Non-Gaussian Source-Filter and Independent Components Generalizations of Spectral Flatness Measure

Shlomo Dubnov

Department of Communication Systems Engineering
Ben-Gurion University
Be'er-Sheva, Israel
dubnov@mail.bgu.ac.il

ABSTRACT

Spectral Flatness Measure is a well-known method for quantifying the amount of randomness (or "stochasticity") that is present in a signal. This measure has been widely used in signal compression, audio characterization and retrieval. In this paper we present an information-theoretic generalization of this measure that is formulated in terms of a rate of growth of multi-information of a non-Gaussian linear process. Two new measures are defined and methods for their estimation are presented: 1) considering a source-filter model, a Generalized Spectral Flatness Measure is developed that estimates the excessive structure due to non-Gaussianity of the innovation process, and 2) using a geometrical embedding, a block-wise information redundancy is formulated using signal representation in an Independent Components basis. The two measures are applied for the problem of voiced/unvoiced determination in speech signals and analysis of spectral (timbral) dynamics in musical signals.

1. INTRODUCTION

In many signal-processing applications, such as compression, modeling, detection or retrieval, one deals with the problem of determining the amount of randomness that is present in a signal. A standard method to measure randomness is based on estimation of the amount of correlation structure by means of a Spectral Flatness Measure (SFM) [1,2,3]. SFM is defined as the ratio of the geometric mean to the arithmetic mean of the power spectral components in every spectral band. Sometimes called also "tonality coefficient", it is used to quantify how much tone-like a sound is, as opposed to being noise-like.

In this paper we consider an information-theoretic view of the SFM by defining a new measure of randomness that depends on the rate of growth of multi-information for every additional sample of the signal. We call this new measure "Marginal Information Redundancy", or "Multi-Information Rate" (MIR). It is shown that MIR equals SFM for Gaussian processes, i.e. for signals that can be described as a Gaussian i.i.d. noise passing through a linear filter (source-filter model). We show that SFM is a maximum entropy estimate of MIR when only second order statistics are available and under assumption of Gaussianity of the residual error, also called innovation process or excitation signal. In this case SFM measures the structure only due to linear or Markov dependencies that are caused by the filter. In case of a non-Gaussian process that "drives" a linear system, we take into account the additional structure or the decrease in entropy of the innovation. A non-Gaussian source has more structure since the Gaussian probability distribution function (pdf) has the maximal entropy among all processes with equal variance. We show that in non-Gaussain case a correction to SFM can be obtained based on MIR approach. This Generalization of SFM (GSFM) is estimated from the Negentropy approximation to the differential entropy of the innovation process.

A second generalization to the SFM is provided by assuming a geometrical embedding of the signal by representing the signal as a sequence of independent linear combinations of *n* dimensional basis vectors. Estimation of a linear independent basis is possible by Independent Component Analysis (ICA) method. This Independent Components (IC) representation models the one-dimensional time signal as a linear combination of basis vectors, so that the expansion coefficients are statistically as independent as possible. Using IC, we define a Independent Components SFM (IC-SFM) that is based on block-wise information redundancy differences. IC-SFM measures the difference between multi-information (MI) of a sequence of blocks and a separate sum of MI of a one block shortened sequence plus MI of the last block. It is proved that this measure equals to the sum of marginal entropies of the independent components, and is independent of the basis (as long as it is IC basis). Moreover, it can be easily estimated using the earlier (generalized) SFM.

2. MARGINAL INFORMATION REDUNDANCY

Given a stochastic process ***x***, the average amount of information that the variable x_1 carries about x_2 is quantified by the mutual information [4]

$$I(x_1,x_2)=H(x_1)+H(x_2)-H(x_1,x_2) \quad (1)$$

Generalization of the mutual information for n variables is

$$I(x_1,x_2,...,x_n)=\sum_{i=1}^{n}H(x_i)-H(x_1,x_2,...,x_n) \quad (2)$$

This function measures the average amount of common information contained in variables $x_1,x_2,...,x_n$. Using the mutual information we define marginal information redundancy to be the difference between the common information contained in the variables $x_1,x_2,...,x_n$ and the set $x_1,x_2,...,x_{n-1}$, i.e. the additional amount of information that is added when one more variable is observed.

$$\begin{aligned}&\rho(x_1,x_2,...,x_n)\\&=I(x_1,x_2,...,x_n)-I(x_1,x_2,...,x_{n-1})\end{aligned} \quad (3)$$

Since in our application we are considering time ordered samples, this redundancy measure corresponds to the rate of growth of the common information as a function of time. It can be shown that the following relation exists between redundancy and entropy

$$\begin{aligned}&\rho(x_1,x_2,...,x_n)\\&=H(x_n)-H(x_n\,|\,x_1,x_2,...,x_{n-1})\end{aligned} \quad (4)$$

This defines redundancy in terms of the difference between the entropy (or uncertainly) about isolated x_n and the reduced uncertainty of x_n if we know its past. In information theoretic terms, assuming stationarity, this measure equals to the difference between the entropy of the marginal distribution of the process x_n and the entropy rate of the process, equally for all *n*.

The estimation of MIR is performed by separate estimation of marginal entropy and entropy rate of the signal *x(*t). Assuming a Gaussian signal, one can show that the marginal entropy of the process is equal to

$$H(x)=\frac{1}{2}\ln(\frac{1}{2\pi}\int_{-\pi}^{\pi}S(\omega)d\omega)+\log_2\sqrt{2\pi e} \quad (5)$$

where $S(\omega)$ is the power spectral density of *x(*t). The entropy rate of Gaussian process, also called Sinai-Kolmogorov Entropy, is given by

$$H_r(x)=\frac{1}{4\pi}\int_{-\pi}^{\pi}\ln S(\omega)d\omega+\log_2\sqrt{2\pi e} \quad (6).$$

MIR is calculated then as the difference

$$\rho=H(x)-H_r(x) \quad (7).$$

Considering a related quantity

$$e^{-2\rho}=\frac{\exp(\frac{1}{2\pi}\int_{-\pi}^{\pi}\ln S(\omega)d\omega)}{\frac{1}{2\pi}\int_{-\pi}^{\pi}S(\omega)d\omega} \quad (8),$$

this expression is known as Spectral Flatness Measure (SFM) [1]. Using this equality we estimate the Marginal IR from the spectral flatness measure as the -0.5log (SFM).

3. NON-GAUSSIAN GENERALIZATION OF SFM

Let us denote by $x_1^n=(x_1,x_2,...,x_n)$ a vector on *n* samples. Writing the multi-information for this vector, we get

$$\begin{aligned}I(x_1^n)&=\sum H(x_i)-H(x_1^n)\\&=J(x_1^n)-H_G(x_1^n)+\sum H(x_i)\\&=J(x_1^n)-H_G(x_1^n)+\sum H_G(x_i)-\sum J(x_i)\\&=J(x_1^n)-\sum J(x_i)+I_G(x_1^n)\end{aligned} \quad (9)$$

with $I_G(x_1^n)=-H_G(x_1^n)+\sum H_G(x_i)$ denoting a

Gaussian MI and $J(x) = H_G(x) - H(x)$ being the Negentropy function [5], which measures the distance of a random variable from a Gaussian distribution. In the above expression, $H_G(x)$ means an entropy of a Gaussian random variable with same covariance matrix as *x*, $H_G(x) = \log \sqrt{(2\pi e)^N \det(R_x)}$.

Assuming a Gaussian auto-regressive (AR) process, the likelihood of observing $x_1, x_2, ..., x_n$ equals to the probability of innovation (error) signal that “drives” the AR process. Let us denote the source (innovation) by $\varepsilon_1, \varepsilon_2, ..., \varepsilon_n$. This results in the probability $p(\varepsilon_n) = p(x_n \mid x_{n-1}, ..., x_1, \vec{a})$, with vector $\vec{a}$ denoting the coefficients of the AR filter model. The log-likelihood for *n* signal samples of the AR process is

$$
\begin{aligned}
&\log P(x_1 x_2 ... x_n \mid AR \quad Model) \\
&= \log P(\varepsilon_1 \varepsilon_2 ... \varepsilon_n) = n \frac{1}{n} \sum_{i=1}^{n} \log p(\varepsilon_i) \\
&\approx nE\{\log p(\varepsilon_n)\}_{p(\varepsilon)} = nH_r(x)
\end{aligned}
\tag{10}
$$

This result suggests that estimation of IR can be done from the entropy of the source signal, which can be estimated using linear prediction (LP) from the entropy of the estimation residual. Given the residual signal ε after LP modeling we get for the Gaussian case

$$
\rho = H(x) - H_r(x) = \frac{1}{2} \log(\frac{\sigma_x^2}{\sigma_\varepsilon^2}) \tag{11}
$$

Now we can generalize SFM for processes other then Gaussian AR. The basic idea is to use the marginal entropy estimate of the innovation process as the estimate of the entropy rate of the original process. Since the innovation is not Gaussian, it must be estimated by non-Gaussian approximations. Using the above relations, it can be shown that

$$
\begin{aligned}
&\rho(x_1^n) = I(x_1^n) - I(x_1^{n-1}) \\
&= J(\varepsilon_n) - J(x_n) + \rho_G(x_1^n)
\end{aligned}
\tag{12}
$$

where we used the relation

$$
\begin{aligned}
&J(x_1^n) - J(x_1^{n-1}) \\
&= \{H_G(x_1^n) - H(x_1^n)\} - \{H_G(x_1^{n-1}) - H(x_1^{n-1})\} \\
&= H_G(x_n \mid x_1^{n-1}) - H(x_n \mid x_1^{n-1}) \\
&= H_G(\varepsilon_n) - H(\varepsilon_n) \\
&= J(\varepsilon_n)
\end{aligned}
\tag{13}
$$

with ε_n denoting the residual error or the excitation. We introduced above a new term $\rho_G(x_1^n)$ to denote the MIR of a Gaussian process, which is a function of the standard SFM derived based on Gaussian assumption. Writing MIR as $\rho(x_1^n) = \rho_G(x_1^n) + \rho_{WNG}(x_n)$ the GSFM is now defined as

$$
GSFM(x_1^n) = e^{-2\rho(x_1^n)} \tag{14}
$$

For the non-Gaussian case, GSFM results in a combination of two factors: 1) part that depends on ρ_G that describes the “structure” due to the linear Gaussian part, and 2) part related to ρ_{WNG} that contains the excessive structure due to the white, non-Gaussian residual. The GSFM of the *process x* can now be expressed as

$$
GSFM(x) = SFM(x) \cdot e^{-2(J(\varepsilon) - J(x))} \tag{15}
$$

where we’ve discarded the time indices from the signal and the innovation processes under the convention that *SFM(x)* is measured based on spectrum (or correlation statistics) of a block of measurements and ρ_{WNG} is calculated using marginal distributions only. This expression can be also given a precise statistical meaning if we view it as equivalence between mean values of relevant statistics assuming we use consistent estimators.

4. GENERALIZATION OF SFM BY INDEPENDENT COMPONENTS LINEARISATION

In this section we consider a geometrical embedding of the signal in a transformed space, and assume that the signal is represented as a sequence of independent linear combinations of *n* dimensional basis vectors. This approach is common to many signal compression schemes such as transform or sub-band coding. Using this representation provides another generalization to the SFM.

Given a multi-variate distribution of vectors

$\overline{x} = (x_1, x_2, ..., x_n)^T$, we want to find a matrix *W* and vector $\overline{s} = (s_1, s_2, ..., s_n)^T$ so that the components of the vector $\overline{s} = W\overline{x}$ are "as independent as possible". Note that we deliberately change the notation, since we are no longer dealing with a block of time samples but actually assume a multivariate process. In other words, it is assumed that exists a multivariate process $\overline{s}$ with independent components and a matrix $A = W^{-1}$, so that $\overline{x} = A\overline{s}$. The representation can be found by applying ICA to one-dimensional time signal that is embedded in *n*-dimensional blocks. We introduce a matrix notation for such a sequence of blocks (vectors)

$$\left[X_1 X_2 ...\right] = A \begin{bmatrix} s_1(1) & s_1(2) & ... \\ s_2(1) & s_1(2) & ... \\ M & M & ... \\ s_n(1) & s_1(2) & ... \end{bmatrix} \quad (16)$$

where $X_i = \left(x_{i\Delta+1}, x_{i\Delta+2}, ..., x_{i\Delta+n}\right)^T$ is *n*-dimensional vector consisting of a consecutive signal samples, with relative shift Δ between the vectors (blocks). For sake of simplicity we shall assume $\Delta = n$ in the following derivations. Note also that in the above notation the index of the sample block appears in subscript, while for expansion coefficients the block index appears in an argument (the subscript is reserved for the coefficient number).

Considering entropy of a linear transformation and using independence assumption of the coefficients, one can write the entropy of a single block as

$$H(X_1) = \sum_{i=1}^{n} H(s_i(1)) + \log|\det(A)| \quad (17)$$

and similarly for two consecutive blocks one can show that

$$H(X_1 X_2) = \sum_{i=1}^{n} H(s_i(1), s_i(2)) + 2\log|\det(A)| \quad (18)$$

Defining a block-wise size *n* marginal information redundancy

$$\begin{aligned} &\rho^n(X_1, ..., X_L) \\ &@ I(X_1, ..., X_L) - I(X_1, ..., X_{L-1}) \\ &= \sum_{i=(L-1)n+1}^{Ln} H(x_i) - H(X_L \mid X_1, ..., X_{L-1}) \end{aligned} \quad (19)$$

we regard the multi-information for sequence of blocks to be the multi-information of the variables that comprise these blocks. Specifically,

$$I(X_1, X_2, ...X_L) = \sum_{i=1}^{Ln} H(x_i) - H(x_1, ..., x_{Ln}) \quad (20)$$

for the case of non-overlapping blocks and with an appropriate correction (smaller number of variables x_i) for the overlapping case. Introducing an auxiliary notion of *n*-block-wise entropy difference for a sequence of *L* blocks

$$\begin{aligned} &H_r^n(X_1, X_2, ..., X_L) \\ &@ H(X_1, X_2, ..., X_L) - H(X_1, X_2, ..., X_{L-1}) \\ &= H(X_L \mid X_1, X_2, ..., X_{L-1}) \end{aligned} \quad (21)$$

one arrives at

$$\begin{aligned} &H_r^n(X_1, ..., X_L) \\ &= \sum_{i=1}^{n} H_r(s_i(1), s_i(2), ..., s_i(L)) + \log|\det(A)| \end{aligned} \quad (22)$$

where $H_r(s_i(1), s_i(2), ..., s_i(L))$ is the usual "scalar" entropy difference for coefficient number *i*. Inserting in the above equation, one arrives at

$$\begin{aligned} &\rho^n(X_1, ..., X_L) \\ &= \sum_{i=(L-1)n+1}^{Ln} H(x_i) - H(X_L \mid X_1, ..., X_{L-1}) \\ &= \sum_{i=(L-1)n+1}^{Ln} H(x_i) - \sum_{i=1}^{n} H(s_i(L)) \\ &\quad + \sum_{i=1}^{n} \rho(s_i(1), ..., s_i(L)) - \log|\det(A)| \end{aligned} \quad (23)$$

Noting that

$$I(X_L) = I(x_{(L-1)n+1}, x_{(L-1)n+2}, ..., x_{Ln}) = \sum_{i=(L-1)n+1}^{Ln} H(x_i) - H(X_L) \quad (24)$$

and using the relation

$$H(X_L) = \sum_{i=1}^{n} H(s_i(L)) + \log|\det(A)|$$

we arrive at the relation

$$\sum_{i=(L-1)n+1}^{Ln} H(x_i) - \sum_{i=1}^{n} H(s_i(L)) = I(X_L) + \log|\det(A)|$$

which gives

$$\rho^n(X_1, ..., X_L) = \sum_{i=1}^{n} \rho(s_i(1), ..., s_i(L)) + I(X_L) \quad (25)$$

This brings us to our final definition of block-wise information redundancy

$$\begin{aligned} &\rho_{IC}^n(X_1, X_2, ..., X_L) \\ &@\rho^n(X_1, X_2, ..., X_L) - I(X_L) \\ &= I(X_1, X_2, ..., X_L) \\ &\quad -\{I(X_1, X_2, ..., X_{L-1}) + I(X_L)\} \\ &= \sum_{i=1}^{n} \rho(s_i(i), ..., s_i(L)) \end{aligned} \quad (26)$$

To sum up, our block-wise information redundancy measure calculates the difference in information between multiple blocks, considering the difference between the multi-information over L consecutive blocks versus the sum of multi-information of the first L-1 blocks and the multi-information in the last block X_L. The convenient property of this measure is that it can be calculated from the marginal entropies of the n independent components, using the standard or the generalized SFM of the previous paragraph.

5. EXPERIMENTAL RESULTS

GSFM and IC-SFM were tested for the problem of voicing determination in speech signals and compared to the standard SFM. The purpose of the experiment was not to develop yet another voicing estimation measure but to try and evaluate the performance of the various measures on real signal. For the experiment we have used speech recordings from the Keele speech database [6]. This database contains simultaneously recorded speech and Laryngograph signals, which are used as a reference signal for describing the true vocal cord activity.

The GSFM was estimated using kurtosis estimate of Negentropy. The spectral estimator used was Burg maximal entropy estimator, with AR filter of order 16. This choice of filter is common to speech processing (speech envelope can be described by as little as 8 or 10 poles filter. Using 16 parameters gives a better approximation. Further increase is not desirable since it might capture spectral properties due to pitch correlations). The IC-SFM was estimated for 8-dimensional embedding. The dimension was chosen so as to give approximately the same amount of parameters (64 instead of 16). Using lower dimensional embeddings gives results that are closer to the scalar GSFM case.

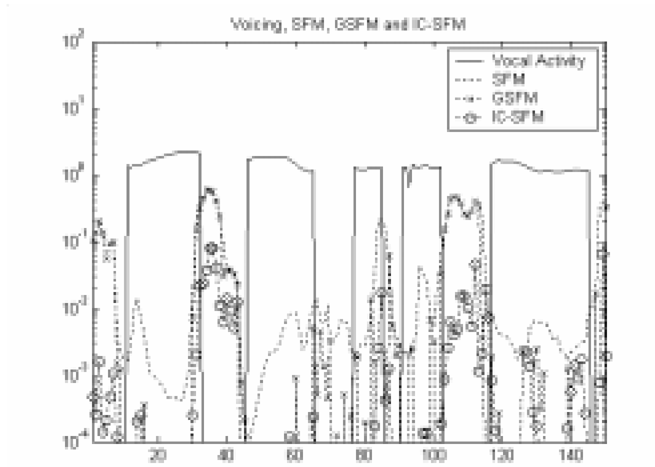

Fig 1. Vocing, SFM, GSFM and IC-SFM

Figure 1 shows the results of applying all three SFM's for a long speech signal. The top graph represents the vocal activity from the Laryngograph. The values of the vocal activity are the estimated pitch in hundred Hertz units. The drops in vocal activity correspond to unvoiced parts. One can observe that SFM increases for the regions where no vocal activity was present. Moreover, GSFM and IC-SFM tend to give correspondingly lower estimates of randomness, thus assigning more structure to areas where the standard SFM is high. On the other hands, the generalized measures have much less “false alarms”, i.e. they almost do not respond (or give low SFM) to areas where vocal activity was present. This is not that case for the SFM, which tends to overestimate randomness and segment speech as being unvoiced even when vocal activity is present. One should note that structure may well exist in speech signal even when no vocal activity occurs. This happens due to plosive or fricative sounds that have a significant amount of structure, even without being considered “voiced”.

Another experiment was conducted on musical sound tracks from various Films. In this case the analysis was done using IC based MIR on sequences of cepstral feature vectors. A detailed retrieval and similarity analysis is still to be performed. Here we bring just a short account of the differences in MIR of the different Musical signals since it seems to demonstrate some of the properties of the IC-MIR analysis. The analysis procedure consisted of several steps:

i) *Preprocessing:*
Cepstral coefficients are calculated over time frames of 600 milliseconds, with a 200 milliseconds overlap. In order not to consider signal aspects that are related to pitch information, the cepstral features were extracted by liftering the cepstral vectors (filtering in cepstral domain) at the first 120 coefficients.

ii) *IC Embedding:*
The cepstral analysis matrix was submitted to IC analysis. A set of coefficient vectors that describe the evolution of each of the basis vector over time was obtained.

iii) *Marginal Information Redundancy:*
The MIR was calculated separately for every column of the coefficients matrix.

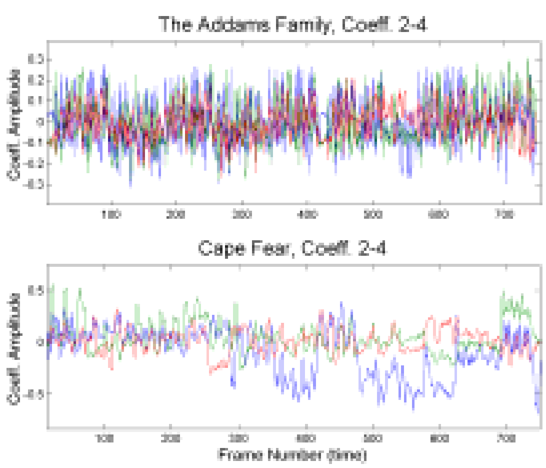

Fig 2. IC for two Musical Film signals

Figure 2 shows that behavior of the time evolution of coefficients 2, 3 and 4 for two distinct film music tracks. The top graph corresponds to “Adams Family” film music track and has a rap-like style, with a lot of rhythmic and color changes. The bottom graph shows “Cape Fear” film music track that has a slowly developing orchestral sound with predominant strong violin section. One can clearly observe the different behavior of the coefficients of these musical tracks. Marginal Information Redundancy quantitatively measures this behavior.

6. CONCLUSION

In this paper we presented several generalizations of the standard spectral flatness measure using an information theoretic formulation of randomness as a marginal information redundancy. It was shown that these new measures are capable of detecting excessive structure due to non-Gaussian properties of the signal. Statistical evaluation of these new generalized measures for practical applications such as voice activity detection, compression or capturing musical mood or style will be explored in the future.

7. REFERENCES

[1] N.S. Jayant, and P.Noll , *Digital Coding of Waveforms*, Prentice-Hall, 1984

[2] J.D. Johnston, *Transform Coding of Audio Signals Using Perceptual Noise Criteria*, IEEE Journal of Selected Areas in Communication, Vol. 6, pp. 314-323, 1988

[3] E. Allamanche, J. Herre, O. Hellmuth, B. Fröba, T. Kastner, M. Cremer, *Content- based Identification of Audio Material Using MPEG- 7 Low Level Description*, in proceedings of International Symposium on Muisc Information Retrieval (ISMIR), Indiana University, 2001

[4] T.M. Cover, and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, New-York, 1991.

[5] A. Hyvärinen, New approximations of differential entropy for independent component analysis and projection pursuit, *Advances in Neural Information Processing Systems*, 10, pp. 273-279, MIT Press, 1998.

[6] F. Plante, G.Meyer, and W.A.Ainsworth, *A pitch extraction reference database*, Eurospeech’95, Madrid, pp. 837-840, 1995