REAL-TIME DISSONANCIZERS: TWO DISSONANCE-AUGMENTING AUDIO EFFECTS

Matthew Hoffman

Dept. of Computer Science,
Princeton University
Princeton, NJ, USA
mdhoffma at cs·princeton·edu

Perry Cook

Dept. of Computer Science,
Dept. of Music,
Princeton University
Princeton, NJ, USA
prc at cs·princeton·edu

ABSTRACT

We present two simple perceptually motivated audio effects designed to increase the perceived sensory dissonance/roughness (a process we call "dissonancization") of audio input. The first involves heterodyning multiple bands of the audio signal at different frequencies to break each sinusoid in each band into two sinusoids separated in frequency by the amount that Kameoka and Kuriyagawa [1] predict will produce a maximally dissonant effect. The second attempts to increase the depth of modulation caused by existing beating partials by exponentiating the amplitude envelope within small bands, enhancing the perceived roughness already present in the signal. The first algorithm can produce very dramatic effects even for very consonant inputs, whereas the second tends to produce a more subtle effect. Both algorithms are quite simple to understand and implement and computationally inexpensive enough to be used in real time, but produce perceptually interesting results. The effects can be selectively applied so as to affect only desired frequency ranges, and can be continuously controlled (e.g. in a performance context) to have more or less impact.

1. INTRODUCTION

Sensory dissonance, put briefly, is the sensation of roughness [1] caused by the rapid amplitude modulation (AM) associated with sinusoids whose frequencies are close enough to produce beating but not so close that the beating is perceived as amplitude modulation per se. It offers a plausible biological origin for traditional Western categorizations of dissonant and consonant intervals and chords [3].

Quantitative models of sensory dissonance have existed for almost half a century ([1, 4], for example), but have rarely been put to use in audio signal processing. (Perhaps the most notable exception is the work of Sethares [5, 6].) One reason may be that, as Leman observes [7], most of the models that have been proposed rely on what he calls "curve-mapping" approaches – that is, they require that one know at each moment the precise amplitudes and frequencies of all partials present in a sound, which one rarely does outside of the context of a sophisticated sinusoidal modeling framework (e.g. [8, 9]). (Leman's own model is a very interesting exception.)

In this paper, we present two simple "dissonancizer" digital audio effects informed by these psychoacoustic models that attempt to augment the levels of sensory dissonance present in audio while causing minimal distortion to the audio's other perceptual qualities. The first relies heavily on Kameoka and Kuriyagawa's model of sensory dissonance [1] and utilizes spectral ring modulation, while the second attempts to increase the depth of existing modulation in the input audio. Neither requires any spectral analysis more sophisticated than a bandpass filter bank.

Although the word dissonance has negative connotations and may seem like a purely undesirable quality, the psychoacoustic phenomenon has a long history of being manipulated to aesthetic effect by performers and composers (modern composers in particular). We believe that these effects, which can be applied in real time, make an interesting and useful addition to the musician's palette.

2. SPECTRAL RING MODULATION

Based on listening tests, Kameoka and Kuriyagawa [1] determined a simple formula that predicts, for a sinusoid at a given frequency f_1, at what frequency $f_2 = f_1 + g(f_1)$ a second sinusoid will produce the maximum sensory dissonance[2]:

$$g(f_1) = 2.27 f_1^{0.477} \tag{1}$$

They also developed a formula describing how the amount of dissonance produced by two interacting partials varies as a function of their frequencies:

$$d(f_1, f_2) = \begin{cases} \frac{2+\log_{10}((f_2-f_1)/f_1)}{2+\log_{10}(g(f_1)/f_1)}, & f_2 - f_1 < g(f_1) \\ 0.9\frac{\log_{10}((f_2-f_1)/f_1)}{\log_{10}(g(f_1)/f_1)} + 0.1, & f_1 > f_2 - f_1 \geq g(f_1) \\ 0, & f_2 \geq 2f_1 \end{cases} \tag{2}$$

The first result is particularly interesting in the context of ring modulation, a classic audio effect where the amplitude of an input signal is modulated by a sine wave of frequency m. In the case where the input signal is a single sinusoid of frequency f_1, basic trigonometric identities show that the result is a signal composed

[1]We use the terms "roughness" and "dissonance" interchangeably throughout this paper to refer to the same phenomenon of sensory dissonance. This is regrettable, since the word "dissonance" has so many other meanings even within a musical context (see Leider's work for a full discussion [2]), but the alternative is awkward.

[2]Note that this assumes that the power p_1 of the first partial is 57dB SPL. The full formula accounting for power is:

$$g(f_1, p_1) = 2.27(1 + (p_1 - 57)/40) f_1^{0.477}$$

In order to use this more precise formula, however, one must guess at what volume level the listener will be playing the sound.

of two sinusoids of frequency $f_1 + m$ and $f_1 - m$. From another perspective, then, equation (1) can be interpreted as dictating at what frequency $m = g(f_1)/2$ to modulate a sinusoid of frequency f_1 so as to produce the strongest sensation of dissonance[3].

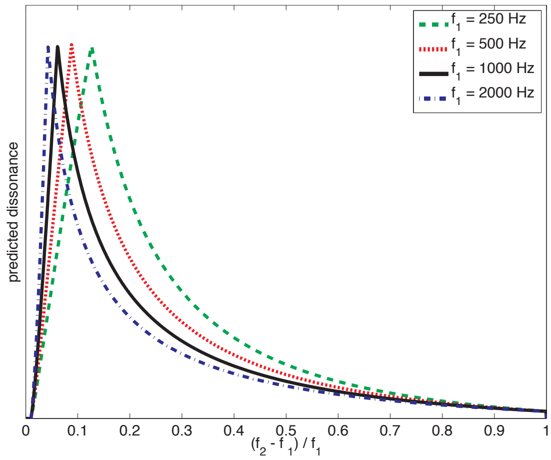

Figure 1: *$d(f_1, f_2)$, the predicted dissonance caused by two beating partials as a function of their frequency difference, shown for several values of f_1. See Equation (2 for details).*

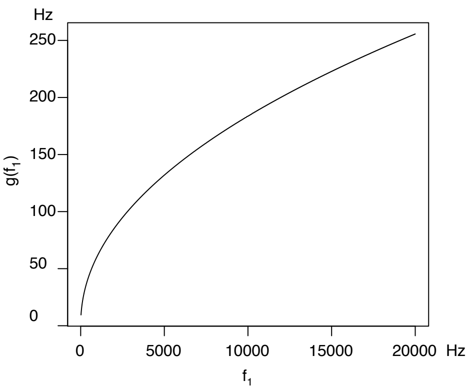

Figure 2: *$g(f_1)$, the frequency difference causing maximum dissonance between two partials, as a function of f_1, the frequency of the lower of the two partials. See Equation (1).*

Of course, in general we are interested in increasing the sensory dissonance produced by complex, broadband audio signals, not just single sinusoids. Ideally we would be able to ring modulate each partial i individually by a sinusoid of the appropriate frequency $g(f_i)/2$. This is somewhat impractical, although it might be accomplished within a sophisticated sinusoidal modeling framework. Fortunately we can implement a simple approximation that produces much the same result.

[3]This is only approximately correct. Technically we should be modulating by $m = g(f_1 - m)/2$, since after ring modulation the lower partial will be at frequency $f_1 - m$. The simpler approximation above overestimates $g(f_1 - m)$ by a maximum of about 1 Hz, which can be more or less safely ignored, especially at high frequencies.

Note that g changes very slowly with respect to f_1, and that modulating a partial with frequency f_1 at a frequency very close to $g(f_1)$ will produce very nearly as much perceived dissonance as modulating it at exactly $g(f_1)$ (see Figure 1). Because $g(f_1)$ changes so slowly, It turns out that for any modulation frequency m, there is a fairly large range of frequencies F_m such that for any $f \in F_m$, $d(f - m, f + m) \approx d(f, f + g(f))$.

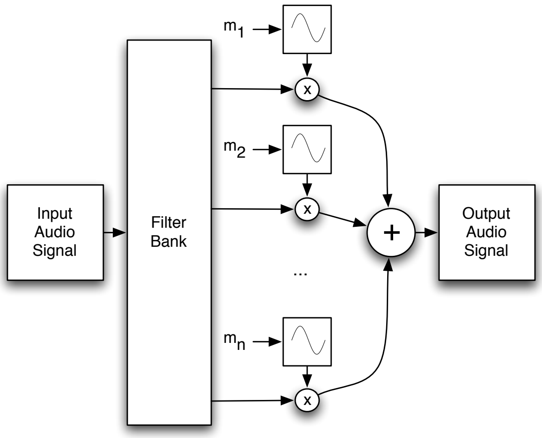

Figure 3: *Spectral ring modulation block diagram. The signal is broken into n subbands, each of which is modulated by a sine wave of frequency m_i chosen to maximize the roughness produced by that band. The bands are then recombined to produce the final output.*

If we use a filter bank to break the input signal into n low-bandwidth signals such that for each band the value of $g(f_{min})$ for partials at the bottom end of that band does not differ dramatically from $g(f_{max})$ for partials at the band's upper end, then ring modulating each of these bands separately will produce much the same perceptual result as the ideal scenario described above. For example, if we use a third-octave filter bank such as is commonly used in graphic equalizers and for each band set $m = 0.5g(0.2f_{min} + 0.8f_{max})$ where f_{min} and f_{max} are the band's lower and upper cutoffs, then equations (2) and (1) predict that modulating a partial with frequency f_i anywhere between f_{min} and f_{max} will be at least 90% of what would be produced by modulating it by the optimal frequency $g(f_i)/2$.

2.1. Modulating the Impact of the Effect

The discussion above has focused on how to produce as much roughness as possible using spectral ring modulation, but the ability to apply a lighter touch with this effect may be desirable. Let $p \in [0, 1]$ be the desired impact of the effect, from no impact to maximum impact. Then, instead of modulating each subband i by a sinusoid $\sin(m_i)$ modulate it by $a = (1 - p) + p \sin(m_i)$. This smoothly mixes the modulating sinusoid with a DC component, so that when $p = 0$ the subband output is unchanged, and when $p = 1$ the maximum roughness is produced. p can be smoothly varied in real time without producing artifacts, for example by an external controller to produce a "dissonance pedal" effect.

p can also be set separately for each subband, allowing the effect to be selectively applied only to the desired frequency ranges. For example, one might not want to apply the effect in the bass range, where it may produce a "muddy" sound.

3. SPECTRAL DYNAMICS EXPANSION

Rather than artificially introduce rapid amplitude modulation as described in section 2, we can also attempt to emphasize any such modulation already present in subbands of the signal. If the perception of roughness is caused by such modulation, then emphasizing AM should emphasize any existing roughness in the input, perhaps resulting in a more subtle effect than that produced by artificially ring modulating the signal.

Our fundamental approach is relatively simple, but effective. We first break the input signal into subbands using a Bark-spaced filter bank [10]. The output of each filter is sent to a spectral expander function described in Figure 4 and below.

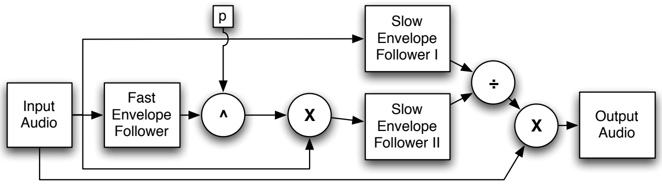

Figure 4: *Subband expander block diagram.*

First, an envelope follower responsive enough to detect high frequency amplitude modulation ("Fast Envelope Follower") tracks the envelope of the input band-limited signal. Its output is raised to the power p (a parameter controlling the strength of the effect), exaggerating the distance between the peaks and troughs of the envelope, and the result is multiplied by the original signal. This results in a version of the original signal with its envelope function raised to power $p + 1$. If $p = 0$, then the signal's dynamics are mostly unchanged. As was the case for the spectral ring modulation algorithm, this technique allows p to be smoothly varied in a real-time performance context.

We now need to normalize the overall signal amplitude back to its original amplitude, since we only want to impact the modulation depth, not overall amplitude. We do this using two envelope followers that cannot track the rapid beats that cause roughness, but follow the more slowly evolving envelope. The first ("Slow Envelope Follower I") tracks the amplitude of the original input signal, while the second ("Slow Envelope Follower II") tracks the overall amplitude of the altered signal. Dividing the output of the first slow envelope follower by the output of the second provides a normalizing signal that can be multiplied by the exaggerated signal to get an output signal with exaggerated high-frequency amplitude modulation whose longer-scale amplitude envelope is unaffected.

3.1. Filter Bank

The passbands of each filter in the filter bank should each allow a range of frequencies no wider than a critical bandwidth to prevent out-of-band signals from interfering with the envelope follower's ability to detect rapid amplitude modulation within critical bands. A Bark-scale filter bank [10] with channel boundaries at 20, 100, 200, 300, 400, 510, 630, 770, 920, 1080, 1270, 1480, 1720, 2000, 2320, 2700, 3150, 3700, 4400, 5300, 6400, 7700, 9500, 12000, 15500, and 22050 is ideally suited to the task. Using finer resolutions may or may not enhance the system's ability to emphasize rapid beating, depending on the input signal characteristics.

Note that, as with the spectral ring modulation in section 2, p can be set separately for each subband, allowing the effect to be applied selectively.

3.2. Envelope Follower Specifications

To extract our envelopes, we use full-wave rectification followed by a simple nonlinear one-pole filter. Each output sample of the envelope $e(t)$ is a weighted average of the absolute value of the input sample ($|x(t)|$ and the previous output sample $e(t-1)$:

$$e(t) = \begin{cases} |x(t)|, & |x(t)| > e(t-1) \\ c_d e(t-1) + (1-c_d)|x(t)|, & |x(t)| \le e(t-1) \end{cases} \quad (3)$$

The envelope responds instantly to rises in signal amplitude, but responds less quickly to drops in signal amplitude.

To determine how best to set the decay time c_f for the fast envelope follower for each band, we tested the envelope follower on pairs of beating sinusoids at the 24 Bark band centers. The first sine's frequency was set to the center frequency b_i of the ith band and the second's frequency was set to $b_i + g(b_i)$, the frequency predicted to produce maximum roughness. We then tested a range of values of c_f and chose the one that produced the highest correlation between the envelope follower's signal and the theoretical envelope of the signal ($2\cos(2\pi \frac{g(b_i)}{2} t)$). We found that, at a sample rate of 44100 Hz, the equation

$$c_f = 1 - 0.00083734\sqrt{b_i} \quad (4)$$

predicts the optimal decay coefficient quite well.

We want to choose a decay time c_s for the slow envelope followers for each band large enough to prevent it from responding to the beating that the fast envelope follower attempts to detect, but not so large that it becomes unresponsive to the larger-scale dynamics of the signal. Setting $c_s = c_f^{1/50}$) produces an envelope that can decay to 20% of its original amplitude in 1,050 samples (for $b_i = 15410$) to 9475 samples (for $b_i = 102$), and blocks 80% of any modulation at $g(b_i)/2$ for all bands.

4. DISSONANCE REDUCTION

An obvious question at this point is whether the algorithms outlined above can be used to *reduce* the sensory dissonance caused by a signal. Unfortunately, a simple scheme based on envelope followers such as the one above is poorly suited to this problem. Recall that the formula for two beating sinusoids $\sin(f_1 t) + \sin(f_2 t)$ is mathematically equivalent to $2\sin(\frac{f_1+f_2}{2}t)\cos(\frac{f_2-f_1}{2}t)$. The term that produces the rapid amplitude modulation, $\cos\frac{f_2-f_1}{2}t)$, changes sign twice every period, an effect that is difficult for our simple envelope follower (which begins by rectifying the signal) to track but is crucial to our perception. If we were to divide out the absolute value of this modulation signal, then instead of modulating by a sinusoid we would be modulating by a square wave, which would not reduce the perceived dissonance of the signal. (And would introduce undesired sidebands.) Furthermore, the situation becomes significantly more complicated if a subband contains signals more complex than a simple pair of sinusoids, since then several overlapping amplitude envelopes must be dealt with.

The ring modulation approach described above suggests an alternative approach to dissonance reduction – if we modulate subbands of our signal with lowpass-filtered noise instead of a sine wave, any pairs of sharp peaks in the spectrum will be smeared together, replacing rapidly beating partials with a single noisy resonance. In fact, this is not so different from what a traditional chorus effect does: adding a large number of detuned duplicates of the spectrum. This perspective suggests that one reason chorus effects are pleasing may be their ability to interfere with the synchronization of rapid beats that otherwise would produce sensory dissonance.

More sophisticated sinusoidal modeling approaches such as the one taken by Sethares [6] may provide the best framework for devising more subtle dissonance reduction effects.

5. EXAMPLES

We applied the ring modulation and spectral expansion effects to a recording of an organ sonata. Below are spectrograms of the results – the ring modulated signal has been clearly altered, but the changes to the spectrally expanded signal are more subtle.

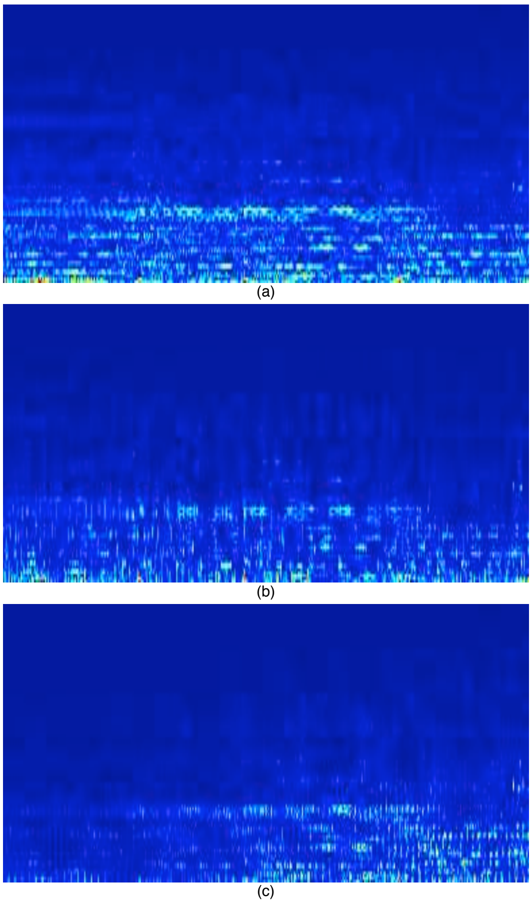

Figure 5: *Spectrograms of (a) the original one-second clip (b) the spectrally expanded clip and (c) the ring modulated clip.*

6. DISCUSSION

As hoped, these two effects accomplish the goal of increasing the roughness present in their input. The first approach based on ring modulation can induce even highly consonant chords to produce large amounts of sensory dissonance in the listener, although the result may sound fairly unnatural. The second approach based on spectral dynamics processing can produce more nuanced results, but relies on the presence of preexisting modulation. Both are well within the range of real-time implementations, and can be used as part of a performer's live signal chain.

Sound examples are available at:
`http://www.cs.princeton.edu/~mdhoffma/dafx08`.

7. REFERENCES

[1] Kameoka, A. and Kuriyagawa, M., "Consonance theory, part II: Consonance of complex tones and its computation method," *Journal of the Acoustical Society of America*, Vol. 45, No. 6, pp. 1460-1469, 1969.

[2] Leider, C. "Dissonance Theory of Sound Objects." Ph.D. diss., Princeton University, 2007.

[3] Huron, D., "Tone and Voice: A Derivation of the Rules of Voice-leading from Perceptual Principles," *Music Perception*, Vol. 19, No. 1, pp. 1-64, 2001.

[4] Plomp, R., and Levelt, W.J.M., "Tonal consonance and critical bandwidth," *Journal of the Acoustical Society of America*, Vol. 38, No. 4, pp. 548-560.

[5] Sethares, W.A., *Tuning Timbre Spectrum Scale*, Second Edition, Springer Verlag, 2004.

[6] Sethares, W.A., "Method and apparatus for dissonance modification of audio signals," U.S. Patent no. 5504270, April, 1996.

[7] Leman, M., "Visualization and Calculation of the Roughness of Acoustical Musical Signals Using the Synchronization Index Model (SIM)," *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-00)*, pp. 125-130, Verona, Italy, December, 2000.

[8] Misra, A., Cook, P.R., and Wang, Ge, "Musical Tapestry: Re-composing Natural Sounds," *Proceedings of the International Computer Music Conference (ICMC 2006)*, pp. 333-340, New Orleans, Louisiana, USA, November, 2006.

[9] Serra, X., and Smith, J., "Spectral Modeling Synthesis: A Sound Analysis/Synthesis Based on a Deterministic plus Stochastic Decomposition," *Computer Music Journal*, Vol.14, no. 4, pp. 12-24, 1990.

[10] Zwicker, E., "Subdivision of the audible frequency range into critical bands," *The Journal of the Acoustical Society of America*, 33, Feb., 1961.