Analysis of musical structure in Audio and MIDI signals using information rate

Shlomo Dubnov
Department of Music, UCSD

Abstract
Information rate (IR) was recently introduced as a novel measure of musical structure that quantifies prediction properties of musical signals. Formulated in information theoretic terms, it quantifies the rate of change of information in time series, and is related to capability of a system to perform information analysis of the data, such as actions involved in music listening. In this paper we extend IR measure to the case of MIDI sequences and perform a comparative analysis of audio signals (acoustic recordings) and symbolic representations (note, i.e. MIDI) of same musical piece. We show that in both cases IR detects similar types of structure, suggesting that IR might capture some inherent properties of musical signals.

Keywords: information rate, entropy, prediction, complexity, musical structure

Introduction

Information rate (IR) was introduced [1] as measure of structure of stochastic process, defined in terms of relative reduction in uncertainty (entropy) due to prediction of the future based on the past. Algorithms for estimation of IR for audio signals, spectra and feature vector sequences were presented. In another work [2], IR was compared to human emotional judgments when listening to music, suggesting that IR might be related to emotions that are evoked in relation to varying extents of human ability to anticipate the future of music material over the course of a musical composition.
In the current paper we extend IR analysis to MIDI sequences by using low order Markov models as predictors of note sequences. We perform comparative analysis of music using both signal and symbolic representations of a Bach Prelude (Prelude in G from Well Tempered Clavier, Book I). We analyze several recordings of different performances of the piece and show that IR detects similar types of structure. This suggests that IR might be capable of capturing some inherent properties of musical signals.

Mathematical background

Mutual information measures the amount of information contained in variables $x_1, x_2, ..., x_n$,

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

where H(.) denotes the entropy of individual variables or of a set of variables [3]. Information rate (IR) is defined as the difference between the information contained in $x_1, x_2, ..., x_n$ versus $x_1, x_2, ..., x_{n-1}$, i.e. lacking the last observation. In other words IR measures the amount of information that is added when next variable is observed.

Definition I:

$$\rho(x_1, x_2, ..., x_n) = I(x_1, x_2, ..., x_n) - I(x_1, x_2, ..., x_{n-1}) \quad (2)$$

It can be shown that IR can be equivalently defined in terms of a difference between entropies of the data before and after prediction

Definition II:

$$\rho(x_1, x_2, ..., x_n) = H(x_n) - H(x_n \mid x_1, x_2, ..., x_{n-1}) \quad (3)$$

and also in terms of mutual information between signal future and its past

Definition III:

$$\rho(x_1, x_2, ..., x_n) = I(x_n, \{x_1, x_2, ..., x_{n-1}\}) = I(x_{past}, x_n) \quad (4)$$

Scalar IR Estimators

For Gaussian linear processes, IR can be estimated using Linear Prediction and Spectral Flatness. Assuming an autoregressive model of order p with innovations ε_n, $\varepsilon_n = x_n - \sum_{i=1}^{p} a_i x_{n-i}$ IR can be expressed as logarithm of the ratio between variance of the signal and variance of the innovation

$$\rho = H(x) - H_{Cond}(x) = \frac{1}{2}\log(\frac{\sigma_x^2}{\sigma_\varepsilon^2}) \quad (5).$$

Another estimate of IR is derived using the Spectral Flatness Measure [6], which is written in discrete case as the ratio of geometric and arithmetic means of the signal spectrum,

$$\exp(-2\rho(x)) = \frac{\exp(\frac{1}{N}\sum_i \ln S(\omega_i))}{\frac{1}{N}\sum_i S(\omega_i)} = \frac{\left(\prod_{i=1}^{N} S(\omega_i)\right)^{\frac{1}{N}}}{\frac{1}{N}\sum_i S(\omega_i)} \quad (6),$$

The IR measure was generalized to linear non-Gaussian processes in [4]. This can be regarded as correction to the standard SFM measure and may be used to detect signal changes when the innovation higher order moments change, even when the prediction filter (or signal spectral envelope) remains unchanged.

Relation of IR to Change Detection

The purpose of change detection is finding points where a current model of a signal no longer "explains" the data and a new model has to be estimated. One of the common methods for comparison of acoustic signals is the Itakura Saito (IS) spectral distance [5], defined as

$$D(S_1, S_2) = \int_{-\pi}^{\pi} \left[e^{-V(\omega)} - V(\omega) - 1\right] \frac{d\omega}{2\pi}, \quad (7)$$

with $V(\omega) = \log\left(\frac{S_1(\omega)}{S_2(\omega)}\right)$, and $S_1(\omega)$ and $S_2(\omega)$ are power spectra of two signal segments being compared. Writing IR as a function of SFM, we get

$$\rho(x) = -\frac{1}{2}\log(SFM(x)) = -\frac{1}{2}[\int \log S(\omega)\frac{d\omega}{2\pi} - \int S(\omega)\frac{d\omega}{2\pi}], \quad (8)$$

which after simple algebraic manipulation becomes?

$$\rho(x) = \frac{1}{2}\int [e^{-V(\omega)} - V(\omega)]\frac{d\omega}{2\pi} + const, \quad (9)$$

using $S_1(\omega) = S(\omega), S_2(\omega) \equiv 1$ in the expression $V(\omega) = \log\left(\frac{S_1(\omega)}{S_2(\omega)}\right)$.

This shows that IR is equivalent, up to a constant factor, to the IS distance between the signal and a process with flat spectrum, that is white noise.

Vector IR

Given a multi-variate distribution of vectors $\bar{x} = (x_1, x_2, ..., x_n)^T$, the multi-information for sequence of blocks is defined as

$$I(X_1, X_2, ...X_L) = \sum_{i=1}^{Ln} H(x_i) - H(x_1, ..., x_{Ln}) \quad (10)$$

This generalizes the definition of IR to the multivariate case

$$\rho_L^n(X_1, X_2, ..., X_L) \triangleq I(X_1, X_2, ..., X_L) - \{I(X_1, X_2, ..., X_{L-1}) + I(X_L)\} \quad (11)$$

Vector IR considers the difference in information over L consecutive signal frames versus the sum of information of the first *L-1* frames and the information in the last frame X_L .

Assuming an expansion of data vectors $X = AS$ in a basis given by columns of the matrix *A,* and independent coefficients *S*, it can be shown [1] that IR becomes sum of IRs of the individual coordinates of the expansion coefficients,

$$\rho_L^n(X_1, X_2, ..., X_L) = \sum_{i=1}^{n} \rho(s_i(i), ..., s_i(L)) \quad (12)$$

IR of spectral sequences

When IR analysis is applied to a sequence of spectral vectors, such as vectors of short time Fourier transform (STFT) or sequences of cepstral coefficients, a rotation of the data matrix is performed in order to satisfy condition (12). In the case of a multivariate Gaussian distribution, independence can be achieved using a decorrelation procedure, such as the Karhunen-Loeve transform (KLT) [7], also known as Principal Components Analysis (PCA). In this paper we are considering sound representation in terms of low order cepstral coefficients [8]. Long term correlations related to pitch structure are removed from this representation by cepstral preprocessing ("liftering" step).

Notes IR estimation

The principles of IR can be applied to symbolic sequences using estimators of entropy and conditional entropy over finite alphabet. By symbolic representation we mean a representation of music as a set of notes, which is equivalent to music notation, or performance actions of a musician (pianist in our case) who performed the piece. This information is stored in MIDI files. In experiments described in this paper we considered note numbers (corresponding to keys on the piano keyboard) sequenced in the order of their appearance but disregarding their exact onset times, durations or dynamics (so called MIDI velocities). Short time estimate of entropy and conditional entropy was performed using blocks of 40 notes, with overlap of 30 notes between successive blocks. The choice of block size was such that the duration of the musical segment corresponded approximately to three seconds in duration, which was the time of analysis of the audio signal. Estimation of the marginal entropy in every block was performed according to the following procedure:

1. Frequencies of appearance of every note in a block were obtained by counting the number of note appearances divided by the total number of notes in the block.
2. Using these frequencies as probabilities $p_i, i = 1..q$, entropy was estimated using the definition

   $H(p_1, ..., p_q) = \sum_{i=1}^{q} p_i \log_2 p_i$ (13), with q the range of different notes.

Conditional entropy was estimated in terms of low order Markov model, separately for every block. In order to maintain a constant number of model parameters across different blocks, a single prediction table was constructed over the whole sequence of the note data (instead of constructing separate table for every block). A

step of constructing a prediction table was as follows:

1. Construct a context table whose entries (columns) are vectors containing all note sequences of length equal to the Markov order used for modeling the piece.
2. For each context construct a list of possible continuations (i.e. single notes that follow that context) throughout the whole piece (table rows) and compute their entropy using the equation (13) using frequencies of appearance of notes given the context.

Given the table, conditional entropy estimation within a block was done as follows:

1. Read the notes in the block from beginning to end using a window of size equal to the Markov order.
2. For every window vector look up the corresponding entropy value form the prediction table.
3. Sum these individual entropies to obtain the conditional entropy for that block.

The note IR is computed as a difference between the entropy and the conditional entropy. The time step of IR was assigned to be the mean time of the note onsets for that block.

Experimental results

We present IR analysis of several performances of a J.S.Bach Prelude in G from the first book of the Well Tembered Clavier. In figure 1 we present the evolution of signal energy, scalar and vector IR and IS distance of a performance of the Bach Prelude by Edwin Fischer. It can be seen that IR and IS graphs show different responses, indicating that they respond to different signal properties. Moreover, scalar IR analysis does not give meaningful segmentation, but the energy has a very prominent profile that outlines musical structure and is close to the IR graph.

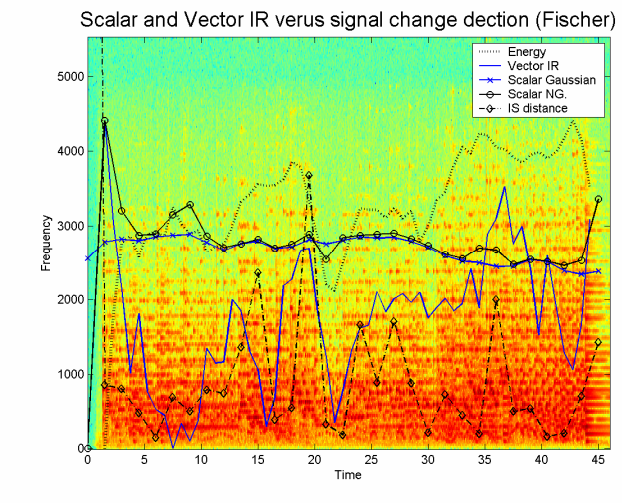

Figure 1. Energy, scalar Gaussian, non-Gaussian and vector IR, and IS distance estimation of the Bach Prelude performed by Edwin Fischer.

Figure 2 presents similar analysis of a performance by Glen Gould. The IR and IS are significantly different compared to Fischer's performance. Moreover, the energy and IR are almost "opposite", with higher IR corresponding to decrease in energy.

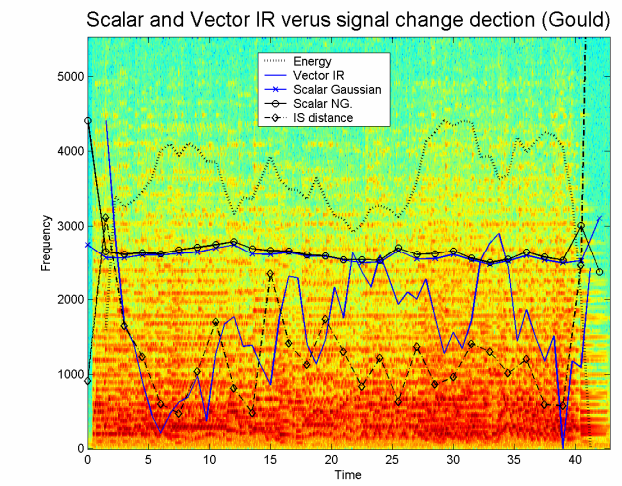

Figure 2. Energy, scalar Gaussian, non-Gaussian and vector IR and IS distance estimation of the Bach Prelude performed by Glen Gould.

A third example is a computer rendering of a MIDI file. The piano sounds are derived from an internal computer synthesizer, recorded back to

the same computer as an audio file. This is the most dynamically and rhythmically "flat" performance. What is also remarkable is that IR and IS profiles are distinct and different from the two live performances.

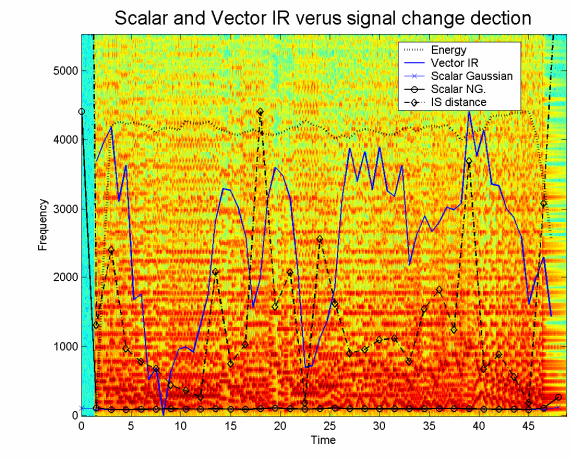

Figure 3. Energy, scalar Gaussian, non-Gaussian and vector IR and IS distance estimation of the Bach Prelude as recorded from synthetic rendering of the music from MIDI file by the computer.

It should be noted that the graphs were scaled so as to fit over the spectrogram in the background. The axes of the plot correspond to the spectrogram units. The scaling of the analysis function in the plot are as follows: energy and spectral anticipations are scaled to 80% of Nyquist frequency, the scalar Gaussian and non Gaussian IR are normalized to be less or equal to 1 and scaled also to 80% of Nyquist frequency, and IS distance is scaled to 20% of Nyquist frequency (without normalization).

Results of IR Analysis of symbolic music representation

Figure 4 shows graphs of note entropy, conditional entropy and their difference (IR), plotted under a graph that shows note occurrences in the Prelude. The y-axis corresponds to MIDI note numbers, and the x-axis is the time. The values of entropy were scaled so as to show conveniently under the score and do not have an absolute meaning in this graph.

It can be seen that entropy of the piece is rather flat throughout its duration, which indicates that statistics such as range and frequency of occurrences of notes in different musical sections do not provide much information about what is "really" going on.

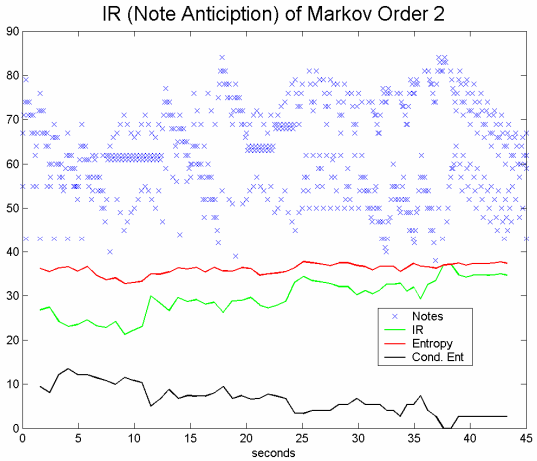

Figure 4. Graphs of note entropy, conditional entropy and their difference (IR), plotted under a graph that shows occurrence of note onsets of the Prelude. The y-axis corresponds to note numbers of the MIDI file.

Conditional entropy decreases towards the end of the piece, indicating that the predictability of the music increases (less bits needed to describe the next note in context of its past).

In order to evaluate the significance of IR for the different music representations, we have compared note and spectral anticipations from MIDI and the synthetically rendered audio files, respectively. The two graphs are shown in figure 5.

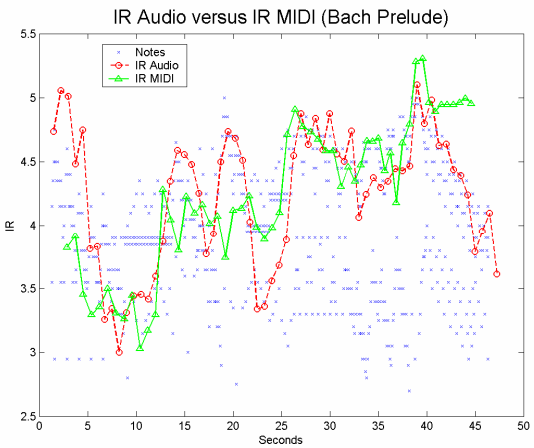

Figure 5. Energy, scalar Gaussian, non-Gaussian and vector IR and IS distance estimation of the Bach Prelude as recorded from synthetic rendering of the music from MIDI file by the computer.

It is interesting to observe that although the signals and their respective statistical models were very different, the IR graphs exhibit significant similarity. The signal IR graph was scaled so as to match the range of the note IR values. The background of the figure shows again the notes onsets, but this time without relation to the y-axis.

References

[1] Dubnov, S. (2004), *Spectral Anticipations*, Proceedings of International Computer Music Conference, Miami.

[2] Dubnov, S. S. McAdams, R. Reynolds (2005). Structural and Affective Aspects of *Music from Statistical Audio Signal Analysis,* to appear in Journal of the American Society for Information Science and Technology, Special Issue on Style 2006.

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

[4] Dubnov, S. (2004). *Generalization of Spectral Flatness Measure for Non-Gaussian Linear Processes*, IEEE Signal Processing Letters, Volume 11, Issue 8, Aug. 2004 Page(s):698 - 701

[5] Gray, R.; Buzo, A.; Gray, A., Jr.; Matsuyama, Y. (1980). *Distortion measures for speech processing*, IEEE Transactions on Acoustics, Speech, and Signal Processing, Volume 28, Issue 4, Aug., 367 – 376.

[6] Jayant, N.S. and P.Noll. (1984). *Digital Coding of Waveforms,* Prentice-Hall Signal.

[7] Hayes, M. (1996). *Statistical Signal Processing and Modeling,* Wiley.

[8] Oppenheim, A.V. and R.W.Schafer (1989). *Discrete Time Signal Processing,* Prentice Hall, New Jersey.