



HAL
open science

Music Design with Audio Oracle using Information Rate

Shlomo Dubnov, Gérard Assayag

► **To cite this version:**

Shlomo Dubnov, Gérard Assayag. Music Design with Audio Oracle using Information Rate. FIRST INTERNATIONAL WORKSHOP ON MUSICAL METACREATION (MUME 2012), Oct 2012, Palo Alto, United States. pp.1-1. hal-01161018

HAL Id: hal-01161018

<https://hal.science/hal-01161018>

Submitted on 8 Jun 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Music Design with Audio Oracle Using Information Rate

Shlomo Dubnov

Music Department
University of California in San Diego
Email: sdubnov@ucsd.edu

G erard Assayag

Ircam - UMR CNRS STMS
Paris, France
Email: assayag@ircam.fr

Abstract

In the proposed demo we will present a method for design of musical composition using the Audio Oracle (AO) - a machine improvisation method that constructs and produces variation from a music recording using a graph of repeated factors found in the recording. The compositional / improvisation use of AO involves controlling the rate of recombination, the length of common history (length of the repeated factors) and selection of regions in the AO where the machine operates. One of the challenges in working with AO is understanding the generative potential of different audio materials and marking and allocating regions in a recording where AO should operate to achieve the desired musical result. Recently we introduced a novel analysis method "on top" of the AO structure that captures aspects of order and complexity that we call Information Rate. This measure, belonging broadly to a field of study know as Musical Information Dynamics, is related to Bense formulation of aesthetics in terms of entropy and compression. In the demo we will briefly explain the theory behind Audio Oracle and Information Rate and demonstrate a process of designing a composition / structured improvisation based on analysis of the AO graph. In contrast to the usual live interaction method where both the audio input to the oracle and the improvisation are done "on the fly", in this presentation the audio analysis will be done prior to the presentation, while during the talk we will show the process of planning a composition and then do a live performance with the AO based on this design.

Introduction

Audio Oracle is a graph structure built on top of features derived from an audio recording that indexes repeating segments (factors) in a sequence of features (Dubnov, Assayag and Cont, 2007). This structure is derived from a string matching algorithm called Factor Oracle (FO) (Allauzen, Crochemore and Raffinot, 1999) and generalizes its exact matching approach to approximate matching over a metric space. For generative applications AO uses the links between common suffixes for recombination of sub-clips of the original recording. Since audio recombinations are done in a smooth manner due to similar past (common suffixes) among the recombined segments, this method effectively

accomplishes "stylistic re-injection" (Assayag, Bloch, Cont and Dubnov, 2010).

Music Information Dynamics is a field of study that considers evolution in information contents of music (Dubnov 2006),(Potter, Wiggins and Pearce 2007),(Abdallah and Plumbley 2009), which is assumed to be related to structures captured by cognitive processes such as forming, validation and violation of musical expectations (Meyer 1956). In particular, a measure called Information Rate (IR) was studied in relation to human judgements of emotional force and familiarity (Dubnov, McAdams and Reynolds 2006)

Information Rate (IR) measures the reduction in uncertainty about a signal when past information is taken into account. It is formally defined as mutual information between past $x_{past} = \{x_1, x_2, \dots, x_{n-1}\}$ and the present x_n of a signal x

$$IR(x_{past}, x_n) = H(x_n) - H(x_n|x_{past}) \quad (1)$$

with $H(x) = -\sum P(x) \log_2 P(x)$ and $H(x|y) = -\sum P(x, y) \log_2 P(x|y)$ being the Shannon entropy and conditional entropy respectively, of variable x distributed according to probability $P(x)$. In (Dubnov, Assayag, and Cont 2011) an alternative formulation of IR using AO was developed

$$IR_{AO}(x_{past}, x_n) = C(x_n) - C(x_n|x_{past}) \quad (2)$$

with $C(\cdot)$ being the coding length obtained by a compression algorithm and measured in terms of the number of bits required to represent each element x_n .

It should be noted that eighty years ago Birkhoff formalized the notion of aesthetic in terms of the relations between order and complexity (Birkhoff 1933), later formulated in terms of information theory by Bense (Bense 1969) where complexity is measured in terms entropy, and order is the difference between uncompressed and compressed representations of the data. In our case past is used for compression, and the whole method bears resemblance to the computational aesthetics approach in the sense that it provides a graph of order (order = uncompressed complexity - compressed complexity), thus motivating us to use these "local aesthetics" measures for choice of materials in our meta-composition design.

In a recent paper (Dubnov, Assayag and Cont 2011) it was shown how Audio Oracle structure can be used to estimate

IR of an audio recording. For the sake of completeness, we will briefly recapitulate this method here. For more details on the compression method itself the reader is referred to (Lefebvre and Lecroq 2001).

Details of AO-IR calculation method

When a new frame ($i+1$) arrives to AO, a new suffix link and the Longest Repeated Suffix (*LRS*) for this frame are computed. In case when no previous repetition was found the suffix link points to state 0. In such case this new frame has to be individually encoded. Otherwise if a suffix link to earlier position is found and the LRS is smaller than the number of steps passed since the last encoding event, then the whole preceding segment is encoded as a pair (*length, position*).

For example, in the symbolic case the word *aabbabbab-bab* will be encoded as follows: The first letter will be encoded using the letter *a* using 1 bit over an alphabet $\{a, b\}$. The next occurrence of *a* can be encoded as a pair (1, 1), but since encoding it will take more bits than encoding *a* individually, we will use the shorter encoding and proceed to encoding next *b* individually and then deciding between representing the following *b* as a pair (1,3) or as a single letter, choosing the later option. The compression advantage appears for the remaining portion of the string. According to the encoding method this segment can be encoded as a pair (8, 2), which will take $\log_2(8) + \log_2(12) = 6.58$ bits, practically offering 1 bit saving compared to encoding of the 8 characters individually.

In the AO case, the symbols are replaced by feature vectors, and in order to define the uncompressed complexity $C(x_n)$ we need to find a quantity that would be equivalent to the log of the size of the alphabet in the symbolic case. We do that by considering the number of forward links originating from state 0, $\delta(0, :)$, which requires $\log_2|\delta(0, :)|$ bits. Accordingly we define $C(x_n) = \log_2|\delta(0, :)|, \forall n$. From the AO structure we find the points where the compression occurs according to the method mentioned above. For every frame that occurs between the encoding events we compute the compression gain in terms of the number of bits required to represent the pair (*length, position*), divided by the length of the encoded segment.

Algorithm 1 Information Rate using AO

Require: Array K containing a list of AO encoding event occurrences, unconditional complexity $C = \log_2(|\delta(0, :)|)$, $M = \max(LRS)$, and sequence length N

- 1: **for** $i = 1$ to $|K| - 1$ **do**
- 2: $L = K(i + 1) - K(i)$
- 3: $IR[K(i) : K(i+1)] = \max(C - \frac{\log_2(N) + \log_2(M)}{L}, 0)$
- 4: **end for**
- 5: **return** Array IR

Compositional Design for Machine Performance

The main goal of the current presentation is to show how AO can be used as a generator (high level instrument) in a

new composition that is designed off-line. For on-line, real-time interaction approach, two main systems have been built : Omax by Assayag, Bloch and Chemillier, an OpenMusic/Max application for co-improvisation with human performers using the Midi, audio and video media ; and MiMi by A. Francois and E. Chew, a multimodal interface for co-improvisation with the computer that gives a visual feedback to the performer, helping him to anticipate the computers reactions. Omax operation is usually done as a trio - human musician-improviser, machine improviser and human operator of the machine. The role of the humans in this setting is quite demanding - the musician creates initial musical materials and then listens and responds to machine improvisation. The human operator listens to the human musician and controls the machine using his own taste and musical skills.

In this demo we separate between the steps of providing musical input to the machine and control of machine improvisation. The recording, analysis and design of the piece will be done off-line. The remaining live element in the demo will be running the oracle according to compositional plan. The process of creating the composition consists of the following steps:

- Using pre-recorded music materials to perform the AO analysis.
- Performing IR analysis of the AO graph structure.
- Manual design of composition based on AO visualization and IR annotations.
- Live rendering of new improvisation according on the compositional design.

To demonstrate the method we use IR analysis of an AO created from a recording of Beethoven Piano Sonata No. 1 in F minor. Comparison to musicological analysis shows that IR successfully captures salient structural aspects of musical form. An additional graph (not shown here) points to the locations in the recording where the repetitions actually occur. A possible improvisation scheme based on this information is as follows:

- Select several short segments from the exposition section as the main operating regions.
- Create an exposition by switching between several generative materials.
- Create a development section by improvising on the exposition starting with low recombination rate and gradually increasing the recombination rate and short repetition length.
- For Cadence uses exposition materials with high minimal repetition length.

It should be noted that during the performance the improvisation can be recorded back into an oracle. This allows reuse of new and surprising materials in a later development. In the demo similar types of analysis will be conducted on various musical materials, with various features and time scales. In terms of the "aesthetic" interpretation, the higher the graph, the more order exists in that segment since it can be "explained" and hopefully perceived by pointing to an earlier repetition.

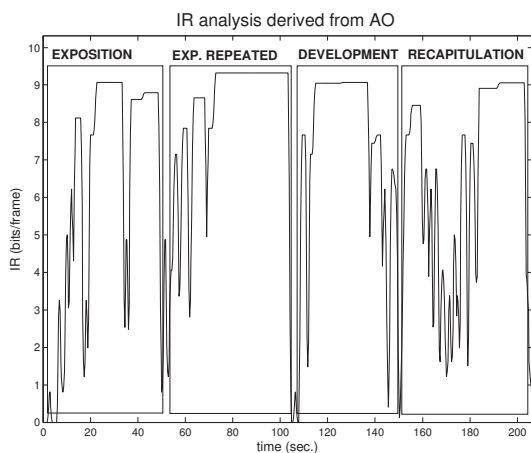


Figure 1: AO IR analysis for Beethoven Sonata No. 1

To get an insight into the segmentation potential of the AO-IR graph we compare it to the first eigenvector of a row-wise normalized pairwise distance matrix (so called Self Similarity matrix) that can be interpreted as a transition probability matrix $P_{i,j}$ between frames i and j . The eigenvector $v = Pv$ gives the stationary distribution under this Markov model, and as we can see, the recurrent segments under AO-IR correspond to frames with higher probability of occurrence under the more simple Markov model.

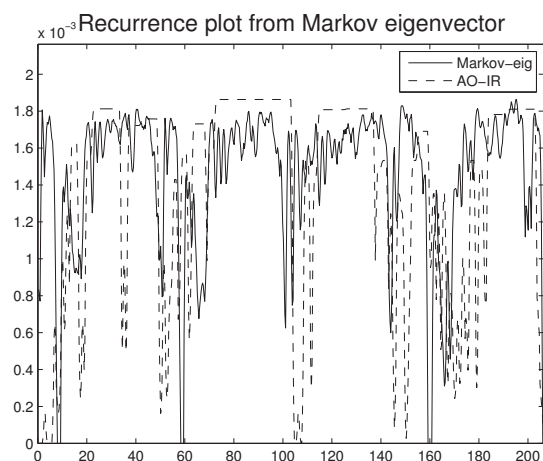


Figure 2: Recurrence profile using the first eigenvector of a Markov Model derived from Cepstral Self Similarity. This graph is compared to the AO based IR measure (dashed)

In (Dubnov, Assayag and Cont 2011) we showed how the sum of IR values over the whole signal can be used to find the optimal threshold for the AO estimate. Generalizing this idea, we consider IRs of different features from AO's constructed from different musical parameters. In our practice, the most common features used for improvisation are pitch

and spectral features, providing two oracles that give different structures for improvisation. Accordingly, IR can be used not only for selection of materials within an oracle, but also for switching between "most informative" or "most aesthetic" oracles at different points in time.

Conclusion

Generating novel music from a recording can be done using AO. The low level repetition structure found by AO can be used effectively for machine improvisation by creating stylistic re-injections into music performed by a human improviser. The re-injections, although stylistically coherent, do not take into account larger musical structures. Using a "compression oracle" we are able to analyze the AO structure in terms of information dynamic properties. We use Information Rate as a meta-feature for improvisation planning in terms of higher level aesthetic categories. In the demo we will present a larger scale planning process for a human operator of a machine improvisation system. Aspects of notating the actions of a human operator and aesthetic criteria for designing such composition will be discussed.

References

- Dubnov, S., Assayag, G. and Cont, A. 2007. Audio Oracle: A New Algorithm for Fast Learning of Audio Structures, In Proceedings of International Computer Music Conference (ICMC).
- Allauzen, C., Crochemore, M., and Raffinot, M. 1999. Factor oracle: A new structure for pattern matching, In Conference on Current Trends in Theory and Practice of Informatics: 295 - 310.
- Assayag, G., Bloch, G., Cont, A. and Dubnov, S. 2010. Interaction with Style, In The Structure of Style: Algorithmic Approaches to Understanding Manner and Meaning, S. Argamon, K. Burns, S. Dubnov, (Eds.), Springer
- Dubnov, S., Assayag, G., and Cont, A. 2011. Audio Oracle analysis of Musical Information Rate, In The Fifth IEEE International Conference on Semantic Computing, Palo Alto.
- Dubnov, S., 2006. Spectral anticipations, *Computer Music Journal*, 30(2):63 - 83.
- Potter, K., Wiggins, G.A., and Pearce, M.T. 2007. Towards greater objectivity in music theory: Information-dynamic analysis of minimalist music, *Musicae Scientiae* 11(2): 295 - 322,
- Abdallah, S. A. and Plumbley, M. D. 2009. Information dynamics: patterns of expectation and surprise in the perception of music, *Connection Science*, 21(2): 89 - 117.
- Meyer, L. B. 1956. *Emotion and Meaning in Music*, Chicago: Chicago University Press.
- Dubnov, S., McAdams, S. and Reynolds, R. 2006. Structural and Affective Aspects of Music From Statistical Audio Signal Analysis, *Journal of the American Society for Information Science and Technology*, 57(11): 1526 - 1536.
- Lefebvre, A., and Lecroq, T. 2001. Compror: Compression with a Factor Oracle, In Proceedings of the Data Compression Conference.

Birkhoff, G.D. 1933, *Aesthetic Measure*, Harvard Univ. Press.

Bense, M. 1969. Introduction to the Information-theoretical Aesthetics, *Foundation and Application to the Text Theory*, Rowohlt Taschenbuch Verlag, (In German).