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

Shlomo Dubnov and Naftali 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 train like functions whose statistical properties resemble those found in real signals. 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 form the 3rd and 4th moments of the residual. Musically speaking we suggest that microfluctuations administer a sense of texture within timbre and these texture properties depend upon the concurrence/non concurrence parameter of the random frequency deviations caused by the jitter.

1 Introduction

The issue 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 [Slawson][Wessel].

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

In this work we show that higher order statistics (HOS) analysis [Mendel][Nikias and Raghuveer][Dubnov et al., 1995b][Dubnov and Tishby 1994] 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 frequencies 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 those 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 [1] and we use this "harmonicity

---

[1] In many sound synthesis programs the pitched input is created by a "buzz" generator which is a band limited version of a pulse train. In the following we shell create a stochastic version of the pulse train by applying a random frequency jitter to the harmonics and thus causing statistical independence among them.

counting" property of HOS for pulse train like signals to investigate the influence of jitter on timbre properties of sound.

2   Finding the excitation

Given a signal, we suggest that the next step beyond analyzing the spectral amplitude distribution characterized by the filter, one should look at the properties of the inversely filtered result, or the so called *residual*.
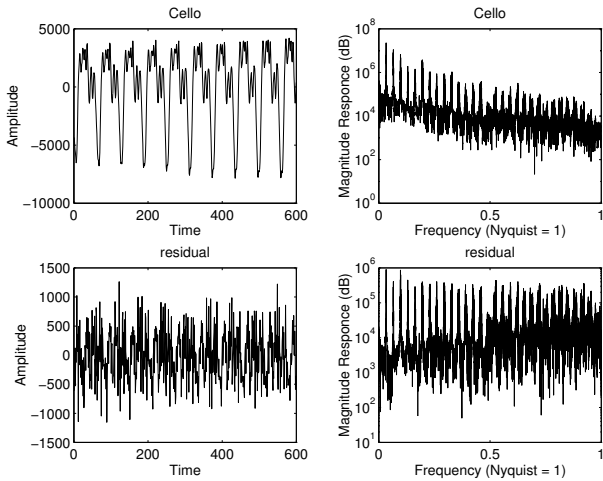

Figure 1: The original Cello signal and its spectrum (Top). The residual signal and its respective spectrum (Bottom). Notice that all harmonics are present and they have almost equal amplitudes, very much like the spectrum of an ideal pulse train. The time domain signal of the residual does not resemble pulse train at all.

The effect of the spectral envelope, which contains the information about the amplitudes of harmonics is removed by inverse filtering of the signal by a filter derived from its lpc model. This action is taken in order to consider only the effect of frequency deviation caused by the jitter upon the excitation signal (residual error) and statistically it amounts to low order decorrelation of the signal. It is interesting to note that investigation of the moments of decorrelated signals was widely used in the analysis of texture in images [Faugeras][Tsatsanis]. In the acoustic case we obtain a statistical interpretation of the moments as probabilities for maintaining harmonicity among groups of partials, that is for the case of pulse train like signals with frequency modulating jitter applied to its partials.

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 plots respectively. The strong presence of the high harmonics in the residual significantly effect the bispectral contents. Notice that Cello residual still has only a few peaks away from the origin.
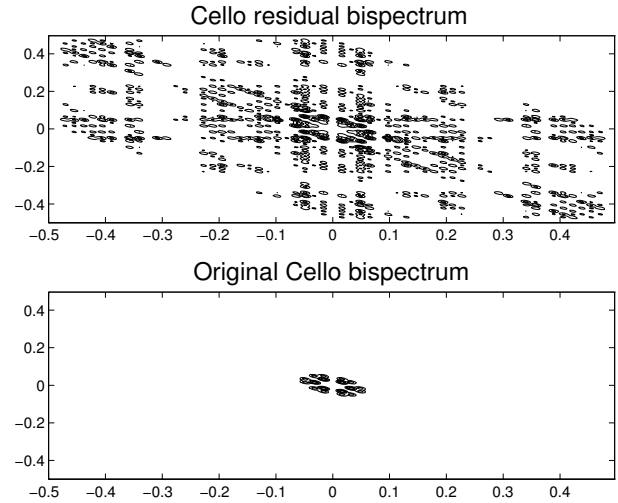

Figure 2: Bispectrum of a Cello residual signal (Top) and the bispectrum of the original Cello sound (Bottom). See text for more details.

How 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 triangular part at the first quadrant only. Similarly, in the trispectrum, we shell 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
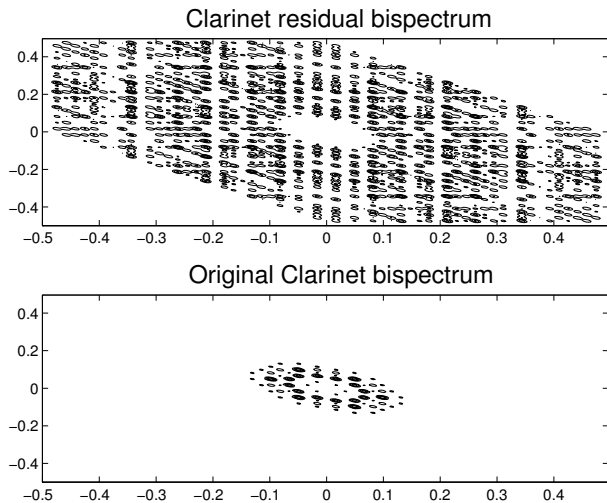
Figure 3: Bispectrum of a Clarinet residual signal (Top) and the bispectrum of the original Clarinet sound (Bottom). See text for more details.

interpretation as a count of harmonicity between triplets (quadruples) 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 [2].

Turning to real musical signals, we evaluate these 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].

---

[2]Briefly we should mention that a common goal in a series of our works was to define a function that would sensibly measure the distance between musical signals [Gray] based on the bispectral information [Dubnov et al. 1995a]. In the current work we attempt to use the bispectral information for resynthesis also.
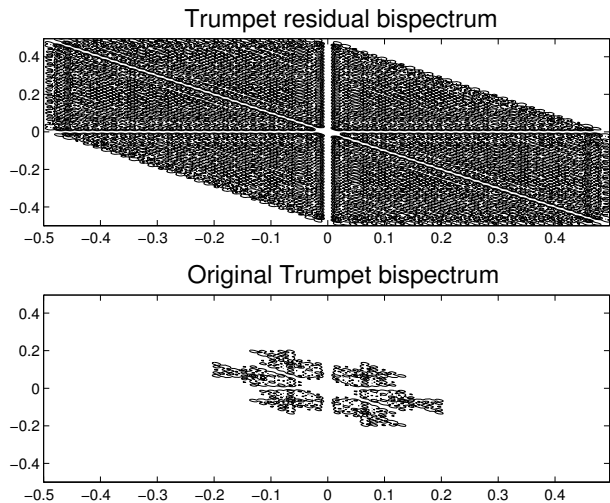

Figure 4: Bispectrum of a Trumpet residual signal (Top) and the bispectrum of the original Trumpet sound (Bottom). See text for more details.

4   Stochastic pulse train model

In the ideal case, the residual is supposed to be a low amplitude gaussian noise, with regularly spaced peaks due to the pitch of the source signal, and is totally characterized by its variance. We assume that instead of the ideal pulse train, we have a sinusoidal model approximation which consist of a sum of equal amplitude cosines, with a random jitter applied to its harmonics.

$$x(t) = \sum_{n=1}^{Q} cos(2\pi f_0 n \cdot t + Jitter(t)) \qquad (1)$$

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* $\gamma_3 = m_3/\sigma^3$ of the signal which is the ratio of the third order moment $m_3 = E(x - Ex)^3$ over the 3/2 power of the variance $\sigma^2 = m_2$ and *kurtosis* $\gamma_4 = m_4/\sigma^4$ which is the variance normalized version of the fourth order moment $m_4 = E(x - Ex)^4$ [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 between har-
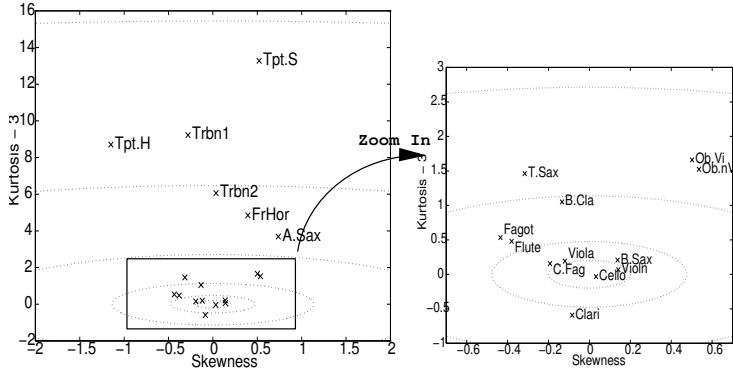
Figure 5: Location of sounds in the 3rd and 4th normalized moments plane. The value 3 is subtracted from the kurtosis so that the origin would correspond to a perfect gaussian signal. Brass sounds are on the perimeter. Strings are in the center.

monic triplets (quadruples) of the signal partials. Basically, the application of frequency modulating jitter to harmonically related partials destroys the harmonicity in case when the jitters are non concurrent (independent) at each partial. Harmonicity is preserved on the contrary in the case when the same jitter (concurrent modulation) is applied to the partials. The vanishing of signal moments indicated that the signal obeys Gaussian statistics. The relations between Gaussianity and harmonicity is discussed at length at the appendix.

Designating the deviation in frequencies by $\Delta_n = \omega_0 n r \xi_n$ where $\xi_n = Mod_n(t)$ is an uniformly distributed random variable between $[-1, 1]$, with r being the modulation depth and n the partial number, we rewrite our stochastic pulse train model

$$x(t) = \sum_{n=1}^{Q} cos((\omega_0 n + \Delta_{i,n}) \cdot t) \qquad (2)$$

.

The extent to which the random jitter causes degradation in harmonicity of the signal is evaluated by counting the number of triplets (quadruples) of partials that retain harmonic relations after the application of jitter.

This count is accomplished by measuring the third (fourth) order moment of the signal

$$m_3 \quad = \frac{1}{T} \int_0^T x^3(t)dt \qquad (3)$$
$$= \quad \frac{1}{(2\pi)^2} \int_{-Q}^{Q\omega_0} \int_{Q}^{Q\omega_0} X_i(\omega)X_i(\omega')X_i^*(\omega + \omega')d\omega d\omega'$$

$$\approx \quad \int\int \sum_{n=1}^{Q} \sum_{m=1}^{Q} \delta(\omega - (n\omega_0 + \Delta_{n,i})) \cdot$$
$$\delta(\omega' - (m\omega_0 + \Delta_{m,i})) \cdot$$
$$\delta((\omega + \omega') - ((n + m)\omega_0 + \Delta_{(n+m),i}))d\omega d\omega'$$

This double integral amounts to the number of harmonic triplets since a contribution of order one is obtained for each harmonically related triplet. A similar evaluation is applicable for the fourth order moment and its respective trispectrum representation in the frequency domain.

5.1 Finding the Effective Number of harmonics

Let us assume that the first Q_{eff} partials of the signal ($n < Q_{eff}$) are subject to concurrent modulation jitter, while the partials above the threshold ($n > Q_{eff}$) are modulated independently. In such a case only partials below Q_{eff} contribute to the HOS [3].

The theoretical calculation of the *skewness* and *kurtosis* is based upon a counting argument for the total number of peaks in the bispectral and trispectral planes that occur due to partial numbers below Q_{eff}. For the bispectral case a lattice of delta functions exists for partials (n, m) over the bifrequency triangle

$$0 < n < Q_{eff}, \quad 0 < m < Q_{eff}, \quad n + m < Q_{eff} \quad (4)$$

in the positive quadrant of the bispectral plane. The area (number of peaks) of this region equals $\frac{1}{2}Q_{eff}^2$.

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

$$0 < n < Q_{eff}, \; 0 < m < Q_{eff}, \qquad (5)$$
$$0 < l < Q_{eff}, \; n + m + l < Q_{eff}$$

equals to $\frac{1}{6}Q_{eff}^3$. In the trispectral 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 $Q_{eff} = 0$ there are still peaks due to cancellations of frequencies on the diagonal planes [4].

---

[3]This assumption is based on empirical observations of bispectrum plots of real musical signals (such as those demonstrated in figure (2)) that demonstrate stronger bispectrum at low bifrequencies and a decay in bispectral amplitude for higher partials

[4]In the trispectrum expression we have the integrand expression $H(\omega_1)H(\omega_2)H(\omega_3)H^*(\omega_1 + \omega_2 + \omega_3)$ which gives a δ function for the pair (ω_1, ω_2), $\omega_1 = -\omega_2$, and there are three choices for such a pair

Eventually, the normalization factor due to the powerspectrum equals $Q^{3/2}$ and Q^2 for the skewness and kurtosis expressions respectively. The resulting equations that relate the skewness γ_3 and kurtosis γ_4 to the effective number of coupled partials Q_{eff} are

$$\gamma_3 \quad = \quad \frac{\frac{1}{2}Q_{eff}^2}{Q^{3/2}} \qquad (6)$$

$$\gamma_4 \quad = \quad \frac{\frac{1}{2}Q_{eff}^3}{Q^2} + 3$$

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 applied to the frequencies of the oscillators. The signal generators were implemented in csound with the parameters set in accordance to the jitter synthesis method reported 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 -10 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 working with 16KHz sampling rate this gives us the total of 30 harmonics (Q=30).
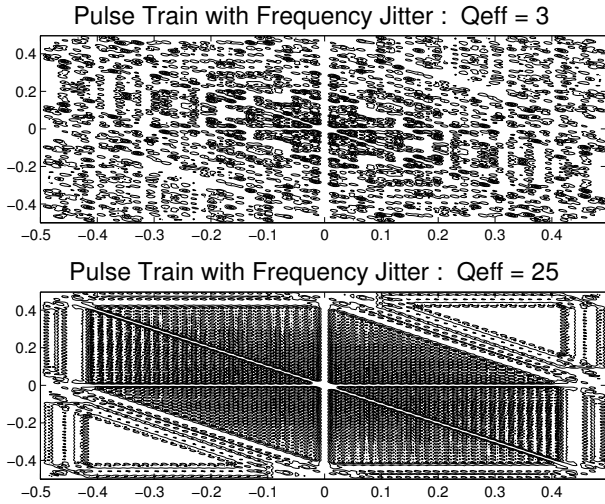

Figure 6: Bispectra of two synthetic pulse train signals with frequency modulating jitter. Top: Qeff = 3. Bottom: Qeff = 25. Notice the resemblance of the two figures to the bispectral plots of the cello and trumpet in figure (2). The Qeff values were chosen especially to fit these instruments according to the skewness and kurtosis values for the instrument in figure (3).

The following table compares the theoretical γ_i and empirical $\hat{\gamma}_i$ values for skewness and kurtosis for different Q_{eff}'s.

Skewness and Kurtosis

Q_{eff}	$\hat{\gamma}_3$	$\hat{\gamma}_4 - 3$	γ_3	$\gamma_4 - 3$
3	0.055	-0.128	0.027	0.015
6	0.150	0.113	0.109	0.120
10	0.413	0.345	0.304	0.555
15	0.731	1.407	0.685	1.875
20	1.037	2.755	1.217	4.444
25	1.579	5.790	1.902	8.680
30	2.859	15.5	2.738	15.0

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 inseparable 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 border between timbre and texture. General verbal characterizations of sounds such as "focused", "synthetic" versus "diffused", "chorused" and etc. are caused by the very same random fluctuations at the microscopic level. A more precise formulation of the phenomenon locates it on the axis between concurrence and nonconcurrence with respect to the random deviations in frequencies of the harmonics. The principles behind this phenomenon : border areas; concurrence and nonconcurrence; fusion/segregation; determinism and uncertainty - are at 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 parameter and its derived schemes that ruled the musical organization in tonal music.

6.1 Concurrence /Non-Concurrence

This term refers to the relation among units and parameters. For instance a perfect concurrence between parameters of pitch and intensity occurs when both change at the same time and with similar trends (such as ascent in pitch concurrent with increase in loudness). Non-concurrence has a plentitude of revelations - it increases the complexity, the uncertainty and even creates a tension and as such becomes an essential parameter in the rules of musical organization and characterizing of style (some of the counterpoint rules of Palestrina refer to the prevention [Cohen71] of non-concurrence and this accordingly to the stylistic ideal of the era. On the other hand, in the music of Bach we find revelation of non-concurrencies of many types).

Here we have treated concurrence and non-concurrence among partials with respect to their defections 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 organization 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 shell 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 msec. 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 occurrences 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 nuances of intonation [Cohen69], "articulatory ornamentations" in non western music and random modulations in electronic music [Tenney 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 characterization of microfluctuations is the "effective number" (Q_{eff}) of coupled harmonics that exists in the sound. For modeling, simulation and resynthesis purpose the coupling was realized by application of concurrent frequency modulating jitter to first Q_{eff} 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 shell 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 harmonicity 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 transitory 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 harmonics pitched signals and show that their statistics approach Gaussianity for large number of partials.
Given a signal $x(t) = \sum_{j=1}^{Q} e^{i\omega_j t}$, the second order time averaged correlation is

$$< x(t) \cdot x^*(t + \tau) > = < (\sum_{j=1}^{Q} e^{i\omega_j t})(\sum_{k=1}^{Q} e^{i\omega_k t}) > \quad (7)$$

$$= lim_{\Theta \to \infty} \; \frac{1}{2\Theta} \sum_{j,k=1}^{Q} (\int_{-\Theta}^{\Theta} e^{i(\omega_j - \omega_k)t} dt) e^{-i\omega_j \tau}$$

$$= \sum_{j=1}^{Q} e^{-i\omega_j \tau}$$

which equals Q for $\tau = 0$ and is zero for harmonically related ω_i's ($\omega_i = i \cdot \omega_0$), but generally is non zero for an arbitrary set of ω_i's. Thus, second order statistics are non zero for both harmonic and non harmonic sounds. The third order correlations though are extremely sensitive to the existence of harmonic relations since

$$< x(t)x(t + \tau_1)x^*(t + \tau_2) >= \qquad (8)$$

$$\lim_{\Theta \to \infty} \; \frac{1}{2\Theta} \sum_{j,k,l=1}^{Q} (\int_{-\Theta}^{\Theta} e^{i(\omega_j + \omega_k - \omega_l)t} dt)$$

$$\cdot \; e^{-i\omega_k \tau_1} e^{-i\omega_l \tau_2}$$

and the bracketed integral expression vanishes for non-harmonic signals since $\omega_j + \omega_k = \omega_l$ 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.
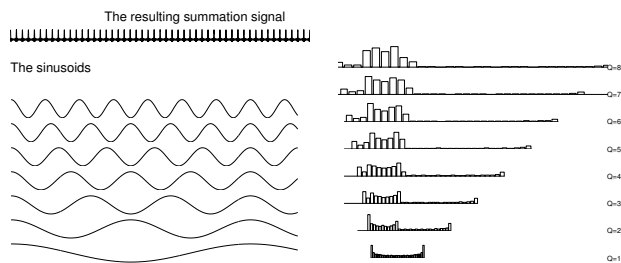

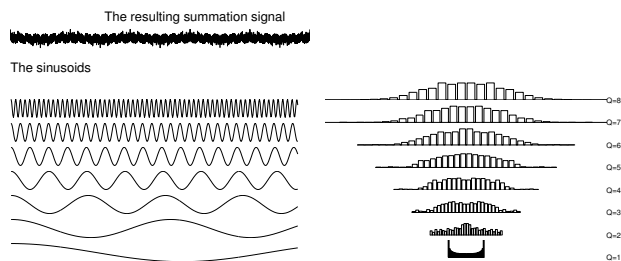Figure 7: Harmonic signal (left) and its histogram (right).

Figure 8: Inharmonic signal (left) and its respective histogram (right).

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

Bibliography

[Adler] S.Adler, *The Study of Orchestration*, Norton and Co., 1989.

[Cohen69] D. Cohen, *Patterns and Frameworks of Intonation*, Journal of Music Theory, 1969.

[Cohen71], D.Cohen, *Palestrina Counterpoint - a musical expression of unexited speech*, Journal of Music Theory, Vol.15, 8-111, 1971.

[Cohen and Dubnov] D.Cohen, S.Dubnov, *Difference and Similarity in Texture: Characterization of Types of Texture as One Manifestation of the Broad Definition of Timbre*, Third International Symposium on Systematic Musicology, Wien, 1995.

[Dubnov and Tishby 1996] S.Dubnov, N.Tishby *Testing for Non linearity and Gaussianity in sustained portion of musical signals*, to be published in the Proceedings of the Journees d'Informatique Musicale, Caen, 1996.

[Dubnov et al., 1995a] S.Dubnov, N.Tishby, D.Cohen, *Clustering of Musical Sounds using Polyspectral Distance Measures*, Proceeding of the International Computer Music Conference, Banff, 1995.

[Dubnov et al., 1995b] S.Dubnov, N,Tishby, D.Cohen, *Hearing beyond the spectrum*, Journal of New Music Research, Vol. 24, No. 4, 1995.

[Dubnov and Tishby 1994] S.Dubnov, N.Tishby, *Spectral Estimation using Higher Order Statistics*, Proceedings of the 12th International Conference on Pattern Recognition, Jerusalem, Israel, 1994.

[Faugeras] O.D.Faugeras, *Decorrelation Mathods of Feature Extraction*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 2, No.4, July 1980.

[Gray] A.H.Gray, J.D.Markel, *Distance Measures for Speech Processing*, IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 24, No.5, October 1976

[Grigoriu] M. Grigoriu, *Applied Non-Gaussian Processes*, Prentice-Hall, 1995.

[ISSM95], Special session on *Sound (timbre) in Eu-*

ropean and non-European Music, Third International Symposium on Systematic Musicology, Wien, 1995.

[McAdams] S. McAdams, *Spectral Fusion, Spectral parsing and the Formation of Auditory Images*, Ph.D. dissertation, Stanford University, CCRMA Report no. STAN-M-22, Stanford, CA., 1984.

[Mendel] J.M. Mendel, *Tutorial on Higher-Order Statistics (Spectra) in Signal Processing and System Theory*, Proceedings of the IEEE, Vol. 79, No. 3, July 1991

[Nikias and Raghuveer], C.L. Nikias, M.R. Raghuveer, *Bispectrum Estimation: A Digital Signal Processing Framework*, Proceedings of the IEEE, Vol. 75, No. 7, July 1987

[Piston] W.Piston, *Orchestration*, Gollancz Ltd., 1989.

[Sandell] G.Sandell, *Concurrent Timbres in Orchestration: A Perceptual Study of Factors Determining "Blend"*, Ph.D. thesis, Northwesten University, 1991.

[Slawson] A.W. Slawson, *Sound Color*, Berkely, CA, University of California Press, 1985.

[Tenney and Polansky] J.Tenney, L.Polansky, *Temporal Gestalt Perception in Music*, Journal of Music Theory 24, 205-41, 1980.

[Tsatsanis] M.K.Tsatsanis, *Object and Texture Classification Using Higher Order Statistics*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 14, No.7, July 1992.

[Wessel] D.Wessel, *Timbre as a Musical Control Structure*, Computer Music Journal 3, No.2, 1979.