

# TIMBRE AS A PSYCHOACOUSTIC PARAMETER FOR HARMONIC ANALYSIS AND COMPOSITION

*John MacCallum, Jeremy Hunt, and Aaron Einbond*  
Center for New Music and Audio Technology (CNMAT)  
University of California Berkeley  
{johnmac, jrmy, einbond}@berkeley.edu

## ABSTRACT

Timbre can affect our subjective experience of musical dissonance and harmonic progression. To this end, we have developed a set of algorithms to measure roughness (sensory dissonance), and pitch correlation between sonorities, taking into account the effects of timbre and microtonal inflection. We proceed from the work of Richard Parncutt and Ernst Terhardt, extending their algorithms for the psychoacoustic analysis of harmony to include spectral data from actual instrumental sounds. This allows for the study of a much wider variety of timbrally-rich acoustic or electronic sounds which was not possible with the previous algorithms. Further, we generalize these algorithms by working directly with frequency rather than a tempered division of the octave, making them available to the full range of microtonal harmonies. The new algorithms, by yielding different roughness estimates depending on the orchestration of a sonority, confirm our intuitive understanding that orchestration affects sensory dissonance. This package of tools presents rich possibilities for composition and analysis of music that is timbrally-dynamic and microtonally-complex.

## 1. INTRODUCTION

Beginning in the 1970's Ernst Terhardt proposed a psychoacoustic model of harmony [8, 9, 10]. Proceeding from Rameau and Helmholtz, he described "musical consonance" as the product of the co-operative perception of "sensory consonance" (the absence of sensory dissonance, or *roughness*) and "harmony" (or *harmonicity*, the similarity of a sound to a harmonic series) [9].

Musicologist Richard Parncutt has further extended and developed Terhardt's theory [3, 4]. In particular, Parncutt describes a measure of roughness of individual sonorities and of pitch commonality between two sonorities. Although other harmonic theories have taken perceptual data into account, a major advantage of Parncutt's algorithms is that they avoid biases towards pre-existing musical styles or techniques, with the exception that they are designed for equally-tempered music. Therefore they are promising tools for the composition and analysis of new, perceptually coherent, post-tonal music.

However, the work of Terhardt and Parncutt accounts for instrumental timbre in only a limited way. Composers

and analysts have become increasingly interested in timbre as a conveyor of musical meaning. It has long been acknowledged that timbre and orchestration have effects on our perception of dissonance and even harmonic relationships [7]. However, no harmonic theory has attempted to quantify these differences. Currently a variety of tools, such as Diphone and AudioSculpt, are available to create acoustical analyses in the versatile Sound Description Interchange Format (SDIF) [11]. Using SDIF data, we extend Terhardt's and Parncutt's measures to take timbre into account.

## 2. ACCOUNTING FOR TIMBRE

### 2.1. Virtual Fundamental

According to Terhardt, our ability to match a sonority to the harmonic series is one of the components of our perception of musical consonance [8]. In the terminology of Parncutt, by matching the *pure tones* of a sonority to an harmonic model, we may sense a *complex tone*, or virtual fundamental. The higher the frequency of the virtual fundamental, and the better its harmonics match the sonority, the more harmonic the sonority. Terhardt's algorithm [10] does not require adjustment to account for instrumental timbre or microtonal frequencies. But we propose inputting to the algorithm not merely an idealized list of pitches, but a list of timbrally-complex sounds each with many pure-tone components. For example, we can take several instrumental notes, each with its own spectrum, and ask what is the virtual fundamental of the spectra together.

### 2.2. Sensory Dissonance

Parncutt built into his model a rudimentary framework for including the effects of timbre, distinguishing between only three types of tones: *pure-tones*, *harmonic-complex tones*, and *octave-spaced (Shepard) tones*. The harmonic-complex tones are meant to model a general instrument timbre—they contain the first 10 partials of the harmonic series (rounded to semitones) with a roll-off that varies as the inverse of the partial number. Although these generalized timbres already give roughness data that correspond to our psychoacoustic experience better than pure

tones, we can include timbre in a more flexible and faithful way. Rather than using a prescribed overtone series to model each pitch of an instrumental chord, we use specific data from spectral analyses of sampled instruments stored in SDIF files. A chord orchestrated for diverse instruments can be simulated by combining the SDIF data from these instruments. We further revise Parncutt’s algorithm to treat the precise frequencies of the partials of complex tones, rather than rounding them to equally-tempered pitches. We also do not limit our calculation to 10 partials; instead we include all partial data available from spectral analysis. One reason Parncutt uses only 10 partials is that the 11th partial is poorly-approximated by semitones [3]. By avoiding equal-temperament, we eliminate this problem. Including an unlimited number of partials allows for a finer measure of the interaction of complex tones. And by using precise frequencies rather than idealized harmonics, we leave open the possibility of analyzing sounds that contain inharmonic spectra, for example bells or electronically-generated sounds.

### 2.3. Successive Pitch Relationships

A major extension Parncutt makes to Terhardt’s theory is in the consideration of successive pitch relationships. In doing this he seeks a perceptual groundwork by which we can understand existing traditions of voice leading and harmonic progression and extend them to post-tonal music. However a limitation of the algorithm is that it is designed for equally-tempered music. Such a theory would be especially useful for microtonal music, as there have been fewer studies of the progression between microtonal harmonies. By revising Parncutt’s chord distance algorithm to take microtonal frequencies into account, we take advantage of a powerful feature latent in the theory.

## 3. IMPLEMENTATION OF ALGORITHMS

### 3.1. Roughness

Roughness (sensory dissonance) is the beating sensation produced by the interaction of two or more components that are sensed within a certain distance in the inner ear. This distance is referred to as the “critical bandwidth” and varies with frequency. Following Parncutt [3], to calculate the degree of roughness between two pitches, we first calculate the critical bandwidth for the area around the mean frequency

$$W_{cb} = 1.72 (f_m^{0.65}) \quad (1)$$

where  $f_m = (f_1 + f_2) / 2$ . We then define the roughness of a sonority as the sum of the roughness of each pair of components

$$\rho = \sum_{j=0}^n \sum_{k=1}^{n-1} \frac{a_j \cdot a_k \cdot g(f_{cb})}{a_j^2} \quad (2)$$

where  $a_j$  and  $a_k$  are the amplitudes of the components and  $f_{cb}$  is the distance between  $f_1$  and  $f_2$  in critical bandwidths.  $g(f_{cb})$  is a ‘standard curve’<sup>1</sup> developed by Parncutt<sup>2</sup> and defined by

$$g(f_{cb}) = \left( e^{(f_{cb}/0.25)} \cdot e^{(-f_{cb}/0.25)} \right)^2, f_{cb} < 1.2 \quad (3)$$

### 3.2. Correlation

In order to measure the harmonic correlation between two chords, using the technique adapted from Parncutt, we must first adjust the chords to take masking into account.

#### 3.2.1. Masking

When two sounds lie within approximately three critical bands, the louder of the two will mask the other. Moore and Glasberg [2] define the following function for *equivalent rectangular bandwidth rate* (ERB-rate) or what Parncutt refers to as *pure-tone height*

$$H_p(f) = H_1 \log_e \left( \frac{f + f_1}{f + f_2} \right) + H_0 \quad (4)$$

where  $f$  is frequency in kHz and  $H_1 = 11.17$ ,  $H_0 = 43.0$ ,  $f_1 = 0.312$  kHz, and  $f_2 = 14.675$  kHz<sup>3</sup>.

The next step is to calculate the *auditory level*  $\Upsilon L(f)$  of each pure-tone component which is defined as the level relative to the threshold of audibility in dB (*sound pressure level*, SPL). This threshold, formulated by Terhardt et al. [10] is

$$L_{TH} = 3.64f^{-0.8} - 6.5e^{-0.6(f-3.3)^2} + 10^{-3}f^4 \quad (5)$$

where  $f$  is frequency in kHz. From there, we can calculate the auditory level as follows

$$\Upsilon L(f) = \max\{SPL(f) - L_{TH}; 0\} \quad (6)$$

where  $SPL(f)$  is the level of  $f$  in dB (SPL) and the max function ensures that the result will not drop below zero<sup>4</sup>.

The degree to which one pure-tone component of a sonority masks another is defined by

$$ml(f, f') = \Upsilon L(f') - k_M |H_p(f') - H_p(f)| \quad (7)$$

where  $k_M$  is the masking gradient which, should be set to a value between 12 and 18 dB. Next, because one maskee could be masked simultaneously by several maskers, we must calculate the overall masking level of a given pure-tone component.

$$ML(f) = \max\{20 \log_{10} \sum_{f \neq f'} 10^{ml(f, f')/20}; 0\} \quad (8)$$

<sup>1</sup> Parncutt’s standard curve approximates the experimental data collected by Plomp and Levelt [5]

<sup>2</sup> This curve is not published, but found in the C code available for download from Richard Parncutt’s website: <http://www-gewi.uni-graz.at/staff/parncutt/>

<sup>3</sup> These parameters were chosen by Moore and Glasberg [2] by fitting experimental ERB estimates using non-linear regression.

<sup>4</sup> In Parncutt’s work, this formula is  $\Upsilon L(P) = \max\{SPL(P) - L_{TH}; 0\}$  where  $P$  is the pitch category in semitones. We have substituted  $f$  (frequency) for  $P$  here and the formulas to follow.

The *audible level* of a pure-tone component is defined as its level above masked threshold

$$AL(f) = \max\{\Upsilon L(f) - ML(f); 0\} \quad (9)$$

which, as it increases, causes the *audibility* of the pure-tone component to saturate (approach 1):

$$A_p(f) = 1 - e^{\left(\frac{-AL(f)}{AL_0}\right)}. \quad (10)$$

In this formula,  $AL_0$  is set to 15dB due to experimental estimation by Hesse [1] and the subscript  $p$  stands for pure-tone component.

The final stage in the simulation of masking effects is to calculate the *complex-tone audibility* (Terhardt’s *spectral pitch weight*). Complex-tone sensations occur when the frequencies of the pure-tone components of a sonority are harmonically related. Parncutt’s model searches for these relationships using a template-matching technique. Like Parncutt, we use as a template an harmonic series with weights ( $W_n$ ) which vary as the inverse of the harmonic number. We also limit the template to 10 partials, although we use an unlimited number of partials elsewhere in our algorithm.

The continuous nature of frequency leaves us without a discrete set of values over which to move our template. This problem can be solved by deciding on a threshold within which the components of the template are said to ‘match’ the pure-tone components. If one or more matches are made, a calculation of complex-tone audibility is made with the following formula, which we have rewritten from Parncutt to use frequency:

$$A_c(f_1) = \frac{1}{k_t} \left( \sum_n \sqrt{W_n A_p(f_n)} \right)^2. \quad (11)$$

$k_t$  is meant to scale the model according to different types of listening and takes on a value between 1 (holistic listening) and 10 (analytic listening). Parncutt sets the value at 3. Finally, if a complex-tone and a pure-tone overlap, we define the audibility ( $A(f)$ ) as the stronger of the two.

### 3.2.2. Pitch Correlation

The measurement of the correlation between sonorities that Parncutt proposes is useful because it takes into account the probability of noticing each pitch of a given sonority based on the effects of masking and the degree to which the pitches are harmonically related. In order to calculate this measure, we must first define two other measures: *multiplicity* and *salience*. The former is the number of tones simultaneously noticed in a sound. This measure takes into account not only the contents of the sonority, but how the listener perceives them as well. An unscaled estimate of multiplicity can be made by assuming that the

probability of noticing a pure-tone component is proportional to the maximally-audible tone in a sonority:

$$M' = \frac{\sum_f A(f)}{A_{\max}}. \quad (12)$$

Salience is defined as the probability of noticing a pure-tone component of a sonority:

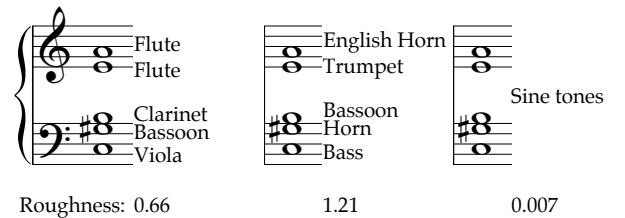
$$S(f) = \frac{A(f) M^{k_s}}{A_{\max} M'}. \quad (13)$$

where  $k_s$  is Parncutt’s “simultaneity perception parameter” which takes on a value between 0 and 1, with 0 for holistic or non-analytical listening and 1 for analytical listening. 0.5 is a typical value that fits the results of Parncutt’s experiments.

We then calculate the pairwise cross-correlation of their respective frequencies, weighted by pitch salience.

## 4. PRACTICAL EXAMPLES

An example of a musical use of the algorithms is taken from Arnold Schoenberg’s *Fünf Orchesterstücke*, Op. 16, “Farben.” The opening alternates between two orchestrations of the same chord, creating a subtly-shifting *klangfarbenmelodie* (tone-color melody). The first orchestration is for Flutes, Clarinet, Bassoon, and Viola, and the second is for English Horn, Trumpet, Bassoon, Horn, and Contrabass (Figure 1). Previous measures of dissonance would not distinguish between these two orchestrations. But by using partials extracted from a library of instrumental samples by AddAn, from the IRCAM package Diphone [6], differences in roughness can be measured. The first orchestration yields a roughness value of 0.66, while the second yields 1.21, on a scale of 0-7. This is in contrast to pure sine tones, which give 0.007. The low value for sine tones reflects the fact that the chord contains intervals mostly greater than a critical bandwidth apart, so roughness comes mostly from the overtones of the written pitches. The fact that the second orchestration is more rough than the first corresponds to our experience of the second chord, containing bright brass instruments, as more dissonant than the first.



**Figure 1.** Orchestrations of the “Farben” chord.

This roughness estimate increases with the number of component frequencies present within a critical band.

While this provides a reasonable estimate for a relatively small number of frequencies, it does not account for the possibility with a larger density of frequencies, such as a tone cluster, that amplitude fluctuations may cancel out causing a sense of “smoothing”. We are currently developing an algorithm that takes this into account.

The same chords are used as an example of the pitch correlation algorithm; unlike Parncutt’s algorithm, the algorithm described above can vary with orchestration, even when the pitch classes remain constant. For two sets of harmonic sounds, like the two orchestrations of Schoenberg’s chord, the correlation is close to unity: 0.9997. However, for more radically-different sounds the effect is much stronger. When we reorchestrate the same chord with spectra of sampled bells from the Sather tower at UC Berkeley, which are inharmonic, we find a much lower correlation: -0.0509. This highlights the fact that even though two musical sounds look the same on paper, they may have little correlation as actual sonorities. Therefore, our algorithm is useful for navigating the wide variety of timbres available to the modern acoustic or electronic composer.

## 5. DISCUSSION

### 5.1. Equal-Temperament

The model Parncutt proposes is constrained by its reliance on the equal-tempered division of the octave. For composers and other musicians interested in spectral music, other tuning systems such as just-intonation, or computer music where temperament need not be considered, this limitation proves problematic. In our revised model, a sonority is input as a list of frequencies thereby allowing the full continuum of frequency space to be used in the analysis of harmony.

### 5.2. Timbre

We propose an environment where actual instrument timbres can be included using SDIF files containing analyses of those instrumental timbres. Unlike the few timbral choices permitted by Parncutt’s algorithm, our algorithms provide a more realistic and flexible model for consideration of instrumental timbre. In particular, it allows distinctions to be made in the measurement of roughness for the same harmonic structures modelled for different instruments.

### 5.3. Compositional Models

We are in the process of implementing our procedure in an automated way using Open Music, such that an input chord with specified orchestration can be used to query a database of SDIF analyses and automatically produce a composite list of frequencies. This tool could be used compositionally, for example, as follows: a vocabulary of sonorities (either instrumental or electronic) could be input and classified according to roughness and virtual

fundamental. Pairwise chord correlations could be calculated. The input sonorities could then be treated as an harmonic space of possibilities for composition or improvisation, where the correlations between chords have psychoacoustic meaning. Novel harmonies could be interpolated between input harmonies and new instrumentations could be suggested based on specified parameters in the space. These measures could therefore prove a vast resource of creative possibilities.

## 6. REFERENCES

- [1] Alex Hesse. Zur ausgeprägtheit der tonhöhe gedrosselter sinustöne. In *Fortschritte der Akustik*, pages 535–38. Deutsche Arbeitsgemeinschaft für Musik (DAGA), 1985.
- [2] Brian C. J. Moore and Brian R. Glasberg. Suggested formulae for calculating auditory-filter bandwidths. *Journal of the Acoustical Society of America*, 74:750–53, 1983.
- [3] Richard Parncutt. *Harmony: A Psychoacoustical Approach*. Springer-Verlag, Berlin, 1989.
- [4] Richard Parncutt and Hans Strasburger. Applying psychoacoustics in composition: “harmonic” progressions of “nonharmonic” sonorities. *Perspectives of New Music*, 32(2):88–129, Summer 1994.
- [5] Reinier Plomp and W. J. M. Levelt. Tonal consonance and critical bandwidth. *Journal of the Acoustical Society of America*, 38:548–60, 1965.
- [6] Xavier Rodet. Synthesis and control of synthesis using a generalized diphone method. *The Journal of the Acoustical Society of America*, 1998.
- [7] Arnold Schoenberg. *Theory of Harmony*. University of California Press, 421-422, 1978/1922.
- [8] Ernst Terhardt. Zur tonhöhenwahrnehmung von klängen (perception of the pitch of complex tones). *Acustica*, 26:173–99, 1972.
- [9] Ernst Terhardt. Ein psychoakustisch begründetes konzept der musikalischen konsonanz. *Acustica*, 36:121–37, 1976.
- [10] Ernst Terhardt, Gerhard Stoll, and Manfred Seewann. Algorithm for extraction of pitch and pitch salience from complex tonal signals. *Journal of the Acoustical Society of America*, 71:671–78, 1982.
- [11] Matthew Wright, Amar Chaudhary, Adrian Freed, David Wessel, Xavier Rodet, Dominique Virolle, Rolf Woehrmann, and Xavier Serra. New applications of the sound description interchange format. *Proceedings of the International Computer Music Conference*, 1998.