Characterizing Musical Signals with Wigner-Ville Interferences

William J. DeMeo
williamdemeo@yahoo.com

Abstract

This paper presents a new characterization of musical signals which may lead to better understanding about how such signals are perceived. We first consider some signal analysis methods which facilitate measures of perceived qualities of music. More precisely, we consider a particular representation of signal energy on which to base a quantitative measure of sensory dissonance*. After defining sensory dissonance in section 2.1, we describe a well known signal decomposition method, the matching pursuit. Thereafter, we consider one aspect of signal energy which has been largely ignored in the literature – the Wigner-Ville interferences. We explain why and how these interferences can be used as the basis for a dissonance characterization of a musical signal.*

1 Introduction

1.1 DSP for Perception Analysis

We begin by stating the two main objectives of this work. Given a musical signal, $x(t)$, we wish to:

1. find useful time-frequency representations for analyzing the information content of $x(t)$, with the goal of characterizing perceptual properties of the signal;

2. find measures of qualities related to human perception of the signal; in particular derive a "dissonance signature" of $x(t)$.

In addressing (1), we use the *matching pursuit* algorithm (Mallat and Zhang 1993) to perform an atomic decomposition of the signal. We then use this decomposition as the basis for an energy characterization of the signal, given by the *Wigner-Ville distribution*. This approach is not new. However, the literature employing this strategy ignores the interference structure of the Wigner-Ville distribution. We retain these interference terms as they are the focus of our approach to the second objective stated above.

The novel contribution of this paper is consideration of how the interference terms of the Wigner-Ville decomposition can be used as the basis for a dissonance measure of a musical signal. For a simple composition of two pure tones, there is a well known relation between the interference terms and the sensory notion of "beating" – i.e. the effect caused by amplitude modulations resulting from the composition of tones. Since some measures of sensory dissonance are motivated by the rate of such beating, this suggests basing our dissonance characterization on the interference structure of a musical signal.

1.2 Sensory Dissonance

The concept of *sensory dissonance* was originally proposed by Helmholtz (1877), and further developed by Plomp and Levelt (1965), and Sethares (1997). What follows is a description of dissonance that motivates our alternative treatment of this concept.

In order to assess the intrinsic dissonance of a musical signal over a small time interval, the aforementioned studies employ a function of the signal's estimated frequency components over that interval. This often provides a useful quantitative measure. However, such a function makes no attempt to account for other widely accepted notions of dissonance. Perhaps the most obvious short-coming results from the point-wise nature of this dissonance function. That is, because it is well localized in time, there is no way for the dissonance function to account for *melodic dissonance* of the signal. The melodic dissonance of a given segment of music depends on that segment's relation to its context. In our present work, we consider signal analysis methods that provide for more dynamic dissonance measures. In particular, we wish to simultaneously account for local, point-wise dissonance, as well as dissonance resulting from the melodic contour of the signal.

2 Measures of Consonance

In later sections, we consider signal analysis methods that are particularly well suited to the type of musical analysis we wish to perform. Therefore, we should first consider some of the musical ideas underlying and motivating our work. The next section presents one such notion – musical *consonance*. We also briefly discuss some existing quantitative measures of this concept; the reader can find a more detailed treatment in Sethares (1997).

2.1 Consonance and Dissonance

According to Tenney (1988) and Sethares (1997), the historical usage of the term *consonance* can be classified according to five distinct categories: *melodic consonance* (CDC-1), *polyphonic consonance* (CDC-2), *contrapuntal consonance* (CDC-3), *functional consonance* (CDC-4), *sensory consonance* (CDC-5).

In this paper, the focus is on CDC-1 and CDC-5, so we describe only these. Briefly, *melodic consonance* applies to successive melodic intervals and describes these intervals as either consonant or dissonant depending on the surrounding melodic context; it refers to relatedness of pitches sounded successively, or the *melodic contour*. *Sensory consonance* equates consonance with smoothness and the absence of beats, and equates dissonance with roughness and the presence of beats.

The definition of sensory consonance is based on the phenomenon of beats. If two pure sine tones are sounded at almost the same frequency, then beating occurs due to the interference between the tones. The beating becomes slower as the two frequencies approach each other and disappears when they coincide. Typically, slower beats are perceived as gentle and pleasant while fast beats are perceived as rough and unpleasant. Observing that any sound can be decomposed into sinusoidal partials, Helmholtz (1877) theorized that the perception of dissonance in a musical tone is determined by the presence and quality of beats among the tone's interacting partials.

The present research effort is directed at the discovery of a measure which might simultaneously quantify multiple notions of consonance. In particular, we would like to exploit the theory of *sensory* and *tonal* consonance (CDC-2 and CDC-5) of Plomp and Levelt (1965) as well as its elaboration in Sethares (1997). Briefly, this theory employs functions called *dissonance curves* which measure the "sensory" dissonance, of a complex tone at each particular instant in time.[1] This provides a useful point-wise measure. However, we would also like a measure that is dynamic and appeals to a melodic sense of consonance, as in CDC-1. For example, a dissonance curve does not account for dissonance due to melodic changes from one complex tone to the next.

3 Energy Distributions

Wavelet and windowed Fourier transforms are computed by correlating the signal with families of time-frequency atoms. The time and frequency resolution of these transforms is thus limited by the time-frequency resolution of the corresponding atoms. Ideally, one would like to define a density of energy in a time-frequency plane with no loss of resolution. This section presents a different class of time-frequency representation (TFR) which is not restricted by the uncertainty principle.

The *Wigner-Ville TFR* is computed by correlating x with a time and frequency translation of itself. (Below we refer to the Wigner-Ville TFR simply as the "Wigner transform.") Though it yields some remarkable properties, the quadratic from of this representation is also considered a drawback which limits its application because of the inevitable cross terms that appear in quadratic forms. An attempt is usually made to attenuate these so-called "interference terms" by performing a time-frequency averaging, but this procedure results in a loss of resolution. It is not hard to show that the spectrogram, the scalogram, and all squared time-frequency decompositions can be written as time-frequency averagings of the Wigner transform; see, e.g., Mallat (1998).

3.1 Wigner Transform

The quadratic form[2]

$$\mathrm{W}_x(t,\nu) = \int x\left(t+\tfrac{\tau}{2}\right) x^*\left(t-\tfrac{\tau}{2}\right) e^{-i2\pi\tau\nu}\, d\tau$$

is known as the *Wigner-Ville distribution*, or *Wigner transform*. It is the one-dimensional Fourier transform of $\phi_x(t,\tau) = x^*\left(t-\tfrac{\tau}{2}\right) x\left(t+\tfrac{\tau}{2}\right)$, with respect to τ. The function ϕ_x has a Hermitian symmetry in τ, so the Wigner transform is real valued. Also, as the two-dimensional Fourier transform of the so called *ambiguity function*,

$$\mathrm{A}_x(\xi,\tau) = \int x\left(t+\tfrac{\tau}{2}\right) x^*\left(t-\tfrac{\tau}{2}\right) e^{i2\pi\xi t}\, dt$$

the Wigner transform satisfies

$$\mathrm{W}_x(t,\nu) = \int \mathrm{A}_x(\xi,\tau) e^{-i2\pi(\xi t+\nu\tau)}\, d\xi\, d\tau \qquad (1)$$

The Wigner transform localizes the time-frequency structures of x. If the energy of x is well concentrated in time around t_0 and in frequency around ν_0 then W_x has its energy centered at (t_0,ν_0), with a spread equal to the time and frequency spread of x.

3.2 Interference Structure

Because the Wigner-Ville transform is a sesquilinear form of the signal, it does not submit to the principle of linear superposition. Instead, as in the quadratic equation, $(a+b)^2 = a^2+b^2+ab+ba$, it is easy to verify that

$$\begin{aligned}\mathrm{W}_{x+y}(t,\nu) &= \mathrm{W}_x(t,\nu) + \mathrm{W}_y(t,\nu)\\ &\quad + \mathrm{W}_{xy}(t,\nu) + \mathrm{W}_{yx}(t,\nu)\end{aligned} \qquad (2)$$

[1] Really, a small interval of time is required to ascertain what pseudo-periodic frequencies are present at a particular instant.

[2] Integrals are over the entire real line unless otherwise noted.

where W_{xy} is the *cross Wigner transform* of the signals x and y, which is defined by

$$\mathrm{W}_{xy}(t,\nu) = \int x\left(t+\tfrac{\tau}{2}\right) y^*\left(t-\tfrac{\tau}{2}\right) e^{-i2\pi\nu\tau}\, d\tau \quad (3)$$

We define the *interference term* of equation (2) by

$$\begin{aligned} I_{xy}(t,\nu) &= \mathrm{W}_{xy}(t,\nu) + \mathrm{W}_{yx}(t,\nu) \\ &= 2\,\mathrm{Re}\left[\mathrm{W}_{xy}(t,\nu)\right] \end{aligned}$$

This real valued function creates non-zero values at interesting locations of the time-frequency plane.

More generally, for any linear combination of signal components,

$$x(t) = \sum_{n=1}^{N} a_n x_n(t)$$

the Wigner transform is

$$\begin{aligned} \mathrm{W}_x = &\sum_{n=1}^{N} |a_n|^2\, \mathrm{W}_{x_n}(t,\nu) \quad (4) \\ &+ 2\sum_{n=1}^{N-1}\sum_{k=n+1}^{N} \mathrm{Re}\left[a_n a_k^*\, \mathrm{W}_{x_n x_k}(t,\nu)\right] \end{aligned}$$

Hence, for a signal with N components, the Wigner transform contains $N(N-1)/2$ additional components. They result from the interaction of different components of the signal, and are called "interference terms" for two reasons. First, the mechanism of their creation is analogous to the usual interferences that can be observed for physical waves. A second reason for this terminology lies in the effect that these terms can have on the time-frequency diagram of the signal energy. As they amount to a combinatorial proliferation of additional, "specious" signal components, they can inhibit our ability to discern "true" signal components in the diagram.

The presence of cross terms in a Wigner transform can be regarded as a natural consequence of its bilinear structure. This very structure is also what leads to most of the good properties of the transform (such as localization). No matter whether one views the cross terms as helpful or hindering, it is important to understand fully the mechanism of their creation. This is indispensable for drawing the correct interpretation from the representation of an unknown signal, and for separating signal component terms from interference terms if desired (Flandrin 1999).

In the present work, we study the cross terms in order to understand how this measure of signal interference relates to "musical interference," i.e. dissonance.

3.3 Examples

For simplicity, suppose that $x \in L(\mathbb{Z}/N)$ represents an elementary signal component, so that x is a discrete periodic function of period N, defined on the group of integers $\mathbb{Z}/N \simeq \{0,1,\ldots,N-1\}$. In this special case[3] the so called *Weyl-Heisenberg* operator is $H : \mathbb{Z}/N \times \mathbb{Z}/N \to L(\mathbb{Z}/N)$, and is defined by

$$\begin{aligned} H(\mathbf{a})x(n) &= x_{\mathbf{a}}(n) \\ &= x(n-a_1)\, e^{i2\pi a_2 n/N}, \quad x \in L(\mathbb{Z}/N) \end{aligned}$$

for any $\mathbf{a} = (a_1,a_2) \in \mathbb{Z}/N \times \mathbb{Z}/N$.

The canonical example used to describe the structure of the cross terms of the Wigner transform begins with a well localized time-frequency atom $x(t)$ centered at $t=0$. From this we construct two atoms which are time and frequency shifted versions of $x(t)$. In particular, let $\mathbf{a} = (a_1,a_2)$ and $\mathbf{b} = (b_1,b_2)$ and consider

$$\begin{aligned} \alpha\, x_{\mathbf{a}}(t) &= \alpha\, x(t-a_1)\, e^{i2\pi a_2 t}, \qquad \alpha \geq 0 \\ \beta\, x_{\mathbf{b}}(t) &= \beta\, x(t-b_2)\, e^{i2\pi b_2 t}, \qquad \beta \geq 0 \end{aligned}$$

The Wigner transform of the composite signal $x_{\mathbf{a}}(t) + x_{\mathbf{b}}(t)$ is

$$\mathrm{W}_{x_{\mathbf{a}}+x_{\mathbf{b}}}(t,\nu) = \mathrm{W}_{x_{\mathbf{a}}}(t,\nu) + \mathrm{W}_{x_{\mathbf{b}}}(t,\nu) + I_{x_{\mathbf{a}}x_{\mathbf{b}}}(t,\nu)$$

The *covariance property* of the Wigner transform ensures that the shifted atoms, taken individually, have Wigner representations given by

$$\begin{aligned} \mathrm{W}_{x_{\mathbf{a}}}(t,\nu) &= \alpha^2\, \mathrm{W}_x(t-a_1, \nu-a_2) \\ \mathrm{W}_{x_{\mathbf{b}}}(t,\nu) &= \beta^2\, \mathrm{W}_x(t-b_1, \nu-b_2) \end{aligned}$$

Since the energy of W_x is centered at $(0,0)$, the energy of $\mathrm{W}_{x_{\mathbf{a}}}$ and $\mathrm{W}_{x_{\mathbf{b}}}$ is concentrated in neighborhoods of $\mathbf{a} = (a_1,a_2)$ and $\mathbf{b} = (b_1,b_2)$, respectively. A direct calculation verifies that the interference term is

$$\begin{aligned} I_{x_{\mathbf{a}}x_{\mathbf{b}}}(t,\nu) &= 2\alpha\beta\, \mathrm{W}_x(t-t_m, \nu-\nu_m) \\ &\quad \times \cos\{2\pi\left[(t-t_m)\Delta\nu - (\nu-\nu_m)\Delta t\right]\} \end{aligned}$$

where

$$\begin{aligned} t_m &= \frac{a_1+b_1}{2}, \quad \nu_m = \frac{a_2+b_2}{2} \\ \Delta t &= a_1 - b_1, \quad \Delta\nu = a_2 - b_2 \end{aligned}$$

This is an oscillatory waveform concentrated in a neighborhood of the point in the time-frequency plane that is the geometric midpoint between the individual components. The frequency of the oscillations is proportional to the Euclidean distance $\sqrt{\Delta\nu^2 + \Delta t^2}$ that separates the points $\mathbf{a}$ and $\mathbf{b}$, where the individual atoms are concentrated. The direction of these oscillations is perpendicular to the line that joins these two center points.

[3]For a more general, lucid treatment of the theory of time-frequency representations on finite abelian groups, see Tolimieri and An (1998).

Physical Interpretation. It is possible to attach physical meaning to the interference structure of the Wigner transform. For the most basic case, in which the signal is a simple superposition of pure frequencies, the cross term can be regarded as a signature of the *beat frequency* resulting from the interaction between the individual frequencies. To see this from the preceeding example, let x be a pure sinusoidal wave, $x(t) = e^{i2\pi\nu_m t}$ at the (mid-point) frequency ν_m, and suppose $\mathbf{a} = (0, -\frac{\Delta\nu}{2})$, $\mathbf{b} = (0, \frac{\Delta\nu}{2})$. Then $x_\mathbf{a}$ and $x_\mathbf{b}$ are the frequency shifted versions of x,

$$x_\mathbf{a}(t) = e^{i2\pi(\nu_m - \frac{\Delta\nu}{2})t},\ x_\mathbf{b}(t) = e^{i2\pi(\nu_m + \frac{\Delta\nu}{2})t}$$

The Wigner transform of the composite signal $x_\mathbf{a} + x_\mathbf{b}$ is given by

$$\begin{aligned} \mathrm{W}_{x_\mathbf{a}+x_\mathbf{b}}(t,\nu) &= \mathrm{W}_{x_\mathbf{a}}(t,\nu) + \mathrm{W}_{x_\mathbf{b}}(t,\nu) + I_{x_\mathbf{a}x_\mathbf{b}}(t,\nu) \\ &= \delta(\nu - (\nu_m - \tfrac{\Delta\nu}{2})) + \delta(\nu - (\nu_m + \tfrac{\Delta\nu}{2})) \\ &\quad + \delta(\nu - \nu_m)\, 2\cos(2\pi\Delta\nu t) \end{aligned} \tag{5}$$

Now let us relate this expression to the physical phenomenon of beats, which are perceived most easily when the distance between signal components is small. To do so, we write the signal as follows:

$$\begin{aligned} x_\mathbf{a}(t) + x_\mathbf{b}(t) &= \tfrac{1}{2}\, e^{i2\pi(\nu_m - \frac{\Delta\nu}{2})t} + \tfrac{1}{2}\, e^{i2\pi(\nu_m + \frac{\Delta\nu}{2})t} \\ &\quad + \cos(2\pi \tfrac{\Delta\nu}{2} t)\, e^{i2\pi\nu_m t} \end{aligned} \tag{6}$$

When the components $x_\mathbf{a}(t)$ and $x_\mathbf{b}(t)$ are close together in frequency – that is, when $\Delta\nu$ is small – the cosine term is slowly varying as compared to the exponential term, and the resulting signal can be viewed as a simple tone of frequency ν_m with a modulated amplitude envelope, with modulation frequency $\Delta\nu$. The term "beating" refers to such amplitude modulations.

Comparing (6) with (5), it is clearly the interference term of the Wigner transform which specifies the existence and nature of beats in the composite signal.

4 Energy Separation

4.1 Matching Pursuit

A *matching pursuit* (Mallat and Zhang 1993) is an iterative algorithm that decomposes the signal over *dictionary* vectors. A dictionary is a family of vectors $\mathcal{D} = \{g_\gamma\}_{\gamma\in\Gamma}$ included in a Hilbert space $\mathcal{H}$, with unit norms $\|g_\gamma\| = 1$. Such a family can be constructed by scaling, translating and modulating a single window function $g(t) \in L^2(\mathbb{R})$. We suppose that $g(t)$ is real, continuously differentiable and $O(\frac{1}{t^2+1})$. We further impose that $\|g\| = 1$, that the integral of $g(t)$ is nonzero, and that $g(0) \neq 0$.

Suppose f is the Gaussian window[4]

$$f(t) = 2^{1/4} e^{-\pi t^2} \tag{7}$$

[4] see, e.g., (Mallat and Zhang 1993)

Our dictionary will comprise periodic, scaled, translated, and modulated versions of f. So we first periodize f. Define $\mathrm{Per}_B\, f \in L(A)$ by the formula

$$\mathrm{Per}_B\, f(a) = \sum_{x\in B} f(a+x), \qquad a \in A$$

and call $\mathrm{Per}_B\, f$ the *periodization* of f over B. $\mathrm{Per}_B\, f$ is B-periodic.

If $f(n)$, for $n \in \mathbb{Z}$, is a discrete version of (7), we define the discrete N-periodic window g as follows:

$$g(n) = \mathrm{Per}_{N\mathbb{Z}}\, f(n) = \sum_{m\in N\mathbb{Z}} f(n+m), \quad n \in \mathbb{Z}$$

Next, for any scale s, let g_s denote the function g scaled by s; that is,

$$g_s(t) = \frac{1}{\sqrt{s}} g\left(\frac{t}{s}\right)$$

For any translation, a_1, and frequency modulation, a_2, let $\gamma = (s, a_1, a_2)$, and define a typical atom in the dictionary Γ by

$$\begin{aligned} g_\gamma(t) = g_{s\mathbf{a}}(t) &= g_s(t-a_1)\langle t, a_2\rangle \\ &= \frac{1}{\sqrt{s}} g\left(\frac{t-a_1}{s}\right) e^{i2\pi a_2 t} \end{aligned}$$

The index γ is an element of the set $\mathbf{\Gamma} = \mathrm{Re}^+ \times \mathrm{Re} \times \mathrm{Re}$. If the original window function $f(t)$ is even, which is generally the case, then the energy of $g_\gamma(t)$ is mostly concentrated in a neighborhood of (a_1, a_2), whose size is proportional to s.

For discrete matching pursuits, in order to describe the dictionary parameters of which the set Γ is comprised, we find the notation of Tolimieri and An (1998) extremely helpful. Suppose the signal of interest, x, has length $N = 2^{K+1}$. Let an arbitrary element of Γ be denoted (s, a_1, a_2). For each $j \in \{1, 2, \ldots, K\}$, set $s = 2^j$, and let successive translation parameters, a_1, be separated by an interval of $L_1 = 2^{j-1}$ samples. Define $M_1 = 2^{K-j+2}$, so that $N = L_1 M_1$. The set of translation parameters is then given by

$$a_1 \in \{0, L_1, 2L_1, \ldots, (M_1 - 1)L_1\} \simeq L_1\mathbb{Z}/N$$

If we let the modulation parameters, a_2, be separated by intervals of M_1 samples, then our parameter set would consist of translation modulation pairs (a_1, a_2) from the following set

$$L_1\mathbb{Z}/N \times M_1\mathbb{Z}/N = L_1\mathbb{Z}/N \times (L_1\mathbb{Z}/N)_*$$

This is a *critical sampling subgroup* of $\mathbb{Z}/N \times \mathbb{Z}/N$. Instead, we choose $M_2 = 2^{K-j}$, and let (a_1, a_2) range over the *integer oversampling subgroup*,

$$\Delta_s = L_1\mathbb{Z}/N \times M_2\mathbb{Z}/N$$

It sometimes simplifies expressions, and their corresponding algorithms, if we write the translation modulation pair as $(a_1, a_2) = (x_1 L_1, x_2 M_2)$ where

$$(x_1, x_2) \in \{0, 1, \ldots, M_1 - 1\} \times \{0, 1, \ldots, L_2 - 1\}$$

For each scale parameter, s, the foregoing describes a *Weyl-Heisenberg (W-H) system*, $\langle g_s, \Delta_s \rangle$. Following Mallat and Zhang (1993), in addition to these K W-H systems, we add to the dictionary of atoms complex exponentials (the Fourier basis) and the set of N discrete Diracs.

The matching pursuit algorithm iteratively decomposes a signal over dictionary vectors as follows. Let $R^0 x(t) = x(t)$, and suppose that we have computed the n^{th} order *residue*, $R^n x$, for $n \geq 0$. We then choose an element, g_{γ_n}, which closely "matches" the residue in the following sense:

$$|C(R^n x, g_{\gamma_n})| = \sup_{\gamma \in \Gamma} |C(R^n x, g_\gamma)|$$

where $C(x, g_\gamma)$ is a correlation function which measures the similarity between x and g_γ. An example is the usual inner product, $\langle x, g_\gamma \rangle$. Next, decompose the residue as

$$R^n x(t) = C(R^n x, g_{\gamma_n}) g_{\gamma_n}(t) + R^{n+1} x(t)$$

which defines the residue for step $n+1$, and fully specifies the algorithm recursion.

With the usual inner product as the correlation function, it can be shown (Mallat 1998) that the magnitude of the residue, $\|R^n x\|$, converges to 0 exponentially as n increases. This yields the following atomic signal decomposition:

$$x(t) = \sum_{n=0}^{\infty} C(R^n x, g_{\gamma_n}) g_{\gamma_n}(t) \qquad (8)$$

4.2 Interference Energy

In this work, we consider the special case in which the correlation function is simply the inner product:

$$C(R^m x, g_{\gamma_m}) = \langle R^m x, g_{\gamma_m} \rangle$$

From the matching pursuit decomposition above, we have

$$x(t) = \sum_{n=0}^{\infty} \langle R^n x, g_{\gamma_n} \rangle g_{\gamma_n}(t)$$

Referring to equation (4), we see that the corresponding Wigner-Ville representation is

$$W_x(t, \nu) = \sum_{n=0}^{\infty} |\langle R^n x, g_{\gamma_n} \rangle|^2 \, \mathrm{W}_{g_{\gamma_n}}(t, \nu) + \sum_{m,n} \langle R^m x, g_{\gamma_m} \rangle \langle R^n x, g_{\gamma_n} \rangle^* \, \mathrm{W}_{g_{\gamma_m} g_{\gamma_n}}(t, \nu) \qquad (9)$$

However, in at least that part of the literature dealing with musical signals, e.g. Gribonval, *et al.* (1996), as well as more generally, we consistently find that the signal energy is reduced to

$$E_x(t, \nu) = \sum_{n=0}^{\infty} |\langle R^n x, g_{\gamma_n} \rangle|^2 \, \mathrm{W}_{g_{\gamma_n}}(t, \nu)$$

As such, only the first term of (9) appears in the definition of E_x, the idea being that this term accounts for the energy of the "true" signal components. Since this is usually the primary concern of signal analyses, the typical definition of a signal's energy leaves out the interference terms.

Denoting the cross terms of (9) by $I_x(t, \nu)$ we can write the Wigner transform as

$$W_x(t, \nu) = E_x(t, \nu) + I_x(t, \nu)$$

where $E_x(t, \nu)$ is the *signal energy* and we call $I_x(t, \nu)$ the *interference energy*.

Computing and retaining a cross Wigner transform among two time-frequency atoms costs roughly the same as computing a single (auto) Wigner transform. However, for a matching pursuit decomposition involving N atoms, there are $N(N-1)/2$ cross Wigner terms required for the computation of I_x. Assuming a fast implementation of the cross Wigner transform based on the FFT, and assuming that our decomposition doesn't involve thousands of atoms, the computational burden is still manageable, as we show for a simple example in the following section.

Example. As a simple example, we construct a signal by adding a constant tone to a tone with a frequency that increases linearly. The latter is called a "linear chirp." We set the starting frequency of the chirp to be equal to the frequency of the constant tone. It then increases linearly until it reaches a frequency double that of the constant tone. Though this is a very simple signal, it makes for a useful example because it shows how a constant, "tonic" tone interacts with a tone that increases continuously from the tonic up through one octave above the tonic.

Figure 1 shows what we have termed the interference energy of the signal. Figure 2 shows the Wigner transform of the signal.

The next section describes how we use the interference energy to derive a dissonance measure of the signal.

5 Dissonance Measures

5.1 A Simple Dissonance Measure

We first consider perhaps the simplest of the many possible dissonance measures based on the information provided by the interference energy of the Wigner transform.

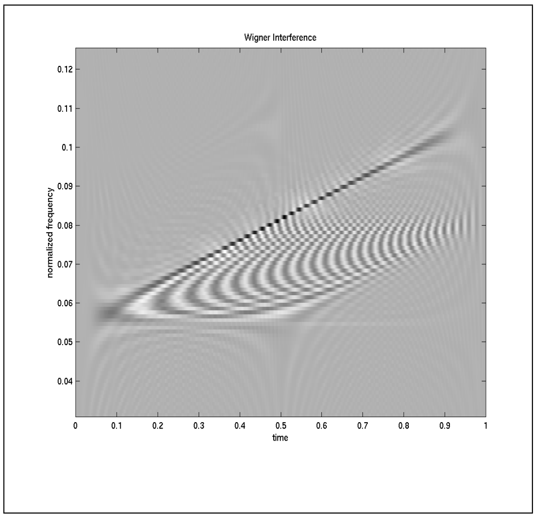

Figure 1: Wigner-Ville interference energy of a constant frequency modulation plus a linear chirp.

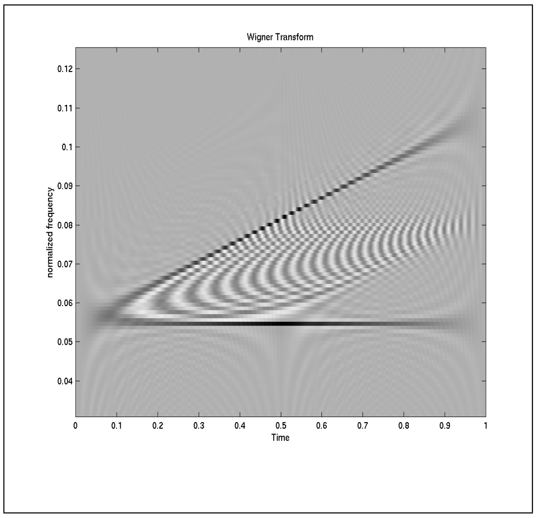

Figure 2: Wigner-Ville transform of a constant frequency modulation plus a linear chirp.

Name	Just	Pythagorean	Equal
m2	16/15	256/243	$2^{1/12}$
M2	9/8	9/8	$2^{2/12}$
m3	6/5	32/27	$2^{3/12}$
M3	5/4	81/64	$2^{4/12}$
P4	4/3	4/3	$2^{5/12}$
Tritone	64/45	729/512	$2^{6/12}$
P5	3/2	3/2	$2^{7/12}$
m6	8/5	128/81	$2^{8/12}$
M6	5/3	27/16	$2^{9/12}$
m7	7/4	16/9	$2^{10/12}$
M7	15/8	243/128	$2^{11/12}$
octave	2	2	$2^{12/12}$

Table 1: Frequency ratios used to delimit the horizontal axis in the figures.

Section 3.1 introduced the function $\phi_x(t,\tau)$. Let's generalize this slightly by defining

$$\phi_{g_{\gamma_m} g_{\gamma_n}}(t,\tau) = g_{\gamma_m}\left(t + \tfrac{\tau}{2}\right) g^*_{\gamma_n}\left(t - \tfrac{\tau}{2}\right)$$

By definition of the cross Wigner transform in equation (3), $\mathrm{W}_{g_{\gamma_m} g_{\gamma_n}}$ is the Fourier transform of $\phi_{g_{\gamma_m} g_{\gamma_n}}(t,\tau)$ with respect to τ. Therefore, the inverse Fourier transform of $\mathrm{W}_{g_{\gamma_m} g_{\gamma_n}}$ is $\phi_{g_{\gamma_m} g_{\gamma_n}}$. That is,

$$\int \mathrm{W}_{g_{\gamma_m} g_{\gamma_n}}(t,\nu) e^{i2\pi\nu\tau}\, d\nu = \phi_{g_{\gamma_m} g_{\gamma_n}}(t,\tau)$$

Suppose that, at any given point in time, we integrate $\mathrm{W}_{g_{\gamma_m} g_{\gamma_n}}(t,\nu)$ over all frequencies, ν. This is equivalent to evaluating $\phi_{g_{\gamma_m} g_{\gamma_n}}$ at $\tau = 0$:

$$\begin{aligned} \int \mathrm{W}_{g_{\gamma_m} g_{\gamma_n}}(t,\nu)\, d\nu &= \phi_{g_{\gamma_m} g_{\gamma_n}}(t,0) \\ &= g_{\gamma_m}(t) g^*_{\gamma_n}(t) \qquad (10) \end{aligned}$$

As a first proposal, we consider measuring the dissonance at time t of the signal x by integrating the interference energy $I_x(t,\nu)$ over all frequencies ν. The result is

$$\mathcal{I}_x(t) = \int I_x(t,\nu)\, d\nu \qquad (11)$$

$$= \sum_{m,n} \langle R^m x, g_{\gamma_m} \rangle \langle R^n x, g_{\gamma_n} \rangle^* \int \mathrm{W}_{g_{\gamma_m} g_{\gamma_n}}(t,\nu)\, d\nu$$

The lower graph in Figure 3 shows how the function $\mathcal{I}_x(t)$ behaves for the constant tone plus linear chirp. The top graph is there for reference and represents the value of the instantaneous frequencies of the signal. Figure 4 also shows the value of $\mathcal{I}_x(t)$ for the example signal. However, in this figure the time axes are delimited by tick marks representing just and Pythagorean tunings. Table 1 presents the frequency ratios corresponding to these tick marks.

We first note that, in all three tuning systems, the perfect 5th falls in roughly the same place, and that this interval consistently corresponds to local minima of $\mathcal{I}_x(t)$. Other significant intervals, such as the major 3rd and the tritone, also correspond to local minima of $\mathcal{I}_x(t)$.

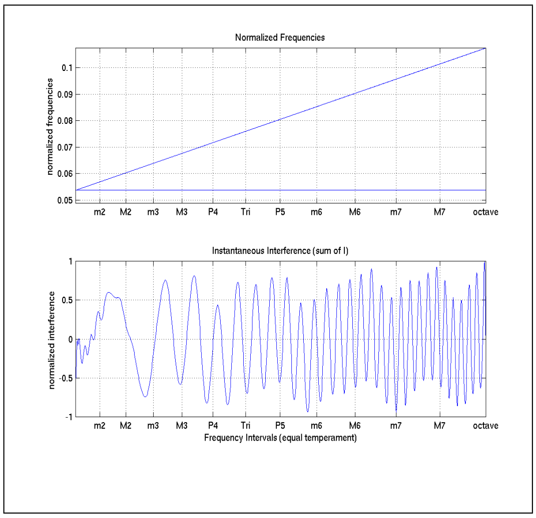

Figure 3: (a) Normalized instantaneous frequencies; (b) Instantaneous interference – the sum of interferences at each point in time; the time axis of (b) is delimited by the ratio of the two frequencies in figure (a), with tick marks illustrating points of an equal tempered scale.

5.2 A General Dissonance Measure

As stated at the outset, we want to find not only a point-wise measure of dissonance, but also a measure that could account for melodic context. The function $\mathcal{I}_x(t)$ is essentially the sum over ν of the function $I_x(t,\nu)$. Since $I_x(t,\nu)$ is a measure of interferences among signal components centered at t as well as those centered at times surrounding t, it might seem as though the function $\mathcal{I}_x(t)$ accounts for melodic context. However, note that

$$\mathcal{I}_x(t) = \sum_{m,n} \langle R^m x, g_{\gamma_m}\rangle \langle R^n x, g_{\gamma_n}\rangle^* g_{\gamma_m}(t) g^*_{\gamma_n}(t)$$

by equation (10). Thus $\mathcal{I}_x(t)$ only measures interferences among signal components at the single time instant t. Still, as the results for our simple example show, it may provide a useful point-wise dissonance characterization of a signal.

We can generalize the foregoing by considering the inverse Fourier transform of the interference energy:

$$\mathcal{I}_x(t,\tau) = \int I_x(t,\nu) e^{i2\pi\nu\tau}\, d\nu$$

$$= \sum_{m,n} \langle R^m x, g_{\gamma_m}\rangle \langle R^n x, g_{\gamma_n}\rangle^* \phi_{g_{\gamma_m} g_{\gamma_n}}(t,\tau)$$

Recall,

$$\phi_{g_{\gamma_m} g_{\gamma_n}}(t,\tau) = g_{\gamma_m}\left(t + \tfrac{\tau}{2}\right) g^*_{\gamma_n}\left(t - \tfrac{\tau}{2}\right)$$

The function $\mathcal{I}_x(t,\tau)$ leads to dissonance measures based on interferences between signal components at different points in time. For instance, a measure of interferences among signal components that are separated by not more than τ_0 units of time is

$$\begin{aligned} \mathcal{I}_x^{\tau_0}(t) &= \int_0^{\tau_0} \mathcal{I}_x(t,\tau)\, d\tau \qquad (12) \\ &= \sum_{m,n} \langle R^m x, g_{\gamma_m}\rangle \langle R^n x, g_{\gamma_n}\rangle^* \times \\ &\quad \int_0^{\tau_0} g_{\gamma_m}\left(t + \tfrac{\tau}{2}\right) g^*_{\gamma_n}\left(t - \tfrac{\tau}{2}\right)\, d\tau \end{aligned}$$

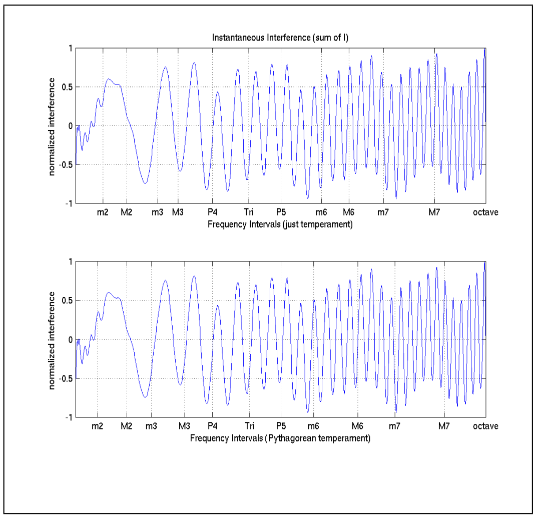

Figure 4: Instantaneous interference for just and Pythagorean tunings. This figure is the same as that of 3 (b), except that the tick marks illustrate points of a just (a) and Pythagorean (b) scale.

Of course, we can vary τ_0 depending on the extent to which we wish to account for interferences among signal components across time.

More generally, put a distribution μ on the domain of time differences among signal components. This distribution describes the relative importance of the interferences across various time intervals. Then define,

$$\mathcal{I}_x^{\mu}(t) = \int_{-\infty}^{\infty} \mathcal{I}_x(t,\tau)\, d\mu(\tau) \qquad (13)$$

The definition of $\mathcal{I}_x$ in equation (11) and $\mathcal{I}_x^{\tau_0}$ in equation (12) are special cases of (13). We arrive at $\mathcal{I}_x^{\tau_0}$ by setting

$$d\mu(\tau) = \chi_{[0,\tau_0)}(\tau)\, d\tau$$

where $\chi_{[0,\tau_0)}(\tau)$ is the characteristic function, equal to 1 when $\tau \in [0, \tau_0)$ and 0 elsewhere. In this case, μ is a uniform distribution of width τ_0. Therefore, μ assigns equal importance to interferences among components separated by at most τ_0 units of time, and zero importance to interferences among components separated by more than τ_0 units. Clearly, by setting $\tau_0 = 0$ in the foregoing, we return to the simplest measure, $\mathcal{I}_x$, with which we began.

Figure 5 shows how the function $\mathcal{I}_x^{\tau_0}$ behaves for our example signal and two values of τ_0. The upper graph shows $\mathcal{I}_x^{\tau_0}$ for $\tau_0 = 5.9$ milliseconds, and the lower graph shows the same function for $\tau_0 = 46.9$ milliseconds.

One interesting aspect of these figures is the behavior they exhibit near the perfect fifth interval. Taken as a measure of dissonance, the figures indicate that dissonance is high when the ratio approaches the perfect fifth and low once it reaches the perfect fifth.

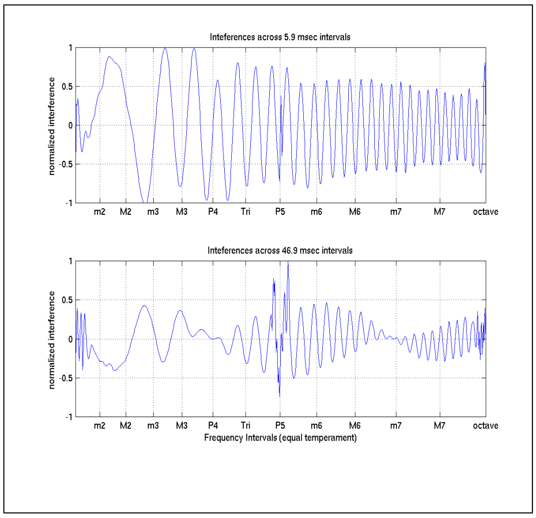

Figure 5: Interference measure $\mathcal{I}_x^{\tau_0}$; the sum of interferences over the given time intervals.

6 Conclusion

We have described the function, I_x, representing the sum of the interference terms of the Wigner transform of a signal. Based on this function we derived a measure, $\mathcal{I}_x^{\tau_0}$, of interference among signal components over a given interval of time, τ_0. Finally, we proposed a general interference measure, $\mathcal{I}_x^{\mu}$, by putting a distribution μ on the domain of time differences between signal components.

The dissonance measure that results from the foregoing depends on the function μ, which represents the relative importance we place on interferences across various time intervals. Generalizing $\mathcal{I}_x$ is this way enables the interference function to account for melodic context, and this provides heuristic justification for the use of $\mathcal{I}_x^{\mu}$ as a measure of melodic dissonance.

We have shown that the measures presented above exhibit interesting behavior for our simple example. However, it is as yet unclear exactly how useful, as measures of dissonance, are such functions. We expect that further research, and experience with these functions in musical situations, will at least demonstrate their utility as a means of characterizing musical signals.

7 Analytic Interference

Reconsider the composite signal $x(t) = x_{\mathbf{a}}(t) + x_{\mathbf{b}}(t)$, where

$$x_{\mathbf{a}}(t) = e^{i2\pi(\nu_m - \frac{\Delta\nu}{2})t}, \; x_{\mathbf{b}}(t) = e^{i2\pi(\nu_m + \frac{\Delta\nu}{2})t}$$

The interference energy is

$$I_x(t,\nu) = \delta(\nu - \nu_m)\, 2\cos(2\pi\Delta\nu t)$$

Integrating,

$$\begin{aligned}\mathcal{I}_x(t) &= 2\cos(2\pi\Delta\nu t) \int \delta(\nu - \nu_m)\, d\nu \\ &= 2\cos(2\pi\Delta\nu t)\end{aligned}$$

To simplify notation, let an atom in the dictionary be denoted g_n. Recall that, for an arbitrary linear combination of atoms,

$$x(t) = \sum_{n=0}^{\infty} \alpha_n g_n$$

the interference energy is given by

$$I_x(t,\nu) = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} \alpha_n \alpha_m^* W_{g_n g_m}(t,\nu)$$

and, under suitable conditions which permit interchanging the summation and integration,

$$\begin{aligned}\mathcal{I}_x(t) &= \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} \alpha_n \alpha_m^* \int W_{g_n g_m}(t,\nu)\, d\nu \\ &= \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} \alpha_n g_n(t) \overline{\alpha_m g_m}(t)\end{aligned}$$

References

Flandrin, P. (1999). *Time-Frequency/Time-scale Analysis*. Academic Press.

Gribonval, R., P. Depalle, X. Rodet, E. Bacry, and S. Mallat (1996). Sound signal decomposition using a high resolution matching pursuit. In *Proceedings of the International Computer Music Conference*. International Computer Music Association.

Helmholtz, H. (1877). *On the Senstations of Tone*. New York, NY: Dover.

Mallat, S. (1998). *A Wavelet Tour of Signal Processing*. Academic Press.

Mallat, S. and Z. Zhang (1993). Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing 41*, 3397–3415.

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

Sethares, W. (1997). *Tuning, Timbre, Spectrum, Scale*. New York, NY: Springer-Verlag.

Tenney, J. (1988). *A History of 'Consonance' and 'Dissonance'*. Excelsior Music.

Tolimieri, R. and M. An (1998). *Time-Frequency Representations*. Boston, MA: Birkhäuser.