Testing for Gaussianity and Non Linearity in the sustained portion of musical sounds.

Shlomo Dubnov and Naftali Tishby
Institute for Computer Science
and
Center for Neural Computation
Hebrew University, Jerusalem 91904, Israel

Abstract

Higher order spectra of a signal contain information about the non Gaussian and non Linear properties of the system that created it. Since the non linearity in musical signal usually originate in the excitation signal while the linear spectral characteristics are attributed to the resonant chambers, we discard the spectral information by looking at the higher order statistical properties of the **residual** *signal, i.e. the estimated input signal obtained by inverse filtering of the sound.*

In the current paper we show that the **skewness** *and* **kurtosis** *values of the residual could be used for characterization of such important sound properties as belonging to families of strings, woodwind and brass instrumental timbres. The* **skewness** *parameter is shown to be closely related to the bicoherence function calculated over the original signal and as such it is succinct to an interpretation as statistical test for the signal conforming to a linear non Gaussian model. The above results are compared to the Hinich bispectral tests for Gaussianity and non Linearity of time series and exhibit a similar classification results. Finally, regarding the higher order statistics of a signal as a feature vector, a statistical distance measure for the cumulant space is suggested.*

1 Introduction

In the current work we try to identify the structure of musical signals based on the information present in the higher order statistics [1] (HOS), of the stationary portion of the sound. Spectral matching of a signal gives a representation of the sound as a white process passing through a linear filter whose role is to enter 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 and moreover, if the relation 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 nonlinear interaction between frequency components in the signal

$$B_X(\omega_1, \omega_2) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} C_X(m,n) e^{-i(\omega_1 m + \omega_2 n)} \tag{1}$$

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 [12],[11] functional expansion, we

[1]We consider high order cumulants in the time domain and their Fourier related polyspectra in the frequency domain

write

$$X_t = \sum_{u=0}^{\infty} h_1(u)U(t-u) + \sum_{u=0}^{\infty}\sum_{v=0}^{\infty} h_2(u,v)U(t-u)U(t-v) + ... \quad . \tag{2}$$

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

$$H_1(\omega) = \sum_{u=0}^{\infty} h_1(u)e^{-i\omega u} \tag{3}$$

while

$$H_2(\omega_1, \omega_2) = \sum_{u=0}^{\infty}\sum_{v=0}^{\infty} h_2(u,v)e^{-i(\omega_1 u + \omega_2 v)} \tag{4}$$

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 fourth 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 then 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 [2].

Schematically, our approach could be summarized by the following graph:
The signal is considered as some kind of stochastic pulse train X_t passing through a linear filter. In case of a pitched signal (i.e. signal with discrete spectrum), Gaussianity of the input signal is obtained either for inharmonic signals [3] or harmonic sounds with statistically independent partials. The independence is obtained for instances such as presence of random independent modulations at each harmonic [3].

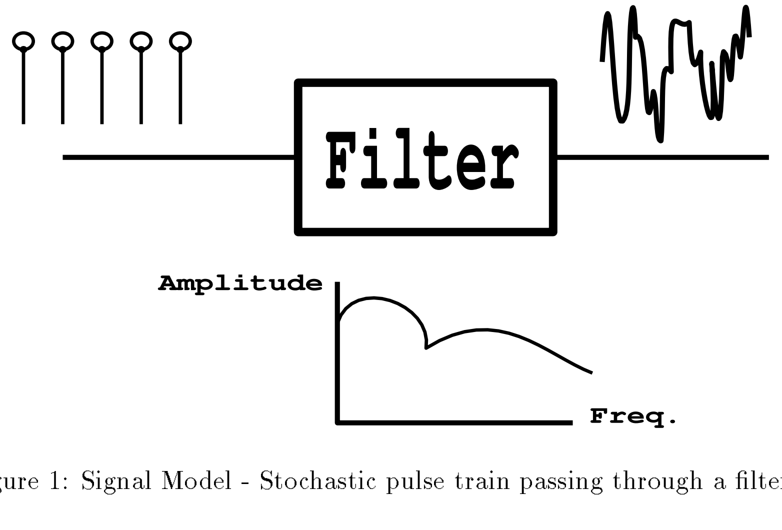

Figure 1: Signal Model - Stochastic pulse train passing through a filter.

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 parameters. Since we have no underlying parametric model for our sounds, this paper is more concerned with the analysis of structure of musical signals with some applications discussed in section 6.

[2]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 the sustained portions of a signal.

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

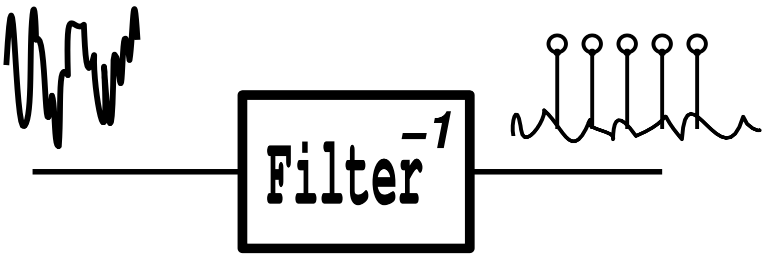

Figure 2: Estimation of the excitation signal (residual) by inverse filtering.

In the following we will make use of the bicoherence index:

$$b(\omega_1, \omega_2) = \frac{|B_X(\omega_1, \omega_2)|}{\sqrt{S_X(\omega_1) S_X(\omega_2) S_X(\omega_1 + \omega_2)}} \tag{5}$$

This is a useful function that compensates for the differences in spectral magnitudes at various frequencies by normalizing the bispectra by the power spectrum

$$S_X(\omega) = \sum_{m=-\infty}^{\infty} C_X(m) e^{-i\omega_1 m} \tag{6}$$

.

2 Higher order moments of the residual signal

Since the non linear properties in musical signal are attributed mainly to the properties of 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 in respect to 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 $h(n)$.

$$X_t = \sum_{u=0}^{\infty} h(u) U(t-u) \tag{7}$$

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

$$\begin{aligned} S_X(\omega) &= \mu_2 |H(\omega)|^2 \\ B_X(\omega_1, \omega_2) &= \mu_3 H(\omega_1) H(\omega_1) H^*(\omega_1 + \omega_2) \end{aligned} \tag{8}$$

(with similar equation for the tri-spectra) with μ_2, μ_3 (and μ_4) being the second, third (and fourth) order cumulants of U_t. For convenience we shell assume $\mu_2 = 1$. Although exact solution of the filter parameters in the non linear case is problematic [11],[12],[4], [5] we assume that it could be effectively estimated based on spectral information only. Taking a spectrally matching filter $\hat{H}(\omega)$

$$S_X(\omega) = |\hat{H}(\omega)|^2 \tag{9}$$

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

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

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

$$B_V(\omega_1,\omega_2) = \frac{B_X(\omega_1,\omega_2)}{\hat{H}(\omega_1)\hat{H}(\omega_2)\hat{H}^*(\omega_1+\omega_2)} \quad . \tag{10}$$

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

The residual signal would be ideally a decorrelated version of the original sound for the case when the excitation is a white noise, i.e. a signal with 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

$$C_V^3(0,0) = \frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} B_V(\omega_1,\omega_2) \quad . \tag{11}$$

so that the cumulants at time zero are integrals of the 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 are integrals over polyspectral planes of various orders (3 and 4 in our work) appear to be important quantities as well, although they can not anymore 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

$$\begin{aligned} C_V^3(0,0) &= \lim_{T\to\infty}\frac{1}{T}\int_0^T V^3(t)dt \\ C_V^4(0,0,0) &= \lim_{T\to\infty}\frac{1}{T}\int_0^T V^4(t)dt - 3\cdot C_V^2(0) \quad . \end{aligned} \tag{12}$$

and these results could be used as estimates of the actual third and fourth order cumulants μ_3, μ_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 [7].

3 Tests for Gaussianity and Linearity

As noted above, if a process is Gaussian, all its polyspectra of order higher then second are zero. Hence, if the process has a non-zero bispectrum this could be due to either:

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.

[6] 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.

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

$$B_X(\omega_1, \omega_2) = \mu_3 H_1(\omega_1) H_1(\omega_2) H_1^*(\omega_1 + \omega_2). \tag{13}$$

The spectral density function of the process is given by

$$S(\omega) = \mu_2 |H(\omega)|^2 \tag{14}$$

hence the bicoherence function

$$b(\omega_1, \omega_2) = \frac{|B_X(\omega_1, \omega_2)|}{\sqrt{S(\omega_1) S(\omega_2) S(\omega_1 + \omega_2)}} = \text{constant} \quad (\text{all} \quad \omega_1, \omega_2) \tag{15}$$

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

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' locates the instrumental groups on 'orbits' with various distances around the origin. This graphical representation also suggest a simple distance measure which could be used for classification of these signals.

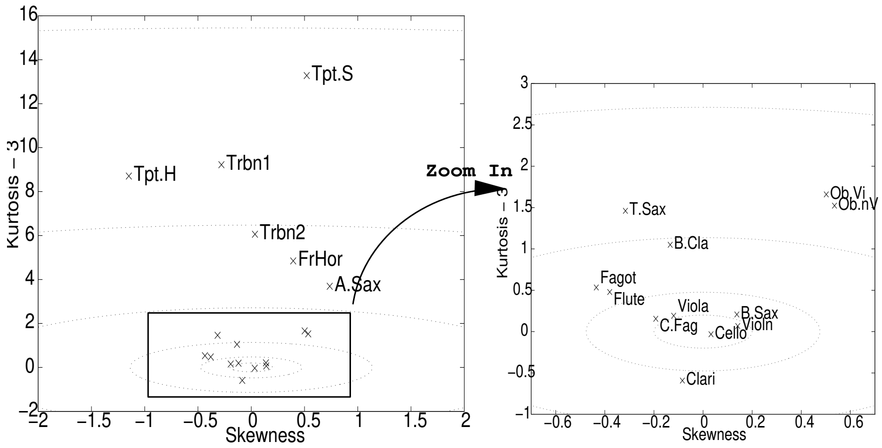

Figure 3: Location of sounds in the 3rd and 4th cumulants plane. The value 3 is substructed from the kurtosis so that the origin would correspond to a perfect gaussian signal. Left: All 18 signals. Brass sounds are on the perimeter. Right: Zoom in towards the center containing strings and surrounded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

Testing for Gaussianity and Linearity [8], [9] gives qualitatively similar results. One must note that the cumulants space representation can not determine 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 though 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 in accepting the alternate hypothesis, that the data have non-zero bispectrum. Usually, the null hypothesis is accepted if PFA is greater than 0.05 (i.e., it is risky to accept the alternate 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 'lambda'. The parameter 'lambda' 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 successfully the earlier Gaussianity test. In the following figure we present the results of Gaussianity and Linearity test plotted so that the X axis represent the Gaussianity PFA value and the Y axis is the difference between the estimated and theoretical ranges.

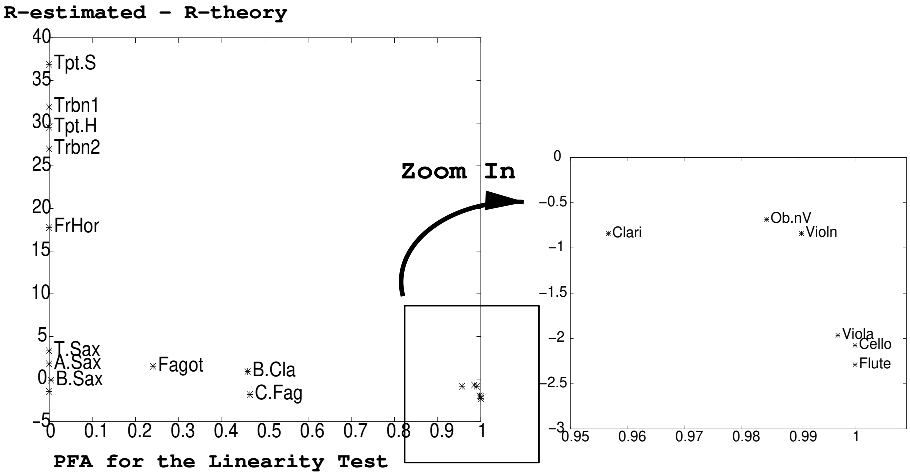

Figure 4: Results of the Gaussianity and Linearity tests. PFA $>$ 0.05 means Gaussianity. R-estimated >> R-theory means non Linearity. The deviations in the Y-axis of the right figure should not be considered since the Linearity test is unreliable for signals with zero bispectrum (Gaussian).

5 Probabilistic Distance 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 for musical instruments. Considering two classes P_1 and P_2, the Bhattacharyya (B) distance between probability densities of features of the two classes is

$$B(P_1, P_2) = -ln \int ([p(\hat{g}|P_1)p(\hat{g}|P_2)]^{1/2} dx \tag{16}$$

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 k-th order statistic $C_V^k(0,0)$ is normal around its true value. For Gaussian densities, the B distance becomes

$$\begin{aligned} B(P_1, P_2) &= \frac{1}{8}(g_1 - g_2)^T (\frac{\Sigma_1 + \Sigma_2}{2})^{-1} (g_1 - g_2) \\ &+ \frac{1}{2} ln \frac{\frac{1}{2}|\Sigma_1 + \Sigma_2|}{(|\Sigma_1|\ |\Sigma_2|)^{1/2}}, \end{aligned} \tag{17}$$

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

$$B(P,N) = \frac{1}{8} g^T \Sigma^{-1} g \tag{18}$$

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

6 Discussion

Our main result is that the cumulant space representation suggests an ordering of the signals according to common practice 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 wood wind instruments which are classified as Linear but Non Gaussian in the Hinich tests. The perimeter orbits are containing 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 creation of a musically meaningful signal with Gaussian or non Gaussian properties which we are currently investigating. In the first approach [2] 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 arbitrary 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 mans approximation to 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 also. 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 [7]. 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 only.

6.1 Musical Significance

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 better 'blend' for instruments of the same family. For combination of timbres S.Adler [1] suggests that *"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, with the idea of using cumulants as *acoustic features* beeing 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 intermediate between woodwinds and brass timbres in the moment space representation. This is very much in accordance with the orchestration handbooks description of the instrument, so to quote W.Piston [10] *"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."*

To conclude the discussion we would like to note that he above results were tested also on a larger (> 50) suite of sounds with comparable results. We note also that since we are dealing with stationary sounds only we neglect any non stationary or transitory phenomena which do not fall in the domain, or could not be considered as microscopic stationary fluctuations at the sustained portion of a sound.

References

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

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

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

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

[5] 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

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

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

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

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

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

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

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