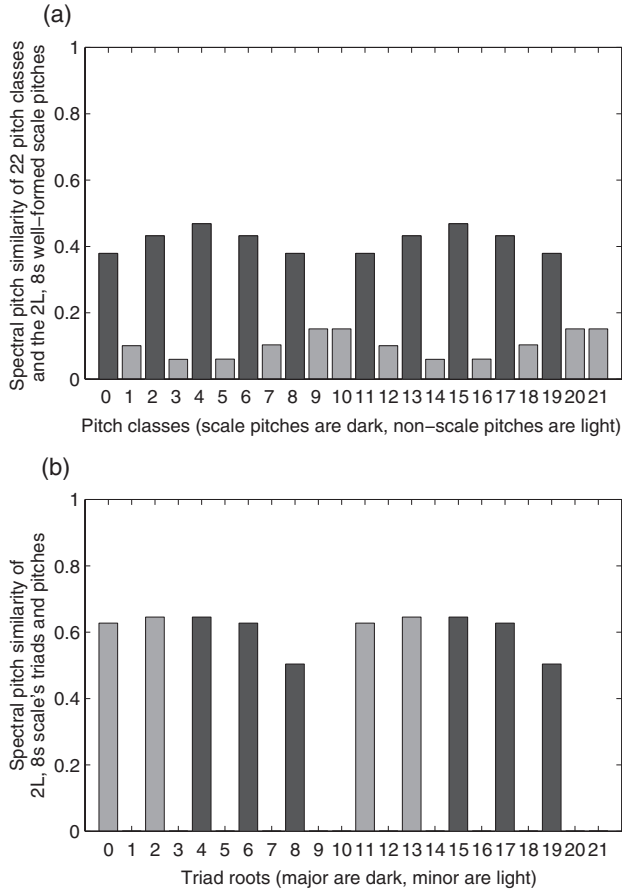

FIGURE 19. Modeled pitch class and chord fit with the srutal 2L, 8s scale.

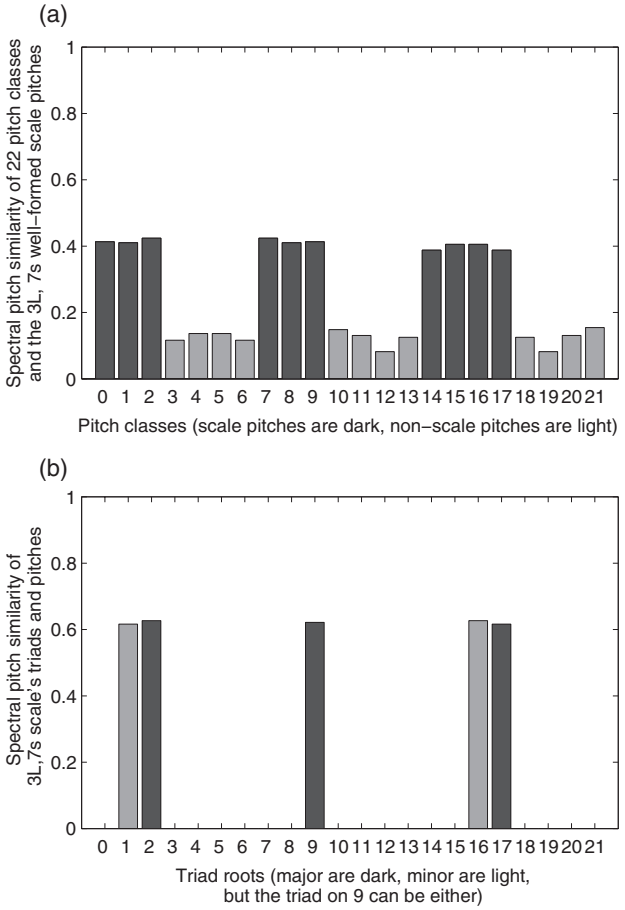

FIGURE 20. Modeled pitch class and chord fit with the magic 3L, 7s scale.

of size 55 cents. As before, the dark bars in Figure 20a indicate the scale pitches. In this scale, every degree that is a root of a major triad is also a root of a minor triad (and vice versa). For this reason, in Figure 20b, only the better fitting (major or minor) is shown on the chart; for the pitch class 9, however, the major and minor triad have equal fit, so this should be borne in mind.

The modeled fits, in Figure 20b, suggest two possible major tonics (with roots on pitch classes 2 and 9) and two possible minor tonics (on pitch classes 9 and 16). Figure 20a shows that, in terms of fit, pitch class 17 looks like a promising leading tone to the root of the minor triad on 16. However, this pitch class is not the third of any triad in the scale. The other leading tone contenders are on 1 and 8, and both of these can be triad thirds. This implies the major chord on 2, and the major or minor chord on 9, may function as tonics in this scale. This suggests effective cadences can be formed with the major chord on 16 (whose third is pitch class 1) proceeding to the major triad on pitch class 2, or the major chord on pitch class 1 (whose third is pitch class 8) proceeding to the major or minor triad on pitch class 9. In Roman numeral notation, relative to their respective tonics, these are VII–I, VII–I, and VII–i. Interestingly, in all these examples the cadences are—in terms of 12-TET—similar to a major chord, whose root is pitched in-between V and ♭VI, proceeding to I or i (the distance between these roots is 764 cents).

Conclusion

We have shown that there at least two types of plausible bottom-up model—Parncutt's virtual pitch class commonality models, and our spectral pitch class similarity models—that can explain why the probe tone data take the form they do. We argue that bottom-up explanations, such as these, are able to account not just for the

existence of fit profiles (as provided by top-down models), but also for the specific form that they take. In light of both theories' ability to explain and predict the data, we suggest that there is now little reason to believe the probe tone data are a function purely of top-down processes. We cannot, on the basis of the probe tone data, determine whether the primary mechanism is spectral pitch class or virtual pitch class similarity. To distinguish between these effects would require novel experiments.

We have also used our model to predict candidate tonic triads for a number of scales that are subsets of the full twelve chromatic pitch classes. The results accord well with music theory. Furthermore, we have suggested some additional mechanisms that may account for strong cadences (a poor-fitting tone moving to the root of a best-fitting triad) and how this, in turn, may cause the diatonic scale to become more oriented to its major (Ionian) tonic rather than its minor (Aeolian) tonic. We also suggested a possible reason for why the seventh degree loses much of its activity (need to resolve) when it is the fifth of the mediant (iii) chord. And, in combination, these two mechanisms support the use of V–I as a cadential chord progression. These latter hypotheses are somewhat speculative because they have not been included in a formal mathematical model, but we feel they are promising ideas that warrant further investigation.

Finally, we have pointed to the way in which microtonal scales can also be analyzed with this technique, and how this may become an important means to explore our general perception of tonality, and to test models thereof. Ideally, any model that purports to explain—from the bottom up—how Western tonality works, should also be able to make useful predictions for the possibly different tonalities evoked by completely different scales and tunings.

Author Note

Correspondence concerning this article should be addressed to Andrew J. Milne, MARCS Institute, University of Western Sydney, Locked Bag 1797, Penrith, 2751, NSW, Australia. E-mail: andymilne@tonalcentre.org

References

Auhagen, W., & Vos, P. G. (2000). Experimental methods in tonality induction research: A review. *Music Perception*, *17*, 417-436.

Barbour, J. M. (1951). *Tuning and temperament: A historical survey*. East Lansing, MI: Michigan State College Press.

Berger, K. (1987). *Musica ficta: Theories of accidental inflections in vocal polyphony from Marchetto da Padova to Gioseffo Zarlino*. Cambridge, UK: Cambridge University Press.

Budrys, R., & Ambrazevičius, R. (2008). 'Tonal' vs 'atonal': Perception of tonal hierarchies. In E. Cambouropoulos, R. Parncutt, M. Solomos, D. Stefanou, & C. Tsougras (Eds.), *Proceedings of the 4th Conference on Interdisciplinary Musicology* (pp. 36-37). Thessaloniki, Greece: Aristotle University.

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*, 219-242.

Carey, N., & Clampitt, D. (1989). Aspects of well-formed scales. *Music Theory Spectrum*, *11*, 187-206.

Deutsch, D. (1997). *The fabric of reality: Towards a theory of everything*. London, UK: Penguin Books.

Erlich, P. (1998). Tuning, tonality, and twenty-two-tone temperament. *Xenharmonikôn*, *17*, 12-40.

Erlich, P. (2006). A middle path between just intonation and the equal temperaments, part 1. *Xenharmonikôn*, *18*, 159-199.

Francès, R. (1988). *The perception of music*. (W. J. Dowling Trans.) Hillsdale, NJ: Lawrence Erlbaum Associates.

Harrison, D. (1994). *Harmonic function in chromatic music: A renewed dualist theory and an account of its precedents*. Chicago, IL: University of Chicago Press.

Helmholtz, H. L. F. von (1954). *On the sensations of tone* (A. J. Ellis, Trans.). New York: Dover. (Original work published 1877)

Huron, D. (1994). Interval-class content in equally tempered pitch-class sets: Common scales exhibit optimum tonal consonance. *Music Perception*, *11*, 289-305.

Hutchinson, W., & Knopoff, L. (1978). The acoustic component of Western consonance. *Interface*, *7*, 1-29.

Kameoka, A., & Kuriyagawa, M. (1969). Consonance theory parts 1 and 2. *Journal of the Acoustical Society of America*, *45*, 1451-1469.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. Oxford, UK: Oxford University Press.

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

Large, E. W. (2011). A dynamical systems approach to musical tonality. In R. Hys & V. K. Jirsa (Eds.), *Nonlinear dynamics in human behavior studies in computational intelligence* (Volume 328, pp. 193-211). Berlin, Germany: Springer.

Large, E. W., & Almonte, F. V. (2012). Neurodynamics, tonality, and the auditory brainstem response. *Annals of the New York Academy of Sciences*, *1252*, E1-E7.

Lee, K. M., Skoe, E., Kraus, N., & Ashley, R. (2009). Selective subcortical enhancement of musical intervals in musicians. *Journal of Neuroscience*, *29*, 5832-5840.

Leman, M. (2000). An auditory model of the role of short-term memory in probe-tone ratings. *Music Perception*, *17*, 481-509.

Lerdahl, F. (1988). Tonal pitch space. *Music Perception*, *5*, 315-350.

Lerdahl, F. (2001). *Tonal pitch space*. Oxford, UK: Oxford University Press.

Lewandowski, S., & Farrell, S. (2011). *Computational modeling in cognition: Principles and practice*. Los Angeles, CA: Sage.

Lindley, M. (2013). Pythagorean intonation. In *New Grove Dictionary of Music and Musicians* (Vol. 15, pp. 485-487). Oxford, UK: Oxford University Press.

Lynch, M. P., Eilers, R. E., Oller, D. K., & Urbano, R. C. (1990). Innateness, experience, and music perception. *Psychological Science*, *1*, 272-276.

Macpherson, S. (1920). *Melody and harmony: A treatise for the teacher and the student*. London, UK: Joseph Williams.

Malmberg, C. F. (1918). The perception of consonance and dissonance. *Psychological Monographs*, *25*, 93-133.

Milne, A. J., Laney, R., & Sharp, D. B. (2015). *A spectral model of melodic affinity.* Manuscript submitted for publication.

Milne, A. J., Sethares, W. A., Laney, R., & Sharp, D. B. (2011). Modeling the similarity of pitch collections with expectation tensors. *Journal of Mathematics and Music*, *5*, 1-20.

Moore, B. C. (1973). Frequency difference limens for short-duration tones. *Journal of the Acoustical Society of America*, *54*, 610-619.

Moore, B. C. (2005). *Introduction to the psychology of hearing*. London, UK: Macmillan.

Moore, B. C., Glasberg, B. R., & Shailer, M. J. (1984). Frequency and intensity difference limens for harmonics within complex tones. *Journal of the Acoustical Society of America*, *75*, 500-561.

Parncutt, R. (1988). Revision of Terhardt's psychoacoustical model of the root(s) of a musical chord. *Music Perception*, *6*, 65-94.

Parncutt, R. (1989). *Harmony: A psychoacoustical approach*. Berlin, Germany: Springer-Verlag.

Parncutt, R. (1994). Template-matching models of musical pitch and rhythm perception. *Journal of New Music Research*, *23*, 145-167.

Parncutt, R. (2011). The tonic as triad: Key profiles as pitch salience profiles of tonic triads. *Music Perception*, *28*, 333-365.

Parncutt, R., & Prem, D. (2008, August). *The relative prevalence of medieval modes and the origin of the leading tone.* Poster presented at International Conference of Music Perception and Cognition (ICMPC10), Sapporo, Japan.

Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, *58*, 193-198.

Piston, W., & Devoto, M. (1987). *Harmony* (5th ed.). New York: Norton.

Pratt, G. (1996). *The dynamics of harmony: Principles and practice.* Oxford, UK: Oxford University Press.

Rimsky-Korsakov, N. (1885). *Practical manual of harmony.* New York: Carl Fischer.

Schellenberg, E. G., & Trehub, S. E. (1999). Culture-general and culture-specific factors in the discrimination of melodies. *Journal of Experimental Child Psychology, 74*, 107-127.

Schoenberg, A. (1969). *Structural functions of harmony* (2nd ed.). London, UK: Faber and Faber.

Smith, A. B. (1997). A "cumulative" method of quantifying tonal consonance in musical key contexts. *Music Perception*, *15*, 175-188.

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, *10*, 89-96.

Temperley, D. (1999). What's key for key? The Krumhansl-Schmukler key-finding algorithm reconsidered. *Music Perception*, *17*, 65-100.

Toiviainen, P., & Krumhansl, C. L. (2003). Measuring and modeling real-time responses to music: The dynamics of tonality induction. *Perception*, *32*, 741-766.

Trehub, S. E., Schellenberg, E. G., & Kamenetsky, S. B. (1999). Infants' and adults' perception of scale structure. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 965-975.

Wilding-White, R. (1961). Tonality and scale theory. *Journal of Music Theory*, *5*, 275-286.

Wilson, E. (1975). *Letter to Chalmers pertaining to moments-of-symmetry/Tanabe cycle* [PDF document]. Retrieved from http://www.anaphoria.com/mos.pdf

Woolhouse, M., & Cross, I. (2010). Using interval cycles to model Krumhansl's tonal hierarchies. *Music Theory Spectrum*, *32*, 60-78.

Appendix A

INTERCORRELATIONS OF MODELS AND DATA

TABLE A1. *Intercorrelations of the Probe Tone Data and the Models.*

	PD	BT	K90*b*	K90*a*	S97	L88	B89	P89	P11*a*	P11*b*	P94	M14*a*	M14*b*	M14*c*
Probe data	1.00	.86	.87	.65	.89	.96	.88	.96	.94	.92	.96	.94	.95	.97
Basic triad	.86	1.00	.80	.57	.76	.88	.82	.84	.84	.93	.86	.91	.91	.89
Krumhansl 90*b*	.87	.80	1.00	.57	.90	.91	.96	.89	.77	.92	.81	.94	.90	.85
Krumhansl 90*a*	.65	.57	.57	1.00	.59	.66	.50	.65	.70	.67	.70	.65	.69	.65
Smith 97	.89	.76	.90	.59	1.00	.90	.87	.87	.83	.93	.85	.93	.89	.89
Lerdahl 88	.96	.88	.91	.66	.90	1.00	.89	.96	.89	.95	.91	.98	.99	.97
Butler 89	.88	.82	.96	.50	.87	.89	1.00	.91	.80	.90	.83	.92	.88	.85
Parncutt 89	.96	.84	.89	.65	.87	.96	.91	1.00	.93	.92	.96	.95	.96	.96
Parncutt 11*a*	.94	.84	.77	.70	.83	.89	.80	.93	1.00	.88	.99	.88	.91	.93
Parncutt 11*b*	.92	.93	.92	.67	.93	.95	.90	.92	.88	1.00	.91	.99	.97	.95
Parncutt 94	.96	.86	.81	.70	.85	.91	.83	.96	.99	.91	1.00	.90	.93	.94
Milne 14*a*	.94	.91	.94	.65	.93	.98	.92	.95	.88	.99	.90	1.00	.98	.96
Milne 14*b*	.95	.91	.90	.69	.89	.99	.88	.96	.91	.97	.93	.98	1.00	.98
Milne 14*c*	.97	.89	.85	.65	.89	.97	.85	.96	.93	.95	.94	.96	.98	1.00

Appendix B

CROSS-VALIDATION CORRELATION

We performed 20 runs of 12-fold cross-validation of the models. Each of the 20 runs utilizes a different 12-fold partition of the probe tone data, each fold containing 2 samples. Within each run, one fold is removed and denoted the *validation set*; the remaining 11 folds are aggregated and denoted the *training set*. All parameters of the model are optimized to minimize the sum of squared errors between the model's predictions and the 22 samples in the training set. For the linear models discussed in this paper, there are only two parameters—intercept and slope. Our spectral models have additional nonlinear parameters. Cross-validation statistics, which measure the fit of the predictions to the validation set, are then calculated. This whole process is done for all 12 folds and this constitutes a single run of the 12-fold cross-validation. The same process is used for all 20 runs of the 12-fold cross-validation—each run using a different 12-fold partition of the data. The cross-validation statistics are averaged over all 12 folds in all twenty runs.

More formally: Let the data set of I samples be partitioned into K folds (the probe tone data comprise 24 values, so $I = 24$, and we use 12-fold cross-validation, so $K = 12$). Let $k[i]$ be the fold of the data containing the ith sample. The cross-validation is repeated, each time with a different K-fold partition, a total of J times. The cross-validation correlation of the jth run of the cross-validation is given by

$$r_{\mathrm{CV}}[j] = 1 - \sqrt{\sum_{i=1}^{I} (y_i - \hat{y}_i^{\setminus k[i]})^2 \Big/ \sum_{i=1}^{I} (y_i - \bar{y})^2}, \quad (1)$$

where $\hat{y}_i^{\setminus k[i]}$ denotes the fitted value for the ith sample returned by the model estimated with the $k[i]$th fold of the data removed, and $\bar{y}$ is the mean of all the sample values y_i. The final cross-validation correlation statistic is the mean over the J runs of the cross-validation (in our analysis, $J = 20$):

$$r_{\mathrm{CV}} = \frac{1}{J} \sum_{j=1}^{J} r_{\mathrm{CV}}[j]. \quad (2)$$

Appendix C

FORMAL SPECIFICATION OF THE SPECTRAL PITCH CLASS SIMILARITY MODEL OF THE PROBE TONE DATA

In this section, we give a formal mathematical specification of our model. The techniques used are based on those introduced by Milne et al. (2011). The MATLAB routines that embody these routines can be downloaded from http://www.dynamictonality.com/probe_tone_files/.

Let a chord comprising M tones, each of which contains N partials, be represented by the matrix $\mathbf{X}_f \in \mathbb{R}^{M\times N}$. Each row of $\mathbf{X}_f$ represents a tone in the chord, and each element of the row is the frequency of a partial of that tone. In our model, we use the first twelve partials (so $N = 12$); this means that, if $\mathbf{X}_f$ is a three-tone chord, it will be a 3×12 matrix.

The first step is to convert the partials' frequencies into pitch class cents values:

$$x_{\text{pc}}[m,n] = 1200\lfloor \log_2(x_{\text{f}}[m,n]/x_{\text{ref}})\rceil \bmod 1200, \quad (3)$$

where $\lfloor\cdot\rceil$ is the nearest integer function, and x_{ref} is an arbitrary reference frequency (e.g., the frequency of middle C). These values are then collected into a single *pitch class vector* denoted $\tilde{\mathbf{x}}_{\text{pc}} \in \mathbb{Z}^{12M}$ indexed by i such that $x_{\text{pc}}[m,n] \mapsto \tilde{x}_{\text{pc}}[i]$, where $i = (m-1)N + n$.

Let each of the partials have an associated weight $x_{\text{w}}[m,n]$, which represents their *salience*, or probability of being perceived. We test three models (a, b, and c). Given model ℓ, where $\ell \in \{a,b,c\}$ denotes the model, the saliences of the tonic triad's partials are parameterized by a *roll-off* value $\rho \in \mathbb{R}$, and a *chord-degree weighting* value $\omega \in [0,1]$, so that

$$\begin{aligned} &\omega^{[m\notin R_\ell]} x_{\text{w}}[m,n] = n^{-\rho} \\ &m = 1,\ldots,M, \text{ and } n = 1,\ldots,12, \end{aligned} \quad (4)$$

where $[m \notin R_\ell]$ denotes an indicator function that equals 0 when tone m is member of the set R_ℓ of tones classed as chord roots in model ℓ, and is otherwise 1. In Model a, all tones are classed as roots, hence all tones have a chord-degree weighting of 1; in Model b, only the conventional roots of the major and minor triads are classed as roots (i.e., pitch class C in the chord Cmaj or Cmin), all other tones have a chord degree weighting of ω; in Model c, the third of the minor triad is also classed as a root (e.g., E♭ in Cmin), the remaining tones have a chord degree weighting of ω. Ignoring the chord degree weighting value, Equation (4) means that when $\rho = 0$, all partials of a tone m have a weight of 1; as ρ increases, the weights of its higher partials are reduced. These values are collected into a single *weighting vector* $\tilde{\mathbf{x}}_{\text{w}} \in \mathbb{R}^{12M}$ also indexed by i such that $x_{\text{w}}[m,n] \mapsto \tilde{x}_{\text{w}}[i]$, where $i = (m-1)N + n$ (the precise method used to reshape the matrix into vector form is unimportant so long as it matches that used for the pitch class vector).

The partials (their pitch classes and weights in $\tilde{\mathbf{x}}_{\text{pc}}$ and $\tilde{\mathbf{x}}_{\text{w}}$) are embedded in a *spectral pitch class salience matrix* $\mathbf{X}_{\text{pcs}} \in \mathbb{R}^{12N\times 1200}$ indexed by i and j:

$$\begin{aligned} &x_{\text{pcs}}[i,j] = \tilde{x}_{\text{w}}[i]\ \delta\big[j - \tilde{x}_{\text{pc}}[i]\big] \\ &i = 1,\ldots,12N, \text{ and } j = 0,\ldots,1199, \end{aligned} \quad (5)$$

where $\delta[z]$ is the Kronecker delta function, which equals 1 when $z = 0$, and equals 0 when $z \neq 0$. This equation means that the matrix $\mathbf{X}_{\text{pcs}}$ is all zeros except for $12N$ elements, and each element indicates the salience $x_{\text{pcs}}[i,j]$ of partial i at pitch j.

To model the uncertainty of pitch perception, these $12N$ delta "spikes" are "smeared" by circular convolution with a discrete Gaussian kernel $\mathbf{g}$, which is also indexed by j, and is parameterized with a *smoothing* standard deviation $\sigma \in [0,\infty)$ to give a *spectral pitch class response matrix* $\mathbf{X}_{\text{pcr}} \in \mathbb{R}^{12N\times 1200}$, which is indexed by i and k:

$$\mathbf{x}_{\text{pcr}}[i] = \mathbf{x}_{\text{pcs}}[i] * \mathbf{g}, \quad (6)$$

where $\mathbf{x}_{\text{pcr}}[i]$ is the ith row of $\mathbf{X}_{\text{pcr}}$, and $*$ denotes circular convolution over the period of 1200 cents; that is,

$$\begin{aligned} &x_{\text{pcr}}[i,k] = \sum_{j=0}^{1199} x_{\text{pcs}}[i,j]g[(k-j) \bmod 1200], \\ &i = 1,\ldots,12N, \text{ and } k = 0,\ldots,1199. \end{aligned} \quad (7)$$

In our implementation, we make use of the circular convolution theorem, which allows (6) to be calculated efficiently with fast Fourier transforms; that is, $\mathbf{f} * \mathbf{g} = \mathcal{F}^{-1}(\mathcal{F}(\mathbf{f}) \circ \mathcal{F}(\mathbf{g}))$, where $*$ is circular convolution, $\mathcal{F}$ denotes the Fourier transform, $\circ$ is the Hadamard (elementwise) product, and $\mathbf{f}$ stands for $\mathbf{x}_{\text{pcs}}[i]$.

Equation (6) can be interpreted as adding random noise (with a Gaussian distribution) to the original pitch classes in $\mathbf{X}_{\text{pcs}}$, thereby simulating perceptual pitch uncertainty. The standard deviation of the Gaussian distribution σ models the pitch difference limen (just noticeable difference) (Milne et al., 2011, Online Supplementary, App. A). In laboratory experiments with sine waves, the pitch difference limen is approximately 3 cents in the central range of frequency (Moore, 1973; Moore, Glasberg, & Shailer, 1984). We would expect the pitch difference limen in the more distracting setting of listening to music to be somewhat wider. Indeed, the

value of σ was optimized—with respect to the probe tone data—at approximately 6 cents.

Each element $x_{\text{pcr}}[i, k]$ of this matrix models the probability of the ith partial in $\mathbf{x}_{\text{pc}}$ being perceived at pitch class k. In order to summarize the responses to all the pitches, we take the column sum, which gives a vector of the expected numbers of partials perceived at pitch class k. This 1,200-element row vector is denoted a *spectral pitch class vector* $\mathbf{x}$:

$$\mathbf{x} = \mathbf{1}'\mathbf{X}_{\text{pcr}}, \tag{8}$$

where $\mathbf{1}'$ denotes a row vector of $12N$ ones. The spectral pitch class similarity of two such vectors $\mathbf{x}$ and $\mathbf{y}$ is given by any standard similarity metric. We choose the cosine:

$$s(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{x}\mathbf{y}'}{\sqrt{\mathbf{x}\mathbf{x}'\mathbf{y}\mathbf{y}'}}, \tag{9}$$

where $'$ denotes the matrix transpose operator that turns a row vector into a column vector (and vice versa). Because $\mathbf{x}$ and $\mathbf{y}$ contain only nonnegative values, their cosine similarity falls between 0 and 1, where 1 implies the two vectors are parallel, and 0 implies they are orthogonal.

We use this model to establish the similarities of a variety of probes with respect to a context. Let the context be represented by the spectral pitch class vector $\mathbf{x}$, and let the P different probes $\mathbf{y}_p$ be collected into a matrix of spectral pitch class vectors denoted $Y \in \mathbb{R}^{P\times 1200}$. The column vector of P similarities between each of the probes and the context is then denoted $\mathbf{s}(\mathbf{x}, \mathbf{Y}) \in \mathbb{R}^P$. For example, the context may be a major triad built from HCTs and the probes may be single HCTs at the twelve chromatic pitches. In this case, the thirty-six harmonics from the context (12 partials for each of the three different chord tones) are embedded into a single spectral pitch class vector $\mathbf{x}$, as described in (3–8). Each of the twelve differently pitched probe tones' 12 harmonics are embedded into twelve spectral pitch class vectors $\mathbf{y}_p$. The similarities of the context and the twelve probes are calculated—as described in (9)—to give the vector of their similarities $\mathbf{s}(\mathbf{x}, \mathbf{Y})$.

Models *a*, *b*, and *c* can now be summarized in mathematical form: Let the vector of probe tone data for both contexts be denoted $\mathbf{d} \in \mathbb{R}^{24}$; let the vector of associated modeled similarities be denoted $\mathbf{s}(\mathbf{x}, \mathbf{Y}; \rho, \sigma, \omega, \ell) \in \mathbb{R}^{24}$, where ρ, σ, ω are the roll-off, smoothing, and chord degree weighting parameters discussed above, and $\ell \in \{a, b, c\}$ denotes the model; let $\mathbf{1}$ be a column vector of 24 ones;

$$\mathbf{d} = \alpha 1 + \beta \mathbf{s}(\mathbf{x}, \mathbf{Y}; \rho, \sigma, \omega, \ell) + \varepsilon, \tag{10}$$

where α and β are the linear intercept and slope parameters, and ε is a vector of 24 unobserved errors that captures unmodeled effects or random noise.

Each model's parameter values were optimized, iteratively, to minimize the sum of squared residuals between the model's predictions and the empirical data; that is, the optimized parameter values for model ℓ are given by

$$\left(\hat{\alpha}, \hat{\beta}, \hat{\rho}, \hat{\sigma}, \hat{\omega}\right)[\ell] = \underset{\alpha,\beta,\rho,\sigma,\omega}{\text{argmin}} \Big((\mathbf{d} - \alpha 1 - \beta \mathbf{s}(\rho, \sigma, \omega, \ell))' (\mathbf{d} - \alpha 1 - \beta \mathbf{s}(\rho, \sigma, \omega, \ell)) \Big), \tag{11}$$

where $\text{argmin}\, f(\boldsymbol{\theta})$ returns the value of $\boldsymbol{\theta}$ that minimizes the value of $f(\boldsymbol{\theta})$.