Polyspectral Analysis of Musical Timbre

A thesis submitted in fulfillment
of the requirements for the degree of
Doctor of Philosophy

Shlomo Dubnov

Submitted to the Senate of the Hebrew University in the year 1996

This work was carried out under the supervision of

Prof. Naftali Tishby and Prof. Dalia Cohen

Acknowledgments

Special thanks are due to:

- Prof. Naftali Tishby and Prof. Dalia Cohen for their help and guidance in carrying out this study.
- Yona Golan for discussions of clustering and for providing some of the routines.
- Itay Gat, Lidror Troyansky, Elad Schneidman and the rest of the members of the lab, for their help and support for the study.
- the late Michael Gerzon and Meir Shaashua for suggesting the subject and providing with many of the original ideas.
- Steve McAdams and David Wessel, for their revision and discussion of parts of this study.
- Eshkol Fellowship of the Ministry of Sceineces and Art for partial support of this study.
- My family.
- My parents.

Contents

List of Figures

Abstract

This study is interdisciplinary in its nature, combining problems in musical research with new techniques of signal and information processing. In the narrow sense, the subject of the research is investigation of a phenomenon of aperiodicities in the waveform or fluctuations (jitter) in frequency that occur during the stable (sustained) portion of the sound. This phenomenon, which is not related to one of the common timbre characteristics such as the spectral envelope, time envelope and etc., has scarcely been investigated in the acoustical domain and its importance for the definition of timbre is now being recognized. The fluctuations in instrumental sounds occur for time scale shorter then 100-200 ms and are beyond the control of the player. Perceptually they bring a sense of separability which, according to the definitions we give in the study, puts the problem in the border area between timbre and texture.

Additional important property is that the fluctuations are strongly related to non-linearities of the sound generating mechanism. Since most of the modeling, analysis and synthesis of sounds is based on linear modeling, it basically uses spectral information or second order statistics (autocorrelation) of the signal. The linearity property obeys the superposition principle - the signal equals to a sum of its components - which from statistical stand point amounts to independence between the spectral components. The linear approach is thus insufficient for the description of the above acoustic phenomena, especially since one of the major properties of the fluctuations that we are about to discuss is the statistical dependence/independence among overtones, or in other words concurrence/ non-concurrence of the jitter.

By application of higher order statistical (polyspectral) methods, we analyze and model these phenomena so as to reveal the properties of the frequency modulating jitter and construct acoustical distance measures that allow for comparison between signals based on the fluctuation/non-linearity properties of sounds. The main sound sources that were examined in this study are timbres of traditional acoustic musical instruments, with some extensions done towards analysis and classification of natural and man-made sounds. From a broader perspective, the method of analysis and its results are significant in many additional musical domains.

The main result of the work is in the correspondence between HOS features and sound timbre properties, specifically the classification into instrument families, that we present in this study. This and other results prove the applicability of our method for the investigation of real acoustical problems and provides an analytical method for the investigation of these acoustic phenomena. Another contribution is in the use of statistical distance measures for assessing the difference between sounds, which is new in musical applications. Combined with new polyspectral source and filter parameter estimation methods, a generalization to existing spectral distance measures is developed. An important correspondence between non-linearity and non-Gaussianity tests of time series analysis and polyspectral contents of a sound is shown. This result is significant for understanding the relationship between models and their acoustic features. Moreover, a simple synthesis method which reconstructs the bispectral and trispectral properties of signals by controlling the concurrence/non-concurrence among jitters of the harmonics is presented. Finally, extensions to more general revelations of timbre and texture, with application to broader class of signals and higher level musical problems are suggested. [1]

[1] Parts of this work were published with some changes in several papers. Please see the bibliography

Part I

Introduction

Chapter 1

The Musical Problem of Timbre and Sound Texture

1.1 Timbre and Sound Texture

Timbre is one of four basic psychoacoustical parameters, but in contrast to the other three - pitch, duration and intensity - which can be relatively simply defined, hierarchically organized and notated, timbre has a very complex definition, it can not be put on a scale and its definition in fact has never been completed. The definition is basically multidimensional (time envelope, spectral envelope, etc.) and here we would like to add another dimension which could be regarded as a border area between timbre and texture.

In the most general sense, timbre can be defined negatively: it is the musical parameter which describes the properties of a sound which are not one of the following: pitch, intensity and duration. This definition, based on the negation of the common musical parameters, [1] has many realisations (it is mainly known as the perceptual property that distinguishes between the sounds of different musical instruments) and it has been given various names such as "sound color", "tone quality", "tone colour" (Slawson 1985, pp.18-20)[123],(Erickson 1975, p.67)[43].

Untill the 20th century, most of the musical theories related to music parameters other than timbre (except for practical orchestration treatises). Since the 19th century the role of the parameter of timbre has gradually increased, reaching a climax in the works of Claude Debussy, who uses the chords more for their timbre than for their harmonic function. [2]

[1] These parameters are the basis for higher musical constructs and learned schemes on which the various musical styles are based in various cultures and different periods, till the 20-th century. Musically, the timbre and the interval are complementary parameters, as we shall see in section (2.3).

[2] An extreme counter example of complete lack of reference to timbre is the "Art of the fugue" by J.S.Bach. The work is a perfect construct of learned schemes/intervals, with no indication about the type of instruments or instrumental writing.

In the 20th century a great deal of the modern music is based on timbre, as well as on *texture* and on various border areas between the two. (One can also speak about border areas between texture and interval. A detailed discussion will be presented in chapter 8.)

Texture is a hyperparameter which relates to a statistical description of the organization of other musical parameters. A more concise definition will be given later. In the meantime it is sufficient to state that we define musical texture as *the principle of organization of sound which is not derived from learned schemes*, [3] and as such it requires a probabilistic approach for its description. The most salient difference between texture and timbre is the perceptual non separability of timbre versus the separability of texture into several sources and segmentation in time.

In the border case, when we have difficulty identifying texture constituents or performing segmentation in time,[4] or when we perceive some internal changes within timbre, an intermediate case between simple timbre and complex organized sound occurs.

Today, in electroacoustic, computer music and in audio applications in general, the basic sound phenomena, or the new "building blocks" for musical composition are defined in a much broader sense than the timbre in the classical period, which related mainly to musical instruments, and it encompasses sound events such as complex musical, environmental or man-made sounds. These sounds are still largely regarded and described as phenomena of timbre.

One could investigate timbre from many aspects: as a maker of musical instruments, as a researcher interested in psychoacoustical and cognitive activity, as a musician interested in timbre oriented composition and many more. Today the focus on timbre is for one of the following reasons:

1. Timbre occupies a central place in contemporary musical composition.
2. Appropriate tools exist for its research and generation.
3. The research on timbre is a part of our general desire to deepen the understanding of music phenomena as a cognitive activity and investigate the role of each constituent in defining the various musical styles.

Due to the complexity of the musical parameter of timbre, the main problems, in our opinion, are the control of timbral properties both for analysis and as a means of its production, with a strong emphasis on the question of "similarity and difference", which is essential for any form of organization and stands at the basis of every cognitive activity (Tversky 1977)[132].

In this work we will extend the research on timbre by the use of a new analytical method - "Higher Order Statistics" (HOS) or "Polyspectra" - and by extending the field of timbre

[3]By learned schemes we mean musical knowledge expressed by the rules of music theory such as harmony, counterpoint and etc., more readily described by rules of syntax and other Artificial Intelligence techniques.

[4]This situation exists due to psychoacoustic constrains.

phenomena to include border areas with timbre. Thus, we have chosen to address the issue mainly from the signal processing point of view, regarding the acoustic signal as a stochastic process and suggesting a new approach for its treatment. As we shall see, the method and the results have interesting implications for musical research and understanding.

1.2 Existing Works on Timbre

1.2.1 Historical Perspective

The research on timbre is relatively young, but something about it, such as the existence of the second harmonic, was known already in Ancient Greece. The existence of harmonics was first reported clearly by the theoretician Mersenne (1636) [90]. He heard at least five partials after the fundamental frequency has decayed, but his explanation was obscure. A thorough explanation of the phenomenon was provided by the French researcher Sauveur (1702) [116],(Scherchen 1950)[117]. He was the first to understand properly the phenomenon of overtones and coined the terms "overtones" and "fundamental frequency". The first book on harmony by Rameau [105] was based on Sauveur's theory; he attempted to explain the rules of harmony based on natural phenomena, as they were understood at the time. As we've mentioned earlier, the research on timbre has advanced mostly in the 20th century. We can not present a broad historical review, but we shall at least stress the importance of research and theory on the shaping of musical practice and musical style.

Many modern studies focus on specific problems and parameters, with the purpose of finding good mathematical models for describing timbre and defining its psychoacoustical dimensions, followed by attempts to find some organizational principles for timbre [73] and to devise methods of synthesis and transformation.

The specific method for analysis that we use in this work - higher order statistics (polyspectra) - has appeared in the statistical and signal processing literature, but it has hardly received any attention in the studies of acoustics and music. The basic issue here is the focus on the steady state of the sound while extending the field to deal with border areas between timbre and texture, and adding a new dimension for the analysis of timbre.

1.2.2 State of the Art in Sound Representation

Our understanding of the world strongly depends on the representation of its objects. The inability to find a satisfactory descriptive vocabulary for sounds or to relate verbal attributes to physical sound parameters had been a central difficulty in sound information handling. Finding high-level means for describing sound holds the promise for arranging on some scale, having means for organization, structuring and handling of sound and much more. The prospect of defining a "representation space" for sound is thus a fairly obvious and attractive idea and indeed several studies have attempted to do so. Most of them have focused on the representation of sounds of musical instruments, while almost no attention has been paid to more complex natural or man made sounds. An important study of differences between musical instruments, based onidentification by human subjects, was done by J.Grey (1977)[55], who suggested a multidimensional representation based on spectral energy distribution, dynamic attributes in the onset portion of the tone and presence of high frequency noise during the early part of the sound. It appears, though, that these dimensions are not enough to characterize all the perceivable differences in sound timbres.

In contrast to these studies of known timbres, other studies deal with describing sound synthesis schemes. One of them, by Ashley (1986)[2], uses a knowledge-based method for mapping between verbal descriptions and FM synthesis parameters. A more recent work by Ethington (1994)[44] uses a series of listening experiments to map verbal descriptions to a synthesis technique. In fact, a strong background in synthesis theory has always been a prerequisite for effective work with synthesizers, where sounds are specified by a large collection of numerical parameters specific to a given synthesis technique.

1.2.3 Modeling and Feature Extraction for the Description of Timbre

Three basic model types are in prevalent use today for musical sound generation: *instrument models, spectrum models,* and *abstract models*. Instrument models attempt to parameterize a sound at its source, such as a violin, clarinet, or vocal tract. Spectral models attempt to parameterize a sound at the basilar membrane of the ear, discarding whatever information the ear seems to discard in the spectrum. Abstract models, such as FM, attempt to provide musically useful parameters in an abstract formula. We shall not deal with the last type of models in this study.

Instrument models can be further classified into two main classes of methods for sound analysis-synthesis. The first one is known as a “Signal Model”. It tries to model the essential properties of some families of signals. It is essentially based on signal processing methods. Source-filter models, as applied first for speech synthesis, are good examples of a signal model.

The second class of models is known as a “Physical Model”. It tries to build a model of a musical instrument by finding and simulating the physical laws that governs its functioning. For instance, models of the vocal tract have been developed for speech synthesis, and many models of musical instruments are being studied. Both classes have advantages and drawbacks, and efforts are being made to build models with the benefit of both.

Naturally, it is very important to discover the relationship between models and synthesis techniques and the psychoacoustical parameters. Concerning the basic paramters such as *pitch, loudness, duration* and the two main dimensions of timbre - *spectral envelope* and *temporal envelope* - plenty of works have been published and we will not explain them here. We would like to mention here briefly some terms that are especially relevant for describing timbre :

- *Brightness.* Brightness is one of the main dimensions in the description of timbre (Grey 1975)[55],and is important for judgments of similarity. It represents the *centroid* of the distribution of spectral energy.

- *Spectral Flux.* This is another dimension in the description of timbre, but it is not well defined. Spectral Flux stands for the synchrony of onset and fluctuations in time of the harmonics. There is, however, no commonly accepted method for calculating this property.

- *Harmonicity.* This parameter distinguishes between "harmonic spectra" (e.g., vowels and most musical sounds), inharmonic spectra (e.g., metallic sounds), and noise (spectra that vary randomly in frequency and time). In other words, this feature represents the degree of definability of pitch due to sound partials being integer multiples of the fundamental frequency. For various reasons, it has not been considered a dimension so far.

- *Attack quality and other features.* Attack quality relates to the initial stage in sound's envelope when the pitch is not well defined, and it represents the degree of noise present at this stage. This and several other specific features that seem to distinguish particular instruments from others, have been reported.

1.2.4 Previous Works on Apperiodicities in Periodic Sounds

Acoustical musical instruments, even chordophonic and aerophonic ones, which are considered to produce a well defined pitch, even when they are unlimited in their duration, still emit waveforms which are never precisely periodic. These aperiodicities supposedly originate in some fundamental mechanism of their sound production. This effect, which for time scales shorter than 100 or 200 ms is beyond the control of the player,[5] is expected to be typical of the particular instrument or maybe of the instrument family.

Recently there has been growing interest in these phenomena due to recognition of their importance for the characterization of the timbre of sound. With the exception of some studies on the periodic fluctuations of bowed string instruments, these aperiodicities have not been quantitatively characterized (McIntyre et al. 1981)[83]. 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, thus neglecting the short time aperiodicities. From the psychoacoustic perspective, the influence of microscopic deviations in frequency and

[5] Vibrato and other expressive musical gestures have a longer time scale.

amplitude of the spectral components on the perception of sound has been investigated by several researchers (McAdams 1984)[86](Sandell 1991)[115]. It has been shown that random frequency deviations influence the perceived sound harmonicity and its coherence contribute to the sense of fusion/segregation among partials. We discuss the psychoacoustic findings further in section 2.3.3.

There are two basic approaches to modeling microvariations, which are models of the aperiodicities. One is routed in chaos theory and attempts to model the microvariations as fractal Brownian motion based on analysis of correlation or fractal dimension (Vettori 1995)([134], (Keefe 1991)[64]. We shall not discuss these models here but we would like to mention that fractal dimension is a non-linear statistic dependent on the source entropy, and as such it is expected to be sensitive to changes in the entropy of a signal caused by the presence of polyspectral factors.

Another approach attempts to investigate the temporal changes in signal contents across successive time frames. Two methods of analysis, one based on comparisons of cycle to cycle waveforms in the time domain, and another based on Fourier transforms of successive sonograms, were suggested by Schumacher (1992)[118]. The Fourier method results in a two dimensional plot, with one axis representing the short time spectral contents of the signal and the second axis revealing the low-frequency (subharmonic) fluctuations of the primary power spectrum. It is interesting to note that this method is closely related to the *correlogram* representation of sound which was investigated in the context of pitch perception and auditory modeling (Slaney 1993)[122]. The correlogram represents sound as a three dimensional function of time, frequency and periodicity. The third dimension measures the correlations between spectral estimates in each time frame, with the time between successive frames allowing for slow temporal changes in the signal. The two dimensional periodicity display (frequency vs. correlation) reveals short temporal events such as glottal pulses and other short-term spectral periodicities. This method has not been applied directly to characterization of aperiodicities or other subtle properties of timbre.

We would like to point out that these methods look for *sub-harmonic*[6] structure in the signal by considering fluctuations of the harmonics in the low frequency range. This significantly differs from our polyspectral analysis method (and the frequency modulating jitter model, presented later in the study), since we emphasize the effect of fluctuations in the high harmonics, specifically dealing with statistical dependence/independence between fluctuations, rather than sub-harmonic effects.

Other works which are concerned with the application of bispectral analysis to musical sounds but have not been able to relate it to aperiodicities or microvariations in the sound components will be presented in section 2.3.2.

[6]This means frequencies that are usually lower than the fundamental frequency and are unrelated to the mechanism of pitch production. These fall in the range from 2 Hz to approximately 100 Hz.

1.3 Overview of the study

1.3.1 Goals

The most general goal of this work is the extension of research on timbre by introducing a new method (polyspectra or higher order statistics) and by extending the field of timbre phenomena to include border areas with texture. Moreover, the methods may have interesting implications in a broader range of musical phenomena, too.

In more detail, the purpose of this work is to investigate acoustic phenomena that originate in the microfluctuations of the sound partials and contribute to the sense of perceived timbre of sound. Although the mechanism responsible for the microfluctuations are physical, of course, we still lack an understanding which would allow us to use Physical Models for the description of this phenomenon. In the course of our presentation we shall use several signal models, such as the source-filter instrument model in chapter 3 and sinusoidal representation of the source excitation, which is usually regarded as a spectral model, in section 6.1. We shall extend these models to include acoustic properties that are not usually accounted for in the traditional methods and provide a mathematical framework (by using higher order/polyspectral signal processing methods) for their analysis. This point is important for understanding the contribution of the present study, and it will be explained in length in the next section.

The issue of timbre analysis of musical signals is extremely complicated due to the multiplicity of factors that compete in the perception of timbre. Various factors such as the formant structure, the waveform of the signal together with its spectral contents and many temporal features, as mentioned in section 1.2, have been investigated in detail both in terms of the technical aspects and with respect to their perceptual (ISSM95)[62] and musical importance (Slawson 1985)[123], (Wessel 1979)[136].

Most of the modeling, analysis, parameter estimation and synthesis of sound is based upon second order statistics or spectral information. This approach seems insufficient for handling the abovementioned types of acoustic phenomena, especially in the case of the very high quality and subtlety required for music. As an example, residual signals from Autoregressive estimation, or from Sinusoids and Noise models need better modeling techniques. These signals, which are closely related to the modeling of the excitation signal in instrument models, and which contain much of the more subtle information about the structure of the signal, are spectrally "uninteresting" (white). The analysis of Higher Order Statistics (HOS) that we use, seems to be the natural next step for their description and it is certainly a field to be explored in priority.

1.3.2 Outline of the Work and Summary of Results

This work is divided into four main parts. The **Introduction** part contains a general introduction into **The Musical Problem of Timbre and Sound Texture**, followed by a more specific presentation of the applications of **Higher Order Spectra for Acoustic Signal Processing**. After introducing the general musical framework of the timbre and texture research, we provide a short overview of the mathematical preliminaries, with an emphasis on music and acoustic applications.

In section 2.2 we shall demonstrate the relevance of HOS analysis for acoustical purposes. We show that several acoustically meaningful bispectral measures could be devised, that are sensitive to properties such as harmonicity and phase coherence among the harmonics. Effects such as reverberation and chorusing are demonstrated to be detected clearly by the above measures.

By using a source-filter model, with non-Gaussian modeling of the source, a maximum likelihood parameter estimation scheme is developed in 3. This scheme also allows us to construct an acoustic distortion function which is the generalization of the Itakuro-Saito distance measure to include polyspectral information.

A better approximation of source excitation is achieved by sinusoidal modeling of the spectrally flat, pulse-train-like input signal. In section 6.1 we suggest a model with frequency modulating jitter "injected" into the frequencies of the harmonics, to simulate their microfluctuations. In this part of the work we show that the presence of jitter is directly related to the non-Gaussianity and non-linearity of the signal. Moreover, it is shown that higher order statistics (HOS) 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. This *effective number of coupled harmonics* serves us for the resynthesis of the "pulse train" signal, with the desired polyspectral properties.

Finally, another important application of HOS is in the classification of a broad range of sounds, musical, natural and environmental (classified by us as sound textures), which are needed by musicians and composers for applications in computer music. In chapter 7 we examine the possibilities for application of our techniques to the classification of man-made noises and other complex sounds. We end this work by presenting an extension of the problem of texture and application of the main principles (such as concurrence/non-concurrence) used here to describe micro phenomena to "macro" musical organization (chapter 8).

To sum up the scientific contribution of the study, we might say that the main acheivements are in the following:

1. Addressing an important acoustical problem of microvariations within sound whose significance is only now being recognized.

2. Using a new mathematical analysis method which has not been applied yet to acoustical

processing and that puts the issue on analytical grounds.

3. Significant results of the correspondance between HOS features and sound timbre properties (such as classification into instrument families) shows the applicability of the method to real acoustical problems.

1.3.3 Musical Contribution

The implications of this work for other fields of music research could be:

1. Probabilistic modeling of other musical parameters. In music many manifestations of uncertainty appear with respect to various parameters on different levels of the musical structure, so that the presence and quantification of uncertainty may serve as an important parameter in the characterization of a musical style.

2. Extension of the term concurrence/non-concurrence to deal with phenomena in the domain of timbre. Concurrence/non-concurrence is a very important variable that contributes to the characterization of the most basic variables for musical organization - similarity/difference, salience/non-salience, certainty/uncertainty - and therefore to the characterization of style.

3. Research of timbre-texture, texture-interval and texture-rhythm border areas, which emphasize the momentary musical events and contribute to a static musical feeling. This border areas are of special importance in contemporary and non western music. Moreover, the relationship between static and dynamic in music is an important characteristic of style.

4. Classification of musical instruments and natural sounds.

5. Understanding of the musical significance of the parameters of timbre, texture and border areas.

The methods and tools developed here may allow for a better understanding and formal treatment of the above musical problems.

Chapter 2

Higher Order Spectra for Acoustic Signal Processing

2.1 Multiple Correlations and Cumulants of Stochastic Signals

How to extract information from a signal is a basic question in every branch of science. The lack of complete knowledge of the signal exists in many physical settings due to the nature of the observed signal or the type of measurement device. In information processing we encounter the inverse problem - given the signal we want to extract information from it in order to perform basic tasks such as detection and classification. We presume that any biological information processing system acts in a similar manner. For instance, our ears analyze the acoustic signal by extracting pitch and timbre information from it. To understand our motivation to study higher order correlations it is worth recapitulating briefly some of the reasons for using the ordinary double correlation. A customary assumption is that our ears perform spectral analysis of the incoming signal. Naturally, not all of the signal information is retained in our ears, and the simplest assumption is that the phase is neglected. It is well known that the amplitude of the Fourier spectrum is equivalent to the Fourier transform of the signal's autocorrelation. Double correlation in time domain is the basic type of information extracted from the signal by our ears. This information has the meaning of the signal's spectral envelope in the frequency domain. Now we intend to widen the scope of acoustical analysis by suggesting the use of triple, quadratic and higher correlations, which are known also as polyspectra in the frequency domain. The kth-order correlation, $h_k(i_1, .., i_{k-1})$ of a signal $\{h(i)\}_{i=0}^{N}$ is defined as

$$h_k(i_1, .., i_{k-1}) = \sum_{i=0}^{N} h(i)h(i+i_1)..h(i+i_{k-1}) \tag{2.1}$$

and in the frequency domain it corresponds to the kth-order spectrum

$$\begin{aligned} H_k(\omega_1,..,\omega_{k-1}) &= \sum_{i_1,..,i_{k-1}=-N}^{N} h_k(i_1,..,i_{k-1}) \cdot e^{-j\omega_1 i_1 ... -j\omega_{k-1} i_{k-1}} \\ &= H(\omega_1)\dots H(\omega_{k-1}) \cdot H(-\omega_1 - \dots - \omega_{k-1}) \end{aligned} \tag{2.2}$$

Under some common assumptions, the time domain correlation converges to the kth-order moment of the process. The kth-order cumulant is derived from the kth and lower order moments, and contains the same information about the process. We prefer to use cumulants in our definition of spectra since for Gaussian processes all higher then second cumulants vanish. For zero mean sequences, the second and third order moments and cumulants coincide. Thus we arrive at an equivalent definition of the kth-order spectrum as the $(k-1)$-D Fourier transform of the respective kth-order cumulant of the process .

2.1.1 Elements System Theory

Let $y(i)$ be the output of an FIR system $h(i)$, which is excited by an input $x(i)$, i.e.

$$y(i) = \sum_{j=0}^{N} h(j)x(i-j) \tag{2.3}$$

Using the definition (1) it is easy to show that

$$y_k(i_1,..,i_{k-1}) = \sum_{j_1,...,j_{k-1}=-N}^{N} h_k(j_1,...,j_{k-1}) \cdot x_k(i_1-j_1,...,i_{k-1}-j_{k-1}) \tag{2.4}$$

where y_k, h_k, x_k are defined as in (1). Further, employing (1) and (2) we arrive at the frequency domain relations

$$Y_k(\omega_1,..,\omega_{k-1}) = H_k(\omega_1,..,\omega_{k-1})X_k(\omega_1,..,\omega_{k-1}) \tag{2.5}$$

An important property of the polyspectra is that if we are given two signals f and g that originate in stochastically independent processes and their sum signal $z = f + g$, then

$$Z_k(\omega_1,..,\omega_{k-1}) = F_k(\omega_1,..,\omega_{k-1}) + G_k(\omega_1,..,\omega_{k-1}) \tag{2.6}$$

This property is important when considering the perception of simultaneously sounding independent signals, as will be discussed later.

2.2 Properties of Bispectrum of Musical Signals

One of the main objectives of research on musical timbre is to identify physical factors that concern our perception of musical sounds. Several such factors have been already discovered by various researchers and some of them, such as spectral envelope, formants, time envelope, etc., are generally accepted as the standard features for describing musical sounds (Grey 1975)[55], (McAdams & Bregman 1979)[88], (McAdams 1984)[86], (Sandell 1991)[115], (Kendall & Carterette 1993)[65].

Below we focus on two more subtle factors, namely harmonicity and the time coherence among the partials, which appear in the mathematical definitions of the bispectrum and its related bicoherence index. We believe that these properties are the ones that make bispectrum acoustically interesting and significant. The manner in which they combine to influence higher acoustical/musical phenomena will be the topic of the next section.

The harmonicity and time coherence factors are not independent factors, however. Psychoacoustics research has pointed out several auditory cues central to the perception of spectral blend. It has been shown that harmonicity of the frequency content and coherence between the spectral elements are the major factors influencing spectral blend (McAdams 1984). [86].

2.2.1 Harmonicity Sensitivity

The "harmonicity measure" concerns the degree of existence of integral ratios among the harmonics of the tone (Sandell 1991) [115]. Various experiments conducted by psychoacousticians indicate that processing of harmonicity as a spectral pattern is a central auditory process. Harmonic tones fuse more readily than inharmonic tones under similar conditions and the degree to which inharmonic tones do fuse is partially dependent on their spectral content. Physical acoustics teaches us also that the presence of harmonicity makes possible the establishment of effective regimes of oscillation which are important for the production of stable, centered tones (Dudley & Strong) [41], (Benade 1976)[4]. In general, musical instruments do not have a single constant set of harmonics ratios throughout the duration of a tone. Thus, the characterization of the degree of harmonicity concerns some overall average behavior of these ratios.

One possible characterization of the harmonicity measure is obtained by means of bispectral analysis of the sound. Presence of a strong bispectral ingredient indicates the existence of a harmonically related triplet of frequencies in the signal spectrum $H_2(\omega_1, .., \omega_{k-1}) = H(\omega_1) \cdot H(\omega_2) \cdot H^*(\omega_1 + \omega_2)$, as directly follows from the definitions. Integrating over the bispectral plane gives a single number that represents the harmonicity measure . In case of a stationary signal, the excitation pattern in the bispectral plane remains constant. In reality the signals are time evolving and the bispectral signature changes with time, so that the above measure gives the instantaneous harmonicity averaged over the time interval chosen for the bispectral analyzer. Averaging again over the whole time span of the signal's existence would result in a single time independent number characteristic of the harmonicity measure of the musical tone.

Naturally, the signal must be normalized in time, frequency and amplitude prior to application of the bispectral analysis. The above procedure in additionally to its advantages in simplicity of application, puts the above question in a rigorous signal processing framework.

2.2.2 Phase Coherence

The coherence of the spectral contents of the signal has been studied with respect to the influence of frequency and amplitude modulation on tone perception. All natural, sustained-vibration sounds contain small-bandwidth random fluctuations in the frequencies of their components. Several experimental results demonstrate the importance of frequency modulation coherence for perceptual fusion. McAdams (1984) explained these results on the basis of an assumption about the existence of mechanisms in our ear system that respond to a regular and coordinated pattern of activity distributed over various cells in the system. In our presentation we suggest yet another lower level mechanism that might be responsible for this phenomenon.

The effect of frequency modulation coherence is best illustrated by an example adopted from Nikias & Raghuveer (1987)[92]. Consider two processes

$$x(n) = cos(\lambda_1 n + \phi_1) + cos(\lambda_2 n + \phi_2) + cos(\lambda_3 n + \phi_3) \tag{2.7}$$

and

$$y(n) = cos(\lambda_1 n + \phi_1) + cos(\lambda_2 n + \phi_2) + cos(\lambda_3 n + (\phi_1 + \phi_2)) \tag{2.8}$$

where $\lambda_1 > \lambda_2 > 0, \lambda_3 = \lambda_1 + \lambda_2$,, i.e. harmonically related triple, and ϕ_1, ϕ_2, ϕ_3 are independent random phases uniformly distributed between $[0, 2\pi]$. It is apparent that in the first signal $x(n)$, the λ_3 is an independent harmonic component because ϕ_3 is an independent random-phase variable. On the other hand λ_3 in $y(n)$ is a result of phase coupling between λ_1 and λ_2. One can verify that $x(n)$ and $y(n)$ have identical power spectra consisting of impulses at λ_1, λ_2 and λ_3. However, the bispectrum of $x(n)$ is zero whereas the bispectrum of $y(n)$ shows an impulse at $(\omega_1 = \lambda_1, \omega_2 = \lambda_2)$. One might notice that the case presented above corresponds to a phenomenon of the so-called *quadratic phase coupling*, which would be due to quadratic non-linearities in the process. The resulting non-zero bispectrum does not depend on this particular mechanism, though, and it will hold for any case of statistical dependence between the phases. An example of such an effect will be demonstrated in succeeding sections.

2.3 Significance for Musical Raw-Material

2.3.1 Timbre versus Interval

In the most general manner we can contrast timbre to the parameter of the musical interval with all its derivatives: scales, rules of harmony (the chords include also a timbral quality), etc. In contrast to the interval, the definition of timbre is very complex, needs many dimensions for characterization, and is not suitable for a simple arrangement on a scale or other clear and complex hierarchical organization (Lerdahl 1987)[73]. Also, in contrast to the interval which is a learned quality, timbre is loaded to a large extent with sensations from the extra-musical world. Furthermore, timbre takes longer time to be percieved, and one cannot precisely remember many rapid timbral changes. In contrast to the interval, whose main meaning is derived from various contexts, timbre is perceived as a single event even on the most immediate level. In its broad sense, timbre will encompass musical/acoustical properties of the register of pitch, intensity, aspects of articulation (various degrees of staccato and legato), vibrato and other microtonal fluctuations, spatialization of sound sources, various kinds of texture, chorusing, etc. However, this extension is a priori limited so as to exclude tonal and rhythmic schemes, contrary to the liberal view of Cogan (1984) [18], who included them as aspects of timbre, too.

Artistically, the timbre might fulfill several roles, such as incorporating extra-musical associations, focusing attention on momentary occurrences, supporting or blurring the musical structure that is otherwise defined by the tonal and rhythmic systems by being in concurrence or non-concurrence with the other musical parameters, and serving as the main subject of the composition. Its role in the composition may be a crucial factor in characterizing the style and the stylistic ideal. Regarding musical timbre as one of the components of the musical raw material, we concentrate, as mentioned above, on a few specific aspects. In the previous sections we discussed the "micro" level, i.e. the notion of bispectrum and some of the related acoustic features. Here we would like to present several higher-level, "musically significant" acoustic phenomena. We will try to show that these phenomena effectively influence the bispectral contents of the signal, and thus might be explained on this basis.

2.3.2 Related Works on Bispectral Analysis of Musical Sounds

The manifestation of the various timbres is most prominent in musical instruments (and in phonemes of speech with which we are not dealing here). Musical instruments are involved in shaping the musical style (Sachs 1940)[114], (Geiringer 1978)[49] for three main reasons:

1. The system of pitches and intervals which may be produced by the instrument.

2. The timbre of a single note. The timbres of various instruments may support or emphasize the structure of the work, which is based in principle on pitch organization;

blur the structure; or serve as an essential parameter for the structure.

3. The possibility of changing the properties of single notes and combinations thereof during performance.

In the west many efforts were made to neutralize the first factor. Musical instruments in the west reflect an optimal selection of timbre types which can be categorized. Thus, in the present study we have chosen to focus first on the contribution of the bispectrum to the characterization of instrumental timbres. Analysis of instrumental timbres based on the known aspects has received plenty of attention and we shall not mention these researches here.

With regards to the bispectrum, the first attempts to use bispectral considerations for sound quality characterization can be traced to Michael Gerzon (1975),[54], to whom we are also indebted for many of the following ideas. The power spectrum, which is generally used for sound characterization, being "phase-blind", cannot reveal the relative phases between the sound components. Although the human ear is almost deaf to the phase differences, the ear can perceive time-varying phase differences. The bispectral analyzer is the generalization of the power spectrum to the third order statistics of the signal. The bispectrum reveals both the mutual amplitude and phase relation between the frequency components ω_1, ω_2. If sound sources are stochastically independent, their bispectra will be the sum of their individual bispectra. In order for a bispectral analyzer to be able to recognize the characteristic signature of the sound in the bispectral plane, the excitation of a given ω_1, ω_2 should be distinguishable from the background noise. Thus, a "good" instrument is supposed to produce the maximum bispectral excitation possible for a given signal energy. Stating the problem as "can we predict the properties of a Stradivarius?" Gerzon claimed that the design requirement for a musical instrument is that "they should have a third formant frequency region containing the sum of the first two formant frequencies." Surprisingly enough, this theoretical criterion seems to be satisfied by many orchestral instruments. For example, specific cases of the Stradivarius violin (435 Hz, 535 Hz, 930 Hz), Contrabassoon (245 Hz, 435 Hz, 690 Hz) and Cor Anglais (985 Hz, 2160 Hz, 3480 Hz). In a later work, Lohman and Wirnitzer (1984)[79] analyzed two flutes by calculating their bispectra. Their results demonstrate that a higher intensity of the phase of the complex bispectra is achieved for a good-quality flute. This also suggests that the intelligibility of speech could be determined by looking at the bispectral signature and might be even enhanced by adding an artificial third formant to the sum of the momentary two lowest formant frequencies. Such a device can easily be constructed by means of a quadratic filter or other non-linear speech clipping system. One must note that such a simple device will modify the spectrum, too, which might be undesirable.

2.3.3 Tone Separation and Timbral Fusion/Segregation

Among the various questions dealing with the timbral characteristics of sounds, the problem of simultaneous timbres (McAdams & Bregman 1979)[88], (McAdams 1984)[86] is basic to musical practice itself, manifestating itself in daily orchestration practice, choice of instruments and the ability to perceive and discriminate individual instruments in a full orchestral sound. Originally treated semi-empirically by the orchestration manuals, vague criteria for evaluating orchestral choices were presented. In recent times more quantitative acoustical studies point out several features in the temporal and spectral behavior of the sounds which are pertinent for instrument recognition and modeling spectral blend (Sandell 1991)[115]. We suggest realizing the power of polyspectral techniques for the analysis of spectral blend.

McAdams (1984) reported about several experimental results that support the notion that frequency modulation coherence contributes to perceptual fusion. He explained his results on the basis of an assumption about the existence of mechanisms in our ear system that respond to a regular and coordinated pattern of activity distributed over various frequency sensitive cells. Several recent works have suggested analyzing sounds of polyphonic music by tracking fluctuations in pitch and amplitude in order to separate the spectrum of a multivoice signal into classes of partials united by a common law of motion (Tanguiane 1993)[127]. Now that we have at our disposal such a powerful tool for detecting coherence between the spectral components of a signal, we claim that the ear performs grouping of the various spectral components present in the sound by relating strong bispectral peaks to a single source. In the following example we demonstrate the above phenomena on a very simple signal constructed of three harmonics at 200, 400 and 600 Hz. In both signals there are random changes in the frequency of the harmonics, but whereas in the first signal these changes are independent, in the second signal there is concurrence among the various harmonics with respect to their instantaneous direction of change. Technically, these signals are similar to the ones described in equations (2.7) and (2.8) in section 2.2.2. Each of the three harmonics in both signals was frequency modulated with a random jitter. The spectrum of the jitter was such that most of its energy was centered around 30 Hz. In the first signal independent jitters were applied to each harmonic (Fig. 2.1-top), while the second one used the same jitter function. In the second signal the frequency modulation, too, was such that the amount of deviation was proportional to the frequency of the harmonic (Fig. 2.1-bottom).

Thus what we have here is almost identical spectral content with a slight random temporal variation between the signal whose sole characteristic is statistical dependence/independence among the frequency deviations. Audition to the two signals reveals two components in the all random signal, while the coherent phases signal sounds like a single source. This clearly non linear effect is easily detected in the bispectral plane. Fig. 2.2 demonstrates the respective bispectra of the two signals. The bispectral peaks are at (200,200)Hz (the half peak on the diagonal) and at (200,400)Hz as expected. [1] It is possible thus that a spectral blend is actually a blend of bispectral patterns where harmonics with strong bispectral components fuse together to form a single sound. To conclude this discussion we should mention that this bispectral mechanism is one of many that influence tone color separation/blending.

[1] Due to symmetries of the bispectrum it suffices to display only one of the eight symmetry regions that exist in the bispectral plane.

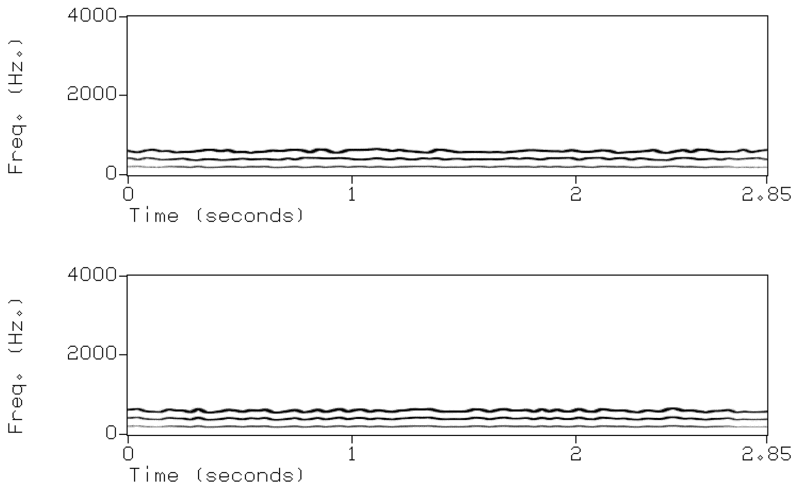

Figure 2.1: **Top:** Spectrogram of a signal with independent random jitters of the harmonics. **Bottom:** Spectrogram of a signal with the same random jitters applied to modulate all three harmonics. One can see the similar instantaneous frequency deviations of all harmonics.

2.3.4 Effects of Reverberation and Chorusing

Other, more subtle problems of intelligibility can be considered by looking at the effects of reverberation and chorusing. As it is an important musical issue, we note, quoting Erickson [43], that "there is nothing new about multiplicity and the choric effect. What is new is the radical extension of the massing idea in contemporary music, and the range of its musical applications; but a great deal more needs to be known before the choric effect is fully understood or adequately synthesized." As mentioned previously, if the sounds are stochastically independent, then their bispectra will simply be the sum of the separate bispectra. Assume a sound source with energies S_1, S_2, S_3 at frequencies $\omega_1, \omega_2, \omega_3 = \omega_1 + \omega_2$ and bispectrum level B at (ω_1, ω_2) subject to reverberation effect. Now let us assume that this effect can be modeled as a linear filter acting as a reverberator added to the direct sound. Suppose that the effect of the reverberation is only to produce a proportionate spectrum energy kS_1, kS_2, kS_3 at $\omega_1, \omega_2, \omega_3$. A plausible model for the linear filter describing the reverberator part alone could be an approximation of its impulse response by a long sample of a random Gaussian process. According to Eq. (10), the bispectral response of such a filter is zero, which results in a zero bispectrum of the output signal. The total resultant signal contains a (stochastically independent) mixture of the direct and the reverberant sound. The spectral energy of the combined sound at $\omega_1, \omega_2, \omega_3$ will be $(1 + k)S_1, (1 + k)S_2, (1 + k)S_3$ at $\omega_1, \omega_2, \omega_3$ and bispectrum level B at ω_1, ω_2. Naturally, the proportion of the bispectral energy to the spectral energy of the signal deteriorated. For a signal with complex spectrum $H(\omega)$, the power spectrum equals $S(\omega) =\mid H(\omega) \mid^2$ and the bispectrum is $B(\omega_1, \omega_2) =$

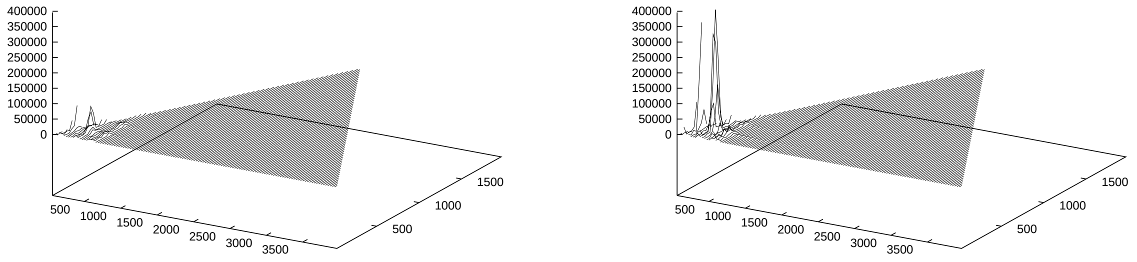

Figure 2.2: Random jitter applied to modulate the frequency of the harmonics. When each harmonic deviates independently, the bispectrum vanishes (left), while coherent variation among all three harmonic causes high bispectrum (right). The bispectral analysis was done over a 0.5 sec. segment of the signal with 16 msec. frames. The sampling rate was 8KHz.

$H(\omega_1)H(\omega_2)H(-\omega_1-\omega_2)$. Taking a bicoherence index

$$b(\omega_1,\omega_2) = \frac{B(\omega_1,\omega_2)}{(S(\omega_1)S(\omega_2)S(-\omega_1-\omega_2))^{1/2}} \tag{2.9}$$

we arrive at a dimensionless measure of the proportionate energy between the spectrum and the bispectrum of a signal. If $b = b_{in}$ for the original signal, then after reverberation $b_{out} = (1+k)^{-3/2} b_{in}$. Thus for a reverberation energy gain k, the relative bispectral level has been reduced by a factor of $(1+k)^{-3/2}$ (Gerzon 1975). Now consider a very similar effect of chorusing. For N identical but stochastically independent sound sources the resultant spectral energies at $(\omega_1, \omega_2, \omega_3 = \omega_1 + \omega_2)$ are (NS_1, NS_2, NS_3) and the resulting bispectra are NB at (ω_1, ω_2). Comparing again the bicoherence indexes we arrive at $b_{out} = N^{-1/2} b_{in}$ giving a relative attenuation of $N^{-1/2}$ due to this chorus effect. It is worth mentioning once again the importance of stochastic independence. The chorusing as described above might be confused with a simple multiplication of the original signal energy by a gain factor N. Such a gain is not stochastically independent and the resulting bispectrum would be augmented by $N^{3/2}$ instead of N. Only a true lack of coherence between the replicated signals will cause the resulting bispectra to be actually NB.

2.3.5 Experimental Results

In order to demonstrate the above effects, we have performed an analysis of various sound examples. The first example is a pair of sampled signals of solo instrument (Solo Viola) and of an orchestral section of the same instruments (Arco Violas). (The signals were recorded from a sample-player synthesizer and are believed to be true recordings of the above instruments.)

The signals have very similar spectral characteristics and the "chorusing" feature, dominantly present in the "Arco Violas" signal, cannot be extracted from the spectral information alone. Nevertheless, it has its manifestation in the signal's bispectral contents. We plotted the amplitude in the bicoherence index for each of the two signals. As we can clearly see from Fig. 2.3, there is a significant reduction of the bispectral amplitude for the "ArcoViolas" signal. Note also that the bispectral excitation pattern is different for the two signals, with the "SoloViola" signal having few clear peaks while the "ArcoViolas" has a much more spread-out and noisy pattern.

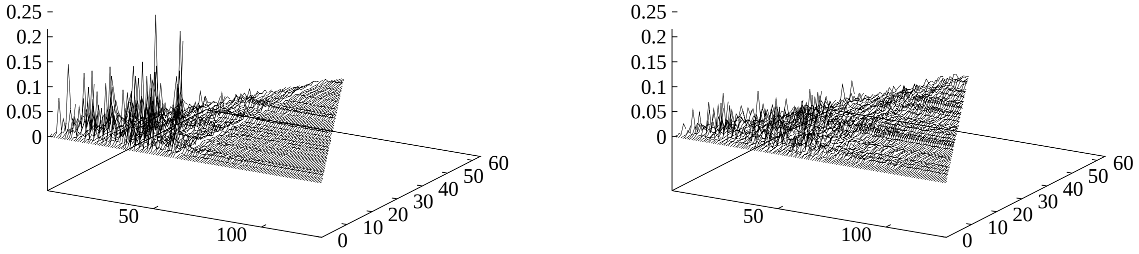

Figure 2.3: Bicoherence index amplitude of Solo Viola and Arco Violas signals. The x-y axes are normalized to the Nyquist frequency (8 KHz).

The second pair of examples was taken from a CD recording of two performances of the opera "Don Carlos" by Verdi. In one performance a cello passage was played by a single instrument (Solo), while in the other it was performed by a section of celli (Tutti).

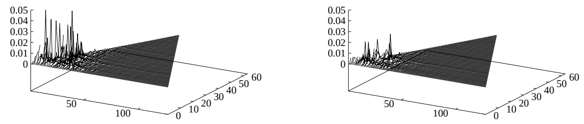

Figure 2.4: Bicoherence index amplitude of Solo Cello and Tutti Celli signals. The indices were calculated with 16 msec. frames, averaged over a 4 sec. segment, with order 8 FFT transform. In order to reveal the high amplitude bispectral constituents both indices were calculated with the spectra denominator thresholded at 10 db.

The last example is taken from a soprano duet versus a women's choir.

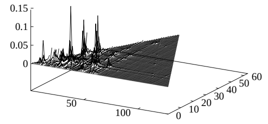
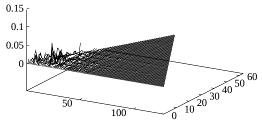

Figure 2.5: Bicoherence of duet soprano singers versus women choir calculated with 16 msec. frames over a 0.5 sec. segment. The indices were thresholded to include approximately just the first two formants.

Before concluding this section, we must remark on an important difference between the voice and string signals above. According to our physical motivation, the resonator-filter remains constant over the duration of the signal. On the other hand, auditory experience tells us that effects such as chorusing are better perceived at the changing portion of sounds, such as short note passages and attacks. Although this might seem surprising, there is no contradiction between the two, since in our formalism we do not actually require signal stationarity, but only filter constancy. Such an assumption is acceptable for all sounds produced by a string instrument, but is only locally applicable to the human voice. Thus, the cello results above were obtained by averaging over (the same) musical passages, while the human voice examples were sampled from short, constant, vowel singing portions.

2.3.6 Example : synthetic all-pass filter

As seen from Eq. (2.5) in section 2.1, the bispectra of the output signal $y(i)$ resulting from passing a signal $x(i)$ through a linear filter $h(i)$ equal the product of their respective bispectra. An equivalent relationship holds for the *linear random process*, i.e. when the output signal results from passing a stationary random signal through a deterministic linear filter. Consider now a device whose impulse response resembles a long segment of a Gaussian process. Although the filter might be deterministic overall, it could be considered a random signal for all practical purposes. Applying, for instance, a bispectral analyzer of final temporal aperture to such an impulse response, would average its bispectral contents to zero, giving us a filter with zero bispectral characteristics. Naturally, the output signal resulting from passing a deterministic signal through such a filter will have a zero bispectrum. Since the impulse response resembles a white noise signal, its spectral characteristics are flat, giving us an all-pass filter. Also, by properly scaling the impulse response we can assure that the filter gain equals 1.
The following figure describes the result of passing the original "Solo Viola" signal through a linear filter whose impulse response was created by taking a 0.5 sec. sample of a Gaussian process. The bispectral analysis of the signal was performed by averaging over 32 frames of 16 msec. each. The subjective auditory result seems to resemble a reverberation device.

Fig. 2.6 shows the bicoherence index of the signal after filtering.

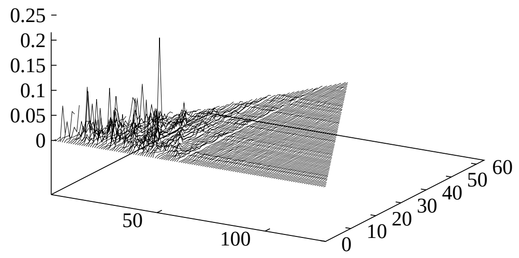

Figure 2.6: Bicoherence index amplitude of the output signal resulting from passing the "Solo Viola" signal through a Gaussian, 0.5 sec. long filter

Part II

The Acoustic Model

Chapter 3

AR Modeling and Maximum Likelihood Analysis

Assuming a linear filter model driven by zero mean WNG noise, we represent the signal as a real pth order process $y(n)$ described by

$$w(n) = y(n) - \sum_{i=1}^{p} h_i y(n-i) \tag{3.1}$$

where $w(n)$ are i.i.d. The innovation (excitation) signals w have an unknown probability distribution function (pdf), with non-zero higher order cumulants. We shall assume a pdf of an exponential type

$$P(w) = \exp(-\sum_{i=0}^{4} \lambda_i w^i) = \exp(-\lambda_0 - \sum_{i=1}^{4} \lambda_i w^i) \tag{3.2}$$

where the parameters $\lambda_i, i = 1..3(4)$ are the parameters of the distribution and

$$\exp(\lambda_0) = Z(\lambda_1..\lambda_4) = \int \exp(-\sum_{i=1}^{4} \lambda_i w^i) dw \tag{3.3}$$

is the normalization integral.

The particular choice of an exponential form pdf can be justified in various ways, such as the Maximum Entropy (ME) principle[130]. The unknown moments of the innovation can be estimated from the higher order statistics of the signal $y(n)$. Under constrains of these moments, the exponential type pdf is obtained as the ME solution.

Given the measurements $Y = \{y_0, .., y_N\}$, the probability of observing this set of samples, given the model (the noise pdf parameters and the AR coefficients $H = \{\lambda_1, .., \lambda_4,\ 1, h_1, .., h_p\}$) can be written. The average (over all samples) log-likelihood, restricted to the first three terms, is then

$$
\begin{aligned}
< \ln P(Y \mid H) >= -N \ln Z(\lambda_1, ..\lambda_3) \\
- \quad N\lambda_2 \sum_{i=0}^{p} \sum_{j=0}^{p} R_2(i-j) a_i a_j \\
- \quad N\lambda_3 \sum_{k=0}^{p} \sum_{l=0}^{p} \sum_{m=0}^{p} R_3(k-l, k-m) a_k a_l a_m \\
= \quad -N \ln Z - N\lambda_2 \int_{-\pi}^{\pi} \frac{d\omega}{2\pi} S_y(\omega) \mid A(\omega) \mid^2 \\
- \quad N\lambda_3 \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{d\omega_1 d\omega_2}{(2\pi)^2} B_y(\omega_1, \omega_2) \cdot \\
\cdot A(\omega_1) A(\omega_2) A^*(\omega_1 + \omega_2)
\end{aligned}
\tag{3.4}
$$

with $a_0 = 1, a_i = -h_i, i = 1..p$, and $A(z)$ denoting the polynomial $a_0 + a_1 z + .. + a_p z^p$, and $R_k(i, j, .., l)$ the k-th order moments tensor of the measurements Y (with $R_1 = 0$). $S_y(\omega)$ and $B_y(\omega_1, \omega_2)$ denote the power spectrum and bispectrum of the signal, respectively.

3.1 Source and Filter Parameters Estimation

If we are given only partial information concerning the source, the remaining parts of the model must be estimated. If, for instance, we know the linear filter coefficients h_i, but the noise source statistics are unknown, the average log-likelihood becomes:

$$
\langle \ln P(Y \mid H) \rangle = -N(\ln Z + \lambda_2 \mu_2 + \lambda_3 \mu_3) \tag{3.5}
$$

where we have assume that the actual noise moments are

$$
\left\langle w^i \right\rangle = \mu_i, \quad \langle w \rangle = \mu_1 = 0 \tag{3.6}
$$

and use the fact that polyspectra of a linear system are provided by

$$
\begin{aligned}
S_y(\omega) = \mu_2 \mid H(\omega) \mid^2 \\
B_y(\omega_1, \omega_2) \quad = \quad \mu_3 H(\omega_1) H(\omega_2) H^*(\omega_1 + \omega_2) \ ,
\end{aligned}
\tag{3.7}
$$

with $H(z) = 1/A(z)$.

Estimation of the λ's is accomplished by maximizing the log-likelihood, namely

$$
\frac{1}{N} \sum_{n=0}^{N} (y(n) - \sum_{j=1}^{p} h_j y(n-j))^k = \quad -\tfrac{\partial \ln Z}{\partial \lambda_k} = \left\langle w^k \right\rangle \ , \tag{3.8}
$$

or in homogeneous form as

$$
\begin{aligned}
-\frac{\partial \ln Z}{\partial \lambda_1} &= \mu_1 = 0 \\
\sum_{i=0}^{p}\sum_{j=0}^{p} R_2(i-j)a_i a_j &= \left\langle w^2 \right\rangle = \mu_2 \\
\sum_{i=0}^{p}\sum_{j=0}^{p}\sum_{k=0}^{p} R_3(i-j, i-k)a_i a_j a_k &= \left\langle w^3 \right\rangle = \mu_3
\end{aligned}
\tag{3.9}
$$

and similar higher order terms if necessary. In this manner we can estimate the excitation moments μ's using the filter parameters and the correlation functions of the signal $y(n)$. These moments determine the probability density function of the noise through their unique correspondence to the parameters λ_i.

Alternatively, we assume that the filter parameters are unknown. The estimation of these parameters is performed similarly by maximizing the log-likelihood expression with respect to the a_i's, yielding the p $(i = 1..p)$ equations

$$
\begin{aligned}
&2 \cdot \textstyle\sum_{j=0}^{p} \lambda_2 R_2(i-j)a_j \\
+ &3 \cdot \textstyle\sum_{j=0}^{p}\sum_{k=0}^{p} \lambda_3 R_3(i-j, i-k)a_j a_k = 0 \; .
\end{aligned}
\tag{3.10}
$$

Notice that the above estimation procedure does not imply the stability of the estimated filter, similar to other parametric methods of higher order spectral estimation[102][95]. This issue can be addressed directly using alternative filter representations which explicitly guarantee the stability (e.g. lpc parameters), but it is beyond our scope here.

3.2 Relations to Other Spectral Estimation Methods

Two important special cases can now be easily derived from our general estimation scheme. In the case of white Gaussian input, the log-likelihood expression reduces to the sum of the first two terms in equation (3.4). Adequately, only the first of the two equations in (3.9) remains, and the filter parameters estimation equation (3.10) becomes equivalent to the standard minimum variance spectral estimation equations. Writing the equations for the Gaussian case explicitly gives

$$
i = 1..p \qquad \sum_{j=0}^{p} R_2(i-j)a_j = 0 \tag{3.11}
$$

$$
\sum_{i=0}^{p}\sum_{j=0}^{p} R_2(i-j)a_i a_j = \left\langle w^2 \right\rangle \; , \tag{3.12}
$$

the last giving the gain factor equation for spectral matching.

An interesting derivation of the bispectral method for filter parameters estimation, which was suggested by Raghuver and Nikias [102], can be derived on a similar basis. Neglecting the Gaussian part in this case, we rewrite equation (3.10) as

$$i = 1..p \quad \sum_{j=0}^{p}\sum_{k=0}^{p} R_3(i-j, i-k)a_j a_k = 0 \ , \tag{3.13}$$

which can be rewritten also as

$$i = 1..p \quad \sum_{j=0}^{p}(R_3(-i,-j) + \sum_{k=1}^{p} R_3(i-j, i-k)a_k)a_j = 0 \ . \tag{3.14}$$

The bracketed part can be written as

$$j = 0..p, i = 1..p \quad R_3(-i,-j) + \sum_{k=1}^{p} R_3(i-j, i-k)a_k = 0 \ . \tag{3.15}$$

If we include the bispectral equation in (3.9) we arrive at

$$\sum_{k=0}^{p} R_3(i,i)a_i = \left\langle w^3 \right\rangle \ , \tag{3.16}$$

which is obtained by taking $j = 0, k = 0$ in the original equation. Combining both equations gives the original Raghuver and Nikias equations

$$j = 0..p, i = 0..p \quad R_3(-i,-j) + \sum_{k=1}^{p} R_3(i-j, i-k)a_k = \left\langle w^3 \right\rangle \delta(i,j) \ . \tag{3.17}$$

3.2.1 Discussion

The log-likelihood expression (3.4) contains both second and higher order information, combined in a unified framework. It requires both an estimation of higher order correlations and the solution of a complex non-linear system of equations. Several comments are in order.

1. Although the non-linear system of equations has no analytic solution, standard numerical techniques could be applied. The drawback of such an analysis is that the stability of the estimated linear filter is not automatically guaranteed. This problem was already present in the cumulant based spectral estimation method [102].

2. In our approach we apply the maximum entropy formalism to infer the source distribution from its moments. For the existence of a maximum entropy solution to the moments problem, the magnitudes of the moments can not be arbitrary [30]. We assume that whenever a solution is obtained these constraints are satisfied.

3. One could suggest estimating the parameters by iterative applications of equations (3.9), (3.10). Such a method requires estimating the γ's from the estimated λ's obtained in each step. Equations (3.9), (3.10) were derived under the assumption of exponential (maximum entropy) pdf of the innovation w and are obeyed at the extrema of the log-likelihood only. Care must be taken during the iteration process not to violate the above moments conditions for the existence of the maximum entropy form. One way which seems to work is to estimate all the parameters simultaneously.

4. The correlation functions, $R_i,\ i = 1..4$, must be estimated from the samples of the process. It has been shown that the conventional estimates are asymptotically unbiased and consistent, but are generally of high variance. Thus a large number of records is required to obtain smooth estimates. Poor estimates might conflict with condition (2) above as well.

In the derivation of the standard estimation methods above, we notice that the cumulant based method provides a way to estimate a_i's, using third order statistical information only, similar to the way that the power spectral methods rely solely upon second order statistics. There is no principled criterion for the choice among the two methods. Comparative experiments with both methods for noisy speech recognition show that the bispectral estimation method gives better results in low signal to noise ratio, for Gaussian noise[95] . We claim here that equation (3.4) provides a statistically consistent approach to the various order spectral estimation methods. The advantage in using the second order method, apart from the efficient and stability of the solution, is the fact that we can easily estimate the innovation pdf parameters from the observed second moments. For Gaussian zero mean random variable x, the second moment (variance) is given by

$$\gamma_2 = \left\langle x^2 \right\rangle = \frac{1}{Z} \int e^{-\lambda_2 x^2} dx \tag{3.18}$$

which gives the simple relationship between the only Lagrange multiplier and the moment $\lambda_2 = \frac{1}{2\gamma_2}$. No such simple relationship holds for non-Gaussian pdf's and in general all moments depend on all Lagrange multipliers.

3.3 Simulation Results

In our simulations we applied standard conjugate-gradient optimization techniques in order to obtain a solution to the likelihood equations. To assure the normalizability of the resulting innovation pdf we have extended the log-likelihood to include a fourth order term as well. We initialized the algorithm by first solving the second order (spectral) part, estimating the initial filter coefficients a_i, and then optimizing over λ_2, yielding a starting point $\{1, a_0, .., a_p, \lambda_2, \lambda_3 = 0, \lambda_4 = 0\}$. The solution was obtained by simultaneously maximizing the likelihood function with respect to all free parameters.

Here we demonstrate the results of maximizing the log-likelihood function for the signal obtained from a CD recording of an aria sang by a female soprano (Figure 3.1). The analyzed signal has a fixed pitch single vowel singing segment.

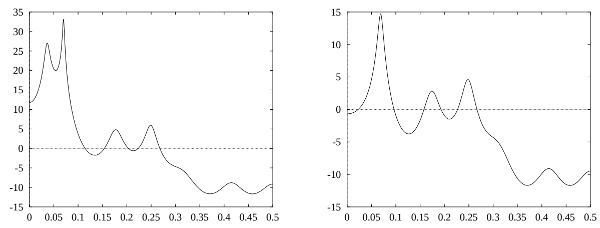

Figure 3.1: Correlation (left) versus cumulant based (right) filter estimates.

The estimated λ's are : $\{\lambda_1 = -0.001696, \lambda_2 = 2.360497, \lambda_3 = 0.520212, \lambda_4 = 0.029856\}$ The estimated pdf can be seen in figure 3.2, with the Gaussian distribution plotted for reference, and the correct $\lambda_2 = 2.945661$ derived by gain matching.

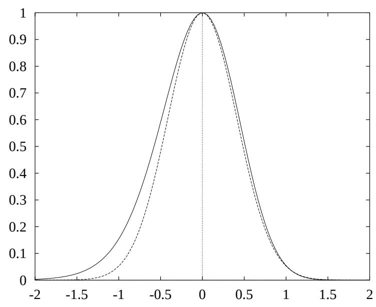

Figure 3.2: Probability distribution function of the non-gaussian input, plotted against the original Gaussian estimate.

A note of caution! The maximum entropy method is known to be anomalous for the three moment constraints, in the sense that almost any value for the third moment can be obtained by taking a normal distribution with a small "wiggle" at an arbitrary distant point [30]. Such a construction leaves the even moments almost unchanged with the largest affect being on the third moment. It might be observed during our process of optimization, that a gradual distortion of the original Gaussian suddenly switches to a solution of a mixture of a Gaussian and a distant "wiggle" as λ_3 grows. This phenomenon occurs for signals with large third order correlations and is controlled in our procedure by the inclusion of the fourth order terms.

Chapter 4

Acoustic Distortion Measures

4.1 Statistical Distortion Measure

The main motivation behind the ideas that we are about to present, from now until the end of the chapter, concerns the construction of functions for signal comparison. The so-called distortion measures give an assessment of the "similarity" between two signals, which is a basic characteristic needed for signal separation and evaluation. Such a separation procedure could be derived on various grounds, and here we chose to adopt tools from information theory which treat signals as stochastic models and give a probabilistic measure of separation between models. By constructing a statistical model we turn the problem of measuring distances between signals into a question of evaluating a statistical similarity between their model distributions. Before proceeding to talk about the construction of distortion function, some words about the models are in place.

As the first step we have chosen a very crude model that treats the power-spectral and the bispectral amplitude planes as one and two dimensional probability distribution functions respectively. The acoustic separation based on this model is the subject of this section. In the next section we shall construct a better acoustical model with a clear underlying physical motivation. This model will serve us for constructing an acoustic distortion measure which combines both the spectral and bispectral parts in a unified, systematic manner.

As mentioned above, in the following analysis we treat the power-spectral and the bispectral amplitude planes as probability functions. We assume that the power-spectral amplitude of each frequency in the spectrum can be viewed as an indication of how likely is it to identify a spectral component at a particular frequency within the whole spectrum. Similarly, the bispectral amplitude distribution indicates the likelihood of having a pair of frequencies whose mutual occurrence probability is proportional to the bispectral amplitude of the signal. In other words, the spectral model assumes that the signal is constructed by a collection of independent oscillators, distributed in accordance with the spectrum. The bispectral model assumes pairwise dependent oscillators, with their respective probabilities given by the bispectral amplitude.

Let us turn now to the distortion measure itself. The idea of looking at signals as stochastic processes and replacing the signal by its stochastic model is a very powerful one. When we receive a new signal, the model enables us to judge how probable the observed sample is within the chosen model. Low probability would mean that the new sound is "far" from the model. Soundwise, distortion measure says how probable it would be for a Flute sound to have been created by a Clarinet source. It is important to note that this 'is not a symmetric quantity, i.e. it is not the same as the probability of having a Clarinet sound played by a Flute.

Calculation of the distortion between signals is achieved by means of the Kullback-Liebler (KL) divergence (Kullback 1968) [68], which measures the distortion between their associated probability distributions. Denoting by $P(Y)$ and $Q(Y)$ the probabilities of having the signal $Y = \{y_0, .., y_N\}$ appear in models P and Q, respectively, the probability of a sample generated by P to be considered as a sample from Q satisfies

$$Q(Y) \propto e^{-ND[P||Q]} \tag{4.1}$$

where $D[P||Q]$ is the KL divergence between the two pdf's and N is the sample size. As explained above, we construct a model for each of the signals that coincides with our "measurements" of the spectrum and the bispectrum of the signal. Statistically this means that we look at a probability distribution that has the same first, second and third order statistics as those of the signal under consideration. Calculation of the KL divergence between the models is accomplished by calculating the average of the loglikelihood ratio of the two models with respect to the probability distribution of one of the signals chosen to be the reference signal.

$$D[P||P'] = \left\langle \ln \frac{P(Y)}{P'(Y)} \right\rangle_P \tag{4.2}$$

The distortion in "spectral sense" is achieved by calculating

$$D_S[S_1, S_2] = \int_{-\pi}^{\pi} S_1(\omega) \ln \frac{S_1(\omega)}{S_2(\omega)} d\omega \tag{4.3}$$

and in "bispectral sense" by

$$D_B[B_1, B_2] = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} B_1(\omega_1, \omega_2) \ln \frac{B_1(\omega_1, \omega_2)}{B_2(\omega_1, \omega_2)} d\omega_1 d\omega_2 \tag{4.4}$$

with the spectra and bispectra normalized amplitudes so as to have a total energy equal to one.

4.2 Acoustic Clustering Using Bispectral Amplitude

The comparative distance analysis was performed on nine instruments obtained from a sample-player synthesizer. The analysis took account only of the steady state portion of the sound, thus ignoring all information in the attack portion of the tone onset. All sounds were one pitch (middle C), the same duration and approximately the same amplitude. These results are summarized in the following tables, with the rows standing for the first, and the columns for the second argument of the distortion function.

Spectral distances matrix

	Cello	**Vla**	**Vln**	**Trbn**	**Trmp**	**FH**	**Cl**	**Ob**	**Fl**
Cello		0.07	0.12	0.40	0.55	0.05	0.29	0.40	0.32
Vla	0.08		0.15	0.27	0.33	0.11	0.18	0.23	0.47
Vln	0.13	0.15		0.20	0.42	0.11	0.44	0.26	0.59
Trbn	0.37	0.22	0.20		0.15	0.32	0.31	0.09	1.09
Trmp	0.54	0.30	0.42	0.15		0.61	0.47	0.18	1.29
FH	0.05	0.09	0.11	0.40	0.62		0.26	0.38	0.31
Cl	0.25	0.16	0.33	0.30	0.48	0.48		0.22	0.35
Ob	0.55	0.49	0.24	0.28	0.09	0.15	0.40		1.11
Fl	0.26	0.36	0.56	1.16	1.26	0.24	0.42	0.99	

Bispectral distances matrix

	Cello	**Vla**	**Vln**	**Trbn**	**Trmp**	**FH**	**Cl**	**Ob**	**Fl**
Cello		0.33	0.48	2.05	2.62	0.08	1.71	1.70	0.70
Vla	0.49		0.68	1.31	1.51	0.40	1.03	0.92	2.14
Vln	0.70	0.68		0.97	1.71	0.50	1.82	0.99	2.71
Trbn	1.96	1.02	1.09		0.47	1.49	1.13	0.30	5.02
Trmp	2.92	1.36	2.01	0.53		2.60	1.47	0.68	5.74
FH	0.10	0.29	0.35	1.70	2.33		1.47	1.38	0.95
Cl	1.85	0.98	1.66	0.77	0.99	0.95		1.63	0.86
Ob	4.69	2.07	0.81	1.25	0.31	0.49	1.06		4.94
Fl	0.32	0.76	1.07	3.50	3.96	0.44	2.69	2.86	

In general, the categorization of musical instruments into similar-sounding groups is performed by both methods in a manner reminiscent of the common orchestration practice. A more detailed analysis shows a clear differentiation into several groups: {**Cello, Viola,**

French Horn} with a related, but a more distant **Violin**; a group of {**Trombone, Trumpet, Oboe**}, with the **Clarinet** and especially the **Flute** remaining relatively detached from all the rest. One could also separate the groups with respect to their distance from the **Flute**, which is larger for the second group and smaller for the first one.

It is important to note that we are looking at the bispectral *amplitudes* only, and thus we neglect the important phase relationships. The phase behavior is actually the major quality that distinguishes between the spectral and bispectral analyses and it will be explored in depth in the following chapters. Thus, the bispectral "amplitudes only" information gives unsurprisingly similar classification to the spectral one, and it is insufficient for providing a truly new and distinct classification.

One should also note the already-mentioned phenomenon of asymmetry of the distance expression with respect to its two arguments. The bispectral distances augment the differences among the signals, especially increasing the asymmetry of the distortion between the pairs. In case of the {**Clarinet, Oboe**} pair, the relative asymmetry magnitudes are converse in the spectral and bispectral cases.

Concerning the asymmetry property, we might propose an intuitive explanation which claims that there exists a better chance to produce a "dull" sound from a "rich" instrument then vice versa. Within this line of thought, the opposite relative asymmetry magnitudes between **Clarinet** and **Oboe** are especially curious - bispectrally **Oboe** is richer then **Clarinet** while spectrally the opposite is the case. It is interesting to note that despite the fact that we ignore the effects of the pitch register and amplitude, this preliminary classification seems to be in agreement with the judgment of several professional musicians. Of course, a complete confirmation of these result calls for extensive and systematic experimentation, yet to be performed.

4.3 Maximum Entropy Polyspectral Distortion Measure

Using the log-likelihood expression that was derived in chapter (3) for the AR acoustic model, we arrive at the primary result of our work, which involves the generalization of the acoustic distortion measure to include higher than second order statistics. In general, acoustic distortion measures are widely accepted in speech processing and are closely related to the feature representation of the signal (Gray & Markel 1976) [51]. Over the years many feature representations were proposed, with the LPC-coefficient-based representation being one of the most widely used (Gray et al. 1980) [52]. One of the distortions, the so-called Itakura-Saito acoustical distortion (Gray et al. 1981)[53], (Rabiner & Schafer 1978)[101] has also been shown to be a statistical, KL divergence between signals represented by their LPC models. By considering the above-described extension of the traditional LPC model we arrive at a new, extended distortion measure, which is a generalization of the Itakura-Saito acoustical distortion measure.

As explained in the previous section, calculation of the KL divergence is accomplished by calculating the average of the loglikelihood ratio of the two models with respect to the probability distribution of one of the signals chosen as the reference signal. Using the average log likelihood expression for our model we arrive at

$$D[P||P'] = \left\langle \ln \frac{P(Y)}{P'(Y)} \right\rangle_P \tag{4.5}$$

which can be shown to be

$$D[P||P'] = -N\frac{1}{2}(\ln \frac{\sigma^2}{\sigma'^2} + 1 - \int_{-\pi}^{\pi} \frac{S_y(\omega)}{S_{y'}(\omega)}) \tag{4.6}$$

for the spectral case, and

$$\begin{aligned} D[P||P'] = -N \ln \frac{\lambda_0}{\lambda_0'} - N(\lambda_2\mu_2 - \lambda_2'\mu_2' \int_{-\pi}^{\pi} \frac{d\omega}{2\pi} \frac{S_y(\omega)}{S_{y'}(\omega)}) \\ -N(\lambda_3\mu_3 - \lambda_3'\mu_3' \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{d\omega_1 d\omega_2}{(2\pi)^2} \frac{B_y(\omega_1,\omega_2)}{B_{y'}(\omega_1,\omega_2)}) \end{aligned} \tag{4.7}$$

for the bispectral case. The μ_i and μ_i''s are the i'th order moments of the two (non-Gaussian) distributions. Note that there exists a convenient inverse relationship between the moments and the pdf parameters in the Gaussian case, which causes a canceling between the second order moment σ_2 and λ_2. This is no longer true for the bispectral case.

4.3.1 Discussion

The above KL divergence depends upon many parameters, such as the μ's and λ's of the excitation source and the spectral and bispectral patterns of the resulting signal. One must note, though, that the μ's and λ's are time-independent parameters whereas all temporal aspects appear in the spectral and bispectral ratios. Although we do not know yet how to estimate these parameters for the bispectral case, several important conclusions can be derived.
Let's assume P to be the complete description of the reference source and let P' be an approximate linear AR model that fits precisely the signal's spectrum.

$$S(\omega) = S'(\omega) = \frac{\mu_2}{| A(\omega) |^2} \tag{4.8}$$

The bispectra $B(\omega_1, \omega_2)$ of the real reference signal will, in general, be different from the bispectra of the AR model, with the linear model's bispectrum given by

$$B'(\omega_1, \omega_2) = \frac{\mu_3}{A(\omega_1)A(\omega_2)A^*(\omega_1 + \omega_2)} \tag{4.9}$$

The resulting distortion between the reference signal and it's AR model will contain no spectral component. The KL expression can now be rewritten as

$$D[P||P'] = C + C' \int_{-\pi}^{\pi} \frac{d\omega_1 d\omega_2}{(2\pi)^2} b(\omega_1, \omega_2) \tag{4.10}$$

where we have replaced the last integrand with the bicoherence index

$$b(\omega_1, \omega_2) = \frac{B(\omega_1, \omega_2)}{(S(\omega_1)S(\omega_2)S(-\omega_1 - \omega_2))^{1/2}} \tag{4.11}$$

and denoted by C and C' the various constants that appear in the equation. Thus we obtain the simple result that indicates that to the first approximation, the distortion between an AR model of a signal, which is spectrally equivalent to the original reference signal, and the reference signal itself, is proportional to the integral of the bicoherence index.

The bicoherence index

The bicoherence index in our model is the ratio of two *different entities*, the bispectrum and the power spectrum estimate derived from the parametric AR model. For high frequencies, the values of the bispectrum and the power spectrum are close to zero, and the quotient might take arbitrary values. Although the AR spectrum is *all pole* and has no zeros, we still need to avoid in the evaluation of the bicoherence index the division by small values of the AR spectrum. This can be achieved by thresholding the AR power spectrum to a minimal value, i.e. raising the low values of the tail of the spectrum to the threshold level. The threshold value is determined to be above the level of the background noise

We shall speak more of the bicoherence index in the next chapter when considering acoustic separation function. For now we shall merely mention that the bicoherence index was already used in chapter 2.3.4 when we demonstrated the effects of chorusing and reverberation on the bispectral contents of an audio signal.

4.3.2 Sensitivity to the Signal Amplitude

Another interesting result is obtained by applying scaling considerations. Multiplying a signal by a gain factor α results in a new model whose parameters will be related to the original model parameters in the following manner :

$$\begin{aligned} Y' &\to \alpha Y' \\ \lambda_0' &\to \lambda_0' + \ln \alpha \\ \lambda_i' &\to \frac{1}{\alpha^i}\lambda_i' \quad for\ i = 1, 2, ... \end{aligned} \tag{4.12}$$

$$\mu_i' \rightarrow \alpha^i \mu_i'$$

and the spectra and bispectra will change to

$$\begin{aligned} S' &\rightarrow \alpha^2 S' \\ B' &\rightarrow \alpha^3 B' \end{aligned} \tag{4.13}$$

The new divergence expressed in terms of the original spectra and bispectra is written now as

$$D[P||P_\alpha'] = C + \ln \alpha - \frac{1}{\alpha^2}\lambda_2'\mu_2' \int_{-\pi}^{\pi} \frac{S(\omega)}{S'(\omega)} - \frac{1}{\alpha^3}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \lambda_3'\mu_3' \frac{d\omega_1 d\omega_2}{(2\pi)^2}\frac{B(\omega_1,\omega_2)}{B'(\omega_1,\omega_2)} \tag{4.14}$$

$$= C + \ln \alpha - \frac{1}{\alpha^2}(OriginalSpectalFactor) - \frac{1}{\alpha^3}(OriginalBispectralFactor)$$

with C containing the gain insensitive constants. The acoustical significance of the above result is rather curious. Basically it says that the "distance" between signals is gain sensitive, with the scaling proportions given by the above equation. At the high gain limit, both the spectral and bispectral factors vanish, leaving the "distance" to be dependent solely upon the excitation source parameters and normalization factors. The "similarity" of any signal to another *very loud* signal is independent of the polyspectral properties of the signals. In the middle range, the behavior of the function is such that a higher weighting is given to the bispectral part at the lower gain values and vice versa the spectra are emphasized when the gain is incremented.

4.4 The Clustering Method

An interesting application of our new distortion measure is an acoustic taxonomy of various musical instrument sounds. Musical sounds are hard to cluster since it is difficult to model sounds as points in a low dimensional space. Though we limit ourselves to the stationary portion of the signal, a grouping of musical instruments, by the above similarity measure, appears in a rather interesting manner, as shown below.

In our approach, acoustic signals are treated as samples drawn from a stochastic source; that is, musical signals are represented by their corresponding statistical models. Each model is chosen to be the most likely in its parameter class to have produced the observed waveform. It is important to note that our relevant observables are the spectrum and the bispectrum of the signal, but in order to evaluate all the parameters of the distortion function, the $\{\mu_i, \lambda_i\}$'s must be estimated as well. Although this can be done in principle and an iterative solution procedure for this was described in the previous chapter, here we prefer to avoid the complete estimation of these parameters (4.7) by using the following arguments.

The distortion function is characterized by two integrals over ratios of the spectra and bispectra of the two signals involved. These are actually the only two quantities that are directly derived from the observed signal. We expect these two observables to contain all the relevant acoustic information and thus we substitute the original distortion function, Eq. (4.7), by the pair $(Sr_{y,y'}, Br_{y,y'})$ where $Sr_{y,y'} = \int_{-\pi}^{\pi} \frac{d\omega}{2\pi} \frac{S_y(\omega)}{S_{y'}(\omega)}$ and $Br_{y,y'} = \int_{-\pi}^{\pi} \frac{d\omega_1 d\omega_2}{(2\pi)^2} \frac{B_y(\omega_1,\omega_2)}{B_{y'}(\omega_1,\omega_2)}$ are the spectral and bispectral integral ratio expressions, respectively.

The idea is to represent a sound y by a collection of all pairs $(Sr_{y,y'}, Br_{y,y'})$, calculated with respect to all signals y' in our data set. This "trace" vector provides a signature of the particular signal y, characterizing it by the values of spectral and bispectral integral ratios, with respect to all the other signals. These signatures are expected to be similar for sounds with similar characteristics and serve as the basis for an acoustic classification, which in turn determines the underlying structure in the musical domain. By this transformation we have turned our data into a collection of points in a high dimensional space, the dimensionality being equal to the length of our signature vectors.[1] Using this vector-signature representation we explore our signal collection by means of a deterministic annealing hierarchical clustering approach. [?] (Rose et al. 1990) As in other fuzzy clustering methods, each point (a signature vector in our case) is associated *in probability* with all the clusters. The "annealed" system consists of a set of probability distributions for associating points with clusters.

The associations probability is controlled by a inverse temperature-like parameter, β. As β gets larger. the temperature is lowered and the associations become less fuzzy. At the starting point $\beta = 0$ and each point is equally associated with all clusters. As β is raised, the cluster centroids bifurcate through a sequence of phase transitions and we obtain an ultrametric structure of associations controlled by the temperature parameter.

4.5 Clustering Results

As described above, hierarchical clustering was performed iteratively by splitting the models (cluster centers) while raising β. This creates a sequence of partitions that give refined models at each stage, with clusters containing a smaller group of signature vectors, which are representatives of the signals themselves.

In order to test in detail the spectral and bispectral components of our data, we performed three separate clusterings: (1) a clustering with the spectral integral ratios vector only, (2) a bispectral integral ratios vector only, and (3) a combined clustering with vectors consisting of the complete $(Sr_{y,y'}, Br_{y,y'})$ set of pairs.

Our data set was taken from a collection of 31 sampled instruments sounds, from the Proteus sample player module. The results of spectral, bispectral and combined clustering are

[1] The signature vector could be calculated with respect to a small representative subset of our signal set. We chose to evaluate the complete distances matrix, i.e. each signal is represented by the distances from all other signals, and the space dimension is thus equal to the size of the data set.

shown in figures 4.1, 4.2, 4.3 respectively. Regarding the spectral and bispectral clusterings, we first note that many similar pairs of sounds, such as **Flutes, Clarinets** and **Trumpets** are grouped together in both clustering methods. It is interesting to notice also that the **Violin** and **Viola** are related (to different extents) in both methods, while the **Cello** is distant from both of them in the two clustering trees. Notice also that the **Bassoon** is close to the **Contra Bassoon** in the spectral tree, while being very distant in the bispectral case. Similarly, but in the other direction, the **Alto Oboe, English Horn** pair are close in the bispectral tree and distant in the spectral tree.

Qualitative analysis of these results suggests some possible interpretations of the behavior of spectral and bispectral clusterings. A rather "trivial" observation is that the spectral tree "grows" in the direction of spectral richness - the Flutes are close to the root, while the deeper branches belong to spectrally richer string and wind instruments. The bispectral tree exhibits clustering that puts at the deep nodes instruments which are normally identified as "solo" instruments, while the sounds closer to the root belong to a lower-register instruments of a more "supportive" character. It may be said that the hierarchical clustering splitting corresponds to some "penetrating" quality of the sounds, with **Flutes** and **Violin** at the extreme.

The combined spectral/bispectral results are difficult to interpret, although traces of the spectral and bispectral trees are apparent there. This finding would also require thorough statistical testing with human subjects, both musicians and non musicians.

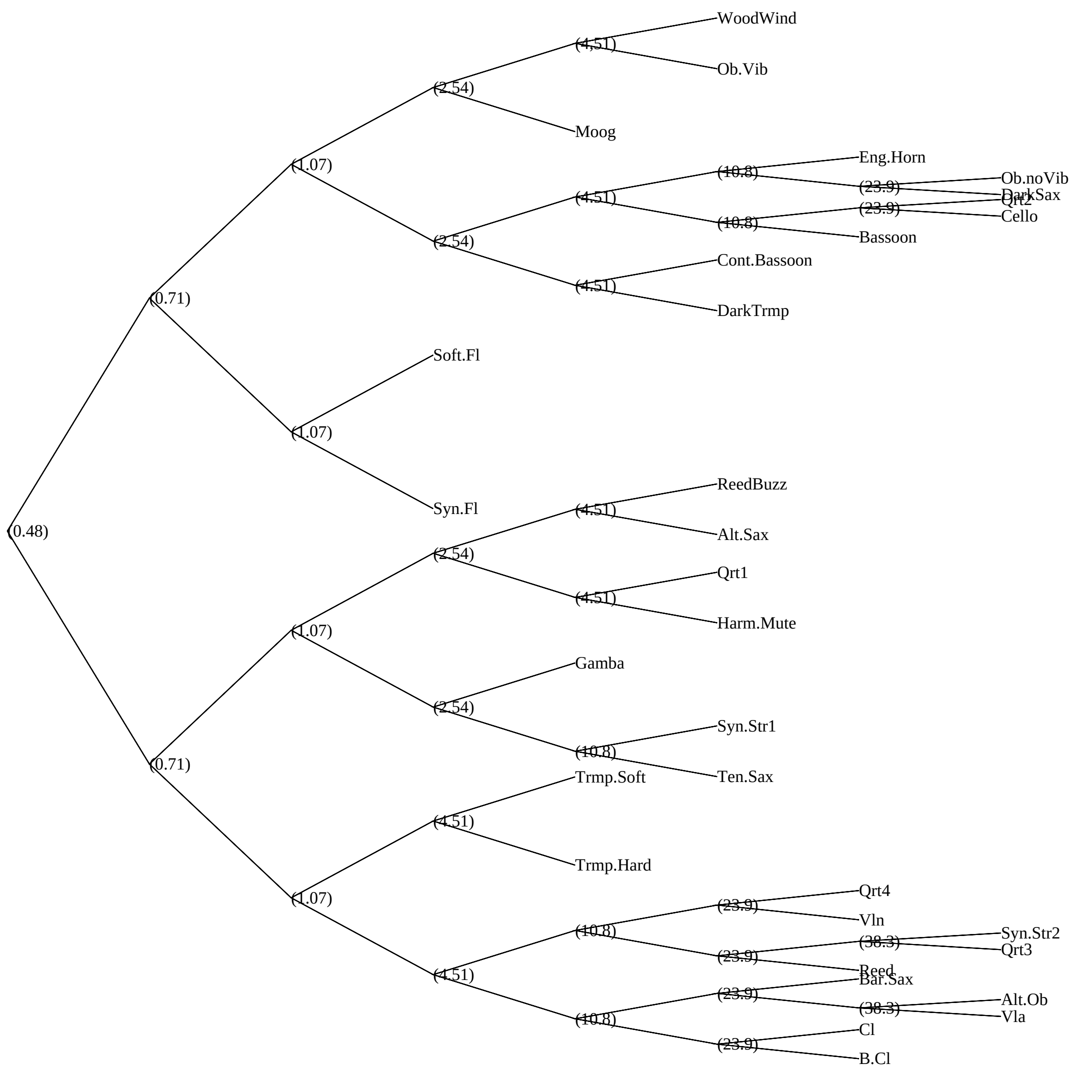

Figure 4.1: Spectral clustering tree. The numbers on the nodes are the splitting values of β.

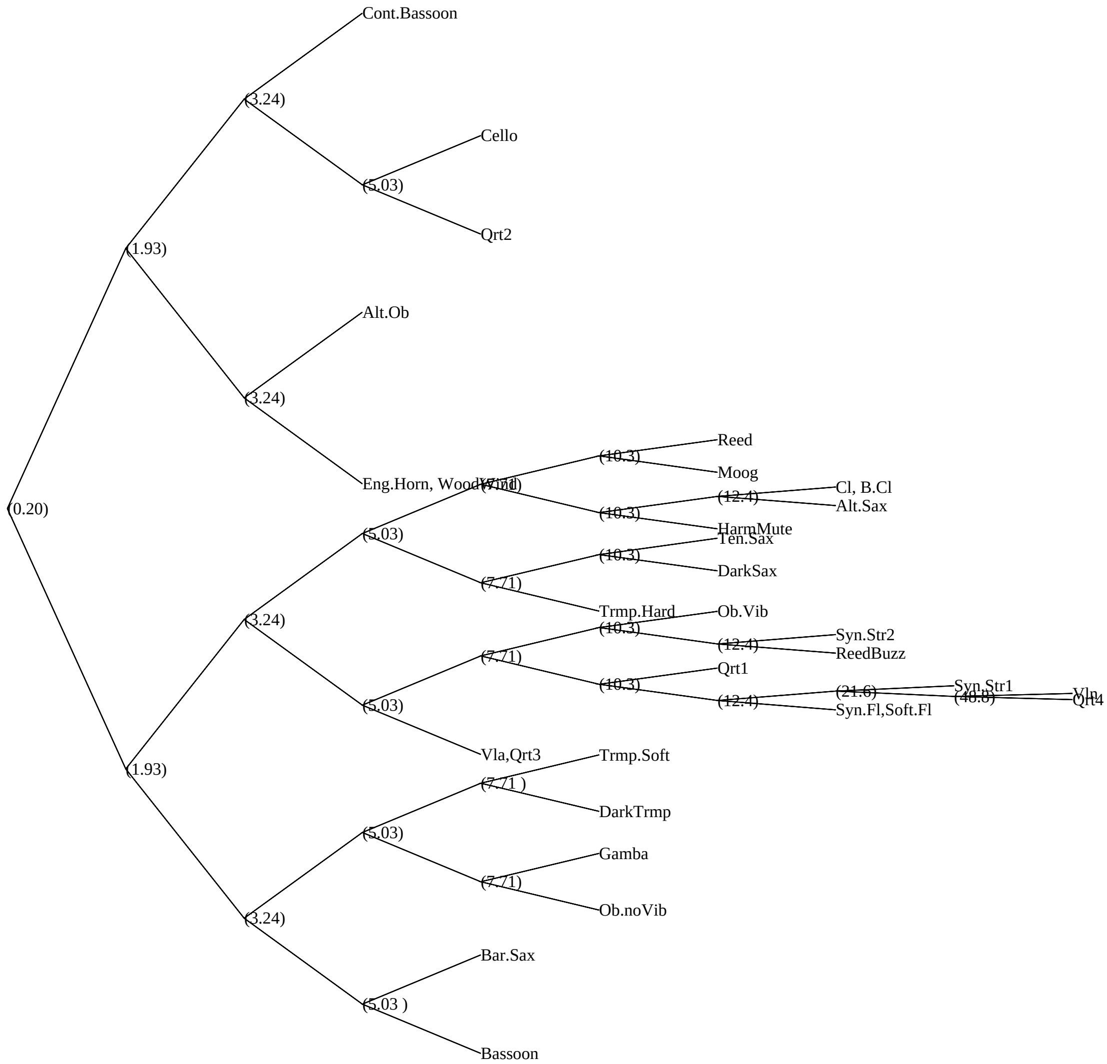

Figure 4.2: Bispectral clustering tree.

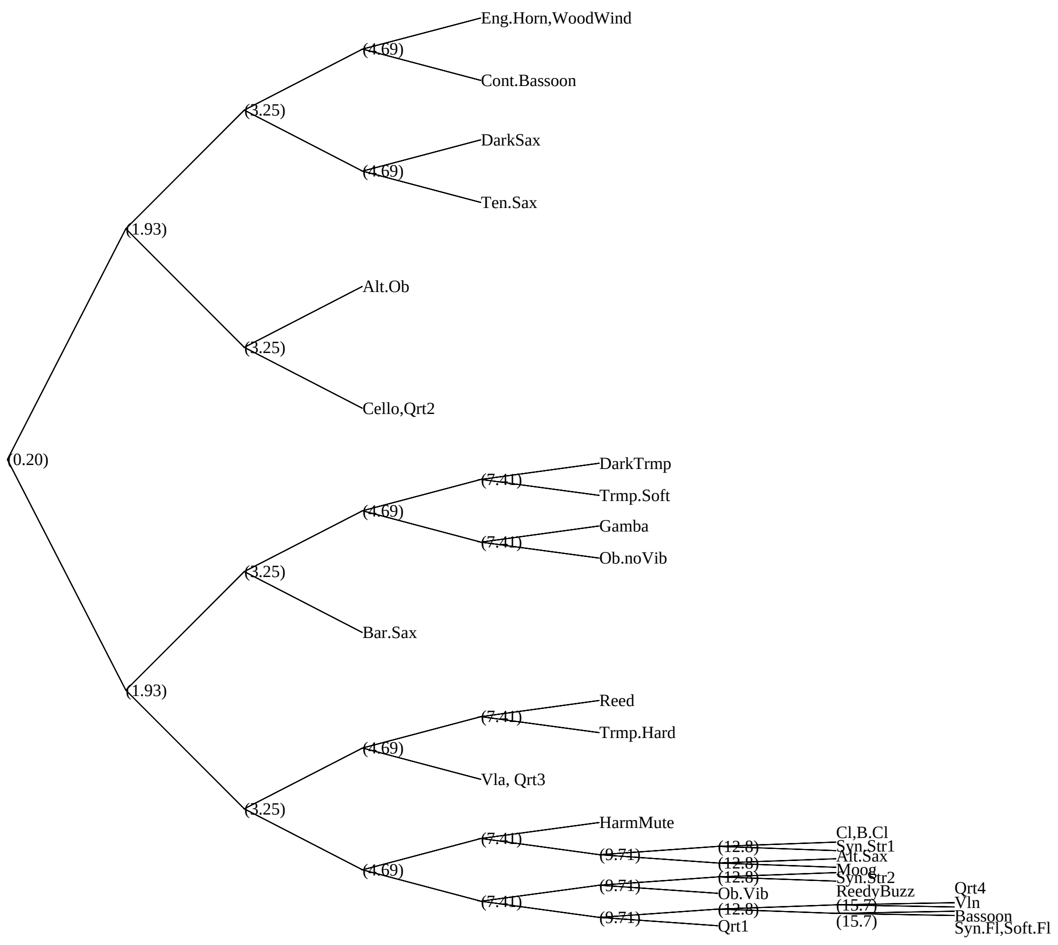

Figure 4.3: A combined spectral and bispectral clustering tree.

Part III

Analysis of the Signal Residual

Chapter 5

Estimating the Excitation Signal

In our work so far we have considered a Signal Model that treated sound as an output signal of a source-filter model, with the source being a non-Gaussian white noise excitation. The main fault of this approach, in our view, is the loose modeling of the sound source by the exponential probability distribution derived from maximum entropy principles.

In this part of the work we take a closer look at the HOS properties of the excitation signal. We consider a Spectral Model for the excitation signal and try to identify the structure of the excitation signals based on the information present in the higher order statistics (HOS) of the stationary portion of the sound.

Spectral matching of a signal a represents the sound as a white process passing through a linear filter whose role is to introduce the second order correlation properties into the signal and thus shape its spectrum. Stable linear systems subject to a Gaussian input are thus completely characterized by the covariance function of their output and hence their power spectrum. In case of a non-Gaussian input signal, higher order cumulants appear in the output. Moreover, if the relationship between the output and input signals is not linear, the statistics of the output signal will not be Gaussian even if the input is normal.

Given a signal X_t, the bispectrum $B_X(\omega_1, \omega_2)$ gives a measure of the multiplicative non-linear interaction between frequency components in the signal

$$B_X(\omega_1, \omega_2) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} C_X(m, n) e^{-i(\omega_1 m + \omega_2 n)} \tag{5.1}$$

with $C_X(m, n) = E(X(t)X(t+m)X(t+n))$ the third order cumulant of zero mean process $X(t)$. Assuming that our signal is generated by a nonlinear filtering operation satisfying a Volterra (Priestley 1989)[100],(Pitas & Venetsanopoulos 1990)[99] functional expansion, we write

$$X_t = \sum_{u=0}^{\infty} h_1(u)U(t-u) + \sum_{u=0}^{\infty} \sum_{v=0}^{\infty} h_2(u, v)U(t-u)U(t-v) + \ldots \quad . \tag{5.2}$$

In the linear case, the model is completely characterized by the *transfer function*

$$H_1(\omega) = \sum_{u=0}^{\infty} h_1(u) e^{-i\omega u} \tag{5.3}$$

while

$$H_2(\omega_1, \omega_2) = \sum_{u=0}^{\infty} \sum_{v=0}^{\infty} h_2(u, v) e^{-i(\omega_1 u + \omega_2 v)} \tag{5.4}$$

represents the kernel that weights the contribution of the components at frequencies ω_1, ω_2 in the input signal U_t to the component at frequency $(\omega_1 + \omega_2)$ in the output X_t (and so forth for higher order kernels).

We say here that a signal X_t is linear if the system that created it has no kernels of order higher than two. This notion of linearity implies that no multiplicative interactions occur between the frequency components in the stationary portion of the signal and the principle of superposition applies in the sense that the resultant signal is obtained as a sum of frequencies with appropriate spectral amplitudes. Although the more common and obvious manifestations of non-linearity appear in the dynamic behavior of the sound (such as the non-linear dependence of the signal properties upon its amplitude), the bispectral properties in the sustained portion can quantify nonlinearities as well. [1]

Schematically, our approach could be summarized by the following Figures 5.1, 5.2. The signal is regarded as some kind of stochastic pulse train X_t passing through a linear filter. In the case of a pitched signal (i.e. a signal with a discrete spectrum), Gaussianity of the input signal is obtained either for inharmonic signals[2] or harmonic sounds with statistically independent partials. The independence is obtained for instances such as presence of random independent modulations at each harmonic [39].

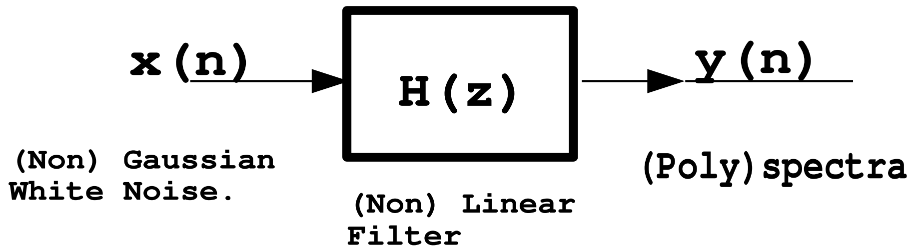

Figure 5.1: Signal Model - Stochastic pulse train passing through a filter.

In the following analysis we shall address the questions of Gaussianity and linearity based on the higher order statistics of X_t without attempting to estimate the kernels or other system

[1] Although it is not clear whether these are the same non-linearities or if the same non-linear mechanisms are active in the transitory and sustained portions of a signal.

[2] Spectrally, the signal is composed of an incommensurate set of sinusoids, i.e. no partial sum of frequencies in the set corresponds to an already-existing frequency.

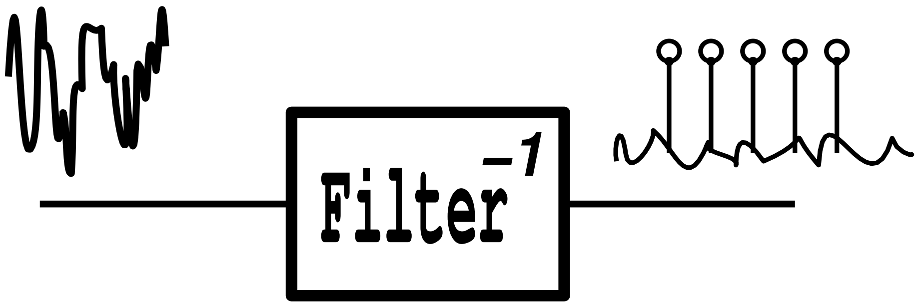

Figure 5.2: Estimation of the excitation signal (residual) by inverse filtering.

parameters. Since we have no underlying parametric model for our sounds, this chapter is more concerned with the analysis of the structure of musical signals; the musical significance discussed in section (5.6).

5.1 Examples of Real Sounds

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

The effect of the spectral envelope, which contains the information about the amplitudes of harmonics is removed by inverse filtering of the signal through a filter derived from its LPC model. This action statistically amounts to low order decorrelation of the signal and leaves all harmonics at equal amplitudes - a situation reminiscent of a pulse train excitation signal[3].

Before going further into analyzing and modeling (in the next chapter) the excitation function, we would like to demonstrate the bispectral signatures of several musical signals and of their respective residuals. In figure **??** 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. The strong presence of the high harmonics in the residual significantly affects the bispectral contents. Notice that the Cello residual has only a few peaks near the origin.

How do we look at these signals? First, we must be aware of the symmetries pertinent in the definition of bispectrum. In the sixfold symmetry it is sufficient to consider a lower triangular 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.

[3]In the next chapter we will show how the HOS properties of the excitation (residual error) are related to frequency deviations caused by a frequency modulating jitter acting on the harmonics of a pulse train like signal. In this case we obtain a statistical interpretation of the moments as probabilities for maintaining harmonicity among groups of partials.

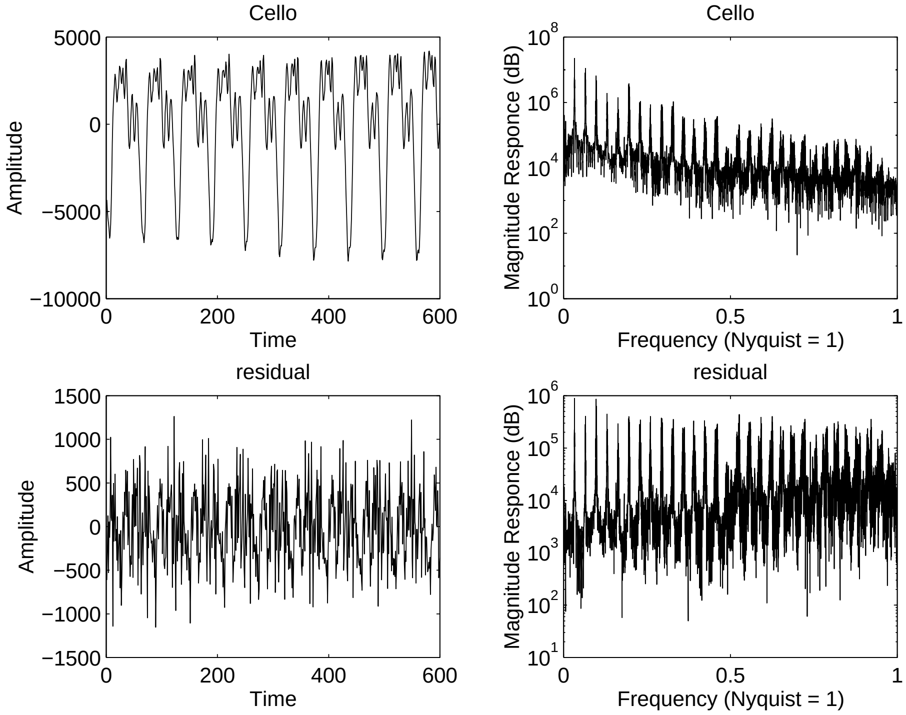

Figure 5.3: The original Cello signal and its spectrum (yop). The residual signal and its respective spectrum (bottom). Notice that all harmonics are present and they have almost equal amplitudes, very much like the spectrum of an ideal pulse train. The time domain signal of the residual does not resemble pulse train at all.

Below 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 chamber spectral shape, but it also has the following important properties to be shown below:

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

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

As could be seen from the plots of the residual bispectrum , the overall area under the three graphs is significantly different. It is interesting to note that investigation of the moments of decorrelated signals were used in the analysis of texture in images [45][131].

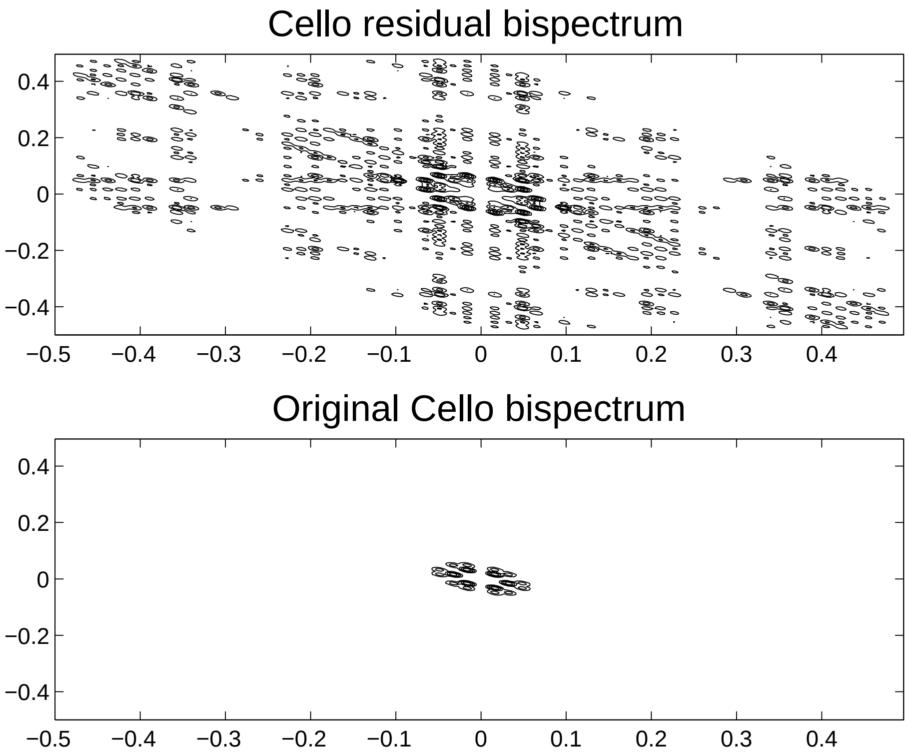

Figure 5.4: The bispectrum of a Cello residual signal (top) and the bispectrum of the original Cello sound (bottom). See text for more details.

5.2 Higher Order Moments of the Residual Signal

Since the non-linear properties of musical signals are attributed mainly to the properties of the excitation signal, the investigation of the higher order statistical properties of the residual seems a natural next step beyond the spectral modeling of the linear resonator. In general, any process X_t could be represented as a sum of deterministic and stochastic components and the Wold decomposition teaches us that any signal can be optimally represented in terms of a filter (linear predictor) acting on a white input noise, with the optimality being with respect to the second order statistics of a signal.[4] Let us assume for the moment that the musical signal X_t could be described by a non-Gaussian, independent and identically distributed (i.i.d.) excitation signal U_t passing through a linear filter $h(n)$.

$$X_t = \sum_{u=0}^{\infty} h(u)U(t-u) \tag{5.5}$$

In such a case, the spectra and bispectra of X_t are equal to

$$S_X(\omega) \quad = \quad \mu_2 |H(\omega)|^2 \tag{5.6}$$

[4] Since in the gaussian case "white" implies independent, the Wold optimality is maintained only then.

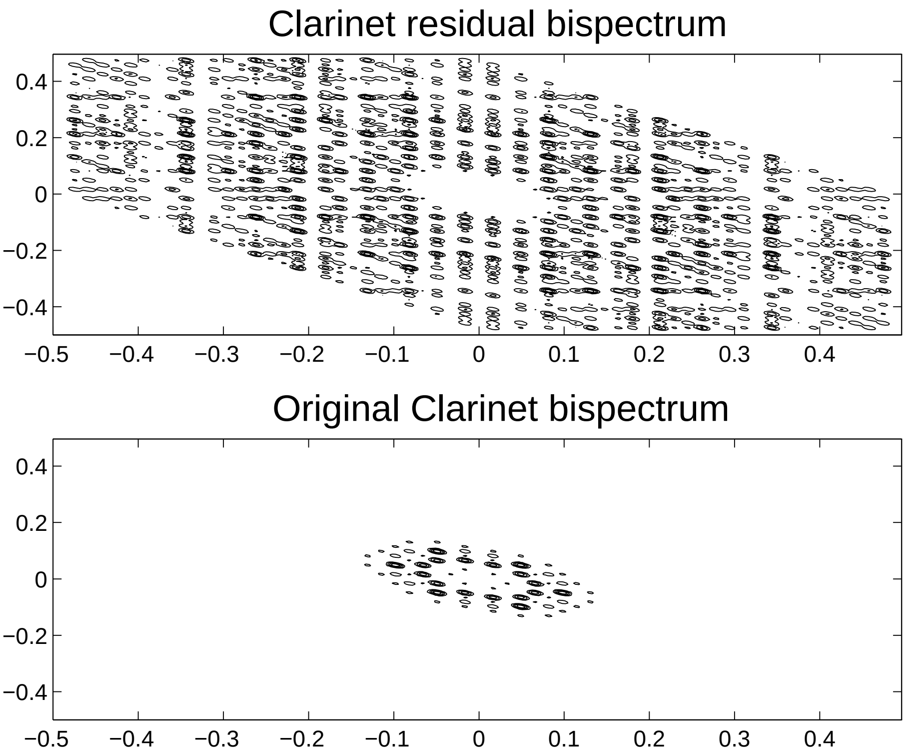

Figure 5.5: The bispectrum of a Clarinet residual signal (top) and the bispectrum of the original Clarinet sound (bottom). See text for more details.

$$B_X(\omega_1, \omega_2) \quad = \quad \mu_3 H(\omega_1) H(\omega_1) H^*(\omega_1 + \omega_2)$$

(with a similar equation for the trispectra), with μ_2, μ_3 (and μ_4) being the second, third (and fourth) order cumulants of U_t. For convenience we shall assume $\mu_2 = 1$. Although exact solution of the filter parameters in the non-linear case is problematic [5][38], we assume that it could be effectively estimated based on spectral information only. Taking a spectrally matching filter $\hat{H}(\omega)$

$$S_X(\omega) = |\hat{H}(\omega)|^2 \tag{5.7}$$

the inverse filtering of X_t with $\hat{H}^{-1}(\omega)$ gives a residual V_t. Since V_t is a result of passing a non Gaussian signal X_t through a filter $\hat{H}^{-1}(\omega)$, its bispectrum is

$$B_V(\omega_1, \omega_2) = \frac{B_X(\omega_1, \omega_2)}{\hat{H}(\omega_1)\hat{H}(\omega_2)\hat{H}^*(\omega_1 + \omega_2)} \quad . \tag{5.8}$$

Taking the absolute value of $B_V(\omega_1, \omega_2)$ and using the spectral matching property of $\hat{H}$, we obtain the equivalence of the right-hand equation to the definition of the bicoherence index for the original signal X_t.

[5]Such a solution must be optimal with respect to the spectral, bispectral and all higher order statistical information and requires complete knowledge of the noise distribution parameters.

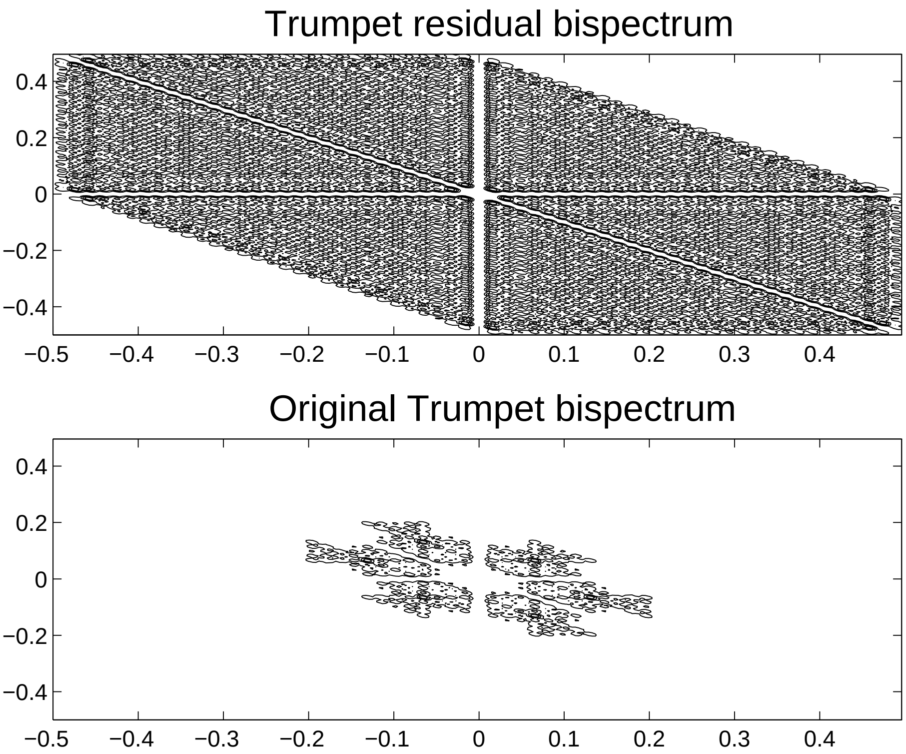

Figure 5.6: The bispectrum of a Trumpet residual signal (yop) and the bispectrum of the original Trumpet sound (bottom). See text for more details.

The residual signal would ideally be a decorrelated version of the original sound for the case when the excitation is a white noise, i.e. a signal with a continuous flat spectrum. In such a case, V_t is an i.i.d. process - its cumulants are zero at all times other then zero. In the case of pitched excitation, the residual signal is not ideally decorrelated[6] and it contains correlations at time lags that correspond to multiples of the pitch period and thus respectively demonstrates peaks in the polyspectra. Inverse Fourier of the bispectra at zero lag times gives

$$C_V^3(0,0) = \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} B_V(\omega_1, \omega_2) \tag{5.9}$$

so that the cumulants at time zero are integrals over the bispectral plane.

We claim that for excitation signals of both the continuous (excitation noise) and the discrete (pitched) spectra types, the cumulants at time zero are significant features. Obviously, for i.i.d. white noise excitation, the first order amplitude histogram of the decorrelated signal is a plausible estimate of the excitation distribution function, and as such it is completely characterized by its moments. In the pitched case, zero time lag cumulants which

[6]The resonant chamber of the musical instrument should be modeled by a low order filter, and its inverse (decorrelating) filter is a high boost that flattens out the spectral envelope but retains the “comb-like” structure of the signal.

are integrals over polyspectral planes of various orders (3 and 4 in our study) appear to be important quantities as well, although they can no longer characterize the excitation distribution function which now contains pairwise and higher order dependencies.

Assuming ergodicity, the cumulants can be calculated more conveniently as time averages

$$\begin{aligned} C_V^3(0,0) &= \lim_{T\to\infty} \frac{1}{T}\int_0^T V^3(t)dt \qquad (5.10)\\ C_V^4(0,0,0) &= \lim_{T\to\infty} \frac{1}{T}\int_0^T V^4(t)dt - 3\cdot C_V^2(0) \ . \end{aligned}$$

and these results could be used as estimates of the actual third and fourth order cumulants $\mu_3,\ \mu_4$ of the input U_t at zero time lag.

Instead of working with the cumulants directly, we 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 $\mu_4 = E(x-Ex)^4$. Since we are dealing with zero mean processes, the third order moment is equivalent to the third order cumulant. For the fourth order we subtract 3 from the kurtosis to get the variance normalized cumulant [56].

5.3 Tests for Gaussianity and Linearity

As noted above, if a process is Gaussian, all its polyspectra of orders higher than second are zero. Hence, a non-zero bispectrum could be due to either of the following factor:

1. The process conforms to a linear model but the input signal statistics are not Gaussian.
2. The process conform to a non-linear model, regardless of the input signal statistics.

Case (1) is examined by testing the null hypothesis that the bispectrum is zero over the entire bispectral plane.
Case (2) is examined by using the expression for bispectra of a signal $X(t)$ assuming it has a linear representation

$$B_X(\omega_1,\omega_2) = \mu_3 H_1(\omega_1)H_1(\omega_2)H_1^*(\omega_1+\omega_2). \qquad (5.11)$$

The spectral density function of the process is given by

$$S(\omega) = \mu_2|H(\omega)|^2 \qquad (5.12)$$

hence the bicoherence function

$$b(\omega_1,\omega_2) = \frac{|B_X(\omega_1,\omega_2)|}{\sqrt{S(\omega_1)S(\omega_2)S(\omega_1+\omega_2)}} = \text{constant} \quad (\text{all}\ \ \omega_1,\omega_2) \qquad (5.13)$$

The test for linearity is based on replacing the $B_X(\omega_1, \omega_2)$ and $S(\omega)$ with their sample estimates and testing the constancy of the sample values of the bicoherence index over a grid of points.

5.4 Results

Returning to real musical signals, we evaluate these moments by empirically calculating the skewness and kurtosis of various musical instrument sounds. These moments were 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 "cumulants space" places the instrumental groups in "orbits" of various distances around the origin. This graphical representation also suggest a simple distance measure which could be used for classification of these signals.

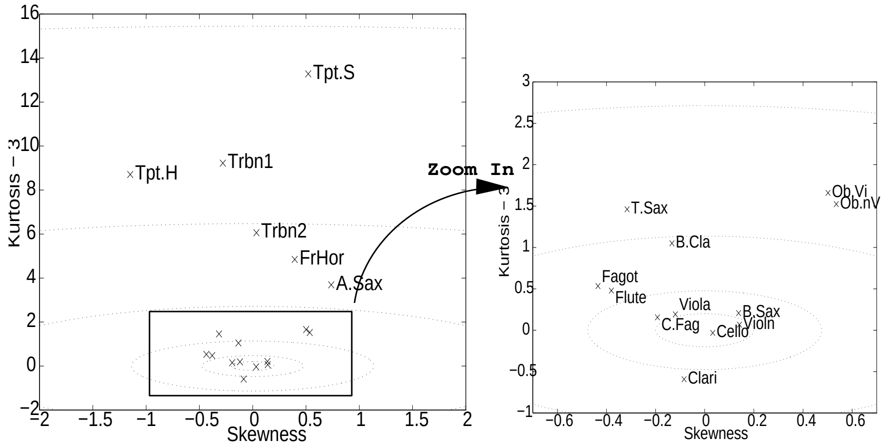

Figure 5.7: Location of sounds in the 3rd and 4th cumulants plane. The value 3 is substructed from the kurtosis so that the origin would correspond to a perfect gaussian signal. Left: All 18 signals. Brass sounds are on the perimeter. Right: Zoom in towards the center containing strings and surrounded by woodwinds.

Testing for Gaussianity and Linearity (Hinich 1982)[61], (Mendel et al. 1995)[85] gives qualitatively similar results. One must note that the cumulants space representation can not resolve the Gaussianity/Linearity problem since the calculation of the cumulants involves integrating over the bicoherence plane and no information about the constancy of the bicoherence values is directly available. One finds, however, an intuitive correlation between the Gaussianity and non-linearity properties and the magnitudes of the cumulants.

In the Gaussianity test, the null hypothesis is that the data have zero bispectrum. Since the estimators are asymptotically normal, our test statistic is approximately chi-square and the computed probability of false alarm (PFA) value is the probability that a value of the chi-squared random variable will exceed the computed test statistic. The PFA value indicates the false alarm probability of accepting the alternative hypothesis, that the data have a non-zero bispectrum. Usually, the null hypothesis is accepted if PFA is greater than 0.05 (i.e. it is risky to accept the alternative hypothesis).

In the Linearity test, the range R_{est} of values of the estimated bicoherence is computed and compared to theoretical range R_{theory} of a chi-squared r.v. with two degrees of freedom and non-centrality parameter λ. The parameter λ, which is proportional to the mean value of the bicoherence, is also computed. The linearity hypothesis is rejected if the estimated and theoretical ranges are very different from one another.

Since the Linearity test assumes a non-zero bispectrum, it should be considered reliable only if the signal passed the earlier Gaussianity test. In the following figure we present the results of Gaussianity and Linearity tests plotted so that the X-axis represents the Gaussianity PFA value and the Y-axis is the difference between the estimated and theoretical ranges.

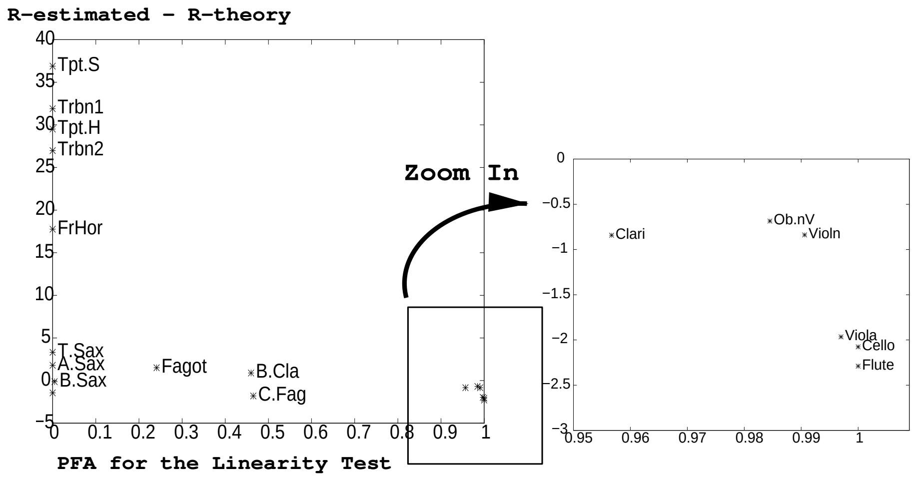

Figure 5.8: Results of the Gaussianity and Linearity tests. PFA > 0.05 means Gaussianity. R-estimated >> R-theory means non-Linearity. The deviations in the Y-axis of the right figure should not be considered since the Linearity test is unreliable for signals with zero bispectrum (Gaussian).

5.5 Probabilistic Distortion Measure for HOS features

In this study we have provided qualitative evidence that the cumulant description of the first-order amplitude histogram of the decorrelated acoustic signal is an important feature of musical instruments. Considering two classes P_1 and P_2, the Bhattacharyya (B) distance between the probability densities of features of the two classes is

$$B(P_1, P_2) = -ln \int ([p(\hat{g}|P_1)p(\hat{g}|P_2)]^{1/2} dx \tag{5.14}$$

where $\hat{g}$ denotes the HOS feature vector with conditional probability $p(\hat{g}|P_i)$ for class i. It is known that asymptotically the distribution of the kth order statistic $C_V^k(0,0)$ is normal around its true value. For Gaussian densities, the B distance becomes

$$\begin{aligned} B(P_1, P_2) &= \frac{1}{8}(g_1 - g_2)^T(\frac{\Sigma_1 + \Sigma_2}{2})^{-1}(g_1 - g_2) \\ &+ \frac{1}{2} ln \frac{\frac{1}{2}|\Sigma_1 + \Sigma_2|}{(|\Sigma_1|\ |\Sigma_2|)^{1/2}}, \end{aligned} \tag{5.15}$$

where g_i and Σ_i represent the feature mean vector and covariance matrix of the classes.

According to our cumulants space representation, the instrumental families are resident on orbits of similar distance from the origin. In such a case, the grouping of timbres into families should be according to the relative B distances of the instruments from a normal gaussian signal $N(0,1)$. Then, the B distances become

$$B(P, N) = \frac{1}{8} g^T \Sigma^{-1} g \tag{5.16}$$

where the covariance matrix must be evaluated for each instrumental class.

5.6 Discussion

Our main result is that the cumulant space representation suggests an ordering of the signals according to common practice for instrumental families of string, woodwind and brass instruments. The grouping is with respect to the distances of the signals from the origin which corresponds to an ideal Gaussian (and hence linear) signal. The string instruments, being the closest to the origin, are the most Gaussian; and similarly, the Linearity test applies to them. The second orbit band is occupied by woodwind instruments which are classified as Linear but Non Gaussian in the Hinich tests. The perimeter orbits contain brass sounds. These sounds also satisfy the non-Linearity tests with increasing level of confidence for signals with stronger moments.

Considering the technical application of these results, there are two approaches to the creation of a musically meaningful signal with Gaussian or non-Gaussian properties, which we are currently investigating. In the first approach (Dubnov et al. 1996) [40] we assume that we can represent our input signal as a sum of sinusoidal components with constant amplitudes. Application of random jitter to these components effectively reduces the correlations between the constituent frequencies, and the resultant signal approaches Gaussianity in the sense that its higher order cumulants vanish. This, of course, does not imply that the signal becomes audibly a noise, since second order correlations remain arbitrarily long and they determine the spectrum and the perceived pitch. Another approach is passing a Gaussian noise through a comb filter, which is a poor man's approximation of a bowing of a Karups-Strong string.

Since Gaussian statistics are preserved under linear transformation, Gaussianity of the output signal indicates that the input signal is Gaussian, too. In the cases where the signal is linear but non-Gaussian, the non-Gaussianity originates in the higher order cumulants of the input signal. In the sum-of-sinusoids interpretation this might be understood as a sum of sinusoids with a constant level of correlation between all components. This effect could be achieved by passing a Gaussian signal through a memoryless non-linear function (Grigoriu 1995) [56]. Strictly speaking, non-Gaussianity is a certain sort of non-linearity where the non-linear properties are effective at a single time instance only while all memory is assigned to the linear component exclusively.

5.6.1 Musical Significance of the Classification Results

For conclusion, it is apt to discuss in brief the musical significance of the orchestral family classifications. Beyond the traditional and instrument builders' reasons for this classification, the categorization into instrumental families has a strong evidence in orchestration practice. Scoring of chords and harmonic (homophonic) writing achieves a better "blend" for instruments of the same family. For a combination of timbres Adler (1989)[1] suggests, *"When considering doubling notes, try to find instruments that have an acoustic affinity for one another."* This very idea of acoustic affinity had been formulated in our work as a problem of representation and statistical distance in HOS feature space; the idea of using cumulants as *acoustic features* is novel in audio signal analysis.

To demonstrate the correspondence between our features and the orchestration practice, it is interesting to note, for instance, that the French Horn is located between woodwinds and brass timbres in the moment space representation. This is very much in accordance with the orchestration handbooks' description of the instrument. To quote Piston (1989) [98], *"The horns form an important link between brass and woodwind. Indeed, they seem to be as much a part of the woodwind section as of the brass, to which they belong in nature."*

One must mention the work of Sandell (1991)[115], who considered temporal and spectral factors that determine blend. Since our method considers the higher order statistical (spectral) properties of the decorrelated signal, it might be worthwhile to consider the relationship between our results and the harmonicity and pitch deviations properties that were considered by Sandell. These factors definitely have an impact on the polyspectral contents of the signal, as was discussed in earlier work (Dubnov et al. 1995)[39].

To conclude the discussion we would like to note that he above results were also tested on a larger (> 50) suite of sounds with comparable results. We note also that since we are dealing only with stationary sounds, we neglect any non stationary or transitory phenomena that do not fall in the domain or could not be considered microscopic stationary fluctuations at the sustained portion of a sound. These findings indicate interesting principles for the classification of music instruments, but as we constantly remark, all of our results require comparative psychoacoustic experimentation.

5.6.2 Meaning of Gaussianity

In order to understand better the influence of jitter on a perfectly periodic sound, we would like to consider briefly the statistical properties of non-harmonic 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

$$\begin{gathered} < x(t) \cdot x^*(t+\tau) > = < (\textstyle\sum_{j=1}^{Q} e^{i\omega_j t})(\sum_{k=1}^{Q} e^{i\omega_k t}) > \\ = lim_{\Theta \to \infty} \frac{1}{2\Theta} \textstyle\sum_{j,k=1}^{Q} (\int_{-\Theta}^{\Theta} e^{i(\omega_j - \omega_k)t} dt) e^{-i\omega_j \tau} \\ = \textstyle\sum_{j=1}^{Q} e^{-i\omega_j \tau} \end{gathered} \tag{5.17}$$

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

$$\begin{aligned} & < x(t) x(t+\tau_1) x^*(t+\tau_2) > = \\ & \lim_{\Theta \to \infty} \frac{1}{2\Theta} \sum_{j,k,l=1}^{Q} (\int_{-\Theta}^{\Theta} e^{i(\omega_j + \omega_k - \omega_l)t} dt) \\ \cdot \quad & e^{-i\omega_k \tau_1} e^{-i\omega_l \tau_2} \end{aligned} \tag{5.18}$$

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. This is easily demonstrated for $\tau = 0$ by looking at the histograms of harmonically and non-harmonically related signals.

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

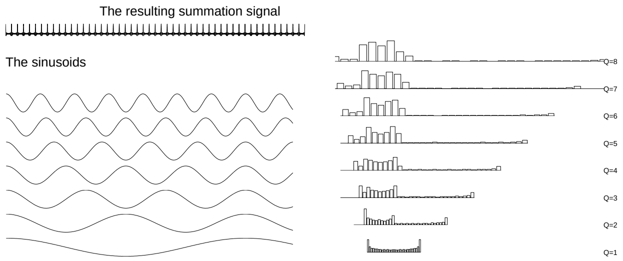

Figure 5.9: Harmonic signal (left) and its histogram (right).

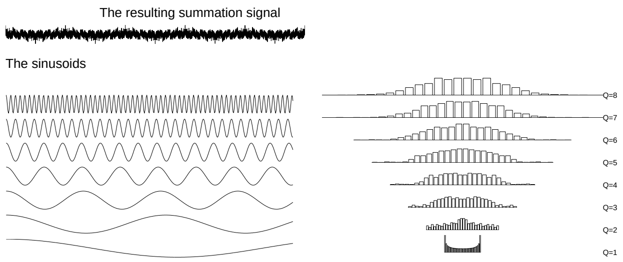

Figure 5.10: Inharmonic signal (left) and its respective histogram (right).

Chapter 6

"Jitter" Model for Resynthesis

6.1 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 consists of a sum of equal amplitude cosines, with a random jitter applied to its harmonics.

$$x(t) = \sum_{n=1}^{Q} cos(2\pi f_0 n \cdot t + Jitter(t)) \tag{6.1}$$

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 1995)[56].

6.2 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 harmonic triplets (quadruples) of the signal partials. Basically, the application of frequency modulating jitter to harmonically related partials destroys the harmonicity when the jitters are non-concurrent (independent) at each partial. Harmonicity is preserved, in contrast, when the same jitter (concurrent modulation) is applied to the partials. The vanishing of signal moments indicate that the signal obeys Gaussian statistics. The relationship 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

$$x(t) = \sum_{n=1}^{Q} cos((\omega_0 n + \Delta_{i,n}) \cdot t) \tag{6.2}$$

.

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

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

$$\begin{aligned} m_3 &= \frac{1}{T}\int_0^T x^3(t)dt \qquad (6.3)\\ &= \frac{1}{(2\pi)^2}\int_{-Q}^{Q\omega_0}\int_{Q}^{Q\omega_0} X_i(\omega)X_i(\omega')X_i^*(\omega+\omega')d\omega d\omega' \\ &\approx \int\int\sum_{n=1}^{Q}\sum_{m=1}^{Q}\delta(\omega-(n\omega_0+\Delta_{n,i}))\cdot \\ &\quad \delta(\omega'-(m\omega_0+\Delta_{m,i}))\cdot \\ &\quad \delta((\omega+\omega')-((n+m)\omega_0+\Delta_{(n+m),i}))d\omega d\omega' \end{aligned}$$

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 to the fourth order moment and its respective trispectrum representation in the frequency domain.

6.2.1 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. [1]

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

[1]This assumption is based on empirical observations of bispectrum plots of real musical signals (such as those demonstrated in figure 5.6) that demonstrate a stronger bispectrum at low bifrequencies and a decay in bispectral amplitude for higher partials.

to partial numbers below Q_{eff}. For the bispectral case a lattice of delta functions exists for partials (n, m) over the bifrequency triangle

$$0 < n < Q_{eff}, \quad 0 < m < Q_{eff}, \quad n + m < Q_{eff} \tag{6.4}$$

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 trickier argument for the trispectrum reveals that the area of the tetrahedron limited by

$$\begin{gathered} 0 < n < Q_{eff},\ 0 < m < Q_{eff}, \\ 0 < l < Q_{eff},\ n + m + l < Q_{eff} \end{gathered} \tag{6.5}$$

equals $\frac{1}{6}Q_{eff}^3$. In the trispectral case one must take also into account the number of possible choices of triplets, which gives a factor 3 to the above. An additive factor of $3Q^2$ also appear because for $Q_{eff} = 0$ there are still peaks due to cancellations of frequencies on the diagonal planes. [2]

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 γ_3 and kurtosis γ_4 to the effective number of coupled partials Q_{eff} are

$$\begin{aligned} \gamma_3 &= \frac{\frac{1}{2}Q_{eff}^2}{Q^{3/2}} \\ \gamma_4 &= \frac{\frac{1}{2}Q_{eff}^3}{Q^2} + 3 \end{aligned} \tag{6.6}$$

6.2.2 Simulation Results

This theoretical result was tested on synthetic signals 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* software, with the parameters set in accordance with the jitter synthesis method reported by McAdams[86]. The jitter depth was taken to be 0.01 of the partial frequency and the jitter spectrum was approximately shaped with a -10 db cutoff at 30 Hz and a second cutoff to zero at 150 Hz. The signals were generated at a pitch of middle C; working with a 16KHz sampling rate we obtain total of 30 harmonics (Q=30).

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

---

[2] 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), $\omega_1 = -\omega_2$. There are three choices for such a pair.

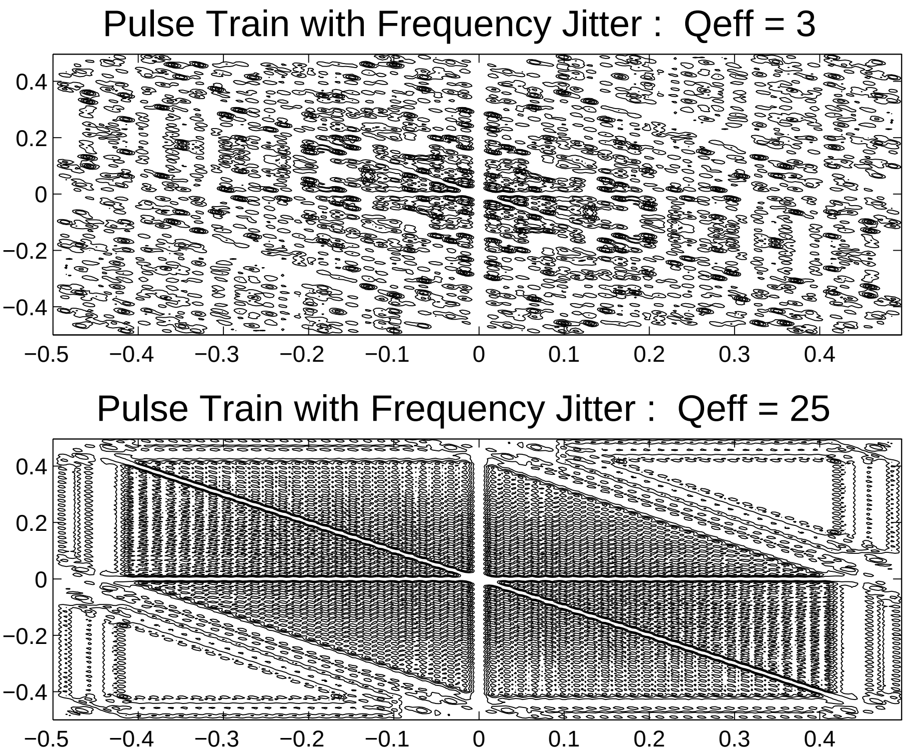

Figure 6.1: Bispectra of two synthetic pulse train signals with frequency modulating jitter. Top: Qeff = 3. Bottom: Qeff = 25. Notice the resemblance of the two figures to the bispectral plots of the cello and trumpet in figure (2). The Qeff values were chosen especially to fit these instruments according to the skewness and kurtosis values for the instrument in figure (3).

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

6.3 Summary

6.3.1 The Findings of the Model

In this chapter 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 showed 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 purposes 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 seen graphically using a cumulant space representation where the groups appear in 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, 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 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 also note that we are dealing with stationary sounds only and neglect any non-stationary or transitory phenomena that could not be considered microscopic stationary fluctuations at the sustained portion of a sound.

6.3.2 Musical Significance

This study, as we have seen, focused on a specific phenomenon that contributes to timbre. The timbre, although extremely complex from the acoustic viewpoint, is perceived by the listener as an inseparable event. Nevertheless, one can detect some microscopic occurences even within the timbre, and their amplification will produce a border area between timbre (with a defined pitch) and noise and a border between timbre and texture. General verbal characterizations of sounds such as "focused", "synthetic" versus "diffused", "chorused", etc. are caused by the very same random fluctuations at the microscopic level. A more precise formulation of the phenomenon places it on the axis between concurrence and non-concurrence with respect to the random deviations in the frequencies of the harmonics. The principles behind this phenomenon - border areas; concurrence and non-concurrence; fusion/segregation; determinism and uncertainty - are at the basis of musical activity on all of its stages and in all levels of the musical material, even in the characteristics of musical style. This research thus shows 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 goals of this work are reciprocal: the abovementioned basic principles help us understand the hidden microscopic phenomena, and on the other hand, the research into these phenomena sheds a new light on the principles. Moreover, this reciprocal relationship is important for musical creation todays, when we have placed emphasis on the momentary events related to timbre and texture instead of the interval parameter and its derived schemes that ruled musical organization in tonal music.

Concurrence /Non-Concurrence

This term refers to the relationship among units and parameters. For instance, 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 numerous revelations - it increases the complexity and the uncertainty and even creates tension; as such it becomes an essential parameter in the rules of musical organization and characterization of style. (Some of the rules of Palestrina counterpoint refer to the prevention (Cohen 1971)[23] of non-concurrence and this accords with the stylistic ideal of the era. On the other hand, in Bach's music we find revelation of non-concurrences of many types.)

Here we have addressed concurrence and non-concurrence among partials with respect to their deflections in frequency.

The Border between Intervals, 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 & Dubnov 1996)[25]. In tonal music texture appears mainly as an aid that may support or contradict the interval organization, while today it has an existence of its own. Actually, most notation systems these days refer to textural phenomena. Without going into the details of texture classification we shell note that the main difference between texture and timbre is that texture is **separable** and usually relates to time scales that are larger than those of timbre. Timbre can be identified for durations of less than 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 in which we can no longer separate the simultaneous occurrences into its components, the texture becomes timbre. In the opposite case, too, 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 (Cohen 1969)[24], "articulatory ornamentations" in non-western music and random modulations in electronic music (Tenney and Polansky 1980)[129].

Part IV

Extentions

Chapter 7

Non Instrumental Sounds

The method of decorrelation and the higher order cumulants of the residual signal seem to be applicable for representing sounds other than musical instruments. In the case of "stationary" complex sounds, such as many natural and industrial sounds, the perceived sound quality could be described verbally as texture - an acoustic phenomenon that holds a certain characteristic variability to its contents. Naturally, sound features such as spectral envelope must play an important role in the description of these sounds as well. Nevertheless, these spectral features are insufficient since they do not account for much of the intrinsic variability present in the sound.

In this part of the study we shall address this problem of sound texture representation and analysis, by adapting the HOS feature extraction and analytical methods for this purpose. Remaining loyal to the source-filter model, we model the spectral envelope by a linear filter. The description of the decorrelated or pre-whitened sound (i.e. the sound source), is possible, from the statistical standpoint, only through better modeling of the white excitation signal, which shall be done again by means of higher order statistics.

Before proceeding with the mathematical investigation, we should note that the basic question of the existence of texture in sound remains open, from the psychoacoustical standpoint, to a great extent. Whether or not there exists a "preattentive auditory system" that, although lacking the ability to process complex sounds on the cognitive level, still has the power to detect differences in "textural" features, remains to be examined by other researchers.

7.1 Bispectral Analysis

For the sound-texture signals, we investigated 11 recordings of sounds such as people talking in a room, engine noises of various kinds, recordings of factory sounds, etc. The sounds were obtained from a Signal Processing Information Base (SPIB) at http://spib.rice.edu/.

Below we show the bispectral amplitudes of several types of signals. These figures visually demonstrate the bispectral signatures for the following sounds: recordings of a people talking (babble), “crashing” factory noise (fac2) and a rather “smooth” buccaneer engine noise (buc1).

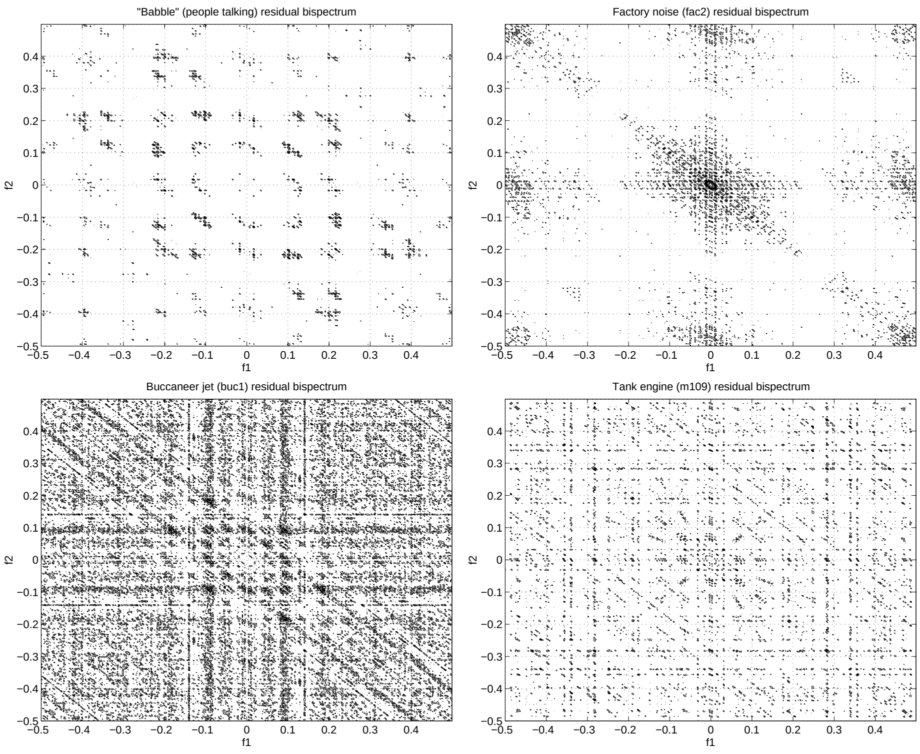

Figure 7.1: Bispectral amplitudes for people talking, car production factory noise, buccaneer engine and m109 tank sounds (left to right).

One can see that the bispectral amplitudes of the signals vary significantly. The cumulant space representation places pulse-like, “gagged” or “rough” sounds, such as talking voices (babble), machinery crashes (fac1, fac2 - factory noises) and noisy engines (m109, dops - destroyer operations room recording) higher in the moments space (i.e. they have higher moments), while the “smoothly” running engines are near the origin, and thus closer to a Gaussian model.

As can be seen from figure 7.1, the bispectral signatures vary significantly for the different sounds. The feature of the volume under the bispectral graph, which equals the skewness of the residual signal in the time domain, is still significant, as can be seen from figure 7.2, but

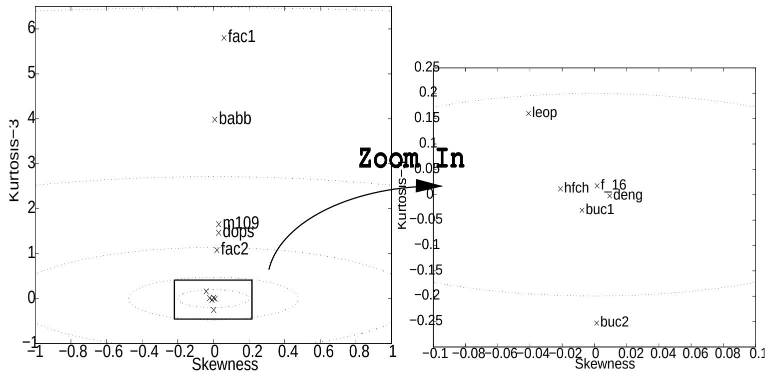

Figure 7.2: Location of sounds in the 3rd and 4th normalized moments plane. The value 3 is subtracted from the kurtosis so that the origin would correspond to a perfect Gaussian signal. Sounds with many random transitory elements are on the perimeter. Smooth sounds are in the center.

it is hardly sufficient for a detailed description of these sounds. By looking at the bispectral signatures it is clear that much structure exists in the decorrelated (residual) signals. Since the bispectrum is not constant, it is clear that these signals are non-linear,[1] and thus they cannot be modeled by a non-Gaussian noise passing through a linear filter. Below we shall try to present a simple model with some non-linear control of the signal properties, which may account for the appearance of bispectral structure.

7.2 "Granular" Signal Model

In order to account for the HOS properties of the decorrelated texture signals and gain some intuition into the sound-generating mechanism, a model for an excitation signal is needed. Although we cannot provide a complete model here, we discuss some possible mechanisms for production of random sound textures and interpret the results based on this model.

One of the most appealing models for random sound texture synthesis is the *granular synthesis* method. A *grain of sound* is a brief acoustical event of time order of several milliseconds, which is near the threshold of human auditory perception. To create a complex sound, thousand of grains are combined. We should note that this method is closely linked to time-frequency sound representation schemes such as the wavelet or short time Fourier transform, which provide a local representation by means of waveforms or grains multiplied by coefficients that are independent of one another. Actually, such a broad definition of *granular synthesis* covers a very significant class of musical and speech sounds, such as the source-filter model. The Formant Synthesis method, which uses a quasi-periodic signal as the

[1]See section 5.3 for a discussion of tests of non-linearity and their relationship to the bispectra of the residual.

excitation of a time-varying linear filter that shapes the spectral envelope of sound, can be seen at times as a variant of granular synthesis and it is often regarded as *Pitch synchronous* granular synthesis (De Polli & Piccialli) [32], (Bennett & Rodet) [5]. When the appearance of the grains is *asynchronous*, i.e. the various grains appear randomly in time, many auditory attributes such as pitch or a precept of a well defined timbre are lost, and we basically get a *sound cloud* made up of thousands of elementary grains scattered about the frequency-time space. Although asynchronous granular synthesis has received much attention in computer music synthesis applications, little analytic work has been done so far (Roads 93) [109].

Below we apply a simple asynchronous model for the description of sound textures. The sound signal is modeled by a sum of time functions (granules) $h_k(t)$

$$x(t) = \sum_k \sum_i h_k(t - t_{k,i}) \tag{7.1}$$

with $\cdot, \cdot, \cdot, t_{k,-1}, t_{k,0}, t_{k,1}, t_{k,2}, \cdot, \cdot, \cdot$, being poisson distributed times of appearances of the k'th granule. This process canalso be viewed as a sum of outputs of a filter bank triggered by several poisson processes $x_k(t) = \sum_i h_k(t - t_{k,i})$. Without loss of generality let us assume that all poisson processes have the same average lag time $E(t_{i+1} - t_i) = \mu$.

For each filter (granule), the power spectrum and bispectrum are given by (Nikias & Raghuveer 1987)[92]

$$\begin{aligned} S_k(\omega) &= \frac{1}{\mu}|H_k(\omega)|^2 \\ B_k(\omega_1, \omega_2) &= \frac{1}{\mu} H_k(\omega_1) H_k(\omega_2) H_k^*(\omega_1 + \omega_2) \end{aligned} \tag{7.2}$$

If the granules have a band limited frequency response, one might end up with a situation where $H_k(\omega_1)H_k(\omega_2)H_k^*(\omega_1 + \omega_2)$ is practically zero for all (ω_1, ω_2). This occurs when the $H_k(\omega)$ has a finite support $[\Omega_1, \Omega_2]$ of less then an octave, since if

$$H_k(\omega) = 0 \quad for \quad \omega < \Omega_1 \quad and \quad \omega > \Omega_2 \tag{7.3}$$

then

$$\omega_1 + \omega_2 < \Omega_1 \quad and \quad \omega_1 + \omega_2 > \Omega_2 \quad iff \quad \Omega_2 < 2\Omega_1 \tag{7.4}$$

We call such granules "bispectrally null". Assuming that our granules are indeed bispectrally null, and that the excitation processes $\{t_k\}$ are statistically independent, the resulting spectrum of the process $x(t)$ is the sum of the spectra $S_k(\omega)$ with zero bispectrum.

Now let us consider a slightly different situation where the same process excites several filters h_k, h_l, h_m, which in other words amounts to synchronous appearance in time of their respective granules. When the granules are "harmonically related", i.e. their frequency supports approximately obey the relationship $\Omega_{k,1} + \Omega_{l,1} \approx \Omega_{m,1}$ and $\Omega_{k,2} + \Omega_{l,2} \approx \Omega_{m,2}$, we have

$$
\begin{aligned}
g(t) &= \sum_i h_k(t-t_i)h_l(t-t_i)h_m(t-t_i) \\
G(\omega) &= H_k(\omega)+H_l(\omega)+H_m^(\omega) \\
S_g(\omega) &= \frac{1}{\mu}|G(\omega)|^2 = S_k(\omega)+S_l(\omega)+S_m(\omega) \\
B_g &= \frac{1}{\mu}G(\omega_1)G(\omega_2)G^*(\omega_1+\omega_2) = \frac{1}{\mu}H_k(\omega_1)H_l(\omega_2)H_m^*(\omega_1+\omega_2)
\end{aligned} \tag{7.5}
$$

We would like to note that the requirement for suboctave filters is interesting because this condition is widely obeyed in auditory modeling. Thus we may conclude that in sounds which can be effectively created/represented by combinations of suboctave granules, the existence of synchronous groups of harmonically related granules causes bispectral peaks to appear. The amplitude of these peaks depends on the amplitudes of the particular granules. In order to give the bispectral plane a meaning of synchronicity detector, proper power spectrum normalization is required. Such normalization would allow interpretation of the peaks as probabilities for synchronous appearances among granules. Decorrelating the signal by our AR spectral matching method only approximately accomplishes this need, and better methods (such as wavelet transform) are required.

Chapter 8

Musical Examples of the Border-State between Texture and Timbre

In this chapter we attempt to understand the manifestations and significance of texture in contrast to other parameters, bearing in mind gestalt principles and the stylistic ideal [1] that varies from culture to culture and from era to era. This statement raises numerous questions, especially because of the tremendous variety of forms texture can take. Texture was generally attributed to pitch-related factors within the tonal system, and its classification was based on the number of parts or strata and the interactions between them: monophonic, homophonic, polyphonic, or heterophonic texture. These terms have been expanded to apply to pitch phenomena in atonal and serial music (Boulez 1971), and even to encompass treatment of other parameters in addition to intervals ("sound texture" as opposed to "harmonic texture" based on the interval system (Goldstein 1974)[50], (Lansky 1974)[70]. Most research on texture as we understand it has focused on twentieth-century music, in which it plays an important role in the organization of the piece. Very little direct attention has been paid to texture in tonal music (Lorince 1974; Ranter 1980; Levy 1982). Indirectly, however, the two (in our opinion) properties of texture - "register" and "contour," sometimes called "design" (Rothgeb 1977)[113] - have received more attention even in tonal music. Few ethnomusicological studies have been done on how texture characterizes a musical culture. Finally, there are studies dealing with the formulation of textural phenomena on the acoustic level in electronic and computer music (Smalley 1986)[?], (Czink 1987)[31]. Nevertheless, no framework has been laid out for understanding texture from all these standpoints. Thus certain questions arise:

[1] Stylistic ideal is characterized, in our opinion, mainly by four variables: (1) connection/lack of connection with the outside world, (2) expression of calmness vs. excitement, (3) focusing on the moment vs. the overall structure, (4) clear or blurred units and directionality

- Is it at all possible to come up with a suitable definition of texture that would encompass all the manifestations alluded to here?
- Can one speak of gestalt phenomena, that is, the existence of units, in texture?[2]
- What is the role of texture in relation to the other parameters in various styles?
- Is texture of significance for the stylistic ideal, including in music that focuses on the moment?
- What tools are needed to analyze textural phenomena?

Here we shall attempt to give general answers to the first four questions, which have to do with the characterization and significance of texture. We will set up a general framework for discussing texture and its significance.

8.1 General Characterization of Textural Phenomena

8.1.1 Texture and its Relationship to Other Parameters; Difference and Similarity and Borderline States

Just as the most general definition of timbre is based on negation of the other parameters, here, too, we shall define texture as a *principle of organization* that is **not** derived from *learned schemes*. In the most general sense we can define texture as a way of distributing the frequencies that represent specific or undefined pitches in the space of frequency, duration, and intensity. This definition also holds true for timbre, but texture is *divisible* and refers to longer durations than does timbre.[3] A simultaneously played chord possesses both timbre properties (such as major/minor) and texture properties, and we can find border areas between the two. These result in special cases when we sense the changes that occur in the timbre or, due to psychoacoustical constraints, we do not separate the texture. Thus texture can be regarded as an expansion of timbre, and, vice versa, timbre can be viewed as a borderline case of texture, with a grey area between them.

Formally, one could define texture as the distribution $P(\omega)$, $\omega \in \Omega$, where Ω is an event space. The events are various musical parameters (pitch, duration, intensity, and etc.) on the lowest level, or operations acting on the parameters on a higher level. In the second case

[2]The existence of units is, of course, a necessary condition for musical organization. In this context we should mention the opinion that texture "cannot exist independently" (Levy 1982, p. 403)[75]

[3]Timbre is defined for a duration of approximately 20 milliseconds and is indivisible when we hear it, despite its extreme complexity. Texture, on the other hand, requires some separability along the frequency or time axis; a single-note event has no texture.

texture is the principle of selection among operations or (grammatical) rules that belong to schemes.

Texture, as we define it, can be viewed as a superparameter that is related to various combinations of other parameters. In the most general sense, textural phenomena *are familiar to us from nature* as psychoacoustic factors that evoke various emotional associations. Another property of texture is that *it is not well defined quantitatively.* In most styles prior to the twentieth century, the parameters (intervals, rhythms, and meters) are not only defined quantitatively but are also subject to systems of learned schemes that serve as building blocks for the musical organization of the particular style.

Difference in learned schemes refers to quantitative difference (for example, the major scale differs from the minor scale in the sizes of three intervals); whereas in texture the important factor is qualitative relationship rather than the quantitative ones, for instance the direction of change and not the exact magnitude of the change. We have termed such a relationship a *type.*[4] We can measure the difference/similarity by using probability separability measures $D(P_1, P_2)$, which measure the "distance" between distribution functions P_1 and P_2. Such measures are widely used in the pattern recognition literature and we will not discuss them here.

Thus, in "stylistic music" there are two kinds of information: one having to do with learned schemes and the other with textural information. Their relationship is one of the characteristics of the style. As in the relationship between timbre and texture, there are also borderline states between texture and intervals and between texture and rhythm. These are the situations in which minor or rapid changes occur in pitch or duration and psychoacoustic constraints prevent clear identification of the pitches, the intervals, or the rhythm. Situations in which intervals turn into texture are intentional in many non-western musical cultures, as well as in contemporary music (Figure 8.1).

The types of events, whose distribution creates texture, may be classified according to separability and definabilty of pitch. Interestingly, one can characterize these types according to computer sound synthesis and algorithmic composition methods :

- White noise: This is a perceptually inseparable sound undefined with respect to all aspects. The methods for its production are by various random number generators.
- Continuous sounds composed of segments that have continuous spectra with a sensation of pitch that is not well defined, such as "fricatives" in speech. These are still on the timbre side of the border. With the computer they are produced by filtering the white input noise.
- Sounds that are composed of discrete set of frequencies such as idiophonic instruments

[4]Type relations can appear in many forms. For example a metric type is poetic meter, which, unlike musical meter, is not well defined quantitatively. "Type of intonation" exists in the scale skeleton of the Arab singing practice.

with continuous sound whose timbre gradually changes.[5] These "complex" sounds are an example of the borderline case with timbre. Such sounds, characterized by inharmonic spectrum, are easily simulated by an FM synthesis method. The separability is sensed due to the lack of a harmonic relationship between the partials.

- Timbres with well defined pitch. When parts of the sound envelope are relatively long (especially during the attack phase), as in the case of art singing, the ear senses the changes and we approach the border between timbre and texture. Interestingly, even in this inseparable timbre we can still classify according to the "concurrence/non-concurrence" of its overtones. The non-concurrent case is the texture within timbre. There are various methods for syntheszing pitched sounds, and the sense of texture is achieved by applying a random fluctuation to the harmonics.[6]

- In the case of very complex interval constructions, as we find, for example, in the music of Ligeti, we arrive at a borderline between the interval and texture (typical of much of modern music).[7] In computer music, many of the algorithmic composition methods are actually concerned with production of textures. A list of parameters organized in a series, as a set or a pitch class, often serves as the only predetermined scheme, while the methods of selection from this list determine the texture.

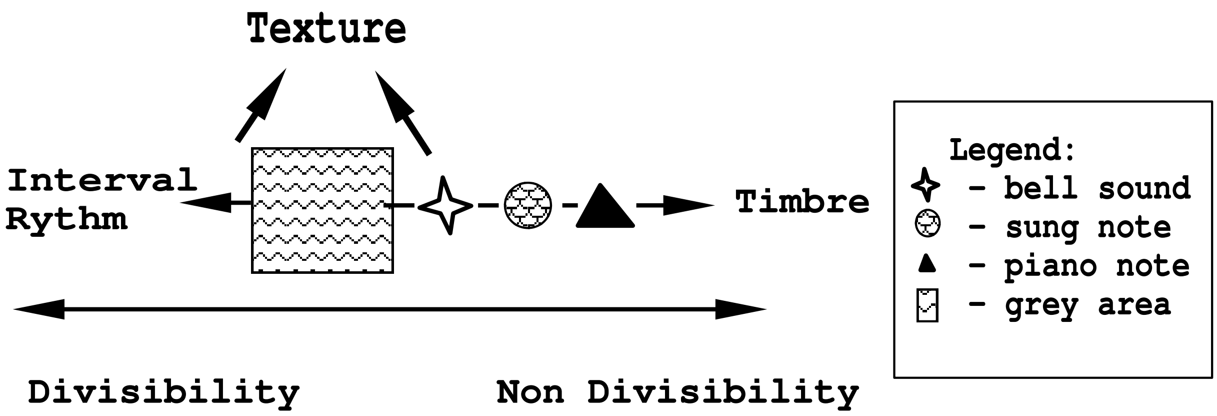

Figure 8.1: The grey area between interval-texture and texture-timbre and some examples of timbres in the borderline between texture and timbre.

[5]Note that in ancient China when one spoke of the meaning of a single note, one was referring to the continuous idiophonic sound.

[6]Applying a random modulation to the frequencies of the sinusoidal components in an additive synthesis method, or randomizing the input pulses in "granular" synthesis, creates smooth transitions between pitch - texture - noise with increasing the amount of randomness.

[7]These textures can turn into noise in some limiting cases or transform into a well defined pitch in others.

The contribution of texture to the static and dynamic poles

One of the characteristics of a piece of music is its attitude to the time axis, which extends between two poles, one representing the sense of "flow" and directionality that accompanies a series of musical events, the other representing the sense of focus on the present moment. Various musical aspects related to these poles have been given various names: vertical/horizontal (Grove 1980)[?]; inside time/outside of time and space/time (Czink 1987)[31]; material/structure (Cage 1966)[13]; local/global; momentariness/overall directionality and complexity (Cohen 1954)[?]. Here we shall call this pair "static/dynamic." In the real world, of course, one cannot appear without the other; but the relations between them vary enormously, reflecting one of the important characteristics of the stylistic ideal.[8] With regard to the different possibilities for simple or complex organization, in the short or long term, the most prominent representative of the static pole is timbre, whereas the most prominent representative of the dynamic pole is the interval (under certain conditions). Particularly noteworthy is the contribution of the borderline states between texture and the other parameters to the static pole. Texture can be placed between timbre and intervals in terms of the static/dynamic and other aspects, as we shall see below.

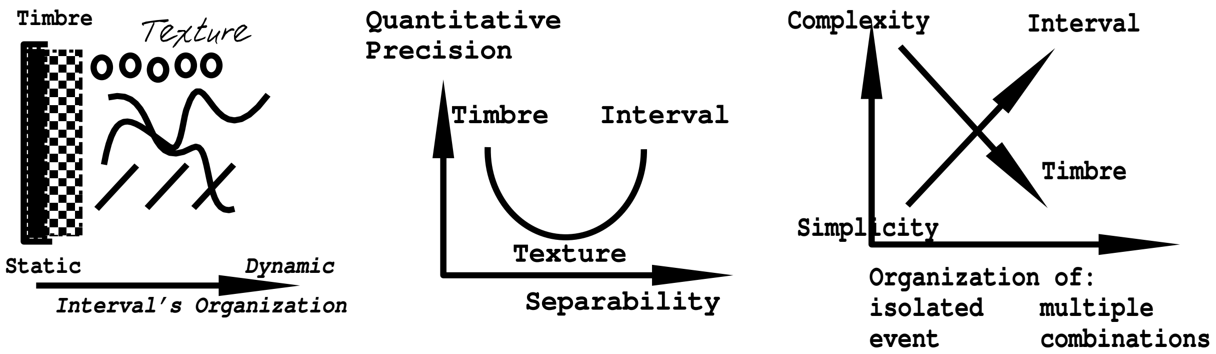

Figure 8.2: Comparison of the three parameters from various aspects.

Classification of textures and textural units

Three main factors may produce units: (1) a scheme or gestalt principle that fuses a collection of events into a unit; (2) a repetition of the collection A, A, A; (3) a substantial change from the surroundings. The learned schemes serve as a unit-forming factor on the various levels. With respect to the texture, too, as we shall see, there are some "natural" schemes that combine to form a single unit (factor 1), but here the important distinctions between units are achieved by different kinds of texture (factor 3). Here we shall try to classify texture

[8]This importance has even been stated explicitly both in non-western cultures (e.g., Chou 1970)[17] and by twentieth-century composers. We should also note that many twentieth-century compositions are based, consciously or otherwise, on the compositional principles of non-western music (Chou 1974)[16].

according to four overall frameworks. The first is derived from the fact that the texture may appear with or without the constraints of learned schemes (especially interval-based tonal schemes). The second is the appearance of texture in "pure" states or in various borderline states. The third is based on segregation, primarily into horizontal layers along the pitch axis, and the fourth is segregation into units along the time axis. Thus, texture may appear in a single layer or in many; without change along the time axis or with a great deal of change producing numerous units of varying degrees of definability. Below we describe the four frameworks for classifying texture.

Connection/lack of connection to a learned scheme, with concurrence or non-concurrence

Texture can refer to the distribution of discrete, well differentiated elements (pitches, intervals, timbres, or intensities) or to continua of these elements with distinct directions of variation. In the first case, in addition to textural information, information is obtained from the distinct elements (intervals, chords, rhythms, and so on); in extreme cases of learned schemes, this information is extremely well defined. In this case, every learned scheme may be realized in a variety of textures, and the texture may be concurrent or non-concurrent with the learned schemes. In a situation of concurrence, textural changes follow changes in the learned schemes, thereby highlighting the structure based on the schemes. In a situation of non-concurrence the texture may blur the structure. In the other extreme case (very common in the twentieth century), the texture is not only unconnected to learned schemes but may be realized in a variety of intervals or durations while retaining the type as expressed in its specific notation. The relationship between texture and learned schemes with respect to their role and importance differs from style to style, as we shall see below.

Borderline states

As mentioned above, for each of the two cases described above (connection and lack of connection to learned schemes) there are borderline states due to psychoacoustic constraints (which set thresholds for our ability to make distinctions). In the first case, where texture refers to discrete elements, there are many degrees of borderline situations. In the second case the borderline situation is when one cannot detect the magnitude and direction of change and can only discern the existence of randomness and distinctions between different kinds of randomness. In the most general terms we will distinguish between relatively stable situations and the various borderline situations.

Ranges of occurrence

These may be related to various parameters that can be arranged on a scale: tempo, pitch, and intensity. The classification is done in accordance with the absolute range and the degree of scatter, described in terms of the three main sizes: low, medium, and high. In some cases the medium size is a normative optimum. Such a classification system is commonly used in various cultures for tempo, tessitura, degree of melismatism, and so on. For example, the chants of Tibetan monks have a one-part layer (monophony) with extremely low scatter in the low range. In Palestrina's polyphonic music, all the indices - the scatter and absolute range of each part and between the parts - are "medium." A focus on the optimum as opposed to the extremes reflects the stylistic ideal.

Curves of change along the time axis

So far we have defined stable texture primitives and used probability functions to describe them. Gradual changes in texture could be described as slow "modulation" acting on the parameters of the probability function. In general there are six types of curves in the various parameters: level, zigzag, ascending, descending, convex, and concave, with concurrence or non-concurrence between parameters. As we shall see, these types of curves are significant in an interesting way.

1. **The flat curve (successive repetitions)**
   This includes various numbers of repetitions, with various degrees of precision, of units of various sizes. It may serve as a background for emphasizing events at another level, to create tension due to uncertainty of continuation, or to increase the clear directionality by a single repetition (two appearances). Hence it is not therefore surprising that the rules of Palestrina counterpoint prohibit various kinds of repetitions, whereas multiple repetitions of small units are extremely common in Baroque music, and the symmetrical periodic phrases (one repetition), are so typical of the Classical period.
   The 2^n scheme
   This scheme is an expansion of the simple symmetry obtained by a single repetition (two occurrences). In the Classical period, symmetry, which is considered to express maximum clarity, generally includes a series of repetitions, with each one containing its predecessor within it, such that a 2^n series is obtained.

   $$1+1 \;+\; 2 \;+\; 4 \;+\; 8 \;+\; 16 \tag{8.1}$$

   This series can also be obtained from an ongoing division of the whole into two parts. The 2^n scheme was common already in Far Eastern (especially ancient Chinese) music, and currently appears mainly in *gamelan* ("orchestral") music in Indonesia, in which 2^n can reach 256 beats, however it is rare in many monophonic musics. Note that the repetitions are not exact; they may include various transformations.
   Transformation
   Another expansion of the single repetition is obtained from the phenomenon of transformation, which is a repetition with the performance of an operation. Interestingly, the operations can be grouped into five categories whose principles govern general cog-

nitive operations (we shall note only a few operations: contrast, which has a wide variety of manifestations (Cohen 1994)[?]; expansion and contraction with regards to various parameters; and shifts in cyclical systems along the time or pitch axis). We can thus see transformations as principles of textural organization that link units.
Because of space limits, we will not elaborate about the rest of curves of change, except for one extremely important curve - the convex curve.

2. **The convex curve**
   In the extreme case, this refers to a gradual rise to a single peak, followed by a gradual decline. It allows for maximum predictability about the next step in the process, and it helps consolidate the units. This curve is prevalent in folk tunes all over the world, except for melodies meant to awaken specific emotions (such as dirges) (Nettle 1977). In art music, it is prominent in the vocal works of Palestrina, whose ideal, as stated above, was tranquility. Here the convexity is expressed on various levels. At the level of the phrase, it appears in the parameters of pitch and duration (see Figure 8.2). The convex curve makes a covert appearance in the soprano line in most works of western tonal music (according to Schenkerian analysis).
   The convex curve and the 2^n scheme are both common schemes that represent types of gestalt rules that contribute to the formation of a single directional unit. A comparison of the two is quite interesting. Both are efficient, extreme methods of achieving predictability, but one (convexity) represents a minimum of divisibility and levels, whereas the other permits a maximum of divisibility and formation of levels in musical organization. Unlike the convex curve, the concave curve does not allow prediction of the direction of the process, because the curve has no upper limit. Concave curves appear in styles that seek to arouse emotions; for example, at the immediate level of performance of Rig Veda chants, which are full of an explicit tension, and at various levels in Romantic music. On the other hand, the rules of Palestrina counterpoint forbid it. Without getting into details, one can see that the convex/2^n curves correspond to stable/self similar "modulation" functions. Of course all six curves may appear with concurrence or non-concurrence between them, in one stratum or in several strata, in a large or small range of variation within each stratum and between the strata, and in different locations within the absolute ranges of the various parameters.

8.2 Further remarks

In light of the above discussion, we would like to add three summarizing remarks concerning the contribution of texture to three topics that have come up repeatedly in various contexts throughout the paper: a comparison of music and speech; symmetry; non-concurrence and uncertainty in music.

8.2.1 Analogy to Semantic/Prosodic Layers in Speech

Comparisons and analogies between music and speech have been offered from various perspectives. Here we wish to propose an additional analogy, between two complementary pairs: learned schemes/texture in music and the lexical stratum/prosody in speech. The stratum of prosody, or "the musical factors in speech" (called "intonation" in linguistics (Bolinger 1972)[8], has two functions: (1) to represent the laws of syntax; (2) to convey the speaker's attitude to the listener and to what he or she is saying that is, to convey the emotions. Texture in music has the same two functions: (1) "syntactic" - to clarify or blur the musical structure based on learned schemes; (2) to convey emotions, that is, to characterize the musical expression in terms of extra-musical factors. In terms of the acoustical appearance, the "musical factors in speech" are represented by the texture, that is, the distribution of loudness, duration, and pitch (without reference to precise quantitative magnitudes). Excitement and calm are, of course, expressed by the same means as texture. Like the texture in music that may be in concurrence or non-concurrence with the learned schemes, the musical factors in speech may be correlated or non-correlated with the message contained in the lexical stratum (try saying "I'd like to see you" twice - the first time emphasizing the semantic content, the second time emphasizing the opposite meaning, in other words, "go to hell!"). They can also appear as an independent message, as in various exclamations. (In studies on bird calls in various situations, we have compared their calls to musical factors in speech, although they have no lexical stratum. Interestingly enough, the universal principles of tension and relaxation that determine the laws of Palestrina counterpoint [and not only them, of course] also apply to bird calls (Cohen 1983)[22].)

8.2.2 Symmetry

One frequently hears today of principles of symmetry that are common to inanimate objects in nature, living organisms, and even works of art (note the existence of an international association for symmetry in nature and in man-made products). Without going into detail, we shall merely state that in music symmetry is expressed both in schemes and in operations (Cohen 1993)[?]. In the present paper we have stressed the two symmetrical textural schemes, the convex curve and 2^n which represent gestalt principles that fuse a collection of events into a unit. Similarly, we mentioned "natural" operations that, by their very nature, represent symmetrical organization in the texture. Finally, there is the common symmetrical organization expressed as ABA. Grouping together all the different phenomena mentioned here, some of them extremely hidden, within the category of "symmetry" is a relatively new idea that reinforces the significance of forms of textural organization in music from another direction.

8.2.3 Non-concurrence and Uncertainty

We have seen that texture, by its very nature, entails some freedom and uncertainty regarding the quantitative expression of the range of its appearance and of its curves of change. The uncertainty is increased by non-concurrence between the curves of change of the various layers and even of the different parameters (for example, an ascent in pitch as intensity decreases). Non-concurrence can occur in an enormous variety of ways, and even within the timbre of a single note (concurrence or non-concurrence between the phases of the overtones), and it contributes to complexity and to various kinds of uncertainty that come up in most musical analyses.

8.3 Summary

To sum up, we have attempted to clarify textural phenomena, taking into account gestalt principles and the stylistic ideal. Below is a brief summary of our answers to the questions raised:

1. In the broadest sense, the definition of texture is suitable for describing the many musical phenomena that we presented in the paper. In special cases, such as in border areas and due to psychoacoustic constraints, some of the definitions are irrelevant.

2. There are indeed texture-based gestalt phenomena. First of all, we should note that the gestalt phenomena that have known formulations (Lerdahl and Jackendoff 1983) and have been corroborated empirically (Deutsch 1980) are textural ones (according to our definition here). Furthermore, one should bear in mind the symmetrical schemes, including the convex curve and 2^n.

3. Texture is a "superparameter" that has three types of relationships with the parameters and their learned schemes: it can support the scheme-based organization, it can blur it, and it has border areas. Moreover, it can serve as a major factor in organization and can be realized in various manners.

4. Texture, by definition, unlike learned schemes, has inherent extra-musical meanings and cannot create complex long-term organization on its own. Texture (as opposed to the learned schemes) contributes directly to states of calm/excitement; certainty/uncertainty; and types of momentary/overall and clear/suspensive directionality. All of these are characteristics of the stylistic ideal. In the different styles one "chooses," whether consciously or unconsciously, to emphasize learned schemes or textural organization, types of learned schemes, and types of textural schemes, including kinds of uncertainty, which form part of the definition of the style.

The research naturally calls for more investigation, but already at this stage we think that texture should be regarded as an indispensible factor in music analysis of all styles.

Bibliography

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

[2] R.Ashley, *A Knowledge-Based approach to Assistance in Timbral Design*, in Proceedings of the International Computer Music Conference,S an Francisco, 1986.

[3] J. Beauchamp, *Brass-Tone Synthesis by Spectrum Evolution Matching with Nonlinear Functions*, Computer Music Journal 3 (2) : 35-43, 1979

[4] A.H. Benade, *Fundamentals of Musical Acoustics*, Oxford University Press, New York, 1976.

[5] G.Bennett, X.Rodet, *Synthesis of the Singing Voice*, in Current Directinos in Computer Music Research, MIT Press, 1989.

[6] W.J.Bernard, *Inaudible Structures, Audible Music: Ligeti's Problem, and his Solution.* Musical Analysis 6: 207-233, 1987.

[7] W.Berry, Structural Functions in Music, Chap. 2. Englewood Cliffs, 1976.

[8] D.Bolinger,(ed.), Intonation: Selected Readings. Penguin Education, 1972.

[9] P.Boulez,*Timbre and Composition – Timbre and Language*, Contemporary Music Review 12: 161-171, 1987.

[10] , P.Boulez, Boulez on Music Today. London: Faber and Faber, 1971.

[11] A.S. Bregman, *Auditory Scene Analysis: The perceptual organization of sound.*, Cambridge, Massachusetts, MIT Press, 1990.

[12] D.R. Brillinger, *An Introduction to Polyspectra*, Ann. Math. Stat., Vol. 36, 1361-1374, 1965.

[13] J.Cage, *Rhythm etc.* in G. Kepes, ed., Module Symmetry Proportion. London: Studio Vista, pp. 194-203, 1966.

[14] J.Cage,Silence. Cambridge, Mass: MIT Press, 1961.

[15] R.Cann, *An Analysis/Synthesis Tutorial*, Computer Music Journal 3(3):6-11; 3(4):9-13, 1979; and 4(1), 36-42, 1980.

[16] Ch.Chou, *Asian Music and Western Composition.* In Dictionary of 20th Century Music, ed. J. Vinton, 22-29, 1974

[17] Ch.Chou, *Single Tone as Musical Entities: An Approach to Structured Deviations in Tonal Characteristics.* American Society of University Composers, Proceedings 3, 1970

[18] R.Cogan, *New Images of Musical Sound*, Cambridge, Massachusetts, Harvard University Press, 1984.

[19] D.Cohen, *Directionality and Complexity in Music.* Musikometrica, vol. 6, (1995 in press).

[20] D.Cohen, *Hierarchical Levels in the Musical Raw Material.* Proceedings of the HISM Conference, 1988.

[21] . D.Cohen, *The Performance Practice of the Rig Veda: A Musical Expression of Excited Speech*, Yuval 6: 292-317, 1986.

[22] . D.Cohen, *Birdcalls and the Rules of Palestrina Counterpoint: Towards the Discovery of Universal Qualities in Vocal Expression*, Israel Studies in Musicology 3: 96-123, 1983.

[23] . D.Cohen, *Palestrina Counterpoint : A Musical Expression of Unexcited Speech*, Journal of Music Theory 15: 68-111, 1971.

[24] . D.Cohen, *Patterns and Frameworks of Intonation.* Journal of Music Theory 13/1: 66-92, 1969.

[25] D.Cohen, S.Dubnov, *Texutre Units in Music that Focuses on the Moment*, Joint International Conference on Systematic Musicology, pp. 51-62, Bruge, 1996.

[26] D. Cohen, R. Katz, *Some timbre characteristics of the singing of a number of ethnic groups in Israel*, Proceedings of the 9-th Congress of the Jewish Study, Division D., Vol II, 241:248, 1986.

[27] D.Cohen, J. Mendel, H. Pratt, and A. Barneah. *Response to Intervals as Revealed by Brainwave Measurement and Verbal Means*, The Journal of New Music Research 23: 265-290, 1994.

[28] L. Cohen, *Time-Frequency Distributions - A Review*, Proceedings of the IEEE, Vol. 77, No. 7, July 1989.

[29] M. Clynes (ed.), *Music Mind and Brain*, New-York, Plenum Press, 1982.

[30] T.M.Cover, J.A.Thomas, *Elements of Information Theory*, Wiley Series in Telecommunications, 1992.

[31] A.Czink,. *The Morphology of Musical Sound*, preprint, 1987

[32] G.De Poli, A.Piccialli, *Pitch-Synchronous Granular Synthesis*, in Representations of Musical Signals, MIT Press, 1991.

[33] D. Deutsch (ed.), *The Psychology of Music*, London, Academic Press, 1980.

[34] D.Deutsch, J. Feroe, *The Inteval of Pitch Sequences in Tonal Music*, Psychological Review 88, 503-522, 1981

[35] R.De Vore, *Silence and its Uses in Western Music*, lecture and abstract in the Proceedings of the Second International Conference on Music Perception and Cognition 51. Los Angeles, California: UCLA, 1992

[36] W.J.Dowling, *Scale and Contour: Two Components of a Theory of Memory for Melodies*, Psychological Review 8, 341-354, 1978

[37] S.Dubnov, N,Tishby, D.Cohen, *Bispectrum of musical sounds: an auditory perspective*, X Colloquim on Musical Informatics, Milano, Italy, 1993.

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

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

[40] S.Dubnov, N,Tishby, D.Cohen, *Investigation of Frequency Jitter Effect on Higher Order Moments of Musical Sounds with Applications to Synthesis and Classification*, in the Proceedings of the Interntational Computer Music Conference, Hong-Kong, 1996.

[41] J.D. Dudley, W.J. Strong, *A Computer Study of the Effects of Harmonicity in a Brass Wind Instrument: Impedance Curve, Impulse Responce, and Mouthpiece Pressure with a Hypothetical Periodic Input*,

[42] J.Edworthy, *Melodic Contour and Musical Structure*, In: Musical Structure and Cognition, eds. P. Howell, J. Cross and R. West. Academic Press, 1985

[43] R., Erickson, *Sound Structure in Music*, Berkely, CA, University of California Press.

[44] R.Ethington, B.Punch, *SeaWave: A System for Musical Timbre Description*, Computer Music Journal, Volume 18, Number 1, 1994.

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

[46] I.Fonagy, K. Magdis. *Emotional Patterns in Intonation and Music.*, In Intonation: Selected Readings, ed. D. Bolinger. Penguin Education, 1963

[47] J.R.Fonollosa, C.L.Nikias, *Wigner Higher Order Moment Spectra: Definition, Properties, Computation and Application to Transient Signal Analysis*, IEEE Transactions on Signal Processing, 41(1):245-266, January 1993.

[48] A.Gabrielsson, *The Multi-faceted Character of Music Experiment*, lecture at the Second International Conference on Music Perception and cognition (abstract in the Conference Proceedings), p. 81, 1992.

[49] K.Geiringer, *Musical Instruments*, Oxford University Press, 1978.

[50] M.Goldstein, *Sound Texture*, In Dictionary of 20th Century Music, ed. J. Vinton, 747-753, 1974

[51] 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

[52] R.Gray, A.Buzo, A.H.Gray, Y.Matsuyama, *Distortion Measures for Speech Processing*, IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 28, No.4, August 1980.

[53] R.Gray, A.H.Gray, G.Rebolledo, J.E.Shore, *Rate-Distortion Speech Coding with a Minimum Discrimination Information Distrtion Measure*, IEEE Transactions on Information Theory, 27 (6), November 1981.

[54] M.A. Gerzon, *Non-Linear Models for Auditory Perception*, 1975, unpublished.

[55] J.M. Grey *An Exploration of Musical Timbre*, Ph.D. dissertation, Stanford University, CCRMA Report no. STAN-M-2, Stanford, CA., 1975.

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

[57] D.Halperin, Contributions to a Morphology of Ambrosian Chant. Ph.D. dissertation. Tel Aviv University, 1986.

[58] W.M.Hartman, *Pitch and Integration of Tones*, Proceedings of the Second International Conference on Music Perception and Cognition, 64. Los Angeles, California: UCLA, 1992.

[59] K. Haselmann, W. Munk, G. MacDonald, *Bispectra of Ocean Waves*, in Proceedings of the Symposium on Time Series Analysis, Brown University, June 11-14, 1962 (ed. Rosenblatt), New-York, Wiley, 1963.

[60] M.Hicks, *Interval and Form in Ligeti's Continuum and Coule*, Perspectives of New Music 31: 172-193, 1993.

[61] M.J. Hinich, *Testing for Gaussainity and Linearity of a Stationary Time Series*, Journal of Time Series Analysis, Vol. 3, No.3, 1982.

[62] ISSM95, Special session on *Sound (timbre) in European and non-European Music*, Third International Symposium on Systematic Musicology, Wien, 1995.

[63] K.,Saariaho. "Timbre and Harmony." Contemporary Music Review, 93-133, 1987.

[64] D.H.Keefe, B.Laden, *Correlation dimension of woodwind multiphonic tones*, J.Acoustic Soc.Am. 90 (4), 1991.

[65] R.A.Kendall, E.C.Carterette, *Verbal Attributes of Simultaneous Wind Insrument Timbres, Part I & II* , Music Perception, 10(4), 1993

[66] H.C.Koch, Introductory Essay on Composition: The Mechanical Rules of Melody, Sections 3 and 4 (transl. from the German by N. K. Baker). New Haven, Conn.: Yale University Press, 1938 [1782].

[67] J.D.Kramer, The Time of Music. New York: Schirmer, 1988

[68] S.Kullback, *Information Theory and Statistics*, New-York, Dover, 1968.

[69] P.Lansky, *Compositional Applications of Linear Predictive Coding*, Current Directions in Computer Music Research, MIT Press, 1989.

[70] P.Lansky, *Texture*, in Dictionary of 20th Century Music, ed. J. Vinton, 741-747, 1974

[71] H.Leichtentritt, Musical Form. Cambridge, Mass.: Harvard University Press, 1951.

[72] D. A.Lentz, The Gamelan Music of Java and Bali. Londong: University of Nebraska Press, 1965.

[73] F. Lerdahl, *Timbral Hierarchies*, Contemporary Music Review 2, 135-160, 1987.

[74] R. Lerdahl, R. S. Jackendoff. A Generative Theory of Tonal Music. Cambridge, Mass.: MIT Press, 1983

[75] J.Levy, *Texture as a Sign in Class and Early Romantic Music* Jams, No. 3, 35: 402-531, 1982

[76] P.Liberman, *Intonation, Perception and Language.* In: The MIT Press Research Monograph, no. 38. Cambridge, Mass, 1967

[77] G.Ligeti, *States, Events, Transformation*, Perspectives of New Music 31: 164-171, 1993

[78] Y.O. Lo, *Towards a Theory of Timbre*, Stanford University, CCRMA Report no. STAN-M-?, Stanford, CA., 198?.

[79] A.Lohmann and B. Wirnitzer, *Triple Correlations*, Proceedings of the IEEE, Vol. 72, No. 7, July 1984.

[80] A.Lomax, *Universals in Song*, The World of Music, No. 1/2, 19: 117-129, 1977.

[81] Lomax, 1968. Folk Song Style and Culture. Washington: American Association for the Advancement of Science.

[82] F.E.Lorince, Jr. A Study of Musical Texture in Relation to Sonata-Form as evidenced in Selected Keyboard Sonatas from C.P.E. Bach through Beethoven, Ph.D. diss., Eastman School of Music, Univ. of Rochester, 1966.

[83] M.E.McIntyre, R.T.Schumacher and J.Woodhouse, *Apperiodicity in bowed string motion*, Acustica 49, 13-32, 1981.

[84] 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

[85] J.Mendel, C.L. Nikias, A.Swami, *Higher-Order Spectral Analysis Toolbox*, MATLAB, MathWorks Inc. 1995

[86] 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.

[87] S.McAdams, *Rhythmic Repetition and Metricality do not influence Perceptual Integration of Tone Sequences*, Proceedings of the Second International Conference on Music Perception and Cognition, 66. Los Angeles, California: UCLA, 1992.

[88] S. McAdams, A.Bregman, *Hearing Musical Streams*, Computer Music Journal 3 (4) : 26-43, 60, 63, 1979.

[89] L.Marks, Unity of Senses. New York: Academic Press, 1978.

[90] M.Mersenne, *Harmonic Universal*, the books on Instruments, 1635, Hague, Martinus Nighoff, english translation of the french edition, 1957.

[91] C.L. Nikias, J.M. Mendel, *Signal Processing with Higher-Order Spectra*, IEEE Signal Processing Magazine, July 1993

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

[93] Nitschke, G. "Ma, the Japanese Sense of Place in Old and New Architecture and Planning," Architectural Design 16(3): 116-156, 1966.

[94] Y. Oura, *Constructing a Representation of a Melody: Transforming Melodic Segments into Reduced Pitch Patterns Operated on by Melodies*, Music Perception 9(2): 251-266, 1991.

[95] K.P.Paliwal, M.Sondhi, ICASSP 1991, (Toronto), Paper S7.4, vol 1, pp 429-432.

[96] E.Patrick, *A Percepual Representation of Sound for Signal Separation*, Proceedings of the Second International Conference on Music Perception and Cognition, 53. Los Angeles, California: UCLA, 1992.

[97] R.B.Pilgrim, *Interval (ma) is Space and Time: Foundation for a Religio-Aesthetic paradigm in Japan*, History of Religions 25(3): 255277, 1996.

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

[99] I.Pitas, A.N.Venetsanopoulos, *Nonlinear Digital Filters*, Kluwer Academic Publishers, 1990.

[100] M.B.Priestley, *Nn-Linear and Non-Stationary Time Series Analysis*, Academic Press. 1989.

[101] L.R.Rabiner, R.W.Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, 1978.

[102] M.R. Raghuveer, C.L. Nikias, *Bispectrum Estimation: A Parametric Approach*, IEEE Transactions on ASSP, Vol. 33, No. 4, October 1985

[103] A.Rakowski, *Intonation Variants of Musical Intervals in Isolation and in Musical Context*, Psychology of Music 18: 60-72, 1990.

[104] A.Rakowski, *Some Phonological Universals in Music and Language*, In: Musikometrica, 1, Quantitative Linguistics, ed. M. G. Broda, 37:1-9, 1988.

[105] J.P.Rameau, *Traite de l'harmonie*, Paris, Ballard, 1722.

[106] L.G.Ratner, *Texture, a Rhetorical Element in Beethoven's Quartets*, Israel Studies in Musicology 2: 51-62, 1980.

[107] B.Reiprich, *Transformation of Coloration and Density in Gyorgy Ligeti's Lontano.*, Perspectives of New Music 16: 167-188, 1978/

[108] C. Roads, *A Tutorial on Nonlinear Distortion or Waveshaping Synthesis*, Computer Music Journal 3 (2) : 29-24, 1979

[109] C.Roads, *Asynchronous Granular Synthesis*, in Representations of Musical Signals, MIT Press, 1991.

[110] K.Rose, E.Gurewitz, G.C.Fox, *A deterministic annealing approach to clustering*, Tech. Rep. C3P-857, California Institute of Technology, 1990.

[111] C.Rosen, The Classical Style. New York: W. W. Norton and Company, 1972.

[112] B.Rosner, L. Meyer, *Melodic Processes and the Perception of Music*, in The Psychology of Music, ed. D. Deutsch. New York: Academic Press, pp. 317-341, 1982.

[113] J.Rothgeb, *Design as a Key to Structure in Tonal Music*, in Readings in Schenker Analysis and Other Approaches, ed. M. Yeston. New Haven and London: Yale University Press, pp. 72-93, 1977.

[114] C.Sachs, *The history of musical intstruments*, New York, W.W.Norton, 1940

[115] G.J.Sandell, *Concurrent Timbres in Orchestration: A Perceptual Study of Factors Determining "Blend"*, Ph.D. dissertation, Evanston, Illinois, 1991.

[116] Sauveur, *Overtones and the construction of organ stops*, Fontenelles report in the Histoire de le Academie des Sciences, 1702.

[117] H.Scherchen, *The nature of Music*, Henry Regenry Comp., 1950, translated from German 1996.

[118] R.T.Schumacher, *Analysis of aperiodicities in nearly periodiv waveforms*, Journal of the Acoustical Society of America, 91 (1), January 1992.

[119] V.Scolnic, *Pitch Organization in Aleatoric Counterpoint in Lutoslawski's Music of the Sixties*, Ph.D. disseration, Hebrew University of Jerusalem, 1993.

[120] M. Schetzen, *Nonlinear System Modelling Based on Wiener Theory*, Proceedings of the IEEE, Vol. 69, No. 12, July 1981.

[121] G.L.Sicuranza, *Quadratic Filters for Signal Processing*, Proceedings of IEEE, 80(8):1263-1285, August 1992.

[122] M.Slaney, R.F.Lyon, *On the importance of time - a temporal representation of sound*, in Visual Representations of Speech Signals, John Wiley & Sons, 1993.

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

[124] , J.A.Sloboda, *Music structure and emotional response*, Proceedings of the First International Conference on Music Perception and Cognition, 377-382. Kyoto, 1989.

[125] K.Stockhausen, *How Time Passes....*, Die Reih (English edition) 3: 10-40, 1957.

[126] , J.Sundberg, *Speech, Song and Emotion*, In Music, Mind and Brain, ed. M. Clynes, 137-148, 1982.

[127] A. S. Tanguiane, *Artificial Perceprion and Music Recognition*, Lecture Notes in Artificial Intelligence 746, Springer-Verlag, 1993, and references therein.

[128] H.Taube, *An Object-Oriented Representation for Musical Pattern Definition*, Journal of New Music Research 24: 121-129, 1995.

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

[130] Y. Tikochinsky, N. Tishby, R.D. Levine, *Alternative Approaches to Maximum Entropy Inference*, Phys. Rev. A 30:2638, 1984.

[131] 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.

[132] A.Tversky, *Features of Similarity*, Psychological Review, vol 84 : 327-352, 1977.

[133] E.Varese, . *The Liberation of Sound*, Perspectives of New Music, 11-19, 1966.

[134] P.Vettori, *Fractional ARIMA Modeling of Microvariations in Additive Synthesis*, Proceedings of the XI Colloquium on Musical Informatics, Bologna, 1995.

[135] J.Vinton, (ed), Dictionary of Twentieth-Century Music, London: Thames and Hudson, 1974.

[136] D.L. Wessel, *Timbre Space as a Musical Control Structure*, Computer Music Journal 3 (2) : 45-52, 1979.

[137] F.Winkel, *Music, Sound and Sensation*, New-York, Dover, 1967, pp. 12 - 23, 112 - 119