Influence of Frequency Modulating Jitter on Higher Order Moments of Sound Residual with Applications to Synthesis and Classification.

Shlom Dubnov and Naftal Tishby
Institute for Computer Science and Center for Neural Computation Hebrew University, Jerusalem 91904, Israel
and
Dalia Cohen
Department of Musicology Hebrew University, Jerusalem 91904, Israel

Abstract
In this paper we provide a simple model for musical sounds that accounts for timbre properties due to microfluctuations in the harmonics of the signal. When considering a sound model that consists of an excitation signal passing through a resonator filter, we find, by means of higher order statistical analysis of the excitation, a grouping of sounds according to common instrumental families of string, woodwind and brass sounds. For resynthesis purposes we model the excitation by a family of stochastic pulse trains like functions whose statistical properties resemble those found in real sounds. By introducing an idea of "effective number of harmonics" that represents the number of coupled, or statistically dependent harmonics among the complete set of partials present in the signal, we show that this number can be calculated directly from the 3rd and 4th moments of the residual. Musically speaking we expect that microfluctuations administer a sense of texture within timbre and that these texture properties depend upon the concurrence/non concurrence parameter of the random frequency deviations caused by the jitter.

1 Introduction
The sense of timbre analysis of musical signals is extremely complicated due to the multiplicity of factors that compete on the perception of timbre. Various factors such as the formant structure, the waveform of the signal together with its spectral contents, many temporal features and others had been investigated in detail both from the technical aspects and with respect to their perceptual [ISSM95] and musical importance [Swenson][Wessel].

Signal models of sound usually describe the behavior of slowly time varying partials or model the gross spectral envelopes of resonant chameleons in musical instruments. Besides these macroscopic characteristics there are microscopic deviations of frequency that contribute to the timbre of sound. These deviations influence the perceived sound harmonicity, it's coherence and contribute to the sense of fusion/ segregation among partials [MeAdams][Sandell].

In this work we show that higher order statistics (HOS) analysis [Mendel][Nikias and Raghuveer][Dubnov et al., 1995b][Dubnov and Tishby 1996] when applied to a residual signal [Dubnov and Tishby96] are directly related to the number of coupled harmonics and that this number could be analytically calculated by considering the average amount of harmonicity apparent among triplets and larger groups of partials in the signal. When frequency of the harmonics (of a perfectly periodic sound source) are randomly disturbed by frequency modulation, the harmonicity relations among the partials are hindered and only these groups of partials which are subject to the same random modulation (i.e. having a concurrent random modulation) retain harmonicity. We believe that the "effective number" of harmonics is an acoustically important factor and we use this "harmonicity

1In many sound synthesizer programs the pitched input is created by a "hotta" generator which is a band limited version of a pulse train. In the following we shall create a stochastic version of the pulse train by applying a random frequency jitter to the harmonics and thus causing statistical independence among them.
3 Some real sound examples

Before going further into modeling of the excitation function we would like to demonstrate the bispectral signatures of several musical signals and of their respective residuals. In figure(2) we present the bispectra of residual signals for three musical instruments: Cello, Clarinet and Trumpet. Their original bispectra (i.e. before the inverse filtering operation for spectrum normalization) are shown under each plot respectively. The strong presence of the high harmonics in the residual significantly affect the bispectral content. Notice that Cello residual still has only a few peaks away from the origin.

Row do we look at these signals? First, we must be aware of the symmetries pertinent in the definition of bispectrum. In the six fold symmetry it is sufficient to consider a lower triangle part at the first quadrant only. Similarly, in the trispectrum, we shall consider only the lower tetrahedron in the positive octant of a three dimensional space.

In the following we shall consider the bispectra (trispectra) of residual signals (although it will not be possible to represent them graphically.) The residuals are not only properly normalized versions of the bispectrum that compensate for the effect of resonance spectral shape, but it also has the following important properties:

- the area (volume) obtained by integrating over the bispectral (trispectral) plane has a statistical
interpretation as a count of harmonicity between triplets (quadruplets) of harmonics.  

- the area (volume) equals to the moments of the signal and thus it can be easily calculated by taking time averages of the signal to the 3rd and 4th power.

As could be seen from the plots of the residual bispectrum, the overall area under the three graphs is significantly different.

Turning to real musical signals, we evaluate three moments by empirically calculating the skewness and kurtosis of various musical instrument sounds. These moments are calculated for a group of 18 instruments and they show a clear distinction between string, woodwind and brass sounds. Representing the sounds as coordinates in 'moments space' locates the instrumental groups on 'orbits' with various distances around the origin, very much according to the traditional orchestration handbook practice [Adler/Piston].

\[ x(t) = \sum_{n} \cos(2\pi f_0 n t + \text{Jitter}(t)) \]  
with \( f_0 \) being the fundamental frequency and \( Q \) the number of harmonics.

The statistical properties of this model are analyzed by calculating the third, fourth and possibly higher order moments of the signal, and specifically we will look at the skewness \( \gamma_3 = \mu_3 / \sigma^3 \) of the signal which is the ratio of the third order moment \( \mu_3 = E(x - Ex)^3 \) over the \( 3/2 \) power of the variance \( \sigma^2 = \mu_2 \) and kurtosis \( \gamma_4 = \mu_4 / \sigma^4 \) which is the variance normalized version of the fourth order moment \( \mu_4 = E(x - Ex)^4 \) [Grigoriu].

5 Influence of frequency modulating jitter on pulse train signal.

The influence of jitter upon higher order moments is considered by its effect on harmonicity betweenhar-
monic triplets (quadruples) of the signal partials.  

Bass sounds are on the perimeter. Strings are in the center.

The double integral amounts to the number of har- 
monic triplets since a contribution of order one is ob- 
tained for each harmonically related triplet. A similar 
evaluation is applicable for the fourth order moment 
and its respective triquad representation in the 
frequency domain.

5.1 Finding the Effective Number of Harmonics

Let us assume that the first $Q_{sij}$ partials of the 
signal ($n < Q_{sij}$) are subject to concurrent modula- 
tion jitter, while the partials above the threshold 
($n > Q_{sij}$) are modulated independently. In such a 

A similar, although more tricky argument for the triquadretrum reveals that the area of the tetrahedron limited by

$$0 < n < Q_{sij}, \quad 0 < m < Q_{sij}$$

$$0 < l < Q_{sij}, \quad n + m + l < Q_{sij}$$

equals to $Q_{sij}^2$. In the trispectrum case one must take 
into account also the number of possible choices of 
triplets, which gives a factor 3 to the above. An additive 
factor of $3Q^2$ appear also due to the fact that for 
$Q_{sij} = 0$ there are still peaks due to cancellations of 
frequencies on the diagonal planes $^6$.

$^6$ The assumption is based on empirical observations of "i'm-
sequence plane of real musical signals (such as those demon-
strated in figure (2)) that demonstrate stronger bispectrum =
low bispectrum and a decay in bispectral amplitudes for higher 
partial.

In the trispectrum expression we have the integrated ex-
pression $H(w) = H(w)H(w)H(w)$ which gives a $t$ 
function for the pair $(w_1, w_2)$, $w_3 = -w_1$, and there are three 
choices for such a pair.
Eventually, the normalization factor due to the power spectrum equals \( Q^{3/2} \) and \( Q^2 \) for the skewness and kurtosis expressions respectively. The resulting equations that relate the skewness \( \gamma_3 \) and kurtosis \( \gamma_4 \) to the effective number of coupled partials \( Q_{eff} \) are

\[
\gamma_3 = \frac{Q_{eff}^2}{Q^2}, \\
\gamma_4 = \frac{Q_{eff}^3}{Q^2} + 3
\]

5.2 Simulation results

This theoretical result was tested on synthetic signals that were created by combination of equal amplitude cosine function oscillators with random jitter added to the frequencies of the oscillators. The signal generators were implemented in Csound with the parameters set in accordance to the jitter synthesis method repeated by McAdams [McAdams]. The jitter depth was taken to be 0.01 of the partial frequency and the jitter spectrum was approximately shaped to have a ±0.5 dB cutoff at 30 Hz and a second cutoff to zero at 150 Hz. The signal were generated at a pitch of middle C and with a 16KHz sampling rate.

The following table compares the theoretical \( \gamma_3 \) and empirical \( \gamma_3 \) values for skewness and kurtosis for different \( Q_{eff} \).

<table>
<thead>
<tr>
<th>( Q_{eff} )</th>
<th>( \gamma_3 )</th>
<th>( \gamma_3 - 3 )</th>
<th>( \gamma_4 )</th>
<th>( \gamma_4 - 3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.065</td>
<td>-0.128</td>
<td>0.007</td>
<td>0.015</td>
</tr>
<tr>
<td>6</td>
<td>0.130</td>
<td>-0.028</td>
<td>0.109</td>
<td>0.120</td>
</tr>
<tr>
<td>10</td>
<td>0.413</td>
<td>-0.045</td>
<td>0.394</td>
<td>0.555</td>
</tr>
<tr>
<td>15</td>
<td>0.791</td>
<td>-1.407</td>
<td>0.685</td>
<td>1.875</td>
</tr>
<tr>
<td>20</td>
<td>1.037</td>
<td>-2.755</td>
<td>1.177</td>
<td>4.444</td>
</tr>
<tr>
<td>25</td>
<td>1.979</td>
<td>-5.790</td>
<td>1.981</td>
<td>8.880</td>
</tr>
<tr>
<td>30</td>
<td>2.559</td>
<td>-15.5</td>
<td>2.728</td>
<td>15.0</td>
</tr>
</tbody>
</table>

6 Musical Significance

This research, as we saw, focused on a specific phenomenon that contributes to timbre. The timbre, although being an extremely complex from the acoustic viewpoint, is perceived by the listener as an inapparent event. Nevertheless one can still notice, even inside the timbre, some microscopic happenings and their amplification will lead to border area between timbre (with a defined pitch) to noise and borders between timbre and texture. General verbal characterizations of sounds such as “focused”, “synthetic” versus “diffused”, “chorused” and etc. are caused by the same random fluctuations at the microspecic level. A more precise formulation of the phenomenon locates it on the axis between concurrence and non-concurrence with respect to the random deviations in frequencies of the harmonics. The principles behind this phenomenon: border areas; concurrence and non-concurrence; fusion/segregation; determination and uncertainty - are the basis of musical activity in all of its stages and in all levels of the musical material, even in characteristics of musical style. This research shows thus that the same principles we utilize for musical analysis in the “macro” level can be found in the “micro”. Putting this into a broad perspective one could state that the goal of this work are reciprocal: the above mentioned basic principles help us to understand the hidden microscopic phenomena and on the other hand, the research into these phenomena shed a new light on the principles. Moreover, these reciprocal relation are important also for musical creation in our days, where we have created an emphasis on the momentary events related to timbre and texture, instead of the interval parameters and its derived schemes that ruled the musical organization in toto music.
6.1 Concurrency /Non-Concurrency

This term refers to the relation among units and parameters. For instance, a perfect concurrency between parameters or pitch concurrent with increase in loudness. Non-concurrency has a plentitude of relations - it increases the complexity, the uncertainty and even creates a tension and as such becomes an essential parameter in the rules of musical organisation and characterising of style (some of the contrary rules of Pyramid refer to the prevention [Cohen71] of non-concurrency and this accordingly to the stylistic ideal of the era. On the other hand, in the music of Bach we find revelation of non-concernences of many types).

Here we have treated concurrency and non-concurrency among particles with respect to their deformations in frequency.

6.2 Texture and the border areas between the interval, texture and timbre.

In contrast to timbre and especially in contrast to the interval the research on texture is scarce, although many contemporary composers refer to it [Cohen and Dubnov]. In tonal music texture appears mainly as an aid that may support or contradict the interval organisation while in our days it has an existence of its own. Actually, most of the notation systems these days refer to texture phenomena. Without going into details of texture classification we shall note that the main difference between texture and timbre is that the texture is separable and usually relates to time-scales that are larger then those of timbre which can be identified for durations of less then 20 notes during which it remains inseparable to the listener. In comparison, texture must contain some sort of separability in the various dimensions - time, frequency or intensity. In extreme cases where we are no longer able to separate the simultaneous occurrence into its components, the texture becomes timbre. Also for the opposite case, when we sense the changes that occur in timbre, timbre becomes closer to texture. There exists then a grey area in the border between texture and timbre and there is a similar border area between pitch (interval) and texture. This applies to wide range of other musical phenomena such as noises of intonation [Cohen69], "articulatory ornamentations" in non western music and random modulations in electronic music [Penny and Polansky].

7 Conclusion

In this paper we presented an analysis-classification-synthesis scheme for instrumental musical sounds. Specifically we focused on the microfluctuations that occur during a sustained portion of single tones and we have shown that an important parameter in the characterisation of microfluctuations is the "effective number" (Qeff) of coupled harmonics that exists in the sound. For modeling, simulation and resynthesis purpose the coupling was realised by application of concurrent frequency modulating jitter to first Qeff partials and non concurrent jitter to the others. We present an analytic formula that relates the higher order moments (actually the skewness and kurtosis) of the sound to the number of coupled harmonics. The classification results locate the sounds in instrumental families of string, woodwind and brass sounds. This is graphically seen using a cumulant space representation where the groups appear on different 'orbits'. The closer the 'orbit' is to the center, the more gaussian is the signal and the greater is the number of non concurrently modulated harmonics that do not contribute to the moments and draw such a signal towards gaussianity.

Although we have used a stochastic version of pulse train, we shall note also that the above considerations are not limited to symmetrical pulse train like signals. Actually, any combinations of sine and cosine functions with equal amplitudes are appropriate for this kind of analysis. The reason that we were looking at kurtosis was that for symmetrical signals, the third moment vanishes, and in real condition the harmony counts are better accomplished by looking at groups of four partials, or equivalently, at the fourth order moment. We note also that we are dealing with stationary sounds only and neglect any non stationary or transient phenomena which could not be considered as microscopic stationary fluctuations at the sustained portion of a sound.

Appendix: Gaussianity of Signal Statistics

Before proceeding to deal with the influence of jitter on a perfectly periodic sound we would like to consider briefly the statistical properties of non harmonic pitch signals and show that their statistics approach Gaussianity for large number of partials. Given a signal \( x(t) = \sum_{i=1}^{N} a_i \cos(\omega_i t) \), the second order time averaged correlation is

\[
< x(t) x^*(t+\tau) > = < \left( \sum_{i=1}^{N} a_i e^{i\omega_i t} \right) \left( \sum_{i=1}^{N} a_i e^{i\omega_i(t+\tau)} \right) > (7)
\]
which equals $Q$ for $\tau = 0$ and is zero for harmonically related $\omega_i$'s ($\omega_i = \omega_j \pm \omega_k$), but generally is non-zero for an arbitrary set of $\omega_i$'s. Thus, second order statistics are non-zero for both harmonic and non-harmonic sounds. The third order correlation, though, is extremely sensitive to the existence of harmonic relations since

$$c(\tau) = c(\tau + \tau_1) = c(\tau + \tau_2)$$

and the bracketed integral expression vanishes for non-harmonic signals since $\omega_1 + \omega_2 = \omega_3$ never occurs.

The vanishing of high order correlations means that the signal statistics are Gaussian, which is easily demonstrated for $\tau = 0$ by looking at the histograms of harmonically and non-harmonically related signals.

![Figure 7: Harmonic signal (left) and its histogram (right)](image)

![Figure 8: Inharmonic signal (left) and its respective histogram (right)](image)

In mixed harmonic/non-harmonic set of frequencies $\omega_i$, the third order moment equals to the effective number of harmonic triplets found in the sounds spectrum.

Bibliography


[ISSM95], Special session on Sound (tremb) in Ex-


