Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music (ESCOM 2009) Jyväskylä, Finland
Jukka Louhivuori, Tuomas Eerola, Suvi Saarikallio, Tommi Himberg, Päivi-Sisko Eerola (Editors)

# A Psychoacoustic Model of Harmonic Cadences: A Preliminary Report

Andrew J. Milne

*Department of Music, Jyväskylä University, Finland*

andymilne@tonalcentre.org


## ABSTRACT

This report presents a psychoacoustically derived computational model of the perceived distance between any two major or minor triads, the degree of activity created by any given pair of triads, and the cadential effectiveness of three-triad progressions. It also provides statistical analyses of the ratings given by thirty-five participants for the "similarity" and "fit" of triads in a pair, and the "cadential effectiveness" of three-triad progressions. Multiple regressions show that the model provides highly significant predictions of the experimentally obtained ratings. Finally, it is argued that because the model is based upon psychoacoustic axioms, it is likely the regression equations represent true causal models. As such, the computational model and its associated theory question the plausibility of theoretical approaches to tonality that use only long-term memory and statistical features, as well as those approaches based upon symmetrical geometrical structures like the torus. It is hoped that the psychoacoustic approach proposed here may herald not only the return of psychoacoustic approaches to tonal music theory, but also the exploration of the tonal possibilities offered by non-standard tunings and non-harmonic timbres.


## I. INTRODUCTION

Psychoacoustic approaches have provided relatively effective explanations for why certain simultaneities of notes (chords) are typically considered "dissonant" while others are considered "consonant" (notably the major and minor triads that are so important in both the theory and practice of Western tonal music) (Plomp & Levelt, 1965; Kameoka & Kuriyagawa, 1969; Sethares, 2004). However, to date, there has been no psychoacoustic explanation for one of the most important and mysterious aspects of Western tonal music—the fact that a succession of consonant chords can induce feelings of "expectation" and "resolution" that are not produced when the same chords are played in isolation, or in a different order.

For example, listeners will typically feel that in the chord progression F major→G major→C major, the second chord sounds particularly expectant whereas the third chord resolves this expectation, thus providing a sense of closure. Chord progressions such as these are called cadences, and they are commonly used in tonal music to mark the ends of phrases, or entire sections. Interestingly, cadences are commonly constructed with only consonant triads (the example above is the familiar IV→V→I cadence; other common cadences using only major and minor triads are ii→V→I, iv→V→I, and ♭VI→V→I). Such cadences imply that the expectation or resolution induced by a chord is not necessarily a function of its inherent (vertical) consonance or dissonance, but rather of its temporal (horizontal) context—most particularly the chords that directly precede and proceed it.

Any theory of harmonic tonality—that form of music using chords (principally triads) to establish a tonic ("home") note or triad (Krumhansl, 1990)—must provide an explanation for these feelings of expectation and resolution that lie at its very heart. In the absence of successful psychoacoustic theories to account for this phenomenon, many contemporary researchers have suggested a statistical (long-term memory) explanation (Bharucha, 1987; Krumhansl, 1990; Tillmann, Bharucha, & Bigand, 2000; Levitin, 2006). These approaches suggest that we are culturally trained, by exposure, to expect certain progressions, and this accounts for the effect produced by the regularities (such as cadences) that are found in tonal music—that is, if we've heard it before, we expect to hear it again. There is little doubt that this is a credible approach, but it has a number of problems if used as the sole explanation for these effects. For example, (a) it implies that the effect induced by a given chord progression—such as a cadence—should be very plastic, but there is little evidence, from either a cultural or historical perspective, that this is the case; (b) short-term memory has been demonstrated to play a significant role in perception of tonality (Leman, 2000); (c) typical cadential progressions have been readily adopted, with no modification, by non-Western cultures (e.g., see Agawu (2003)).

Statistical approaches undoubtedly play an important part in the cognition of harmonic cadences, but I propose there are important psychoacoustic processes that underlie them. In this report I present a psychoacoustically derived model designed to explain the flow of expectation and resolution induced by a succession of chords. The model is built in MATLAB, and is currently relatively simple (it calculates only root position major and minor triads), and can be substantially developed. I also present preliminary analyses of recently conducted experiments (collecting human ratings of "similarity", "fit", and "cadential effectiveness" of a variety of chord progressions) designed to test the model. The preliminary analyses of the experimental data strongly support the model. I finish with a discussion of some of the implications of this proposed psychoacoustic approach to tonal music theory.

Supplemental information relevant to this report (including additional figures and mathematical proofs) will be accessible from www.tonalcentre.org/deepertheory/psycad.html. But before proceeding to the next section, a quick explanation of the notation used in this report: for the sake of brevity I will refer to major triads in upper case, minor triads in lower case—so "A" is an A major triad, "g" is a g minor triad. Furthermore, without qualification, all triads are considered to be major or minor and in root position.

## II. THE THEORETICAL FOUNDATIONS OF THE MODEL

The underlying theory assumes the presence of five latent variables, which may be thought of as psychoacoustic or cognitive components of the listener's auditory system. The model contains a simulation of each of these latent variables and their interactions.

I presume that the first three variables, *pitch distance* (*pd*), *fundamental response distance* (*frd*), and *spectral response distance* (*srd*) are a function of psychoacoustic data (tone frequency, timbre, and frequency difference limens). Each of the variables is a different type of metric used to assess the distance (level of difference) between any two chords (or tones).

I hypothesise that pitch distance and fundamental response distance are the main variables responsible for the value of a latent cognitive variable called *voice-leading distance* (*vld*), and that spectral response distance is the main variable responsible for the value of a latent cognitive variable called *spectral distance* (*sd*). Furthermore, I hypothesise that *voice-leading distance* (*vld*) and *spectral distance* (*sd*) together determine the level of *tonal activity* (*act*) induced by a pair of chords.

These relationships are summarised in the path diagram of Figure 1; note that the model attempts to replicate each of these latent variables, and the relationships between them.

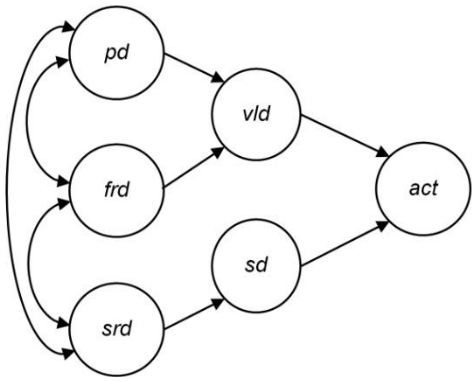


**Figure 1. A path diagram showing the proposed flow of causation from the psychoacoustic variables pitch distance (*pd*), fundamental response distance (*frd*), and spectral response distance (*srd*), to the cognitive variables of voice-leading distance (*vld*), spectral distance (*sd*), and tonal activity (*act*). Error terms are not shown.**

I explain the psychoacoustic metrics (*pd*, *frd*, and *srd*) in the next two subsections, and the cognitive variables (*vld*, *sd*, and *act*) in section III.

### A. Pitch Distance (*pd*)

The pitch distance between two tones is approximated by the logarithm of their pitch ratio. I also assume octave equivalence; so all intervals are reduced (by octave inversion) to be no greater than six equally tempered semitones. So, if it is assumed that the tones have harmonic spectra, pitch distance can be approximated accordingly:

$$pd = \min\left(\left|\{\log_2(f_1/f_2)\}\right|, 1 - \left|\{\log_2(f_1/f_2)\}\right|\right), \qquad 1$$

where $f_1$ and $f_2$ are the fundamental frequencies of the two tones, and {} denotes the fractional part.

When calculating the pitch distance between two chords, the pitch distance moved by each voice is separately calculated to enable each to be separately analysed. Pitch distance (in conjunction with the fundamental response distance discussed later) is intended to give an indication of the voice-leading distance between two chords.

### B. Fundamental and Spectral Response Distances (*frd* and *srd*)

The two response distance measures are novel metrics based upon the tenets of signal detection theory. Given a signal with a specific frequency, the auditory system is assumed to produce an *internal response* that may be characterised as consisting of both signal plus noise; furthermore, the noise component is assumed to have a Gaussian distribution. So the internal response to a sine wave with a specified frequency may be characterised as a Gaussian centred on that frequency. It is this noise component that makes the *frequency difference limen* (*frequency DL*) greater than zero—that is, when two sine waves of similar frequency are played successively, the listener may, incorrectly, hear them as having the same pitch.

In a two-alternative forced-choice (2-AFC) experiment, the frequency DL is normally defined as the value at which the true positive and false positive rates correspond to a $d'$ of approximately one. Because $d'$ is equivalent to the distance between the means of two distributions divided by their standard deviation, the standard deviation of the internal response is equal to the frequency DL. This enables the internal frequency response to a sine tone to be modelled using experimentally obtained measurements of frequency DL, such as those obtained by Moore, Glasberg, and Shailer (1984).

The response distance between any two sine tones is the distance between their (Gaussian) internal responses. Although there may be many suitable metrics to measure this distance, I have chosen cosine distance because it is relatively easy to express in functional form, and because it makes intuitive sense—being the normalised cross-correlation between the two Gaussians. (A possible alternative metric would be the area under the ROC curve produced by two such Gaussian distributions.) The cosine distance $d_{\cos}(f_1, f_2)$ between two sine tones, as function of their frequencies $f_1$ and $f_2$, is given by:

$$d_{\cos}(f_1, f_2) = 1 - \frac{\sqrt{2}}{\sqrt{\left|\frac{\mathrm{DL}(f_1)^2 + \mathrm{DL}(f_2)^2}{\mathrm{DL}(f_1)^2\,\mathrm{DL}(f_2)^2}\right| \left|\mathrm{DL}(f_1)\mathrm{DL}(f_2)\right|}} \exp\left(\frac{-(f_2 - f_1)^2}{2\left(\mathrm{DL}(f_1)^2 + \mathrm{DL}(f_2)^2\right)}\right), \qquad 2$$

where $\mathrm{DL}(f_i)$ is the difference limen (standard deviation) at the frequency $f_i$. For a full derivation of this equation, see the above-mentioned website. Equation 2 shows that the cosine distance between two Gaussians is a function of their standard deviation and their frequency difference. Given two sine tones with independent frequencies, the spectral distance gives an indication of the probability that they are distinguishable: when they are identical in frequency, their spectral distance is zero; when they are far apart in frequency, their spectral distance approaches 1.

*1) Spectral Response Distance.* Given two successive complex tones, or chords comprising a number of complex tones, the spectral response distance is the sum of the response distances between all possible pairings of partials where each pair contains a partial from the first chord (or tone) and a partial from the second chord (or tone). The partials in any given complex tone may have different amplitudes (typically the higher the partial the lower its amplitude), so the product of their respective amplitudes weights the cosine distance for any given partial pair. For two chords, one with *m* partials of frequency $f_i$ and amplitude $a_i$ (for $i = 1$ to $m$), the other with $n$ partials of frequency $f_j$ and amplitude $a_j$ for $j = 1$ to $n$, the total spectral response distance can be expressed accordingly:

$$srd = \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n} a_i a_j d_{\cos}(f_i, f_j). \quad 3$$

Spectral distance is, therefore, calculated in a manner similar to dissonance algorithms such as Sethares' (2004). Spectral response distance is intended to give a measure of the perceived spectral distance between two chords.

According to Moore, Glasberg, and Shailer (1984), the frequency DLs for harmonics within a complex tone vary according to their harmonic number (harmonics lower than five typically have a frequency DL of approximately 0.5%, harmonics higher than seven typically have a frequency DL of approximately 3%). At the time of writing, the psychoacoustic model does not allow for different widths to be chosen for different harmonics, so a compromise value of ERB/13, which corresponds to a frequency DL of approximately 1%, was chosen. ERB denotes the equivalent rectangular bandwidth, and has the value ERB($f$) = 0.108$f$ + 24.7 (Glasberg & Moore, 1990).

*2) Fundamental Response Distance.* The response distance can also be applied to just the fundamentals of each tone. This fundamental response distance (in conjunction with pitch distance) is intended to give an indication of the voice-leading distance between two chords.

According to Moore, Glasberg, and Shailer (1984), the frequency DL for a complex harmonic tone, as a whole, is smaller than that for any of its partials, and generally approximates 0.2%. This is approximated by ERB/66 which is the value used by the model to calculate fundamental response distance.

## III. APPLYING THE MODEL TO MUSICAL SYSTEMS

Let it be assumed that the musical system under analysis uses a number of independent tones each of which are composed of dependent (i.e. approximately harmonically related) spectra. This is a fair description of the majority of western music, where voices move with some degree of independence, and the majority of these voices are harmonic complexes with a clear sense of pitch. This type of musical system creates strict constraints upon movements within the continuum of all possible spectral tunings. For example, imagine we are able to create any possible spectrum, containing 16 independently tuned partials. Any point in this 16-dimensional spectral continuum is a specific spectral tuning, and it would be possible to move from any arbitrary spectral tuning within this space to any other. But in conventional music, with the above-mentioned constraints, we can control—and are accustomed to hearing—the movement of a limited number of tones (complexes of harmonic partials). This means that the range of musically possible spectral tunings, and possible paths between them, is substantially constrained.

*Voice-leading distance* is the cognitive distance between two spectral tunings under these constraints. The distances along these constrained paths are mediated by the pitches (fundamentals) of each tone; so it makes sense to hypothesise that voice-leading distance is a function of the pitch distance and fundamental response distances (because, these two distances are concerned only with the frequencies of the fundamentals).

*Spectral distance*, on the other hand, is the unconstrained distance between all available partials. Because it is a function of the tuning of all partials in each chord, it makes sense to hypothesise it be a function of the spectral response distance.

A corollary of having two independent distances is that it is possible for a pair of triads to be voice-leading close but spectrally distant; or for a pair of triads to be voice-leading distant but spectrally close. This has a very important consequence: given two pairs of chords that are voice-leading close (e.g., the two triad pairs D♭↔G and D♭↔g are voice-leading close because G and g differ by just one semitone in one voice), such that one pair is spectrally more distant than the other (in reference to the triad D♭, the triad G is spectrally more distant than the triad g), the spectrally distant pair will tend to be heard as if it were a voice-leading alteration of the spectrally closer pair. I hypothesise that this sense of alteration (i.e., of a more "complex", or "difficult", choice made as a substitute for a "simpler", more "straightforward", choice) is the origin of tonal activity, or expectation. I also hypothesise that this activity is resolved by allowing the altered tone to continue in the direction of its alteration to a triad that is spectrally close (has a "simple", "straightforward") relationship to, preferably both of, the preceding two chords.

The next three subsections discuss voice-leading distance, spectral distance, and tonal activity in more detail.

### A. Voice-leading distance

When assessing the perceived distance between two chords, it is common to measure the overall pitch distance between them. This is typically calculated as the city block, Euclidean or other Minkowski, distance between the semitone values of the notes in two chords. It seems reasonable to assume this is a good measure for pairs of tones, or other simple stimuli. But when it comes to measuring the distance between triads, or between any voice-leading involving three or more parts, is it reasonable to expect a listener to individually track the degree of movement of every voice before summing them?

For standard musical tuning systems, the fundamental response distance is effectively binary—it has a value of 0 for a common tone, a value of 1 for anything else (see the next paragraph for a fuller explanation). It seems plausible that, due to the simplicity of this binary measure, the fundamental response distance may also play a part in determining the voice-leading distance for more complex stimuli (such as three, or more, part voice-leadings).

The Gaussian noise component of the internal frequency response is relatively narrow compared to the smallest musical interval used in common practice (the semitone). This means that the fundamental response distance effectively acts as a counter for the number of non-unisons between two chords. That is, it gives a distance of zero to two identical triads, a distance of approximately 1/3 to two triads sharing two tones (e.g., parallel triads like C and c, relative triads like C and a, leading tone exchange triads like C and e), a distance of approximately 2/3 to two triads sharing one common tone (e.g., dominant triads like C and G), and a distance of approximately 1 to two triads with no common tone (like C and D). This is clearly in accord with Riemannian and neo-Riemannian music theory, which treats the above-mentioned common-tone transformations as being especially close (e.g., Kopp (2002)).

We might, therefore, expect a listener to judge the voice-leading distance between two chords to be a function of pitch distance and fundamental response distance. Furthermore, we might expect that the pitch distance of the most salient note (or notes), such as the bottom note, top note, or root, may be more important than the pitch distances between less salient notes.

### B. Spectral Distance

In the same way it seems unreasonable to expect a listener to track the movement of every single tone in a three, or more, part voice-leading, it is even more unreasonable to expect a listener to track the distance moved by every single partial found in one chord to the partials found in a second chord. Furthermore, in normal listening even those partials that can be resolved are not actually "heard out"; instead they are subsumed into the unified perceptions of virtual pitch and timbre. Furthermore, even if they were actively heard out, it would be almost impossible to know in which direction any given partial "moves"—does it "go" to the partial that is closest in pitch, or the partial that has the same position in a frequency-ranked stack of partials? The spectral response distance does not attempt to "track" any supposed motion of partials, it simply scores every coincident pair as 0, every non-coincident pair as 1. Every pair that is almost coincident is given a score between 0 and 1—its precise value determined by the width of the underlying Gaussian internal response curve.

For this reason, we might expect the spectral distance between any two triads to be strongly correlated to their spectral response distance.

In tonal terms, spectral distance acts as a type of weighted counter, in that a smaller distance is given to those intervals (dyads) whose tunings approximate simple numerical frequency ratios (such as the perfect fifth, which approximates 3/2). This is because simple ratio intervals have more coincident partials. (For two tones, with harmonic partials, that have a frequency ratio of $p/q$, the ratio of coinciding partials to all partials is $2/(p+q)$). This means that the greater the number of approximately simple ratios (and the greater their simplicity) between two triads, the lower the spectral distance between them. This results in smaller spectral response distances for melodic dyads that are conventionally considered harmonically consonant (perfect fifths and fourths, and thirds and sixths), than for intervals like seconds and sevenths and the tritone.

### C. Tonal Activity, Resolution, and Cadential Structure

*1) Tonal Activity.* I hypothesise that tonal activity is the result of the interplay between voice-leading distance (which is a function of pitch distance and fundamental response distance) and spectral distance (which is a function of spectral response distance).

Given a musically presented triad pair, let any other pair against which that musically presented pair is mentally compared be denoted a *comparison* pair. I hypothesise that when a triad pair has a higher spectral distance than a comparison triad pair that is voice-leading close, the former chord pair is heard as an alteration of the comparison triad pair. This can be stated more formally: Let there be a pair of triads, $t_1 \leftrightarrow t_2$, with a spectral distance of $sd_{t1\leftrightarrow t2}$. If a triad $t_3$ is used in place of triad $t_2$, we get a pair $t_1 \leftrightarrow t_3$ with a spectral distance $sd_{t1\leftrightarrow t3}$. Let the voice-leading distance between $t_2$ and its comparison triad $t_3$ be denoted $vld_{t2\leftrightarrow t3}$. The activity, $a_{t1\leftrightarrow t2\,|\,t1\leftrightarrow t3}$, of $t_1 \leftrightarrow t_2$ due to the voice-leading proximity of the comparison pair $t_1 \leftrightarrow t_3$ is given by:

$$a_{t1\leftrightarrow t2\,|\,t1\leftrightarrow t3} = (sd_{t1\leftrightarrow t3} - sd_{t1\leftrightarrow t2}) / vld_{t2\leftrightarrow t3}. \quad (4)$$

Similarly, the pair $t_1 \leftrightarrow t_2$ might be compared to the pair $t_4 \leftrightarrow t_2$, giving $t_1 \leftrightarrow t_2$ an activity, due to the voice-leading proximity of $t_4 \leftrightarrow t_2$, of $a_{t1\leftrightarrow t2\,|\,t4\leftrightarrow t2} = (sd_{t4\leftrightarrow t2} - sd_{t1\leftrightarrow t2}) / vld_{t1\leftrightarrow t4}$. Or $t_1 \leftrightarrow t_2$ might be compared with $t_4 \leftrightarrow t_3$, giving $t_1 \leftrightarrow t_2$ an activity, due to the voice-leading proximity of $t_4 \leftrightarrow t_3$, of $a_{t1\leftrightarrow t2\,|\,t4\leftrightarrow t3} = (sd_{t4\leftrightarrow t3} - sd_{t1\leftrightarrow t2}) / (vld_{t1\leftrightarrow t4} + vld_{t2\leftrightarrow t3})$.

A result of this definition is that $a_{t1\leftrightarrow t2\,|\,t3\leftrightarrow t4} = -a_{t3\leftrightarrow t4\,|\,t1\leftrightarrow t2}$. Assuming the absolute level of activity is more than negligible, I hypothesise that the triad pair with positive activity is heard as an alteration of the comparison pair with negative activity. This is because the two triads in a pair with high positive activity have a more distant (complex) spectral relationship than the voice-leading close pair with negative activity. I presume that triad pairs with negative activity sound "passive", "stable", and "at rest", while triad pairs with positive activity sound "active", "unstable", and "restless".

Let me illustrate this concept with a relatively straightforward example. The psychoacoustic model predicts that the two root-position triad pairs C↔d and C↔D are voice-leading close (they have two common tones, and the bass note does not move). It also predicts that C↔d is spectrally closer than C↔D (the latter replaces the former's low distance perfect fourth—between the root of the first chord and the third of the second chord—with a high distance tritone). So the former pair has negative activity, the latter has positive activity, which means that the latter is heard as an alteration of the former.

(If that example still seems difficult to understand, consider just two successive tones. We might consider melodic intervals of a semitone and tritone to be active because they can be mentally compared to the voice-leading close melodic intervals of the unison and perfect fourth/fifth, respectively. The process described above is simply an extension of this concept to a higher-dimensional tone space—illustrations of which are given in Section IV, and Figure 2 and Figure 3.)

When considering only those comparisons that are voice-leading close—and so maximise activity (see Equation 4)—the psychoacoustic model of tonal activity has a plausible asymmetry. For the following explanation and examples, I will consider just the parallel comparison—i.e., let $t_2 \leftrightarrow t_3$ be parallel transformations (e.g., $t_2$ = d, and $t_3$ = D), and $t_1 \leftrightarrow t_4$ be parallel transformations (e.g., $t_1$ = C, and $t_4$ = c). The reason for favouring the parallel comparison is because it is reasonable to surmise that, for root position triads, the parallel transformation will be judged to have the smallest voice-leading size. This is because it has two common tones; the moving tone uses the smallest possible pitch distance (one semitone); the moving tone is not the salient root (bass note) of the two chords. The parallel transform is the only one that has all three of these characteristics. Equation 4 shows that the absolute value of activity is maximised by having a comparison pair that is voice-leading close (i.e., $vld_{t2\leftrightarrow t3}$ is small), so by choosing the parallel comparison, we are likely to be exploring those tonal activities that are most important to our perception of music (see Section V.B. for a further discussion of this issue).

For more compact notation, let the activity of the pair $t_1\leftrightarrow t_2$ due to comparison with its parallel transform pair $t_1\leftrightarrow t_3$, be denoted $a_{t1\rightarrow t2\,|\,\mathbf{P}}$ (the arrow points to the chord that is transformed, and the bold letter indicates the type of transform: **P** is the parallel transform, though **R** and **L**, etc. could be used to denote the relative and leading tone exchanges respectively). Similarly, the activity of the pair $t_1\leftrightarrow t_2$ due to comparison with its parallel transform pair $t_4\leftrightarrow t_2$ is denoted $a_{t1\leftarrow t2\,|\,\mathbf{P}}$ (in this case, the arrow points to the first chord, because it is this chord that is being compared to its parallel transform).

Generally, $a_{t1\rightarrow t2\,|\,\mathbf{P}} \neq a_{t1\leftarrow t2\,|\,\mathbf{P}}$. For example, the model calculates that $a_{\mathrm{C}\rightarrow \mathrm{D}\,|\,\mathbf{P}} > a_{\mathrm{C}\leftarrow \mathrm{D}\,|\,\mathbf{P}}$; indeed $a_{\mathrm{C}\rightarrow \mathrm{D}\,|\,\mathbf{P}} > 0$, while $a_{\mathrm{C}\leftarrow \mathrm{D}\,|\,\mathbf{P}} < 0$. In words, given the pairing of chords C and D, the D is heard as altered, rather than the C. A natural consequence of this is are that a cadential progression proceeding from C to D to some resolution triad is likely to be more cadentially effective than a progression from D to C to some resolution triad. Hence the tonal asymmetries that are a vital aspect of our cognition of tonality (e.g., see Dahlhaus' discussion of the order of "functions" within cadences (1990), or Toiviainen and Krumhansl (2003)) are a natural consequence of the proposed activity function. Such asymmetries cannot be explained by inherently symmetrical structural models, such as Lerdahl's (2001), without the addition of a separate layer of theory.

*2) Resolution and Cadential Structure.* When a triad is heard as an active alteration of another triad, we expect it to resolve. This means that it needs to move to a triad whose pairings with preceding triads have negative activity. If this not the case, the final chord is less likely to feel as if it is a successful resolution—indeed it may feel that it requires a further resolution.

This structure gives a template for harmonic cadences formed with three triads denoted *antepenult* ($A$), *penult* ($P$), and *final* ($F$) (i.e., the putative cadential progression is $A\rightarrow P\rightarrow F$, and so there are six different activity values to be considered: $a_{A\rightarrow P\,|\,\mathbf{P}}$, $a_{A\leftarrow P\,|\,\mathbf{P}}$, $a_{P\rightarrow F\,|\,\mathbf{P}}$, $a_{P\leftarrow F\,|\,\mathbf{P}}$, $a_{A\rightarrow F\,|\,\mathbf{P}}$, and $a_{A\leftarrow F\,|\,\mathbf{P}}$).

I hypothesise that an effective cadence requires the pairing between $A$ and $P$ to have positive activity, and for the pairings between $P$ and $F$, and between $A$ and $F$, to have negative activity. An ideal template for a three-triad cadence is, therefore, $a_{A\rightarrow P\,|\,\mathbf{P}} > 0$, $a_{P\rightarrow F\,|\,\mathbf{P}} < 0$, $a_{A\rightarrow F\,|\,\mathbf{P}} < 0$ (the classic IV→V→I and ii→V→I cadences (e.g., C→D→G and d→G→C, respectively) have precisely this pattern of activities. Generally speaking, we might expect to see a positive correlation between $a_{A\rightarrow P\,|\,\mathbf{P}}$ and $a_{A\leftarrow P\,|\,\mathbf{P}}$ and cadential effectiveness, and a negative correlation between $a_{P\rightarrow F\,|\,\mathbf{P}}$, $a_{P\leftarrow F\,|\,\mathbf{P}}$, $a_{A\rightarrow F\,|\,\mathbf{P}}$, and $a_{A\leftarrow F\,|\,\mathbf{P}}$, and cadential effectiveness.

Furthermore, we might expect the most effective resolution to be one where the active note in the penult resolves to the root of the final triad. This aspect of cadential structure is likely very important, but is beyond the scope of this preliminary report.

# IV. PLOTTING THE MODEL

The model is easier to understand and interpret when plotted. To simplify things, the plots show response distance in relation to pitch distance. This means that a single dimension (axis) is required for each voice, and one additional dimension for the response distance. When considering triad pairs (with three-part voice-leading), this means that a four-dimensional plot is required. This is clearly impracticable, but if we consider only root-position triads, the root and the fifth of the chords no longer need to move independently and so can be concatenated into a single root + fifth dimension. This results in the form illustrated by Figure 2 and Figure 3. The *x*- and *y*-axes of these figures can be considered to represent a single 2-D plane extracted from the full 3-D tone space created by three fully independent voice. The *x*-*y* distance corresponds to the Euclidean pitch distance between any two triads; the *z*-axis represents the response distances from the central reference triad, and is indicated with a lit surface.

There are two plots: the spectral response distances of all possible triad tunings (with a root and fifth) from a 12-TET major triad (Figure 2), and from a 12-TET minor triad (Figure 3). The major and minor triads run up the two diagonal lines, with major triads located vertically above their minor parallels (I have labelled a few examples to help locate the reader). Let me give two examples. In Figure 3, note how, in relation to c, the triad D is more distant (it is "higher" on the z-axis) than the triad d. This suggests that in the progression c→D, the latter chord will be heard as an alteration of d, and hence is active and seeks resolution (e.g., to g). In Figure 2, note that, in relation to C, the triad E is more distant ("higher" on the z-axis) than e, hence C→E is likely to be heard as an alteration of C→e, and seek resolution to a.

It is also interesting to note that the major and minor reference triad charts are 180° rotations of one another. This is a graphic visualisation of the duality of harmonic functions noted by authors such as Riemann and Harrison (1994).

# V. DESIGN OF THE EXPERIMENTS

The cognitive variables (voice-leading distance, spectral distance, and tonal activity) cannot be directly measured. A participant is likely to somewhat confound the two variables, as well as be influenced by unplanned factors. Despite that, I hoped that with careful experimental design, it would be possible to get a fairly good indication of the value of these latent variables, such that regressing the ratings with respect to pitch distance, fundamental response distance, spectral response distance, and tonal activity, would produce a useful test of the psychoacoustic model and its underlying theory.

Two experiments were conducted with a total of 35 participants. The experimental interface (see the above-mentioned website for some screenshots) was created with Max/MSP. The chords were stored as MIDI files and played through a sampler to emulate a string quartet (the synthesizer was Dimension Pro playing a sample set from Garriton). A string quartet was chosen because, after discussions with colleagues, it was felt to be more pleasant than listening to a purely synthetic sound, and because it lends itself to the hearing out of four independent melodic parts. The music was played on headphones, and the individual instruments were panned to provide a naturalistic stereo image.

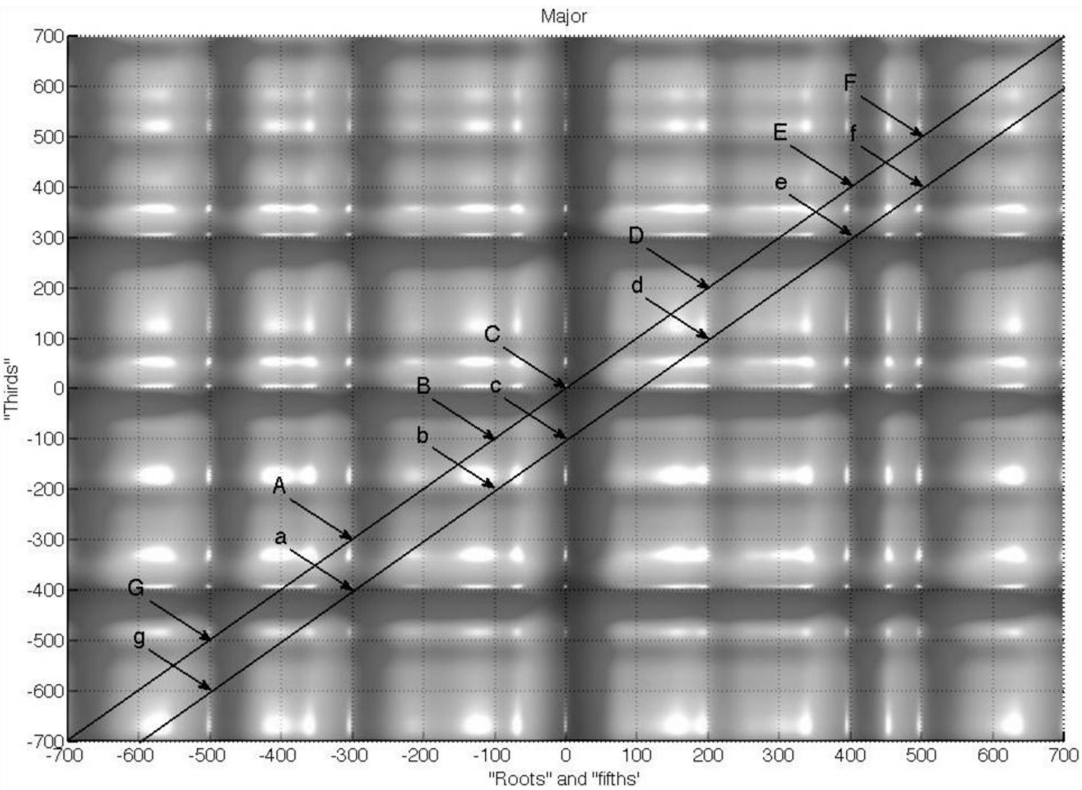


**Figure 2. The *z*-axis shows the spectral response distance of a continuum of triad tunings relative to a 12-TET major triad. The *x*-axis shows the pitch distance (in cents) between the "roots" and "fifths" of the continuum triads and the root and fifth of the reference triad. The *y*-axis shows the pitch distance between the "thirds" of the continuum triads and the third of the reference triad. The *x*- and y-axes have been scaled to ensure that all *x*-*y* distances are Euclidean. A selection of specific continuum triads are labelled.**

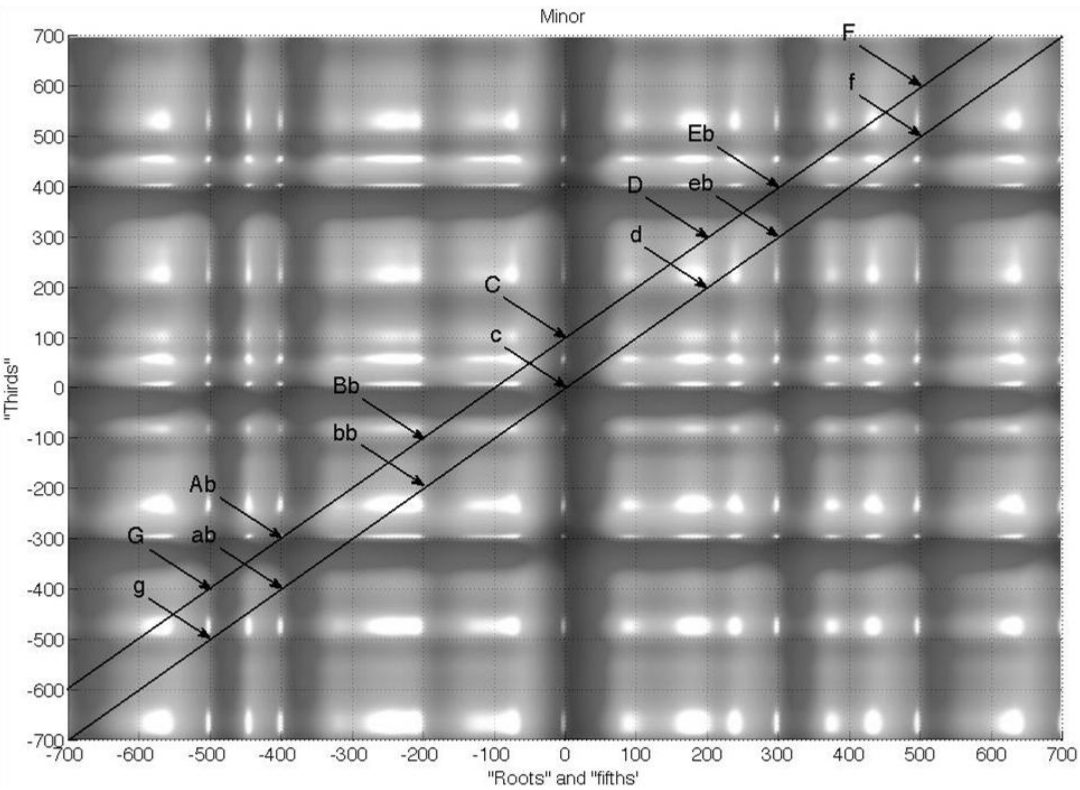


**Figure 3. The *z*-axis shows the spectral response distance of a continuum of triad tunings relative to a 12-TET minor triad. The *x*-axis shows the pitch distance (in cents) between the "roots" and "fifths" of the continuum triads and the root and fifth of the reference triad. The *y*-axis shows the pitch distance between the "thirds" of the continuum triads and the third of the reference triad. The *x*- and y-axes have been scaled to ensure that all *x*-*y* distances are Euclidean. A selection of specific continuum triads are labelled.**

For each chord progression, voice-leadings were chosen according to standard rules of harmony: there were four parts; common tones and steps were used rather than leaps; parallel fifths and octaves were completely avoided; hidden fifths and octaves were avoided when possible and, when unavoidable, approached by step in one part (given some of the very unusual chord pairings required, hidden fifths cannot always be avoided without creating unpleasant leaps). The scores for every progression can be seen at the above-mentioned website.

The order of presentation was separately randomised for each participant, the tuning was conventional twelve-tone equal temperament (12-TET), and the precise pitch of every chord progression was randomised (in 12-TET steps) over an octave. In between each progression, a short sequence of randomly generated chords was played to lessen the possibility of the previous progression colouring the response to the next.

After the test, participants were briefly interviewed to get their general impressions and response strategies, an approximate rating of their confidence in the accuracy of the answers they gave, their familiarity with different musical genres, and their playing and music theory experience.

### A. Chord Pairs—"Similarity" and "Fit"

In the first part of the experiment, each participant was asked to rate all possible pairs of 12-TET triads (when disregarding order and overall transposition, there are just 26 different pairs of 12-TET triads) for their "similarity" and "fit". Each triad pair was played as a loop—going from chord 1 to chord 2 to chord 1 to chord 2, and so on. Each chord had a minim (half-note) length, and the tempo chosen was 100 beats per minute.

The ratings were made on two separate five-point scales marked at the bottom and top with "similar" and "dissimilar", and "good fit" and "bad fit", respectively. A value of 1 was given to a rating of maximal similarity or fit (i.e., minimal distance), and a value of 5 to a rating of minimal similarity or fit (i.e., maximal distance). In the instructions, "similar" chords were defined as being those "you might inadvertently think the same"; "dissimilar" with "their difference is obvious and easy to hear"; "good fit" was likened to a chord transition that was "straightforward", "elegant", "easy"; "bad fit" to "clumsy", "awkward", "difficult".

The aim of the "similar/dissimilar" question was to get a rating for voice-leading distance. The aim of the "good fit/bad fit" question to get a rating of spectral distance. It was expected that there would be some confounding of the two concepts, as well as some confounding with other variables (such as activity). But I hoped the ratings would give some indication of the two types of distance.

### B. Chord Triples—"Cadential Effectiveness"

Ignoring transposition, there are 1,152 different order-dependent triples of 12-TET triads, so it is unfeasible (in a single experiment) to obtain ratings for all of them. It is, however, possible to take a subset of 72 triad-triples to test how tonal activity, due to a single type of comparison, impacts upon cadential effectiveness. The comparison chosen was the parallel transformation (e.g., comparing the spectral distance of triad pair C↔E with the spectral distance of the triad pair C↔e). As discussed above, this particular comparison was chosen because it is likely to have the smallest possible voice-leading distance, and so should maximise the absolute value of the activity produced by the two pairs (see Eq. 4). The parallel, relative, and leading-tone exchange parings are typically considered to have the smallest voice-leading distances (they each have two common tones) and, of these three, the parallel also has the same root note making it a natural candidate for having the smallest possible perceived voice-leading distance. Furthermore, even a cursory examination of the spectral distances generated by the model shows that the implications of this particular type of comparison provides a highly effective explanation for most of the triadic cadences commonly used in Western music.

The selection of triad triples was made in the following way. The antepenult was either C major or c minor, this makes two possible one-triad "progressions".

The penult was each of the 24 different triads in 12-TET (i.e. the major and minor chords on each degree of the chromatic scale), making a total of 48 different progressions. If the theory behind the psychoacoustic model is correct, then for each pair that is a parallel of another (e.g., c→A♭ compared to c→a♭) one pair will have positive activity (e.g., c→a♭), the other negative (e.g., c→A♭). Hence the latter should be heard as an alteration of the former.

The final was the same for each parallel pair of antepenult→penult pairs (so c→D and c→d get the same final). The root of the final was chosen to be the resolution of the active tone of the active penult (resolution of an active tone is made by a semitone step in the same direction as the alteration); the mode of the final was chosen in order to make its relationship with the antepenult have negative activity; this gives pairs of progressions such as c→D→g and c→d→g.

This selection method is a way of controlling the sign of the activity for each of the following three pairs $a_{A\to P\,|\,\mathbf{P}}\,|\,a_{P\to F\,|\,\mathbf{P}}\,|\,a_{A\to F\,|\,\mathbf{P}}$ (the reverse pairs $a_{A\leftarrow P\,|\,\mathbf{P}}$, $a_{P\leftarrow F\,|\,\mathbf{P}}$, and $a_{A\leftarrow F\,|\,\mathbf{P}}$ are not controlled for). It provides, therefore, four groups with the following patterns of activity for each of the above controlled pairs: Group 1 = + | + | –, Group 2 = – | + | –, Group 3 = + | – | –, and Group 4 = – | – | –. For every member of Group 1 there is a member of Group 2 that has exactly the same triads (ignoring transposition) except for the penult, which has a different mode. The same holds for Groups 3 and 4. The value of having paired groups is that it helps to reduce the degree to which uncontrolled variables contaminate the experiment. Each group contains essentially the same elements, but with the variable of interest ($a_{A\to P}$) being changed. Note also that all the progressions have a negative $a_{A\to F\,|\,\mathbf{P}}$, lessening the impact of this variable on the analysis.

When these 48 progressions are transposed to give a final major or minor triad with the same root (e.g., C or c), there are just four different penult→final endings (G→C, G→c, g→C, g→c, B♭→C, B♭→c, b♭→C, and b♭→c). Of these, only the progressions ending with G→C, G→c, b♭→C, or b♭→c have an active penult (i.e., they are members of Groups 1 and 3, which have a positive $a_{A\to P\,|\,\mathbf{P}}$). The final 24 progressions (making a of 72) use these endings but substitute all possible antepenults that give a negative value for both $a_{A\to P\,|\,\mathbf{P}}$ and $a_{A\to F\,|\,\mathbf{P}}$. This provides two more groups: Group 5 = – | + | –, and Group 6 = – | – | –. For every member of Group 1 there is a member of Group 5 that has (ignoring transposition) exactly the same penult, a final that has the same root (but not necessarily the same mode), and an antepenult of the opposite mode. The same holds for groups 4 and 6. Note also, that all these progressions have a negative $a_{A \to F \mid \mathbf{P}}$, thus ensuring the impact of this variable on the analysis is still lessened.

These related groups are intended to provide an effective way to estimate the impact of the sign (and magnitude) of $a_{A \to P \mid \mathbf{P}}$ upon cadential effectiveness, but this is by no means the only way to select a manageable, but useful, subset of triad triples. However, it does provide a systematic and, therefore, unbiased method to select those triples that should effectively test the model.

Each triad triple was played once through in full, but the participant could repeat play after a two-second delay. Each chord had a minim (half-note) length, and the tempo chosen was 80 beats per minute

The rating of cadential effectiveness was made on a seven-point scale marked "cadentially effective" at the top and "cadentially ineffective" at the bottom. The instructions gave the following explanation of "cadential effectiveness": "how effectively does the third chord give a feeling of 'closure' or 'finality'? For example: If the progression is 'cadentially effective', the third chord gives a clear and definite sense of closure, and would be an effective and unambiguous ending for a piece of music; if the progression is 'cadentially ineffective', the third chord suggests or implies that another chord, or chords, should follow; if the progression is 'neutral', the third chord may give no feeling of closure, but neither does it imply a need for any more chords to follow."

## VI. ANALYSIS OF THE EXPERIMENTS

### A. Similarity

A correlation matrix for the 35 participants' ratings of all 26 triad pairs was created. One participant had three negative correlations with other participants and a low average correlation level (0.16), and so was removed as an outlier. The overall average correlation level for the remaining participants was 0.49 (per participant averages ranging from 0.30 to 0.70), with no negative values between any pairs of participants.

For each triad pair, the ratings of similarity were averaged over the 34 participants to create a variable called *sim*. A stepwise multiple linear regression on *sim* was performed using the four variables: bass pitch distance (*bas*), tenor + alto + soprano + pitch distance (*tas*), fundamental response distance (*frd*), activity (*act*). *Tas* drops out due to insignificant correlation, with the three remaining variables giving a highly significant $R^2 = 0.943$ ($R^2_{\text{adj}} = 0.935$). Coefficients and their significance for this regression are summarised in Table 1, and a scatter plot is shown in Figure 4.

**Table 1. Regression coefficients and significance for multiple regression of *sim* on *frd*, *bas*, and *act*.**

| Model | B | Std. Error | Beta | t | Sig. |
|---|---|---|---|---|---|
| (Constant) | 3.382 | .104 | | 32.497 | .000 |
| *frd* | .121 | .012 | .667 | 9.977 | .000 |
| *bas* | .118 | .023 | .271 | 5.041 | .000 |
| *act* | .081 | .019 | .277 | 4.333 | .000 |

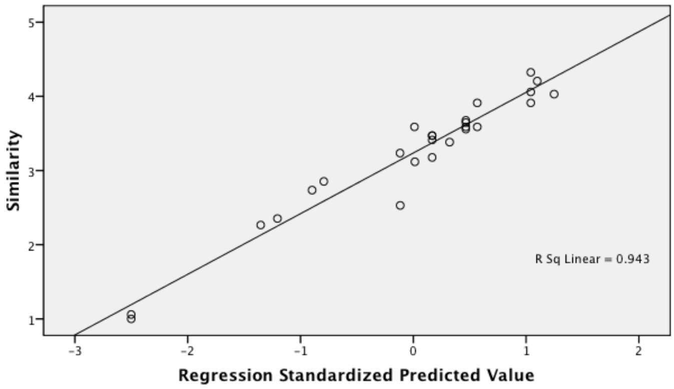


**Figure 4. Multiple regression of *sim* on *bas*, *frd*, and *act*.**

### B. Fit

A correlation matrix for the 35 participants' ratings of all 26 triad pairs was created. Three participants had low average correlation levels (−0.02, 0.02, and 0.07), and so were removed as outliers. The overall average correlation level for the remaining participants was 0.38 (per participant averages ranging from 0.19 to 0.51), with nine negative values between pairs of participants.

Clearly, the responses for fit were less consistent than those for similarity. Indeed, in the interviews following the test, many participants mentioned that they were using familiarity as a strategy—if they recognised a particular progression they would give it a higher fit. This suggests that these ratings are somewhat affected by each participant's musical taste and familiarity—in other words, a *long-term memory* (*ltm*) component.

For each triad pair, the ratings of similarity were averaged over the 32 participants to create a variable called *fit*. A multiple linear regression on *fit* was performed using the two variables of spectral response distance (*srd*), and activity (*act*), giving a highly significant $R^2 = 0.672$ ($R^2_{\text{adj}} = 0.644$). Coefficients and their significance for this regression are summarised in Table 2, and a scatter plot is shown in Figure 5.

**Table 2. Regression coefficients and significance for multiple regression of *fit* on *srd* and *act*.**

| Model | B | Std. Error | Beta | t | Sig. |
|---|---|---|---|---|---|
| (Constant) | 4.118 | .411 | | 10.012 | .000 |
| *srd* | .022 | .006 | .580 | 3.883 | .001 |
| *act* | .021 | .010 | .328 | 2.195 | .038 |

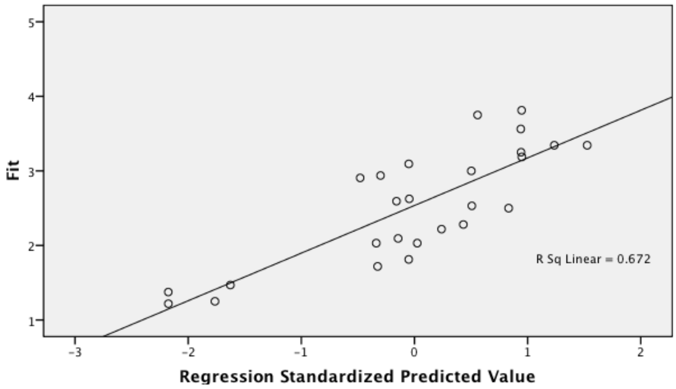


**Figure 5. Multiple regression of *fit* on *srd* and *act*.**

An attempt was made to simulate the effects of the long-term memory component by finding all tested cadences that contained a given chord pair. The cadence with the highest-rated cadential effectiveness transferred this rating to an *ltm* rating for that chord pair. The assumption being made here is that if a chord pair is found to be cadentially effective, it is likely to play a prominent and familiar role in music. An example is the progression C↔F♯ (and its transpositions), which was given a much higher rating for *fit* than is predicted from its high spectral response distance and tonal activity. However, this progression is part of a cadence, C→F♯→b (and its transpositions), that was rated as being highly effective (it is the familiar Neapolitan ♭II→V→i cadence).

Regressing fit with this additional *ltm* variable, significantly increased the regression coefficient—giving $R^2$ = 0.797 ($R^2_{\text{adj}}$ = 0.770). This suggests not only that fit is influenced by long-term memory, but also that the long-term memory component can be endogenously modelled using calculated values for cadential effectiveness (but, at the time of writing, this has not yet been done).

### C. Similarity and Fit

The results of this first experiment suggest the following path diagram, as illustrated in Figure 6.

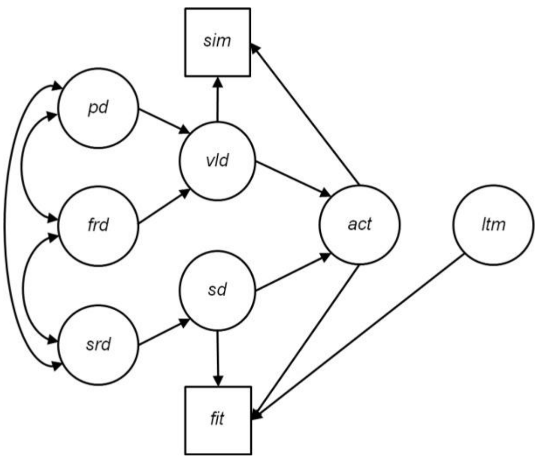


**Figure 6. A path diagram showing the proposed relationships between the cognitive variables discussed above—including a long-term memory (*ltm*) component—and the measured variables similarity (*sim*) and fit (*fit*). Error terms are not shown.**

If the path diagram is correct, these results suggest that not only can similarity be predicted with great accuracy using *pd*, *frd*, and *act*, but also that the latent voice-leading distance variable can be accurately predicted with just *pd* and *frd*. This is important because voice-leading distance is required as an input for the cadential effectiveness model.

### D. Cadential Effectiveness

The procedure for determining cadential effectiveness from the psychoacoustic model is complex, because it should take into account not just the presence of an active penult (due to a positive $a_{A\rightarrow P\,|\,\mathbf{P}}$), but also the successful resolution of any such active note. This is a preliminary report, and at this stage only a very simple model has been created. Cadential effectiveness (*eff*) was regressed against the following five variables: $a_{A\rightarrow P\,|\,\mathbf{P}}$, $a_{A\leftarrow P\,|\,\mathbf{P}}$, $a_{P\rightarrow F\,|\,\mathbf{P}}$, $a_{P\leftarrow F\,|\,\mathbf{P}}$, and $a_{A\leftarrow F\,|\,\mathbf{P}}$, and these values have been made equal to either +1 (for any positively valued *a*) or 0 (for any negatively valued *a*), which is a gross simplification. (Due to the selection method described in section, $a_{A\rightarrow F\,|\,\mathbf{P}}$ always has a negative value, and so was not entered into the regression equation.) Calculations of whether or not active tones are resolved in the final triad, and to which tone (root, third, or fifth), have not yet been made.

However, despite these simplifications (which are responsible for the vertical bands seen in Figure 7), a highly significant correlation is still obtained—$R^2$ = 0.593 ($R^2_{\text{adj}}$ = 0.562). Coefficients and their significance for this regression are summarised in Table 3 (note that all of the statistically significant variables have parameters with the expected sign, see section IV.2), and a scatter plot is shown in Figure 7.

**Table 3. Regression coefficients and significance for multiple regression of *cadential effectiveness* on $a_{A\rightarrow P\,|\,\mathrm{P}}$, $a_{A\leftarrow P\,|\,\mathrm{P}}$, $a_{P\rightarrow F\,|\,\mathrm{P}}$, $a_{P\leftarrow F\,|\,\mathrm{P}}$, and $a_{A\leftarrow F\,|\,\mathrm{P}}$.**

| Model | B | Std. Error | Beta | t | Sig. |
|---|---|---|---|---|---|
| (Constant) | 4.544 | .184 | | 24.754 | .000 |
| $a_{A\rightarrow P\,|\,\mathbf{P}}$ | 1.537 | .266 | .507 | 5.777 | .000 |
| $a_{A\leftarrow P\,|\,\mathbf{P}}$ | .215 | .296 | .065 | .727 | .470 |
| $a_{P\rightarrow F\,|\,\mathbf{P}}$ | −.771 | .250 | −.267 | −3.078 | .003 |
| $a_{P\leftarrow F\,|\,\mathbf{P}}$ | −2.233 | .248 | −.730 | −9.021 | .000 |
| $a_{A\leftarrow F\,|\,\mathbf{P}}$ | .259 | .312 | .073 | .831 | .409 |

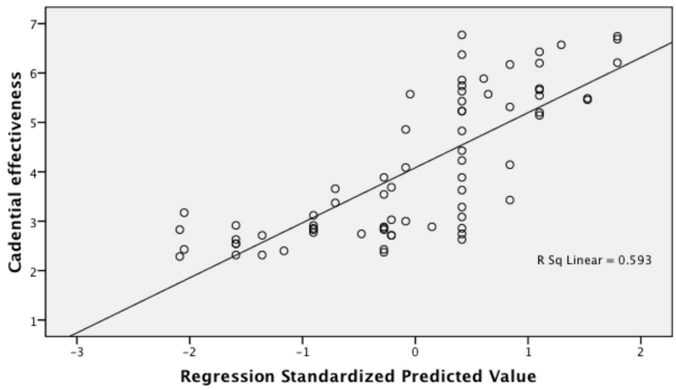


**Figure 7. Multiple regression of *cadential effectiveness* on $a_{A\rightarrow P\,|\,\mathrm{P}}$, $a_{A\leftarrow P\,|\,\mathrm{P}}$, $a_{P\rightarrow F\,|\,\mathrm{P}}$, $a_{P\leftarrow F\,|\,\mathrm{P}}$, and $a_{A\leftarrow F\,|\,\mathrm{P}}$.**

## VII. DISCUSSION AND CONCLUSION

The experimental data support the conclusion that the model effectively explains how successive triads induce feelings of expectation and resolution. Experimental testing cannot prove a model and the causal assumptions it makes. But because this model is based upon empirical observations of frequency difference limens, and follows a logical path from these direct observations to measures of cognitive distance, it seems a reasonable conclusion to make.

At the time of writing, the model for cadential effectiveness has not yet been fully completed, but the preliminary results seem promising. Furthermore, the model can still be substantially developed to account for memory effects, non-root position chords, more complex chords, and other factors.

To conclude, I would like to discuss a number of features that an effective theory of tonality should possess, and assess the current model against them.

*1) Testability*. Any successful theory should be able to make testable hypotheses. As described in this report, the model has already been tested. Furthermore, the model has no restrictions on the underlying tonal or spectral tunings used, allowing it to be tested against non-standard tunings such as those described by Erlich (2006) and Sethares (2009). Such non-standard tunings are likely to eliminate the possibility of contamination from long-term memory (cultural learning), making the psychoacoustic basis of this model highly testable. This is an area of future research I intend to pursue actively.

*2) Historical Tunings.* The effects of tonal music are robust over the range of tunings used throughout the common practice period (such as meantones, just intonation, and 12-TET (Barbour, 1951)). Any successful theory should be similarly robust. The predictions of the model are, indeed, broadly similar over all the above-mentioned historical tunings. Furthermore, the model does not (like so many others, such as Lerdahl's (2001), or Woolhouse's (2007)) rest upon an implicit assumption of twelve-tone equal temperament—a point that is crucial given that tonality was born a time when the most common tuning was quarter-comma meantone, not 12-TET.

*3) Privileged Ionian and Aeolian Modes.* Modal music, prior to the seventeenth century, gave no privileged status to any particular mode. Tonal music, on the other hand, privileges the major (Ionian) and minor (Aeolian) modes. Their privileged status is a natural consequence of the model: Given a diatonic scale, the only chord pairs with positive activity are those containing both members of the tritone (e.g., in the "white note" diatonic scale, activity is present only if one triad contains the note *f* and the other triad contains the note *b*). If both tritone notes are to resolve within the scale, there are only two triads that contain both resolution notes (*e* and *c*)—the root and third of the Ionian final, and the third and fifth of the Aeolian final.

*4) Historical Development of Tonality.* It is interesting to observe that the birth of tonality in the 17th century coincided with the birth of triadic harmony. Balzano (1980) speculates that the privileging of the Aeolian and Ionian modes and the use of triadic harmony are mutually dependent. The model presented here has a similar dependency—it is only when pairs of triads, rather than individual tones or dyads, are used that the conventional effects of tonality are predicted. This mirrors the historical development and demonstrates a causal dependency of tonality upon triadic harmony.

*5) Tonal Asymmetries.* As mentioned earlier, purely structural models, such as the tonal toroid proposed by many contemporary researches (e.g., see volume 15 of Tonal Theory in the Digital Age: Computing in Musicology) and Lerdahl's tonal pitch space, are inherently symmetrical and so cannot capture the asymmetries that are an important part of our perception of tonality. In the model presented here, tonal asymmetries are a function of the comparison chords used—no extra theory needs to be tacked on to account for them.

I hope it is evident that the psychoacoustic approach to tonality proposed in this report holds great promise. Indeed, I hope it may herald a return to psychoacoustic approaches in music theory, as well as act as a launch pad for the exploration of the tonal possibilities opened up by non-standard tunings and spectra (Sethares, Milne, Tiedje, Prechtl, & Plamondon, 2009).

## ACKNOWLEDGEMENTS

Bill Sethares for being such an available and patient sounding board, for profound discussions about musical perception and mathematics, and in particular for supplying me with the MATLAB routines that served as the starting point for my model.

Tuomas Eerola for welcoming me with open arms into academia, and allowing me to follow my own path for my Master's thesis (upon which this report is based) when a safer choice would have been much, mmm, safer for all of us. Also to Petri Toiviainen for continuing this policy of openness, and in particular for suggesting the use of the cosine distance. And to Vinoo Alluri for lighting up the darker corners of MATLAB's functionality. MMT rules!

Jim Plamondon for being Jim, for making the contacts and introductions that made this project possible, and for his indefatigable enthusiasm and optimism, and belief in me.

Christine and Astra for graciously giving me the time to follow my own private obsession.

## REFERENCES

Agawu, V. K. (2003). *Representing African Music: Postcolonial Notes, Queries, Positions.* New York and London: Routledge.

Balzano, G. (1980). The group-theoretic description of 12-fold and microtonal pitch systems. *Computer Music Journal*, *4* (4), 66–84.

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

Bharucha, J. J. (1987). Music cognition and perceptual facilitation: A connectionist framework. *Music Perception*, *5*, 1–30.

Dahlhaus, C. (1990). *Studies on the origin of harmonic tonality.* (R. O. Gjerdingen, Trans.) Oxford: Princeton University Press.

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

Glasberg, B. R., & Moore, B. C. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, *47* (1–2), 103–138.

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

Kameoka, A., & Kuriyagawa, M. (1969). Consonance theory, part I: Consonance of dyads. *The Journal of the Acoustical Society of America*, *45* (6), 1451–1459.

Kopp, D. (2002). *Chromatic transformations in nineteenth-century music.* Cambridge: Cambridge University Press.

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

Leman, M. (2000). An auditory model of the role of short-term memory in probe-tone ratings. *Music Perception*, *17* (4), 481–509.

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

Levitin, D. J. (2006). *This is your brain on music: Understanding a human obsession.* New York: Dutton.

Moore, B. C., Glasberg, B. R., & Shailer, M. J. (1984). Frequency and intensity difference limens for harmonics with complex tones. *The Journal of the Acoustical Society of America*, *75* (2), 500–561.

Plomp, R., & Levelt, W. J. (1965). Tonal consonance and critical bandwidth. *The Journal of the Acoustical Society of America, 38*, 548–560.

Sethares, W. A. (2004). *Tuning, timbre, spectrum, scale* (2nd ed.). London: Springer-Verlag.

Sethares, W. A., Milne, A. J., Tiedje, S., Prechtl, A., & Plamondon, J. (2009). Spectral tools for Dynamic Tonality and audio morphing. *Computer Music Journal*, *33* (2).

Tillmann, B., Bharucha, J. J., & Bigand, E. (2000). Implicit learning of music: A self-organizing approach. *Psychological Review*, (107), 885–913.

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

Woolhouse, M. (2007). Interval cycles and the cognition of pitch attraction in Western tonal-harmonic Music. *PhD Dissertation*.