Lossless Audio Compression as a Metric for Music Complexity

Pierre Labendzki

Université Gustave Eiffel
pierre.labendz@gmail.com

Abstract

Using lossless audio compression algorithmic to measure information content of musical messages. This approach base on Minimum Length Description offers a quantitative measurement for musical complexity that can analyses micro, meso and macro-scale redundancy in music. Visualisation of complexity over time of a given musical message can be use as a tool for musicology study. Cognitive constructivism presents a good explanation for this analysis.

Keywords: music complexity, lossless compression, Kolmogorov complexity, audio perception

1. Introduction

Trying to measure music complexity may seem foolish due to music highly subjective aspect, but we will try to show that the compressibility of a music piece can give hint to it's perceive complexity. One of the major goal of information theory is to quantify information content in a given message. The medium of the message can differ from acoustic information to digital representation on a hardrive or electric impulses in the brain. The information content, whether in verbal communication, in ambient noise, or in music can be analyse using lossless compression (LLC) measure, special attention will be given to music.

By using lossless audio compression made for music on different musical stimulus both artificial and naturalistic we aim to demonstrate the wide range of possible observation. We will first address the motivation for such measurement in the musical context. Previously made previsions on a *hedonic value* of music relative to *complexity* [8] and *Value* relative to *originalité* [9] Then we will present previous metric trial from different works on audio signal complexity, coupled musicology approach to complexity evaluation. We will explain the LLC method and it's relation to the constructivist model of understanding and representation of the environment. At last will will present the data from the LLC analysis of popular music from different genres

2. Motivation

One of the first publication on the mathematics of aesthetic can be trace back to Birkhoff aesthetic measure [3] of 1933. According to Birkhoff, experiencing a musical piece or any work of art is set in three consecutive stages

> "… (1) a preliminary effort of attention, which is necessary for the act of perception, and which increases in proportion to what we shall call the complexity (C) of the object; (2) the feeling of value or aesthetic measure (M) which rewards this effort; and finally (3) a realisation that the object is characterised by a certain harmony, symmetry, or order (O), more or less concealed, which seems necessary to aesthetic effect."[3]

Birkhoff metric for aesthetic is $M = \frac{O}{C}$ a function of the ratio between Order and Complexity. Considering music as a message, being subject to message analysis such as content retrieval, intelligibility and appreciation is not new idea. however information theory and it's statistical tool, paired with the increase in processing power can given us new ways to read a communication. The fist address of information content date back to telegraph usage, but since Shannon seminal publication in 1948 information oriented research boomed. But these publications centered more on a symbolic aspect of communication, putting aside the potential aesthetic value of analog message such as music. Abraham Moles formulated an hypothesis linking information content "originalité" and a "value" (Figure 1)

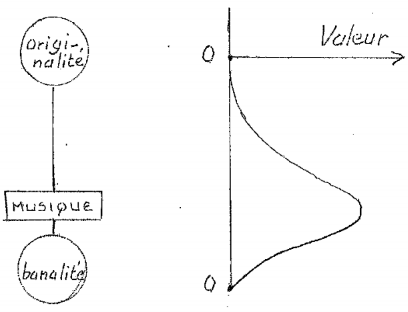

Figure 1: *Value as a function of originality, drawing from Moles [9]*

Music is suppose to be between complexity and simplicity, a subtle balance between trivial and incomprehensible [9]. Even though Moles speaks of redundancy and bits, computer science is not yet able in 1964 to process audio signals and it's lack the metric to evaluate music complexity.
Other model like the Wundt curve, predict a similar relation between the stimulus complexity and hedonic value [8].
Many critics can be made toward an absolute and objective measure of aesthetic value or hedonic value. The goal of our metric is not to asses such a hypothetical value, but simply to measure what previous aesthetic theory called complexity or originality.

3. Previous works

A popular metric for complexity analysis of a audio message is the "1/f power law". This power law measurement is based on the observation that there exist a preference of neural coding for signal that show a 1/f relation in there spectrum analysis [7]. Conclusions concerning the music predictability can be made by computing the slope of the 1/f power law, the steeper the slope, the more predictable the music. However such metric need a long audio segment to perform accurately the slope

analysis and thus cannot be used to discuss complexity as a function of time within a musical piece. Furthermore even thought this metric is reliable for repetition and periodic event like rhyme, it lack vision on other predictable patterns and one direction musical phrases like scales.
Aesthetic measure based on information theory and especially Kolmogorov complexity is already establish for paintings analysis [10]. Using image algorithm compression to compute a natural compressibility factor, the paintings of Mondrian can be said to be less complex than the ones from van Gogh.
Hudson suggest that compression algorithms can also be applied for music complexity analysis but the pilot study only address global complexity of a whole musical piece[6]. What we want from our metric is to be sensitive to change in complexity over time.

4. Lossless audio compression

4.1. Kolmogorov complexity

Kolmogorov complexity is the quantitative aspect of the principle of Minimum Length Description. Kolmogorov complexity $K(s)$ of a finite binary sequence s is the length of the shortest program that generate s [4]. Such program, that can describe exactly s in $K(s)$ symbols is the Minimum length description of s and can be think of as representation of the original message s but reduce to it's core without any redundancy. A simple example not involving musical message : the sequence "01010101010101010101" (20 symbols) can be express as "ten times 01" (11 symbols). Whereas the sequence "00011010000111001011" can hardly be compress because of it's randomness and unpredictability.
Among all binary sequences of length n, less than one sequence in 2^n has a Kolmogorov complexity inferior to $n - k$. For instance, for a sequence of length 100, less than one sequences among 1024 has Kolmogorov complexity inferior to $100 - 10 = 90$. Meaning that most of the random sequences cannot be shorten by a more efficient description, this result linked to the pigeon hole problem is also the reason why an universal lossless compression algorithm is not computable. However all the music than we ear is highly compressible (cf Results). Kolmogorov complexity is not computable, no program that can take as input any sequence s and give as output $K(s)$

4.2. Basics principles for audio lossless compression algorithms

Even thought Kolmogorov complexity is not computable, one can approach it by using lossless compression (LLC) algorithms [4]. Unlike lossy compression algorithms, like mp3 or jpeg, which compress the original audio file to a fix bit rate by removing information not perceived or not relevant to our sens, lossless compression conserve every information. LLC takes advantage of the natural compressibility of human communication, such as repetition, regularity, predictable patterns, symmetry , and compress it (encode) to a shorter description that will be decompressed (decode) to be read.
Different LLC algorithms usually follow similar steps in order to compress an audio file. In this paper, Flac (Free Audio Lossless Codec) will be use for it's general compression power and good documentation, but other lossless format like Monkey's Audio, WavPack or Alac can be used.
For our purpose, the flac compression algorithm process can be break down in these 5 consecutive stages :

1. Break the file in chunks between 2000 and 6000 samples, default mode is 4096 samples.
2. Compute a model based of previous chunks that can best fit the data by using simple expressions like polynomials or general linear predictive coding.
3. Evaluate the difference between the prediction and the real data, this difference is the residual part which is usually small for accurate predictions and thus takes less bits to encode.
4. Efficiently encode the predictions and the residual by using entropic coding based algorithms, meaning assigning a short code for events that occur more often.
5. At the end, each chunk is the sum of the modeled signal and the residual.

The size of the compressed file is a good approximation of Kolmogorov complexity, but depends on the size of the original file. To correct bias on complexity due to the file original characteristics : duration, sampling frequency, bit depth, the compressed size is divided by the original size.

$$Complexity\ (audio\,file) = \frac{size\ (compressed\ audio\ file)}{size\ (original\ audio\ file)}$$

For a given LLC algorithm, the size of the analyse window affect the measured complexity :

- With a too short window, the predicted signal is accurate and the compression effective but the header size relative to the total size is to important and result in an overestimation of the complexity. However the size allows for a detail analysis of the content complexity.
- A too large window will result in poorer prediction of the signal within but with a smoother estimation of complexity over time.

For a given LLC algorithm and given size, the natural compressibility depends on the complexity of the audio file content. A very simple music piece will have a high natural compressibility, meaning that Complexity(simple music) ≈ 0, and a very complex one, like white noise, will not be compressible, meaning Complexity(complex music) ≈ 1.

4.3. Sliding window analysis

This measurement of complexity over time represent the short term memory and echoic perception. Indeed the size of the sliding window can be set for 1 up to 10 seconds and even longer time (with a decrease in complexity precision at a precise time).

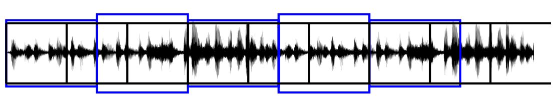

Figure 2: *scheme of the sliding window process*

4.4. Cumulative window analysis

The cumulative window analysis is window that expend in time but still uses data from the beginning of the track. The only parameter is the size of the incremented length. Unlike the sliding window, this method can given hints to the building of complexity over time.

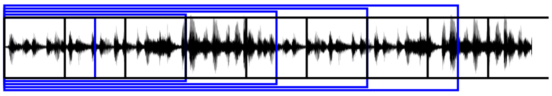

Figure 3: *scheme of the cumulative window process*

4.5. Comparative results for synthetic audio

The sliding and cumulative windows differ in usage, scope and their results have to be analyse in different fashion. Both techniques have advantages and drawbacks, and for an in depth analysis of complexity, they need to be read jointly. As discuss earlier the sliding window reassemble our short term memory with a range of 3 to 6 seconds, on the other hand, the cumulative window mirror our understanding of a musical piece from start to finish and with a sensibility to the musical events order. In this sens, the sliding technique is more on an objective measurement at a precise time whereas the cumulative one reassemble more the experience on complexity that one can have when listening to music.

Audio description	Time (seconds)	Complexity
silence	5	0.049
silence	60	0.004
440Hz sin wave	5	0.199
440Hz sin wave	60	0.146
white noise	5	0.961
white noise	60	0.914

Table 1: *Complexity of silence, 440Hz sin wave and white noise for 5 and 60 seconds. Complexity varies from* $\sim$ 0 *(simple, predictable, compressible) to* $\sim 1(complex, unpredictable, incompressible)$.

4.6. Compression and constructivism

An analogy can be made between LLc algorithms and the constructivsm model of cognition ignited by Piaget. These two ways of processing data share similarities :

- An input window with a limited span of attention in time. The buffer in the case of the LLC algorithms and the echoic memory for the brain.
- Processing of the incoming data in respect to the previously acquired data already store in a model. It's the search for already existing symbols in the dynamic dictionary of the LLC algorithm or prediction made using previous chunks. This is similar to the assimilation phenomenon of environmental stimuli that are confronted to the mental representation of reality, either accepted as part of the model or rejected.
- The update of the existing model base on the new data. For algorithms it's the adjustments of the dictionaries to make sure that the more frequent inputs are encoded in the shortest code, or the tweaking of the parameter for the prediction based LLC algorithms. In the cognitive level, this stage is called accommodation, meaning changing the mental representation to best fit the perceive data in order to make better future predictions.
- At the end of the process of each new chunk of data, the system is at a stable state, the incoming symbols are represented using general law that best fit them. These data are incorporated to the model, making it more complete and more suitable to describe future data or stimuli.

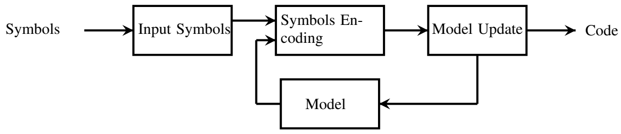

Figure 4: *Scheme of a adaptive general compression algorithm*

Music seen as a stream of acoustic information that can be process by a LLC algorithms or heard by a human is especially well suited for such comparison between compression and cognitive constructivsm. Indeed music listener usually make assumption concerning the unveiling of the music. Predictions on a micro-scale concerning the continuity of sound and note. Meso-scale predictions are made when a musical patterns like scales or phrases, melodic leitmotiv or even rhythms is presented, it take fewer assumptions and description to expect the completion of the pattern. A common feature in music is the repetition of large macro-scale events, like verses and chorus, it helps listeners to follow the piece or musicians to play along. These types of repetitions and predictability can be use by a LLC algorithm to reduce the overall redundancy of a files.

5. Result

On the basis that complexity takes several form in the music, like background noise from the recording, rich spectrum of instruments and effect or complex composition effect, this metric might be more relevant for a comparative musical analysis than for an absolute one. The absolute levels of complexity on figure 5 depend on the size of the window, but an overall evolution of complexity overtime emerges with the wider windows.

6. Conclusions

On of the failure of this kind of analysis is the lack of an inter-musical compression, a process that greatly simplify human experience of music. A cover a song will be easy to process because of the overall resemblance with the original. This can also be said of a 12 bar blues pattern that structures different songs and created and inter-musical redandancy. Kolmojorov complexity and LLC can even

7. Acknowledgements

Note to authors: Authors should not use logos in the acknowledgement section; rather authors should acknowledge corporations by naming them only.

8. References

[1] Bagot, J.D . *Information, sensation et perception*, Masson et Armand Colin Editeurs, Cursus, Paris, 1996.

[2] Berlyne, D.E. . "Novelty, complexity, and hedonic value". Perception Psychophysics 8, 279–286. `https://doi.org/10.3758/BF03212593`, 1970.

[3] Birkhoff, G.D . Aesthetic Measure. Harvard University Press, 1933.

[4] Delahaye, J-P . *Complexité aléatoire et complexité organisée*, Quæ Édition, Versailles, 2009.

[5] Flac documentaion, `https://xiph.org/flac/index.html`.

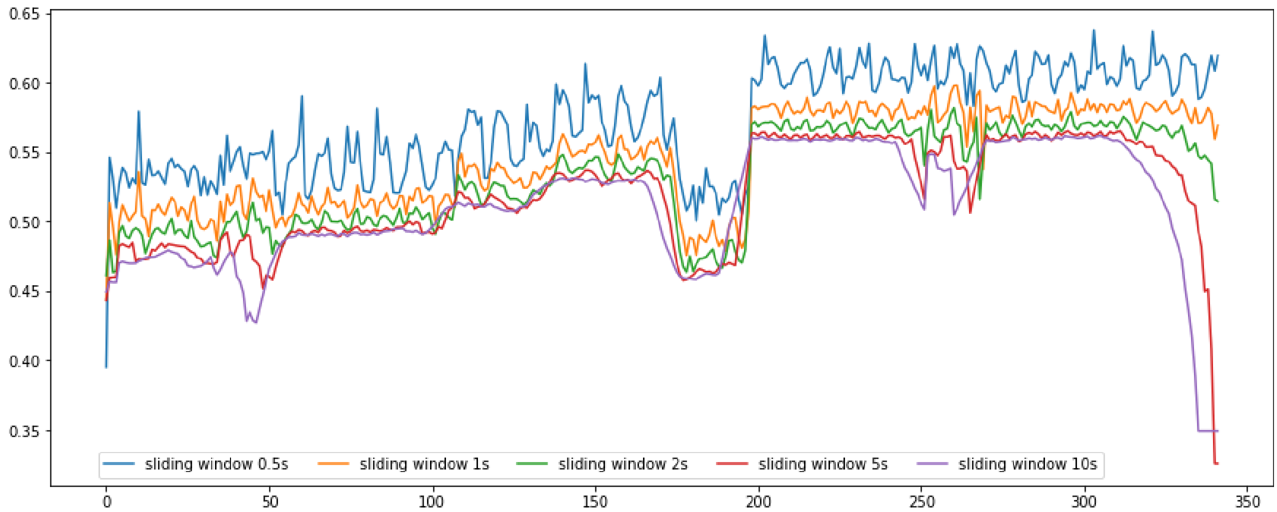

Figure 5: *Complexity over time, for the sliding window with various window sizes, the score is Veridisco by Daft Punk (2001). The wider the windows of 5 and 10 seconds smooth the complexity over time and efficiently convey different part of the song but cannot describe time fine structures. The piece consist of a repeating musical phrase with different electronic instruments. The pattern is first simply presented with minimal effect and then played over different beats and background chords.*

[6] Hudson, N.J . "Musical beauty and information compression: Complex to the ear but simple to the mind?". BMC Res Notes 4, 9. https://doi.org/10.1186/1756-0500-4-9, 2011.

[7] Levitina, D.J and Chordiab,P and Menonc, P . "Musical rhythm spectra from Bach to Joplin obey a 1/f power law", Proceedings of the National Academy of Sciences, 2012.

[8] Margulis, E.H . The Psychology of Music, A Very Short Introduction, Oxford University Press, 2019.

[9] Moles, A.A . "Théorie informationnelle de la musique", Bulletin n°5 du Groupe d'Acoustique Musicale, 1964.

[10] Rigau, J and Feixas, M and Sbert, M . Informational Aesthetics Measures. IEEE computer graphics and applications. 2008.

[11] Schmidhuber, J . "Driven by Compression Progress: A Simple Principle Explains Essential Aspects of Subjective Beauty, Novelty, Surprise, Interestingness, Attention, Curiosity, Creativity, Art, Science, Music, Jokes", ArXiv abs/0812.4360, 2008.

[12] Snyder, B . *Music and Memory: An Introduction*, MIT Press, 2001