Lossless Audio Compression as a Metric for Music Complexity

Pierre Labendzki

Université Gustave Eiffel

pierre.labendz@gmail.com

Abstract

Lossless audio compression algorithm can be used to measure information content of musical messages. This approach base on Minimum Length Description and Kolmogorov complexity offers a quantitative measurement, close to entropy, for musical complexity that can analyse micro, meso and macro-scale redundancy in music. Cognitive constructivism presents a good explanation for this analysis. Visualisation of complexity over time of a given musical message or of a musical corpus can be used as a tool for musicology studies or music information retrieval.

Keywords: music complexity, lossless compression, Kolmogorov complexity.

1. Introduction

Trying to measure music complexity may seem foolish due to music highly subjective aspect, but we will try to show that the natural compressibility of a music piece can give hint to it's perceived 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 a musical performance, once recorded, can be analysed using lossless compression (LLC) measure, special attention will be given to music.

By using LLC algorithms, design for music, on different musical stimuli 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 an "*hedonic value*" of music relative to "*complexity*" [9] and "*value*" relative to "*originalité*" [10]. Then we will present previous metrics trials from different works on audio signal complexity or on algorithmic complexity of art. We will explain the LLC method and it's relation to the constructivist model of understanding and mental representation of the environment. At last we will present the data from the LLC analysis of popular music from different genres, coupled with a musicology approach to complexity evaluation.

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 $M = \frac{0}{C}$, 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 increased processing power can given us new ways to read a communication. The first address of information content dates back to telegraph usage, but since Shannon seminal publication in 1948, information oriented research boomed. The word entropy was coined to describe the amount of information that can be convey by a source. 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 "originality" and a "value" (Figure 1)

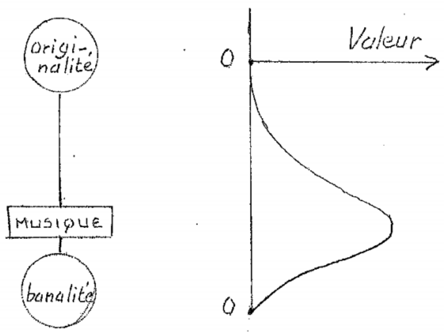

Figure 1: *Value as a function of originality, drawing from Moles [10]*

Music is suppose to be between complexity and simplicity, a subtle balance between trivial and incomprehensible [10]. 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 an "hedonic value" [9].

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 assess such a hypothetical value, but simply to measure what previous aesthetic theory called complexity, originality or high entropy.

3. Previous works

A popular metric for complexity analysis of an 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 [8]. 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 needs long audio segments to compute accurately the slope analysis, and thus cannot be used to discuss complexity over time within a musical piece. Furthermore even thought this metric is reliable for repetition and periodic events, like rhyme, it lack vision on other predictable patterns and uni-directional musical phrases, like scales.
For Edgar Morin "is complex what cannot be pin down to a law, what cannot be reduce to a simple idea" [11], this definition of complexity can be understood as a simplify version of Kolmogorov complexity (cf 4.1) and a good start to what we will call complexity. Aesthetic measure based on information theory and especially Kolmogorov complexity is already established for paintings analysis [12]. Using image LLC algorithm to compute a natural compressibility factor, the paintings of Mondrian can be said to be less complex than the ones from van Gogh. Other study uses similar LLC algorithms for music analysis in order to create a phylogenetic map of genre and subgenre down to the compositor [13]. By processing the LLC factor of a track based on data from a another track, an information distance can be computed. 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 changes 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. The Kolmogorov complexity $K(s)$ of a finite binary sequence s is the length of the shortest program that generates 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 could be 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 a 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 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 approximate it by using lossless compression (LLC) algorithms [4]. The more powerful the compressor, the closer to the Kolmogorov complexity and indirectly to the entropy of the message. LLC algorithms seem to converge on a compression limit :

> [...] existing algorithms come within a few percent of theoretical limits in terms of compression efficiency. No-one knows exactly the maximum compression which could be achieved on a given file, but the way the various algorithms all approach a particular percentage for a particular piece of music, and how those percentages vary widely according to the nature of the music, seems to indicate the best of the algorithms is getting close to some unknown theoretical limit. [17]

This convergence indicates that LLC complexity is a good approximation for Kolmogorov complexity. 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 used for it's general compression power and good documentation. But other lossless formats 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 this 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 analysis window affects 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. This results in an overestimation of the complexity. However the size allows for a detail analysis of the content complexity.
- A too large window will results in poorer prediction of the signal 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 represents 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).

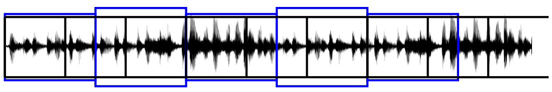

Figure 2: *scheme of the sliding window process*

4.4. Cumulative window analysis

The cumulative window analysis is a 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, relative to past events.

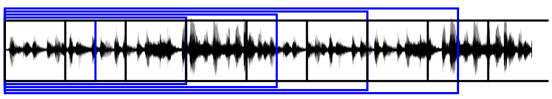

Figure 3: *scheme of the cumulative window process*

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, with a sensibility to the order of musical events. 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* ~ 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.

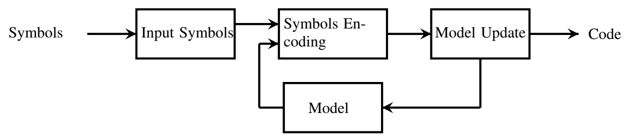

Figure 4: *Scheme of a adaptive general compression algorithm*

Music, when considered 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 LLC algorithms to reduce the overall redundancy of 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. It seems to appear like volume , or sound level is plotted overtime, but a comparative study of different levels and complexity reveals no significant relation. For instance instance figure 6 shows complexity and sound level calculation using energy level of a sequences x of n data point,

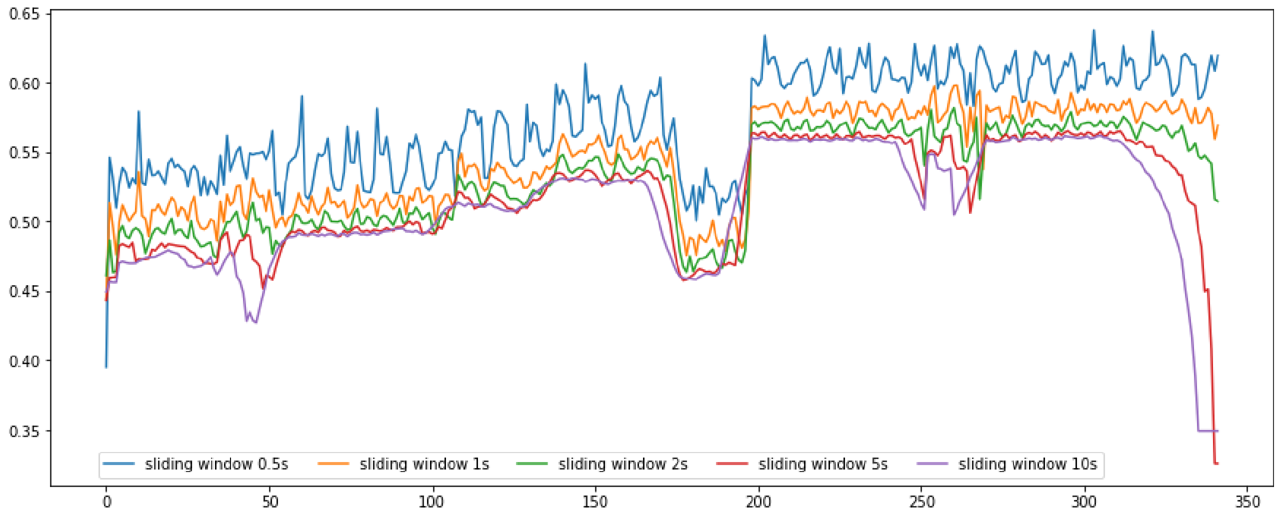

Figure 5: *Complexity over time, for the sliding window with various window sizes, the score is Veridisco by Daft Punk (2001). The wider the windows of 5 and 10 seconds smooth the complexity over time and efficiently convey different part of the song but cannot describe time fine structures. The piece 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.*

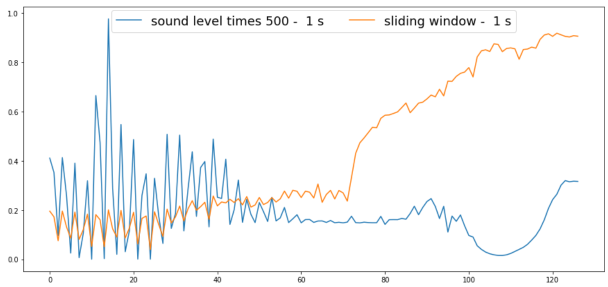

Figure 6: *energy levels (multiplied by 500 for visual purpose) and complexity over time. The difference between simple sound with a complexity below 0.3 and noise with increasing band width from 0.3 to 1.*

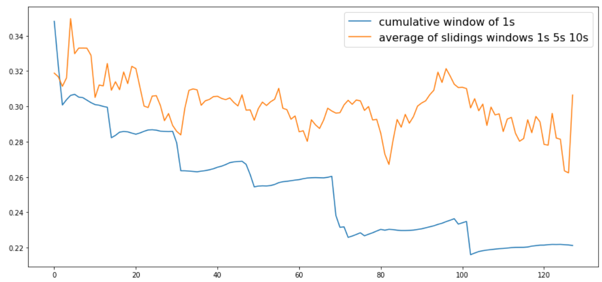

Figure 7: *Complexity over time, for the cumulative window with an incremental of 1 second and an averages of the sliding windows of 1, 5 and 10 seconds. The track is The Well Tempered Clavier by Bach (1744)*

with $Energy(x) = \sum_{i=1}^{n} x[i]^2/n$, for a music piece created for the demonstration. For this particular example, energy level and complexity are not correlated (Pearson correlation coefficient = -0.24). First pure tone sin wave in increasing level with sin wave envelope, then a speed up of these note and a randomization of the sequence of note and finally a white noise with increasing spectral width, from [2000 to 6000Hz] to [0 to 20000Hz]. The file end up with maximal entropy and complexity ≈ 1 but sound level is still low. All the code necessary to reproduce these results can be found on a GitHub repository [7].

5.1. Analysis with the sliding window

The absolute levels of complexity on figure 5 depend on the size of the window, but as windows sizes widen the measured complexity seems to converge toward the absolute level. An overall evolution of complexity over time emerges with the wider windows. The crescendo of complexity with a staircase shape from 0 to 170 seconds is clearly visible from the data point, especially the 5 and 10 seconds windows. Starting with a simple mono-instrumental presentation of the melodic line followed by a rhythms part from around 55 seconds in, the theme is backed with ambient chords from 60 to 110 seconds. Around 120 second, a melodic instrument is brought to the mix, followed by increasing effects. Between 175 and 200 seconds, a drop in complexity occurs in the track, resulting in complexity level matching the beginning of the song. At 200 seconds a "musical drop" happens and the instrumentation is even richer than before.

5.2. Analysis with the cumulative window

Cumulative window results are often less dynamic over time, apart from very informative step down. Toward the end of the file, any changes in the immediate complexity of the file does not influence much the total complexity due to the increasing "inertia" of the file. This is why apart for the first 60 seconds and apart from down stairs drop, the variation usually does not changes much. But even a stationary cumulative complexity can mean an increase in instantaneous complexity of the data that still reassembles prior data. This is due to the fact that if new unknown data are "presented" to the LLC algorithm, an upward step will occur due to the necessity to encode new information. Down steps, as seen in figure 7, are too steep to be cause by a natural decrease, meaning a stationary instantaneous complexity but a increasing file size. These down steps are due to sudden compression cause by new efficient way to compress data. For instance the completion of an already processes part, "123456" can be store as "A" (1 code for 6 input symbols) but when the track goes "A12345" the compression is not that efficient. But finally when the input string becomes "A123456" the compressor can output "2*A".

5.3. Complexity over genre

This stereotypical categorization of "classical" versus "rock" is meant to simplify and demonstrate the difference of complexity. In a symphonic or solo piano tracks versus rock music featuring distorted guitars and voices.

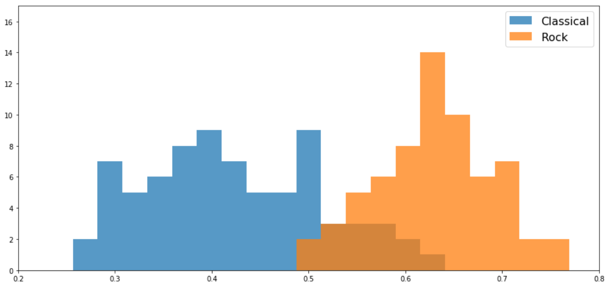

Figure 8: *histogram of the complexity for "classical" and "rock". The "classical" corpus contains : 32 tracks from Bach' Goldberg Variations and 43 tracks from Tchaikovsky' Swan Lake. The "Rock" corpus includes : 13 tracks from The Jimi Hendrix Experience, Axis bold as love, 26 tracks from Led Zeppelin' Remasters, 8 tracks from Oasis' What's the Story, 10 tracks from The Doors' Best Of. All wav files were ripped from CD at 44100kHz, 16 bits depth.*

A more detail representation can even highlight the compression similarities within composer and artiste. This analysis could be useful to measure listening habits on complexity and present complex music accordingly.

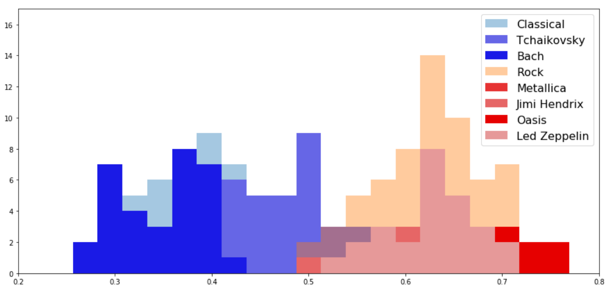

Figure 9: *histogram of the complexity for "classical" and "rock", with detailed label for each composer.*

The two genres clearly belong to two distribution of complexity, with Rock (mean = 0.62 ,sd = 0.06) being more complex than Classical (mean = 0.42 ,sd = 0.09). Some tracks (Spanish Dance from Swan Lake, complexity = 0.63) from the "classical" corpus are more complex than ones from the "rock" one (Spanish Caravan, The Doors, complexity = 0.49). The results shown in figure 7 polarizes musical complexity into simple and complex. However the complexity scale is from 0 to 1 and musical complexity is often around 0.5, suggesting like for language [15] that communication tool are a balance between information (high complexity) and redundancy (low complexity).

6. Discussion and Conclusion

A more comprehensive explanation of LLC compression measure could involve simple synthetic audio. In order to identify even more accurately what kind of information, temporal or spectral, are measure by the algorithm. Improvements could also be made by applying a low-pass filter to the audio before submitting it to the LLC metric. Indeed high frequency above 15kHz, not perceivable by most people, may account for a non negligible part of the information measured.

A metric based on LLC algorithm and it's application such as analysis over time or over a corpus, could be used for different purposes. This kind of global analysis could be useful for music information retrieval tool to label complexity of musical piece, for recommendation or classification. Musicological studies could also gain insight on a particular music piece complexity. One 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 redundancy. Methods using conditional Kolmogorov complexity [13] could be an answer to inter-musical complexity.
Proof could be brought by syncing physiological data that indicates attention, such as frontal EEG recording or pupil dilatation, with the complexity over time. Human perception of musical complexity could also be assess by answering questions concerning different stimuli. Lossless audio compression based techniques for music analysis is yet to be explore further.

7. Acknowledgements

This article was discuss with Kevin Dahan from Université Gustave Eiffel and Antoine Chevreuil from Laboratoire Gaspard Monge, who provided insight and expertise that assisted the research.

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.

[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] Labendzki, P . MetricForComplexity, GitHub repository of the project, https://github.com/pierrelabendzki/MetricForComplexity, 2021.

[8] 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.

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

[10] Moles, A.A . ”Théorie informationnelle de la musique”, Bulletin n°5 du Groupe d’Acoustique Musicale, 1964.

[11] Morin, E . Introduction à la pensée complexe, Edition du Seuil, Paris, 2005.

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

[13] Rudi, C and Vitányi, P and Wolf, R . Algorithmic Clustering of Music Based on String Compression. Computer Music Journal, 2004.

[14] 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.

[15] Shannon, C.E . A Mathematical Theory of Communication, Bell System Technical Journal, 1948.

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

[17] Whittle, M . ”Why I am no longer pursuing this field”, `https://www.firstpr.com.au/audiocomp/lossless/`, 2003.