



HAL
open science

Designing Sound Representations For Musicology

Pierre Couprie

► **To cite this version:**

Pierre Couprie. Designing Sound Representations For Musicology. Sound and Music Computing Conference (SMC), Université de Saint-Etienne; INRIA; GRAME, 2022, Saint-Etienne, France. pp.563-569. hal-03683837

HAL Id: hal-03683837

<https://hal.science/hal-03683837>

Submitted on 9 Jun 2022

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.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

DESIGNING SOUND REPRESENTATIONS FOR MUSICOLOGY

Pierre Couprie

RASM-CHCSC, Paris-Saclay University
pierre.couprie@univ-evry.fr

ABSTRACT

This article describes the use of representations based on data extracted from the audio signal (sonogram, chromagram, audio descriptor) in musicology. A first part gives the historical context of these representations and discusses their transfer from the field of exact sciences to musicology. In the second part, we propose three new types of representation particularly effective in musical analysis: visualizations based on waveform and amplitude, a multilayered sonogram and chromagram, and an interface to assist the interpretation of self-similarity matrices for the analysis of musical forms and microstructures.

1. INTRODUCTION

There are many forms of representation of the sound signal: from the waveform to higher-level representations such as FFT visualization, audio descriptors or the chromagram. Some of them are commonly used by musicians or musicologists, others are used more rarely. Some representations are easily applicable in musicology, others are more adapted to musical acoustics and their interpretation is often difficult for a researcher in human sciences¹. For the last fifteen years, we have been working on the improvement of these visualization techniques and on the development of representations dedicated to musicology. The aim of this article is to present some representations and layouts particularly effective for musicological research and more specifically for musical analysis. These representations have been produced using the iAnalyse version 5 software².

In the first part of the present paper, we will present the context related to the use of these representations in a musicological framework. In the second part, we will introduce three types of representation allowing actual interaction between the knowledge handled by the musicologist, musical perception and data extracted from the audio signal.

2. CONTEXT

2.1 Signal Representations in Musicology

In musicology, representations of the audio signal and the information it contains are often used in the practice of

musical analysis. This practice, which is largely empirical [2], is based on the use of a visual support enabling the formalization of musical data in order to study them. However, a significant part of the musical repertoire has no visual support - such as a transcription or a score - or an incomplete support - as in the case of music mixing acoustics and technologies³, for example electroacoustics [3] or popular music. Creating this visual medium has therefore become indispensable. The study of complex sound materials, often used in experimental creation, presents problems of memorization. In the case of long works, musical analysis can be particularly fastidious and the use of sonograms or self-similarity matrices allows the analyst to overcome this problem. In the same way, a visual support greatly facilitates the discovery of a work by establishing a hierarchy for the listening process or by facilitating identification of significant events, moments of stability or singularities. The use of detailed representations also enables a more specific study of microevolutions of the material⁴ such as inflections and variations of the musical interpretation of a melodic line, differences in pitch, modifications of timbre by effects of granularity, variations between channels reflecting effects of space or phase shifting etc. Finally, one of the major difficulties for the musicologist working on audio recordings resides in the communication of the work. If the use of score excerpts is a good way to illustrate a demonstration, in the case of music without score or the analysis of non-notated musical characteristics, visualization has become an indispensable support.

2.2 Three Types of Audio Signal Data Used in Musicology

The objective is not to provide an exhaustive state of the art but to present the most important research activities in the field of musicology⁵.

2.2.1 Waveform and Amplitude

As early as the 1960s, Pierre Schaeffer used a representation of the evolution of the sound amplitude produced using a bathygraph [5]. This graph allowed him to observe the permanence of the intensity curve of certain sounds even after having removed the attack.

³ This trend has increased over the last 30 years with the development of studies on repertoires that were previously rarely studied, such as popular or traditional music.

⁴ For a thoroughly detailed use of signal visualization techniques for musicology, see Michèle Castellango's book [4].

⁵ There are in particular numerous research projects applied to musical acoustics, to the study of musical instruments or to creation. They are not presented in this paper, which focuses on musicological uses.

¹ Some books are even exclusively designed for musicologists with a solid background in exact sciences, for example the Handbook of Systematic Musicology [1].

² <http://ianalyse5.pierrecouprie.fr>.

In the representation of *Rosace 5*, an extract from *Vibrations composées* by François Bayle (1973), produced by Dominique Besson, the waveform is drawn in two superimposed solid curves [6]. Each curve represents the amplitude of a channel colored in red for the right channel and in green for the left channel. The superposition of the two amplitude curves facilitates the analysis of the correlations between audio channels. This representation of the sound amplitude associated with a transcription of the various sounds is similar to the transcription of Karlheinz Stockhausen's *Studie II* (1954) performed by the composer [7].

The waveform - or an equivalent representation derived from the RMS amplitude - is the representation usually used in audio software to visualize intensity and manipulate the sound. It is therefore the first representation of the signal available to the musicologist. It facilitates visualization of the formal elements related to amplitude. Thus, in her analysis of Christine Groult's *La condition captive* (2003), Ana Dall'Ara-Majek uses the waveform as a background for the representation of formal segmentation [8]. In this work, the waveform underlines the different parts through opposition of intensities or morphological variations.

Although the waveform may not provide information on the timbre parameter, it does highlight the balance of intensities, breaks and abrupt variations that are often rhythmic or even formal markers.

2.2.2 Sonogram

Emile Leipp is probably one of the first to adopt spectrum analysis for sound recognition [9]. He presents the sonogram as the ideal tool for sound analysis, pointing out that the acoustic image corresponds exactly to the mental image suggested by perception. His examples are most instructive and convincing, but is this technique applicable in a musicological context? In 1984, Robert Cogan proposed a method of musical analysis based on the study of spectral graphs [10]. From these graphs, the author deduced color registrations of timbre representing a schematized spectral distribution that allows an analysis of the musical form and also an identification of types of sound and their transformation. However, their representation was poor quality and the method of description did not really result in an accurate analysis of spectromorphologies, but rather in broad spectral block opposition. In spite of its limitations, this approach had the merit of raising the question of the use of the sonogram in musical analysis.

After Cogan, Martha Brecht [11] and Mara Helmuth [12] were among the first researchers to propose an analytical method adapted to electroacoustic music and based on a multidimensional representation including the sonogram. Brech uses a representation combining three elements: the sonogram, musical structures on several levels and the amplitude. In addition, the author often superimposes textual comments on the sonogram. Helmuth's device combines five elements stacked from bottom to top: a sonogram, pitches on a staff, a reading of musical phrases above the staff, a graph of amplitude and a line of textual comments. According to the author, this

layout provides a clear visualization of the elements essential to the analysis of the work.

In 2010, Michael Clarke proposed the use of an interactive stereophonic sonogram to visualize the frequency balance between channels [13]. This stereophonic sonogram offers an additional level of visualization of sound characteristics for musical analysis.

In 1990, Waters and Ungvary list the potential uses of spectral representations for music [14]: from composition to performance, including analysis and music teaching, the sonogram is presented as a representation especially suited to these musical practices.

However, although the waveform or the amplitude representation prove to be relatively reliable, two limitations must be underlined. The first limitation, as Curtis Roads points out [15] is that the sonogram is only an estimation of the spectrum analysis by the machine and not an image of what our perception analyses. The second concerns the adjustments and coloring of the sonogram which, far from being trivial, make it particularly suitable even when it concerns recordings that are difficult to visualize because of complex sounds or those close to noise, use of extreme amplitudes, continuity of the texture or very slow variations of the material, etc. These are all parameters of musical analysis that are barely visible when using the default settings of the sonogram.

2.2.3 Audio Descriptors

There are three types of descriptor [16]: low-level descriptors (spectral brightness, spectral dispersion, noise level in the signal, etc.), medium-level descriptors offering an estimation of some musical parameters (fundamental frequency, inharmonicity, beats, etc.) and high-level descriptors (recognition of harmony, tonality, emotion, musical genre, etc.). Only the first-level descriptors use robust algorithms. The results obtained with second and third level descriptors often include errors and are adjusted to specific repertoires, so their generalization is not possible in musicology.

As early as 1957, Charles Seeger used a device to extract the fundamental frequency and the intensity per range in order to facilitate the transcription of melodic lines [17]. It is only since the 1990s that audio descriptors have been developed to allow the classification of large digital sound databases. Although their use in musicology is still struggling to become common practice, there are many examples of their application. Thus, Kerstin Neubarth et al. list a group of musicological practices that benefit from the use of audio descriptors, from transcription to musical analysis, including the study of perception of works [18].

In 2009, Tae Hong Park et al. presented a MATLAB toolkit (EASY) for the analysis of electroacoustic music using audio descriptors [19]. For example, they propose the use of 3D timbregram representation based on the association of several descriptors and clustering techniques to analyze the timbre.

2.3 Technology Transfer to Musicology

Even if some technologies are now commonly used in musicology, the transfer of methods from exact sciences to human sciences poses serious challenges.

The first difficulty lies in understanding their methods and the way in which the computer algorithms that implement them function. The musicologist uses the sonogram in order to find answers to questions that emerge during analytical perception. They do not seek to demonstrate the obvious - what any expert listener is able to perceive - but to complete their analytical knowledge, even to question it. The sonogram will therefore generally be used to analyze complex textures, mixtures of timbre that are difficult to identify even for an expert, interfering noises that allow the recording to be located, etc. These will often be non-standard recordings that require specific FFT computation settings and coloring. In addition to understanding the calculation of this FFT, it must also be situated in the musical domain, not only in the domain of sound or acoustics. The transfer of these technologies must therefore be accompanied by a transfer of method. Using a sonogram in musicology does not only mean using technology coming from acoustics and computer science, but also means using a digital method of musical analysis.

The second difficulty is that musicologists are generally limited to using tools developed for other sciences that are basically inadequate when applied to their scientific methods. Extracting an audio descriptor and visualizing it in the form of a graph, as is possible with several software programs, is not enough to enable its scientific use in music. It is necessary to develop visualizations of the descriptor adapted to its musical interpretation. In other words, a simple graph is perfect for understanding the mathematical characteristics of the data, but does not give any clues regarding a possible musical interpretation - for example in terms of morphologies, formal transitions, musical variation, etc.

The third difficulty is related to the previous one. Many research projects conducted in exact sciences would be a precious help in musicology. Unfortunately, the production of software is often limited to a few Python scripts that are difficult to use for a musicologist. On the one hand, the researcher in social sciences and humanities is not trained to use them and, on the other hand, it requires a high level of constant scientific monitoring in fields for which the musicologist does not have the time.

The fourth difficulty concerns the error transmitted by the data. The sonogram is only an estimation of a spectral content that is often far from our musical perception. Similarly, the audio descriptors often generate many errors, even for low level descriptors - for example the spectral centroid values when the sound amplitude decreases. These errors are unavoidable and often seem incompatible with the scientific methods of the musicologist. What may seem an important step forward for a researcher in exact sciences - for example the extraction of data from a vast musical collection, even with a margin of error - is unacceptable for the

musicologist who cannot be satisfied with approximations when analyzing a corpus.

The application of these techniques is therefore not to be found in a simple transfer between the exact sciences and musicology but in the development of new musicological methods. This is the role of digital musicology that differs from computational musicology, which is often based on the transfer of methods and technologies. Digital musicology is interdisciplinary, which is not without its problems, since even if the subject of research, music, is also interdisciplinary, its study remains deeply rooted in certain disciplines of the humanities - generally history, analysis, aesthetics, philosophy, and anthropology.

For all of these reasons, we have developed the software iAnalyse which allows the musicologist to use complex technologies with representations adapted to their field of research⁶. We have already discussed the basic functions and possible utilizations in several publications [20]. In the third part, we will focus on some new forms of spectrum representation or descriptors that are particularly effective for music analysis.

3. EXAMPLES OF REPRESENTATION

These three types of representation are the result of several recent research projects in musical analysis and are integrated in the iAnalyse software.

3.1 Amplitude and Augmented Waveform

As we have discussed previously, the representation of the sound amplitude or the waveform is probably the most commonly used visualization in the musical domain. It facilitates identification of the temporal relationship between sound characteristics such as attacks, musical intensity, morphologies or occurrence of silences. In the iAnalyse software, we have experimented with little used or forgotten representations. Figure 1 shows three examples of waveform or sound amplitude visualization. The top representation displays the two channels and their deviation⁷ (red curve). The middle representation displays the two channels on different layers, it facilitates the visualization of their differences. Finally, the bottom representation provides an overview of three audio descriptors: the spectral centroid (y-axis) which is a good estimate of the brightness of the sound, the RMS amplitude (height) and the zero-crossing rate (colors) which measures the noise level (and therefore the inharmonicity) of the signal. This last type of representation is based on the brightness standard deviation (BStD) proposed by Mikhail Malt and Emmanuel Jourdan for the study of timbre [22].

⁶ Technologies implemented in iAnalyse 5 are derived from previous work on another software program, EAnalysis, developed in partnership with the MTT² of De Montfort University [21].

⁷ The deviation is computed from the RMS amplitude as for the other graphs in the figure.

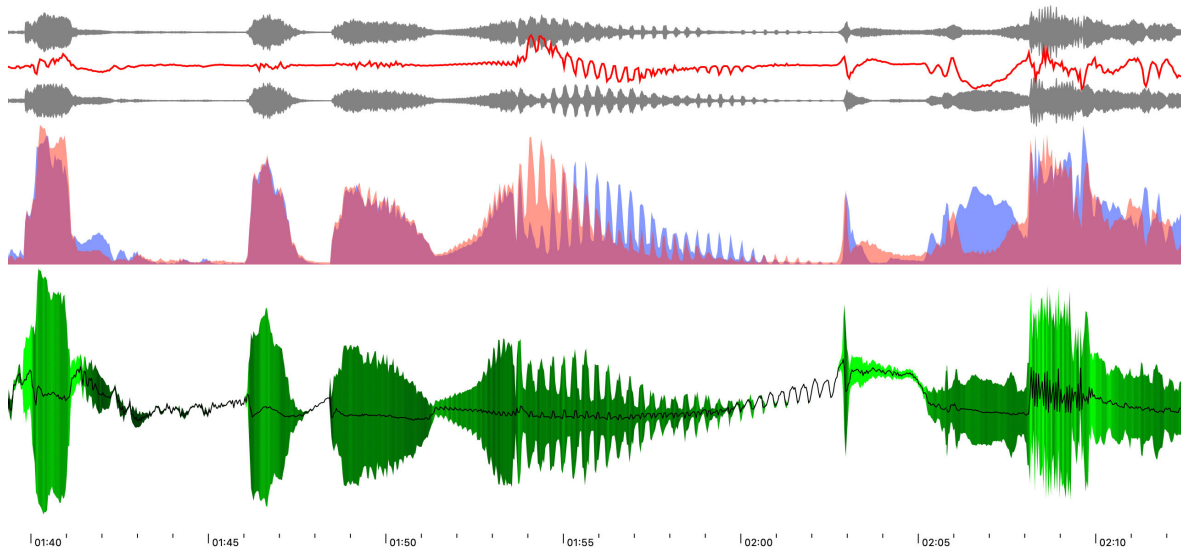


Figure 1. Three examples of amplitude visualization applied to an excerpt from *Ah Dulcinée* by Alain Savouret (from top to bottom): waveform (gray) and channel deviation (red curve), RMS amplitude on two layers, a combination of three audio descriptors (y: spectral centroid; height: RMS amplitude; colors: ZCR)

These three figures demonstrate the impact of the visual element on the interpretation of the data. For example, the central figure presents an immediate view of the balance between the channels and the lower figure illustrates the melodic profile. This immediacy is very useful for the musician or researcher who, at a single glance can identify moments on which they should focus particular attention.

3.2 Multilayered Sonogram and Chromagram

In 2019, the composer and musicologist Lin-Ni Liao initiated a research project⁸ on the sheng in contemporary music. This instrument, which has a strong presence on the Asian continent, has been the object of much enthusiasm from contemporary composers for over 20 years. During the summer of 2019, we carried out a free improvisation session with the musician Li Lin Chin and the composer and musician Benjamin Levy at Ircam using the improvisation software OMax. Our goal was to study the behavior of the software with the sheng improvisations. We recorded the audio in multitrack and then developed representations to visualize interactions between the musician and the OMax software. Two visualizations have proved to be particularly relevant: the multilayered sonogram (figure 2) and the multilayered chromagram (figure 3).

The first figure simply consists of a superposition of transparent logarithmic sonograms. Each sonogram is tinted with a color enabling a comparison of the musician's playing (in red) and the three iterations of OMax⁹ (here only two iterations are active in green and

blue). The processes used by the improvisation software are thus immediately recognizable: transposition, recording-playback with a delay, densification by duplication and their combinations.

The second figure (figure 3) uses the same principle of superposition from a chromagram¹⁰. In this case, as the sheng is principally a harmonic instrument¹¹, this type of representation proves to be most suitable for recording. In order to improve readability, we decided to draw the dynamic profile (amplitude RMS) directly on the pitches, which permits correlation of data between the different tracks (red, green, blue and orange). Here also, the interaction between the musician (in red) and the software (the other colors) is clearly visible whereas the spectral texture is somewhat complex: response of the musician on some morphologies, morphological differentiation of pitches, improvisation strategies between the musician and each iteration of the software, software fragmentation of sounds, etc.

These two representations therefore provide a tool for music analysis using visualizations, known to acousticians but adapted to musicology.

⁸ <http://www.tpmc-paris.com/sheng-research-2019-2020/>.

⁹ OMax is an automatic improvisation software program developed at Ircam. It has generated many variations such as SoMax or, more recently, DyCI2 (see: <https://www.ircam.fr/projects/pages/omax/>).

¹⁰ The chromagram represents the intensity of each note computed from the FFT and plotted for each octave.

¹¹ Some blowing effects and key noises were also used by the musician, but they are not relevant in the excerpt in Figure 3.

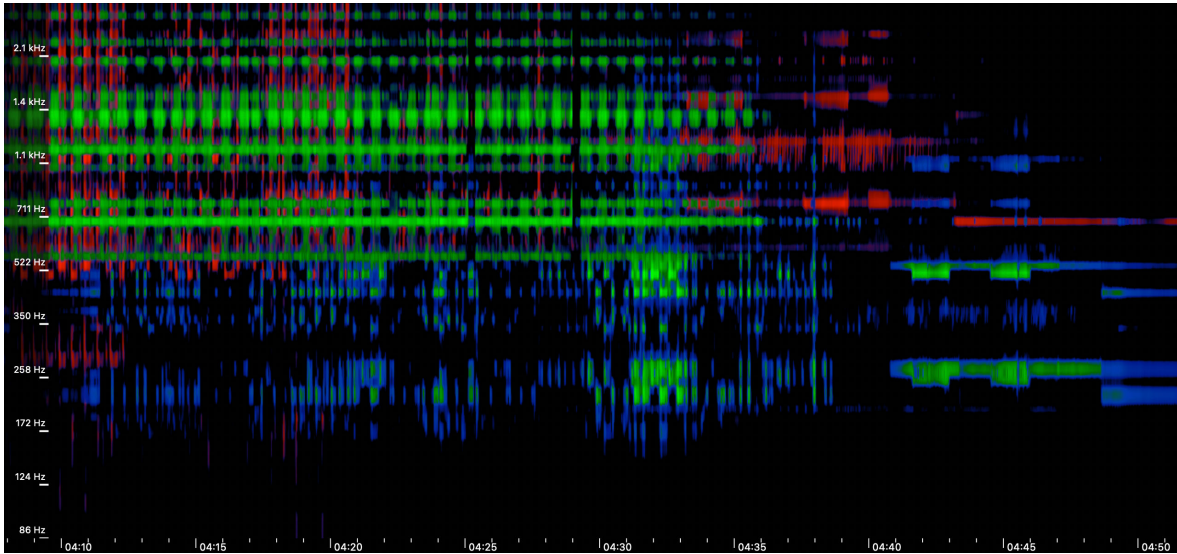


Figure 2. Multilayered sonogram to visualize correlations between recorded audio tracks.

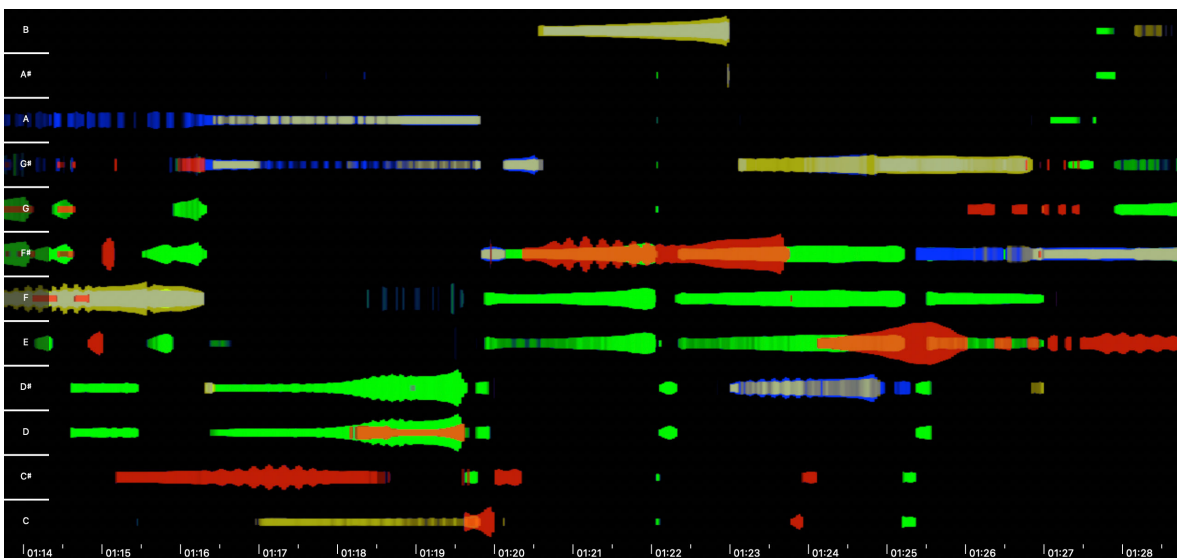


Figure 3. Multilayered chromagram to visualize correlations between recorded audio tracks.

3.3 Self-Similarity Matrix and Distance Estimation

Self-similarity matrices are computed from the FFT data or from a set of audio descriptors. The calculation is performed in three steps (figure 4). The first step (A) consists in creating a matrix (table) having the same values on the x-axis and y-axis and calculating distances between the values. The second step (B) consists in

mapping the numerical values against the color gradient values. Finally, the last step (C) consists in coloring the matrix in false colors in order to enhance interpretation. The abscissa and ordinate values being the same, the matrix is symmetrical in its diagonal and the two triangles are therefore symmetrically identical.

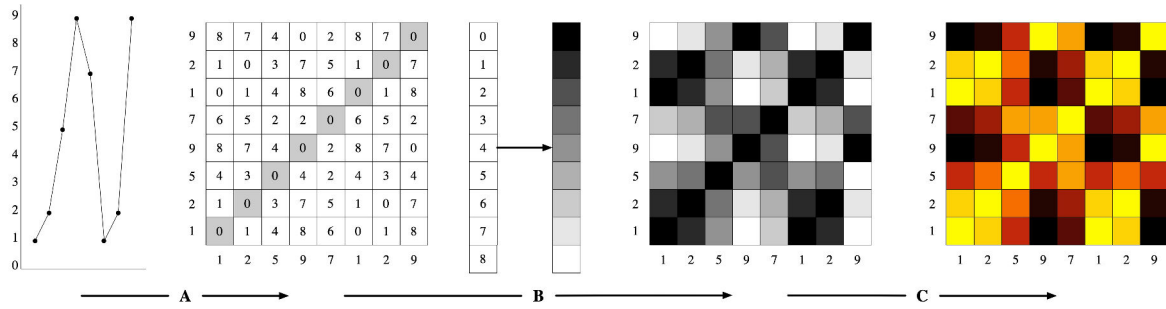


Figure 2. The steps involved in producing a self-similarity matrix.

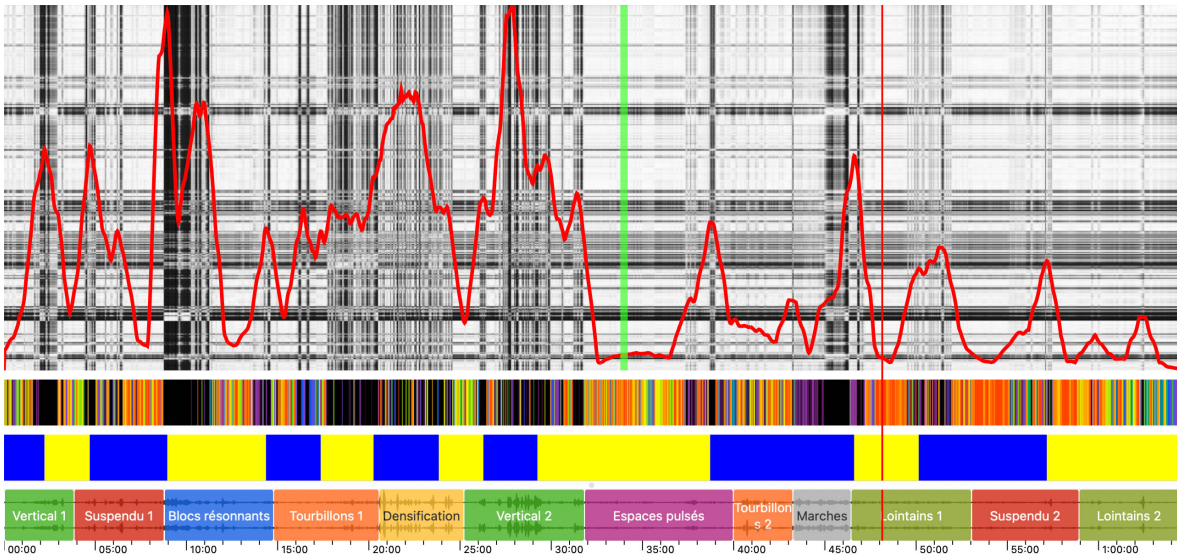


Figure 4. The interface for assisting the reading of a self-similarity matrix for the study of musical form in *Burning bright* by Hugues Dufourt.

Self-similarity matrices are particularly relevant in the analysis of musical form or microstructures¹² [23, 24, 25, 26, 27]. As true macroscopies¹³, they also facilitate the exploration or navigation of long audio files. However, interpretation is difficult for the musicologist, which is why we have developed an interpretation-assistance system. This consists of 3 representations. The first one (top of figure 4) includes a self-similarity matrix computed from the FFT or from audio descriptors as well as a curve representing the novelty parameter. From this parameter, a segmentation is extracted (yellow and blue graph) using the technique of Foote and Cooper [30] based on peak recognition. A playback head (green) can be moved to display the result of a distance computation just below it (warm colors represent small distances). Finally, a main playback head (red) is used to navigate through the audio playback. The work represented in this figure is *Burning bright* by Hugues Dufourt. This is a 64-

minute piece for percussion composed in one movement. Figure 4 also presents the segmentation (at the bottom) into tracks proposed in the recording. The aim is to facilitate the analysis of the musical form, which is difficult to grasp over such a long duration, but also to study concordances with the segmentation selected for the disc edition by the musicians and composer.

This interface is only in an experimental phase, but makes it possible to consider improvements for the analysis of long works or collections of works.

4. CONCLUSION

These three types of representation highlight what we underlined in the first part of the article. They are based on visualizations quite common in musical acoustics and their adaptation for the study of music makes them suitable tools for musicology. The development of these new methods is not based on a simple transfer but on the adoption of techniques specific to one scientific domain by another. It is in this sense that we have conceived our research into the study of 20th and 21st century music. This research obviously requires interdisciplinary skills, and these are indispensable to achieve as close a study as possible of a complex object such as music.

¹² The application of a similarity matrix for the comparison of two different versions of a work is described by Laura Zattra and Nicola Orto on *Stria* by John Chowning [26].

¹³ The term macroscopy, in reference to the macroscope [27], is used here to designate representations offering a global view of the information extracted from the signal while keeping a trace of the details.

5. REFERENCES

- [1] R. Bader, Springer Handbook of Systematic Musicology. Springer, 2018.
- [2] N. Cook, E. Clarke, “Introduction : What is Empirical Musicology?,” in Empirical musicology Aims, methods, prospects, Oxford University Press, 2004, pp. 3- 14.
- [3] P. Couprie, “Methods and Tools for Transcribing Electroacoustic Music,” in Proc. of the International Conference on Technologies for Music Notation and Representation, Montréal, 2018, http://tenor-conference.org/proceedings/2018/02_Couprie_tenor18.pdf.
- [4] M. Castellengo, Ecoute musicale et acoustique. Eyrolles, 2015.
- [5] P. Schaeffer, Traité des objets musicaux. Seuil, 1966.
- [6] D. Besson, Les musicographies. INA-GRM/38es Rugissants/Mois du graphisme d’Échirolles, 1995.
- [7] K. Stockhausen, No 3 Elektronische Studien Studie II. Universal Edition, 1997.
- [8] A. Dall’Ara-Majek, “Symbolism in Acousmatic Music through an Analysis of La condition captive by Christine Groult,” eOREMA 2, 2014, <http://www.orema.dmu.ac.uk/?q=content/symbolism-acousmatic-music-through-analysis-la-condition-captive-christine-groult>.
- [9] E. Leipp, Acoustique et musique. Mines ParisTech, 2010[1971].
- [10] R. Cogan, New Images of Musical Sound. Harvard University Press, 1984.
- [11] M. Brech, Analyse elektroakustischer Musik mit Hilfe von Sonagrammen. Peter Lang, 1994.
- [12] M. Helmuth, “Multidimensional Representation of Electroacoustic Music,” Journal of New Music research 25(1), 1996, pp. 77-103.
- [13] M. Clarke, “Wind Chimes: An Interactive Aural Analysis,” in Portrait polychrome Denis Smalley, INA-GRM, 2010, pp. 35-57.
- [14] S. Waters, T. Ungvary, “The Sonogram: A Tool for Visual Documentation of Musical Structure,” in Proc. of the International Computer Music Conference, Glasgow, 1990, pp. 159-162.
- [15] C. Roads, Musical Signal Processing. Routledge, 1997.
- [16] X. Serra et al., Roadmap for Music Information Research. 2013, <http://hdl.handle.net/10230/21766>.
- [17] C. Seeger, “Toward a Universal Music Sound-Writing for Musicology,” Journal of the International Folk Music Council 9, 1957, pp. 63-66.
- [18] K. Neubarth, M. Bergeron, D. Conklin, “Associations Between Musicology and Music Information Retrieval,” in Proc. of the International Society for Music Information Retrieval Conference, Miami, 2011, pp. 429-434.
- [19] T. H. Park, Z. Li, W. Wu, “EASY does it: The Electro-Acoustic music analysis toolbox,” in Proc. International Society for Music Information Retrieval Conference, 2009, pp. 693-698.
- [20] P. Couprie, “iAnalyse 5 : le développement d’un logiciel pour la musicologie numérique,” in Journées d’informatique musicale, Bayonne 2019, <https://jim2019.sciencesconf.org/resource/page/id/3>.
- [21] P. Couprie, “EAnalysis: Developing a Sound Based Music Analytical Tool,” in Expanding the Horizon of Electroacoustic Music Analysis, Cambridge University Press, 2016, pp. 170-194.
- [22] M. Malt, E. Jourdan, “Le BStD – une représentation graphique de la brillance et de l’écart type spectral, comme possible représentation de l’évolution du timbre sonore,” in L’analyse musicale aujourd’hui, Delatour, 2015, pp. 111- 128.
- [23] J. Foote, M. Cooper, “Visualizing Musical Structure and Rhythm via Self- Similarity,” in Proc. of the International Computer Music Conference, Havana, 2001, pp. 419-422.
- [24] J. Paulus, M. Müller, A. Klapuri, “Audio-Based Music Structure Analysis,” in Proc. of the International Society for Music Information Retrieval Conference, Utrecht, 2010, pp. 625-636.
- [25] G. Peeters, X. Rodet, “Signal-based Music Structure Discovery for Audio Summary Generation,” in Proc. International Computer Music Conference, Singapore, 2003.
- [26] M. Ramstrum, “Philippe Manoury’s Opera K...,” in Analytical Methods of Electroacoustic Music, Routledge, 2006, pp. 239-274.
- [27] P. Couprie, “Analytical Approaches to Taurhiphanie and Voyage absolu des Unari vers Andromède by Iannis Xenakis,” in From Xenakis’s UPIC to Graphic Notation Today, Hatje Cantz Verlag/ZKM, 2020, pp. 434-456.
- [28] L. Zattra, N. Orio, “ACAME – Analyse comparative automatique de la musique élec- troacoustique,” Musimédiane 4, 2006, <http://www.musimediane.com/numero4/LZattra/index.html>.
- [29] J. de Rosnay, The macroscope: A new world scientific system, Harper & Row, 1979.
- [30] J. T. Foote M. L. Cooper, M. L., “Media segmentation using self-similarity decomposition,” in Storage and Retrieval for Media Databases 5021, 2003, pp. 167-175.