NONLINEAR TIME SERIES ANALYSIS OF MUSICAL SIGNALS

J. D. Reiss and M. B. Sandler

Queen Mary, University of London
Department of Electronic Engineering
Mile End Road, London, N52LL, United Kingdom
josh.reiss,mark.sandler@elec.qmul.ac.uk

ABSTRACT

In this work the techniques of chaotic time series analysis are applied to music. The audio stream from musical recordings are treated as representing experimental data from a dynamical system. Several performance of well-known classical pieces are analysed using recurrence analysis, stationarity measures, information metrics, and other time series based approaches. The benefits of such analysis are reported.

1. INTRODUCTION

Data from a multidimensional nonlinear system is often difficult to analyse. Structure may be masked by the inherent complexities of the system. Thus a specialised set of techniques have been designed specifically for the analysis of time series from nonlinear systems.

Extracting and investigating the musical content in an audio signal presents unique challenges. Although much of the relevant information is based in rhythmic and harmonic structures, even frequency domain analysis may fail with polyphonic, multivoice recordings. Furthermore musical structures may be based on nonlinear correlations in the data. Traditional signal processing techniques are not designed to identify such relationships. Nor do they provide quantitative measurement of the complexity or information content in the signal.

There have been few previous attempts to apply nonlinear time series analysis techniques to audio. Witten[1] used the reverse approach, attempting to understand well-known dynamical systems as musical compositions. Boon's studies [2,3] are quite similar to ours, with the primary exception being that he considered the symbolic representation, not the raw audio. Hemenway [4] attempted to calculate the fractal dimension of compositions by Mozart and Bach, again using the symbolic representation. Gibiat [5] approached audio as being a dynamical system, but he concentrated on the structure of notes and chords, not on the dynamics of rhythms. In some sense, those approaching music and audio analysis using a hidden Markov model[6] or neural network[7] approach are most similar to ours because they involve the identification of rhythmic structures as a result of underlying dynamical processes.

In this work, we use time series analysis to reveal hidden musical structures that represent long-term rhythms and repeating patterns. Furthermore, we can use this to quantify the complexity of the signal. Several examples are presented and the results are reported. We also comment on the challenges presented in this type of analysis, and how they might be overcome.

2. METHOD

We base our analysis around several well-known classical compositions. This was done partly because classical recordings typically have little distortion and also because of their familiarity. These recordings, freely available and in the public domain (www.intelliscore.net), are described in Table 1 and will be referred to in the text by their abbreviations.

We should note here that we do not limit ourselves in any way to monophonic or single voice recordings. However, the samples that were analysed were relatively simple recordings from a synthesiser. This was to eliminate some pitch and rhythm errors that would occur in live performances on acoustic instruments.

Table 1. A list of the four audio samples analysed, with data concerning their duration and quality, as well as the abbreviation used to refer to them in the text.

Abbreviation	*Composer*	*Title*	*Sample Rate (kHz)*	*Channels*	*Duration (sec)*	*Sample size (bits)*
E	Elgar	Pomp and Circumstance	44	1	38	16
B	Bach	Jesu, Joy of Man's Desiring	22	2	13	16
R	Ravel	Bolero	44	1	39	16
T	Tchaikovsky	Swan Lake	44	1	27	16

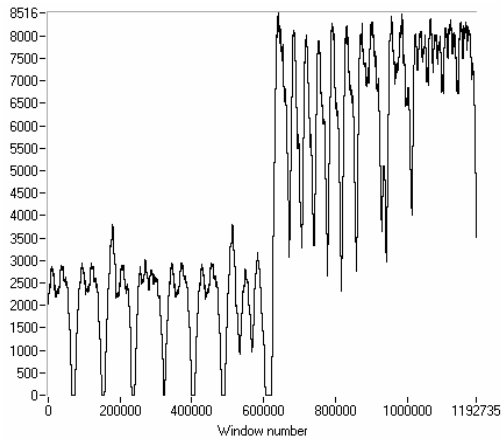

Figure 1: Frame variances from recording ***T***.

The analysis techniques used were as follows.

2.1. Sliding window analysis

This type of analysis is typically performed in order to identify long-term dynamics, parameter drift, and other forms of nonstationarity in data sets. Sliding windows of varying length are applied to the data sets and some quantitative measure of each window is plotted as a function of time. In this case, standard deviation was thought to be an appropriate measure because it would capture rapid changes in frequency and amplitude. The results of this analysis are depicted in Figure 1. The entire musical excerpt lasts about 30 seconds, so we can identify repetitive rhythms on a long time-scale.

This measure also clearly identifies the dramatic change at approximately 20 seconds into the music. The rhythm changes slightly, and the amplitude increases. In fact, we can characterize this as symbolic dynamics, with a transition from a low variance to high variance repetition. It is our goal to more accurately identify the dynamics of music that typify this piece. It should be expected that we find certain time series analysis-based measures which are robust to changes in timbre, and hence isolate relevant aspects of harmony, rhythm and melody.

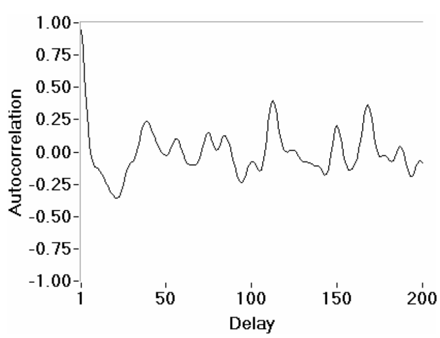

Figure 2. The autocorrelation of recording ***B*** as a function of the delay used in a 2 dimensional embedding.

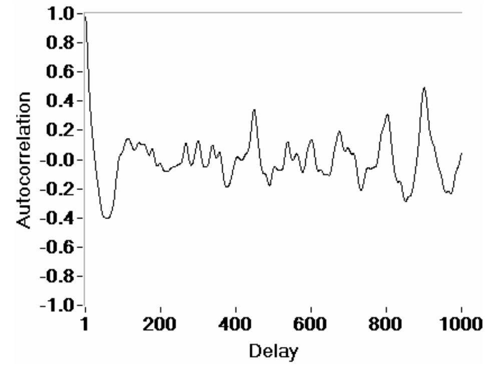

Figure 3. The autocorrelation of Elgar's *Pomp and Circumstance* as a function of the delay t used in a two-dimensional embedding.

2.2. Delay Coordinate Embeddings

A common first step in the analysis of time series data from a nonlinear dynamical system is to construct a delay coordinate embedding. A delay coordinate embedding is typically used as a method to reconstruct multidimensional data from a one dimensional time series[8].

Consider a time series of the form $X_1, X_2, ...X_N$. From this series, vectors are created using a delay t and embedding dimension D:

$$\begin{aligned}\overline{Y_1} &= (X_1, X_{1+t}, X_{1+2t}, ...X_{1+(D-1)t}) \\ \overline{Y_2} &= (X_2, X_{2+t}, X_{2+2t}, ...X_{2+(D-1)t}) \qquad (1) \\ &...\end{aligned}$$

Various methods have been suggested for the determination of an appropriate delay [9]. In this section we concentrate on two alternatives; the first zero-crossing of the autocorrelation function; (as a function of the delay as it is increased from zero), and the first minimum of the mutual information function.

Both these choices serve to represent vectors where the individual components share minimal information due to the dynamics, but are still closely related in time. The autocorrelation function is often used to identify periodicities in data. However, the autocorrelation function only measures linear correlations, whereas the mutual information function describes nonlinear relationships between coordinates. The mutual information may give a better value because it takes nonlinear correlations into account. In this section, we use an efficient method [10] of calculating mutual information in order to identify and characterize nonlinear, delayed correlations in the data

Figure 2 depicts the (normalized) autocorrelation as a function of the delay in recording ***B***. The delay is given in units of the sample rate, or 0.045 msec. The first zero-crossing of the autocorrelation occurs at t=7, or 0.315 msec. For recording ***E***, the autocorrelation function is given in Figure 3.

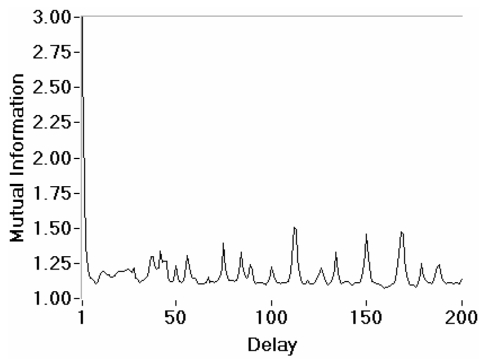

Figure 4. The mutual information of recording ***B*** as a function of the delay used in a 2 dimensional embedding.

By comparing Figure 2 and Figure 3, we can see that both reveal strong correlations at t>100. However, these correlations are still at time durations less than a second. Thus the autocorrelation function is revealing the correlations within a note, and not from long-term rhythmic structures in the audio. This does not invalidate the method however, since it simply gives us a choice of delay which does not *overemphasize* the short-time correlations.

The mutual information, depicted in Figure 4, suggests a similar value for t. However, the mutual information is more difficult to interpret, since it is more susceptible to noise, data complexity, and finite data set size. Coupled with the fact that computation of mutual information from a large data set is computationally intensive, and it becomes clear that when successful, the autocorrelation function is preferred over the mutual information as a method of determining an appropriate delay to use in embedding an audio signal.

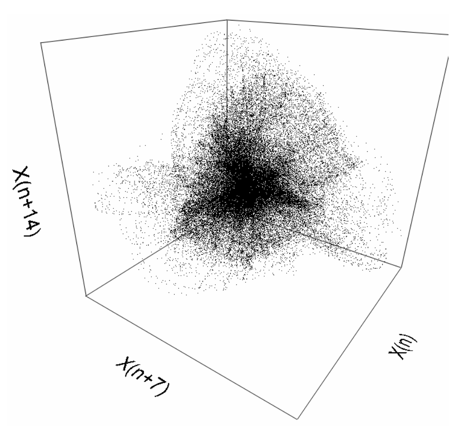

Figure 5. A three dimensional embedding of Bach's *Jesu, Joy of Man's Desiring* using the delay τ=7, as suggested by the autocorrelation function.

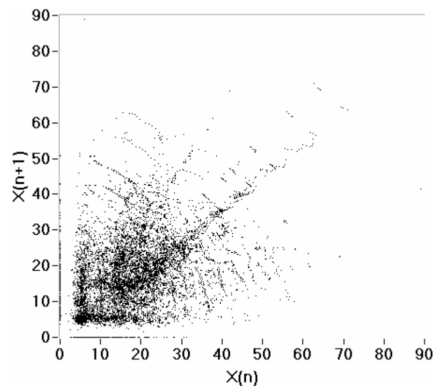

Figure 6. A two-dimensional return map of a Poincare section of recording ***B***.

In Figure 5, a three-dimensional delay coordinate embedding of recording ***B*** is provided. Here, we used the delay suggested by the first zero-crossing of the autocorrelation function. Note that this procedure, which incorporates multidimensional temporal structure into a simple graph, reveals a nonrandom distribution to the data. Certain directions appear favored for the 3 dimensional delayed vectors. This would not be revealed from the time domain waveform or the frequency domain spectrum.

2.3. Poincare sections using peak detection

If a dynamical system contains a term that is periodic in time then it can be sampled at that period and hence reduced in dimensionality. The act of taking a Poincare section, as it is known, reduces a continuous time system to a return map and is a useful tool in nonlinear dynamics. However, with audio signals it is not clear what period, if any can be used.

Here, we can borrow a technique that is used in the application of nonlinear time series analysis to biological systems[11-13]. For data from neurons, both experimental and simulated, the dynamics consist of both slow and very fast processes. This has the effect of creating sharp spikes in the time domain waveform. The times between these spikes represent the times between piercings of a Poincare surface of section. This suggests that vectors consisting of successive interspike intervals will reconstruct the multidimensional dynamics in a set of dimension exactly one less than the attractor dimension.
Although a musical signal does not consist of spike trains, it does have fast (the period of a single note) and slow (the times between notes) processes. Thus we can apply a similar technique to audio.

Using a polynomial interpolation, the time intervals between peaks were extracted for recordings ***R*** and ***B***. These were used to form multidimensional vectors which are depicted in Figure 6 through Figure 8. Figure 6 shows a structured, nonGaussian distribution to the peak intervals. A simulated noisy waveform would exhibit a Gaussian distribution focused on zero. Instead, there appears to be well-defined allowable regions, for the embedded inter-peak times. These should correspond to certain notes, and the peak interval times represent the period of the noteor notes, i.e., pitch.

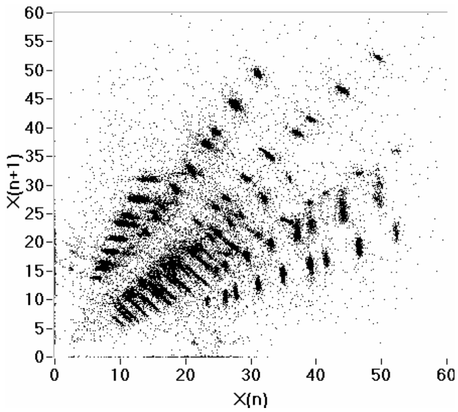

Figure 7. A two-dimensional return map of a Poincare section of *Bolero*, created from the time intervals between peaks. Units are given in terms of the sample period.

This behaviour is even more pronounced in Figure 7. The allowable regions (ignoring points that seem to represent a noisy background) form several sets of discrete stairs ascending linearly away from the origin. These paths overlap, thus indicating that more than two dimensions are required to unfold the structure.

This is done through the use of a three dimensional plot, as depicted in Figure 8. The figure has been rotated to reveal approximately 11 paths. In three dimensions, the paths are independent. Again we believe that these regions, represent the harmonics of a note or chord. By incorporating additional temporal information into these graphs, such as through a windowing technique to create short time Poincare sections, it may be possible to use this method in a transcription scheme.

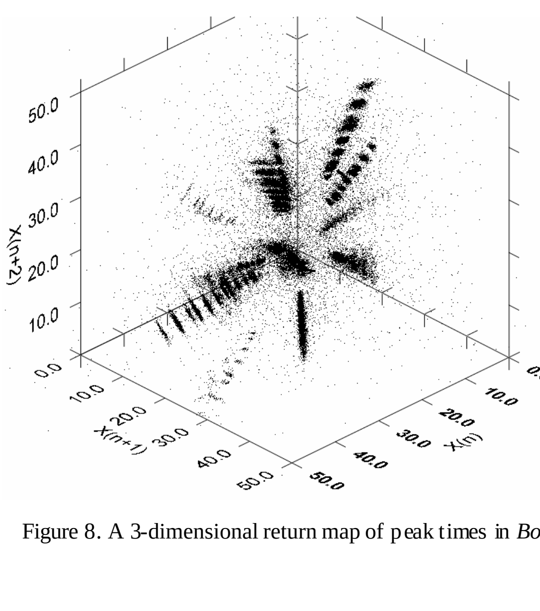

Figure 8. A 3-dimensional return map of peak times in *Bolero.*

2.4. Recurrence plots

The recurrence plots were first suggested by Eckmann, et. al.[14] as a means of identifying nonstationarity as well as aperiodicity. They have since been found to be a useful tool in visualising a wide variety of behaviour.

The phase space is reconstructed by taking the time series $X_1, X_2 \ldots X_N$ and creating D-dimensional vectors using the embedding method mentioned in Equation (1).

The recurrence plot is formed by comparing all embedded vectors with each other. The distance between vectors *i* and *j*, $\left\| \overline{Y_1} - \overline{Y_2} \right\|$ is then plotted (using a grayscale intensity) at the coordinate value (*i, j*). This is depicted in Figure 9. Note that the recurrence plot is symmetric since the distances are symmetric between *i* and *j*. The horizontal and vertical white bands are associated with rare states, and the dark bands indicate states that remain roughly unchanged for a long duration.

Of further interest is the relative lack of diagonal lines. This implies a lack of long-term periodicity, and hence the rhythm must be represented in a different manner. Furthermore, the change in amplitude is visualized both by the abrupt rise in variance in Figure 1(for *Swan Lake*) and also by the increased density in the region to the right of Figure 9(for *Bolero*).

3. CONCLUSIONS

The investigations reported here were speculative since there has been very little previous work on the application of nonlinear time series techniques to raw audio music. These techniques, although powerful and extensible, are not designed for use with audio.

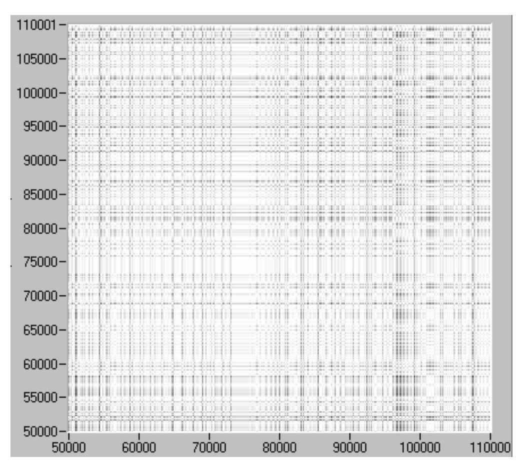

Figure 9: A recurrence plot depicting the dynamics in *Bolero.*

In particular, without modification, they have a tendency to identify short term dynamics and miss the rhythmic structures. Despite this shortcoming, nonlinear time series analysis techniques are well-suited to musical audio signals. This was evidenced by the fact that these techniques allowed us to observe rhythmic structures occuring over very long time spans. Such structures would not have been evidenced using many frequency domain approaches.

We also showed that the times between peaks in a peak detection routine can be used to identify notes. By incorporating additional temporal information, this can be used to represent a symbolic dynamics. Thus this work relates directly to the previous work on time series analysis of music in the symbolic representatiuon[3,4].

4. REFERENCES

[1] M. Witten, *The Sounds of Science: II. Listening to Dynamical Systems. Towards a Musical Exploration of Complexity*, Computers and Mathematics with Applications **32** (1), 145 (1996).

[2] J.-P. Boon, A. Noullez, and C. Mommen, *Complex dynamics and musical structure*, Interface **19**, 3 (1990).

[3] J. P. Boon and O. Decroly, *Dynamical systems theory for music dynamics*, Chaos **5** (3), 501 (1995).

[4] J. Hemenway, *Fractal Dimensions in the Music of Mozart and Bach*, The Nonlinear Journal **2**, 86 (2000).

[5] V. Gibiat, *Phase space representations of acoustical musical signals*, J. Sound Vib. **123**, 529 (1988).

[6] A. Durey. *Melody Spotting Using Hidden Markov Models*. 2nd Int. Symposium on Music Inform. Retriev. (ISMIR), Bloomington, IN, USA, October 2001.

[7] A. Robel. *Neural network Modeling of Speech and Music Signals*. Neural Information Processing Systems 9 (NIPS 96)1996.

[8] T. Sauer, J. A. Yorke, and M. Casdagli, *Embedology*, J. Stat. Phys. **65**, 579 (1991).

[9] A. M. Fraser, *Reconstructing Attractors from Scalar Time Series: A Comparison of Singular System and Redundancy Criteria*, Physica D **34**, 391 (1989).

[10] J. D. Reiss, N. Mitianoudis, and M. B. Sandler. *Computation of Generalized Mutual Information from Multichannel Audio Data*. 110th Convention of the Audio Engineering Society, Amsterdam, The Netherlands, 2001 May 12-15 2001.

[11] D. M. Racicot and A. Longtin, *Interspike interval attractors from chaotically driven neuron models*, Physica D **104**, 184 (1997).

[12] T. Sauer, *Reconstruction of dynamical systems from interspike intervals*, Phys. Rev. Lett. **72**, 3811 (1994).

[13] R. Castro and T. Sauer, *Reconstructing chaotic dynamics through spike filters*, Phys. Rev. E **59** (3), 2911 (1999).

[14] J.-P. Eckmann, S. O. Kamphorst, and D. Ruelle, *Recurrence plots of dynamical systems*, Europhys. Lett. **4**, 973 (1987).