Synopsis Seriation: A computer music piece made with time-frequency scattering and information geometry

Vincent Lostanlen, Florian Hecker

SYNOPSIS SERIATION: A COMPUTER MUSIC PIECE MADE WITH TIME–FREQUENCY SCATTERING AND INFORMATION GEOMETRY

Vincent Lostanlen
LS2N
CNRS
Nantes, France

Florian Hecker
Edinburgh College of Art
The University of Edinburgh
Edinburgh, UK

RÉSUMÉ

Cet article présente *Synopsis Seriation* (2021), une création musicale générée avec l'aide de l'ordinateur. L'idée centrale consiste à ré-organiser des fragments de pistes dans une œuvre multicanal pré-existante afin de produire un flux stéréo. Nous appelons "sériation" la recherche de la plus grande similarité de timbre entre fragments successifs dans chaque canal ainsi qu'entre canal gauche et canal droite. Or, puisque le nombre de permutations d'un ensemble est la factorielle de son cardinal, l'espace des séquences possibles est trop vaste pour être exploré directement par l'humain. Là contre, nous formalisons la sériation comme un problème d'optimisation NP-complet de type "voyageur de commerce" et présentons un algorithme évolutionniste qui en donne une solution approximée. Dans ce cadre, nous définissons la dissimilarité de timbre entre deux fragments à partir d'outils issus de l'analyse en ondelettes (diffusion temps-fréquence) ainsi que de la géométrie de l'information (divergence de Jensen–Shannon). Pour cette œuvre, nous avons exécuté l'algorithme de sériation sur un corpus de quatre œuvres de Florian Hecker, comprenant notamment *Formulation* (2015). La maison de disques Editions Mego, Vienne, a publié *Synopsis Seriation* en format CD, assorti d'un livret d'infographies sur la diffusion temps-fréquence conçu en partenariat avec le studio de design NORM, Zurich.

1. INTRODUCTION

In mathematics, the seriation problem seeks to arrange elements of a finite set $\mathcal{U}$ into a sequence $\boldsymbol{u}_1 \dots \boldsymbol{u}_N$ in such a way that distances $d(\boldsymbol{u}_i, \boldsymbol{u}_j)$ are small if and only if $|i-j|$ is also small [12]. Seriation bears a resemblance with the traveling salesperson problem (TSP), which aims to minimize the average distance $d(\boldsymbol{u}_i, \boldsymbol{u}_{i+1})$ between adjacent elements in the sequence.

Drawing inspiration from these mathematical ideas, the piece *Synopsis Seriation* (2021, see Figure 1) consists of a sequence of musical parts whose ordering in time reflects similarity in timbre. The set $\mathcal{U}$ corresponds to an unstructured collection of musical material: in our case, various pre-existing creations gathered under the name of *Seriation Input*. *Seriation Input* amounts to 283 minutes of audio in total, comprising hundreds of musical parts.

Figure 1. Album cover of *Synopsis Seriation*, released in March 2021 by Editions Mego, Vienna. The CD imprint represents the time–frequency scattering transform of the piece, which serves as a feature for the segmentation and structuration of the piece. Graphical design by NORM, Zurich. Website: https://editionsmego.com/release/EMEGO-256

The search space of all possible sequences is too vast to be explored manually. Indeed, the number of possible arrangements of $\mathcal{U}$ is equal to $N! = N \times (N-1) \times \dots 2$. This number is over one million for $N > 10$ and over one billion for $N > 13$. Coping with such a combinatorial explosion thus requires the help of the computer.

In this article, we describe the algorithmic workflow which has led to the synthesis of *Synopsis Seriation*. On a conceptual level, the worfklow involves a virtual agent which "listens" to *Synopsis Input*, segments it into temporal parts, and ultimately rearranges those parts to maximize the auditory similarity between adjacent parts.

One originality of our approach is that the virtual agent operates purely in the audio domain, without resorting to an external notation system such as MIDI or MusicXML. Furthermore, the agent does not assume that the input follows a traditional structure of repeated sections, such as verse-chorus or AABA forms. Lastly, the agent assigns parts of *Seriation Input* to either a stereophonic output by optimizing a joint objective of temporal consistency and binaural (left-right) consistency.

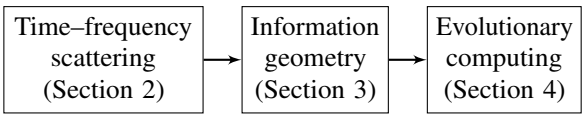


Figure 2. Flowchart of the computational stages involved in the synthesis of *Synopsis Seriation*. Time–frequency scattering is the acoustic frontend, information geometry performs sequential changepoint detection, and evolutionary computing solves a variant of the traveling salesperson problem (TSP). See paragraph below for details.

Our proposed procedure of seriation is akin to a family of digital audio effects known as concatenative synthesis [24]. Generally speaking, concatenative synthesis operates by assembling short audio segments which are taken from a large corpus so as to achieve a certain similarity objective. In this sense, our choice of audio descriptor (time–frequency scattering) and segmentation algorithm (generalized likelihood ratios) could potentially apply to real-time concatenative synthesis frameworks, such as CataRT [23]. However, we note that CataRT produces sounds according to a local target specification that is expressed in terms of sound descriptors or via an example sound. On the contrary, *Synopsis Seriation* does not rely on a predefined target; instead, it formulates a global problem of combinatorial optimization (the TSP) and arranges all segments of *Synopsis Input* accordingly. This formulation guarantees a one-to-one mapping between audio material in *Synopsis Input* and *Synopsis Seriation*.

The flowchart in Figure 2 summarizes the technical components of *Synopsis Seriation*. To begin with, Section 2 presents the acoustic frontend of the virtual listening agent: namely, time–frequency scattering. Time–frequency scattering is an operator whose architecture resembles spectrotemporal receptive fields (STRF) in auditory neurophysiology and convolutional neural networks (convnets) in deep learning. Section 3 presents the algorithm which segments the *Synopsis Input* audio stream into parts. This algorithm is a numerical application of information geometry, a field of research at the intersection between statistical modeling and differential geometry. Section 4 presents the algorithm which rearranges the segments parts and produces the *Synopsis Seriation* stereophonic piece. This algorithm is massively parallel and converges by evolutionary optimization. Section 5 presents the CD booklet of *Synopsis Seriation*, containing computer-generated visualizations of time–frequency scattering as well as creations of graphical design which summarize the functioning of the virtual listening agent. Lastly, Section 6 discusses the link between *Synopsis Seriation* and prior works on the spatiotemporal structuration of music, notably Iannis Xenakis's *Diatope*.

2. TIME–FREQUENCY SCATTERING

Time–frequency scattering comprises three stages. The first stage is a constant-Q transform (CQT) followed by pointwise complex modulus. The second stage is a convolutional operator in the time–frequency domain with wavelets in time and log-frequency, again followed by pointwise complex modulus. The third stage is a local averaging of every scattering coefficient over the time dimension.

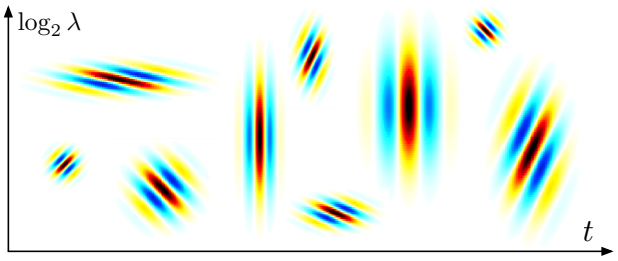


Figure 3. Interference pattern between wavelets $\boldsymbol{\psi}_\alpha(t)$ and $\boldsymbol{\psi}_\beta(\log_2 \lambda)$ in the time–frequency domain $(t, \log_2 \lambda)$ for different combinations of amplitude modulation rate α and frequency modulation scale β. Darker shades of red (resp. blue) indicate higher positive (resp. lower negative) values of the real part.

2.1. Constant-Q wavelet transform

We build a filter bank of Morlet wavelets of center frequency $\lambda > 0$ and quality factor $Q = 12$ via the equation

$$\boldsymbol{\psi}_\lambda(t) = \lambda \exp\left(-\frac{\lambda^2 t^2}{2Q^2}\right) \times (\exp(2\pi \mathrm{i} \lambda t) - \kappa), \quad (1)$$

where the corrective term κ guarantees that each $\boldsymbol{\psi}_\lambda$ has one vanishing moment, i.e., a null average. We discretize the center frequency variable as $\lambda = \xi 2^{-j/Q}$ where j is integer and ξ is a constant. In this way, there are exactly Q wavelets per octave in the filterbank. To make sure that the filter bank covers the Fourier domain unitarily, the center frequency of the first wavelet $(j = 0)$ should lie at the midpoint between the center frequency of the second wavelet $(j = 1)$ and the center frequency of the complex conjugate of the first wavelet, hence:

$$\xi = \frac{1}{2}\left(2^{-1/Q}\xi + (f_s - \xi)\right) = \frac{f_s}{3 - 2^{-1/Q}} \quad (2)$$

where f_s denotes the sampling frequency. The CD standard $f_s = 44.1\,\mathrm{kHz}$ yields $\xi = 21\,448\,\mathrm{Hz}$. We set the number of wavelets to $J = 96$, hence a range of $J/Q = 8$ octaves below ξ. The minimum frequency is $2^{-8}\xi = 84\,\mathrm{Hz}$.

Let the asterisk symbol $(*)$ denote the convolution product. Given a signal $\boldsymbol{x}(t)$ of finite energy, we define its CQT as the following time–frequency representation:

$$\begin{aligned} \mathbf{U}_1 \boldsymbol{x}(t, \lambda) &= |\boldsymbol{x} * \boldsymbol{\psi}_\lambda|\,(t) \\ &= \left| \int_{\mathbb{R}} \boldsymbol{x}(\tau) \boldsymbol{\psi}_\lambda(t - \tau)\, \mathrm{d}\tau \right|, \end{aligned} \quad (3)$$

indexed by time t and wavelet center frequency λ.

2.2. Spectrotemporal receptive field

For the second layer of the joint time–frequency scattering transform, we define two wavelet filterbanks: one over the time dimension and one over the log-frequency dimension. In both cases, we set the wavelet profile to Morlet (see Equation 1) and the quality factor to $Q = 1$. With a slight abuse of notation, we denote these wavelets by $\boldsymbol{\psi}_\alpha(t)$ and $\boldsymbol{\psi}_\beta(\log \lambda)$ even though they do not have the same shape as the wavelets $\boldsymbol{\psi}_\lambda(t)$ of the first layer, whose quality factor is equal to $Q = 12$.

Frequencies α, hereafter called amplitude modulation *rates*, are measured in Hertz (Hz) and discretized as $2^{-n\frac{2}{5}\xi}$ with integer n. Frequencies β, hereafter called frequency modulation *scales*, are measured in cycles per octave (c/o) and discretized as $\pm 2^{-n\frac{2}{5}Q^{-1}}$ with integer n. The edge case $\beta = 0$ corresponds to $\boldsymbol{\psi}_\beta(\log \lambda)$ being a Gaussian low-pass filter $\boldsymbol{\phi}_F(\log \lambda)$ of bandwidth F^{-1}.

For each rate–scale pair (α, β), we define the spectrotemporal receptive field (STRF) of $\boldsymbol{x}$ as the following time–frequency representation:

$$\begin{aligned}\mathbf{U}_2\boldsymbol{x}(t,\lambda,\alpha,\beta) &= \left|\mathbf{U}_1\boldsymbol{x} \overset{t}{*} \boldsymbol{\psi}_\alpha \overset{\log_2 \lambda}{*} \boldsymbol{\psi}_\beta\right|(t,\lambda)\\ &= \left|\iint \mathbf{U}_1\boldsymbol{x}(\tau,s)\boldsymbol{\psi}_\alpha(t-\tau)\boldsymbol{\psi}_\beta(\log_2\lambda - s)\,\mathrm{d}\tau\,\mathrm{d}s\right|,\end{aligned} \tag{4}$$

that is, stacked convolutions in time and log-frequency with all wavelets $\boldsymbol{\psi}_\alpha(t)$ and $\boldsymbol{\psi}_\beta(\log_2 \lambda)$ followed by complex modulus [1]. Thus, $\mathbf{U}_2\boldsymbol{x}$ is a four-way tensor.

Figure 3 shows the interference pattern of the product $\boldsymbol{\psi}_\alpha(t-\tau)\boldsymbol{\psi}_\beta(\log_2\lambda - s)$ for different combinations of time t, frequency λ, rate α, and scale β. We denote the multiindices (λ, α, β) resulting from such combinations as scattering *paths* [17].

2.3. Temporal averaging

Lastly, we define a Gaussian low-pass filter $\boldsymbol{\phi}_T$ of width equal to $T = 372\,\mathrm{ms}$, i.e., 2^{14} samples at a sampling frequency of $f_s = 44.1\,\mathrm{kHz}$. We apply this low-pass filter separately on each scattering path of the CQT of $\boldsymbol{x}$, yielding

$$\mathbf{S_1}\boldsymbol{x}(t,\lambda) = \left(\mathbf{U_1}\boldsymbol{x} \overset{t}{*} \boldsymbol{\phi}_T\right)(t,\lambda). \tag{5}$$

Likewise, we apply $\boldsymbol{\phi}_T$ separately on each second-order path of the spectrotemporal receptive field of $\boldsymbol{x}$, yielding

$$\mathbf{S_2}\boldsymbol{x}(t,\lambda,\alpha,\beta) = \left(\mathbf{U_2}\boldsymbol{x} \overset{t}{*} \boldsymbol{\phi}_T\right)(t,\lambda,\alpha,\beta). \tag{6}$$

3. INFORMATION GEOMETRY

3.1. Energy conservation

We restrict the set of modulation rates α in $\mathbf{U_2}\boldsymbol{x}$ to values above T^{-1}, so that the power spectra of the low-pass filter $\boldsymbol{\phi}_T(t)$ and all wavelets $\boldsymbol{\psi}_\alpha(t)$ cover unitarily the Fourier domain: at every frequency ω, we have

$$\left|\widehat{\boldsymbol{\phi}}_T(\omega)\right|^2 + \frac{1}{2}\sum_{\alpha > T^{-1}}\left(\left|\widehat{\boldsymbol{\psi}}_\alpha(\omega)\right|^2 + \left|\widehat{\boldsymbol{\psi}}_\alpha(-\omega)\right|^2\right) \lessapprox 1, \tag{7}$$

where the notation $A \lessapprox B$ indicates that there exists some $\varepsilon \ll B$ such that $B - \varepsilon < A < B$. Likewise, in the Fourier domain associated to $\log_2 \lambda$, one has $\sum_\beta |\widehat{\boldsymbol{\psi}}_\beta(\omega)|^2 \lessapprox 1$ for all ω. Therefore, applying Parseval's theorem on all three wavelet filterbanks (respectively indexed by λ, α, and β) yields the inequality:

$$\|\mathbf{S_1}\boldsymbol{x}\|_2^2 + \|\mathbf{U_2}\boldsymbol{x}\|_2^2 \lessapprox \|\mathbf{U_1}\boldsymbol{x}\|_2^2, \tag{8}$$

where the squared ℓ^2 norm of a scattering representation is the sum of squared ℓ^2 norms of its coefficients.

Furthermore, we neglect the DC component of $\boldsymbol{x}$, which is inaudible and thus can be calibrated to zero without affecting auditory perception. Therefore, because wavelets $\boldsymbol{\psi}_\lambda$ cover the audible range unitarily, the operator $\mathbf{U_1}\boldsymbol{x}$ preserves the energy in $\boldsymbol{x}$: $\|\mathbf{U_1}\boldsymbol{x}\|_2^2 \lessapprox \|\boldsymbol{x}\|_2^2$.

Lastly, we consider that, for T not too large (below 500 milliseconds), the averaging operation in Equation 6 involves a negligible loss of energy: $\|\mathbf{S_2}\boldsymbol{x}\|_2 \lessapprox \|\mathbf{U_2}\boldsymbol{x}\|_2$. This is a consequence of Waldspurger's theorem of exponential decay of scattering coefficients [26]. We conclude with a property of approximate energy conservation of joint time–frequency scattering:

$$\begin{aligned}\|\mathbf{S_1}\boldsymbol{x}\|_2^2 + \|\mathbf{S_2}\boldsymbol{x}\|_2^2 &\lessapprox \|\mathbf{S_1}\boldsymbol{x}\|_2^2 + \|\mathbf{U_2}\boldsymbol{x}\|_2^2\\ &\lessapprox \|\mathbf{U_1}\boldsymbol{x}\|_2^2\\ &\lessapprox \|\boldsymbol{x}\|_2^2.\end{aligned} \tag{9}$$

Let us now shift perspective from continuous time to discrete time. We denote by $\mathbf{S}\boldsymbol{x}[t,k]$ the concatenation of first- and second- order scattering coefficients, where the multiindex k encapsulates singletons (λ) in $\mathbf{S}_1\boldsymbol{x}(t,\lambda)$ and triplets (λ,α,β) in $\mathbf{S}_2\boldsymbol{x}(t,\lambda,\alpha,\beta)$, and ranges from 1 to K. The energy conservation property described in Equation 9 implies that, for every t, the sum of squared coefficients $\mathbf{S}\boldsymbol{x}[t,k]^2$ roughly correspond local energy $E[t]^2$ of $\boldsymbol{x}[t]$ within a Gaussian temporal window of duration T.

3.2. Sequential changepoint detection

Dividing $\mathbf{S}\boldsymbol{x}[t,k]^2$ by its row-wise ℓ^1 norm yields renormalized scattering coefficients $\widetilde{\mathbf{S}}\boldsymbol{x}[t,k]$ which are nonnegative and sum to one, and can thus be interpreted as the source parameters of a categorical probability density function on the discrete set $\{1 \dots K\}$:

$$\widetilde{\mathbf{S}}\boldsymbol{x}[t,k]^2 = \frac{\mathbf{S}\boldsymbol{x}[t,k]^2}{\sum_{\kappa=1}^K \mathbf{S}\boldsymbol{x}[t,\kappa]^2} \tag{10}$$

By virtue of the one-to-one correspondence between density functions in the exponential family and Bregman divergences [3], we deduce a sequential changepoint detection algorithm in which parameters both before and after change are unknown [6].

Figure 4. Installation view of Florian Hecker's exhibition *Synopsis* at Tramway Glasgow, 6 May – 30 July 2017. Photography © Keith Hunter, reproduced with permission.

The algorithm begins by setting $t_0 = 0$ and $t_1 = 1$, and $n = 0$. At every clock tick, it increments t_{n+1} and tests for the presence of a changepoint t for every t between $(t_n + 1)$ and t_{n+1}, by comparing the binary logarithms of generalized likelihood ratios $\mathrm{GLR}(\boldsymbol{x}, t_n, t_n + 1)[t]$ to some fixed threshold. Let

$$H(\mathbf{X}) : t \longmapsto -\sum_{k=1}^{K} \mathbf{X}[t,k] \log_2 \mathbf{X}[t,k] \qquad (11)$$

be the Shannon entropy of a stream of histograms $\mathbf{X}[T,k]$. The binary logarithm of the GLR is defined as

$$\begin{aligned} \log_2 \mathrm{GLR}(\boldsymbol{x}, t_n, t_{n+1})[t] = \\ (t_{n+1} - t_n) \times H(\mathbf{X}_{\mathrm{P}\cup\mathrm{F}}(t_n, t_{n+1})) \\ - (t - t_n) \times H(\mathbf{X}_{\mathrm{P}}(t_n, t_{n+1}))[t] \\ - (t_{n+1} - t) \times H(\mathbf{X}_{\mathrm{F}}(t_n, t_{n+1}))[t], \end{aligned}$$

where the sufficient statistics before change ($\mathbf{X}_{\mathrm{P}}$ where P stands for "past") and after change ($\mathbf{X}_{\mathrm{F}}$ where F stands for "future") are respectively equal to

$$\mathbf{X}_{\mathrm{P}}(t_n, t_{n+1}) : (t,k) \longmapsto \frac{1}{t - t_n + 1} \sum_{\tau=t_n}^{t} \widetilde{\mathbf{S}}\boldsymbol{x}[\tau,k]^2,$$

$$\mathbf{X}_{\mathrm{F}}(t_n, t_{n+1}) : (t,k) \longmapsto \frac{1}{t_{n+1} - T} \sum_{\tau=t+1}^{tn+1} \widetilde{\mathbf{S}}\boldsymbol{x}[t,k]^2, \text{ and}$$

$$\mathbf{X}_{\mathrm{P}\cup\mathrm{F}}(t_n, t_{n+1}) = \frac{\mathbf{X}_{\mathrm{P}}(t_n, t_{n+1}) + \mathbf{X}_{\mathrm{F}}(t_n, t_{n+1})}{t_{n+1} - t_n + 1}. \qquad (12)$$

If $\log_2 \mathrm{GLR}(\boldsymbol{x}, t_n, t_{n+1})[t]$ exceeds some threshold $\Delta H^{\ddagger}$, then n is incremented, t_{n+1} is set to t_n, and t_n is set to t. By analogy between information theory and thermodynamics, we propose to name the constant $\Delta H^{\ddagger}$ the *activation entropy* of the changepoint. We refer the reader to [14, Chapter 2] for a detailed explanation of the algorithm.

3.3. Application to *Synopsis Seriation*

Synopsis Input, the raw audio material of *Synopsis Seriation*, proceeds from the exhibition *Synopsis*, which was presented at the Tramway Arts Center in Glasgow, Scotland, UK [19]. Figure 4 presents two photographs of the installation in the exhibition space (gallery "Tramway 2"), which has an approximate area of 1011 square metres [1]. *Synopsis* is an immersive installation: over the duration of 25 minutes and 20 seconds, visitors may navigate freely inside Tramway 2 and experience changes in auditory spatialization, depending on their chosen position.

Ch.	Title	Year
1		
2	*Formulation*	2015
3		
4		
5	*Formulation DBM Self*	2015–2017
6		
7		
8	*Formulation As Texture [hcross]*	2017
9		
10		
11	*Formulation Chim 111 [hcross]*	2017
12		

Table 1. Audio contents of *Synopsis Input*. Channels 1 through 12 correspond to the loudspeakers in the *Synopsis* exhibition. These loudspeakers are grouped into four different auditory environments, each playing a different three-channel piece.

The installation comprises twelve loudspeakers which are suspended at ear's height and arranged at different locations of Tramway 2, thus forming an irregular 2-D pattern. Eight vertical panels with sound-absorbing surfaces are also in suspension between pillars of the building. These panels divide the total area of Tramway 2 into four auditory environments, creating distinct but overlapping spatial layouts for the audience to experience.

Table 1 summarizes the audio contents of the *Synopsis* installation, which is reused as input to *Synopsis Seriation*. Each the four auditory environments in the exhibition space

[1] Official website of Tramway Arts Center: www.tramway.org

play a different three-channel piece. These four pieces comprise a source material, named *Formulation* (2015), as well as three variants which were obtained by computer analysis and resynthesis.

Formulation is rich in dynamics and abrupt changes of heterogeneous sonic content stemming from diverse synthesis processes. Amongst others, *Formulation* features material generated with the "Sound Texture Synthesis Toolbox" [2] of Josh McDermott and Eero Simoncelli [20]. *Formulation DBM Self* (2015–2017) employs a "Deep Boltzmann Machine Sparse Decomposition" algorithm provided by Bob Sturm [25]. *Formulation Chim 111 [hcross]* (2017) features an "Auditory Chimera" algorithm by Bertrand Delgutte and Jayaganesh Swaminathan [22]; and together with *Formulation As Texture [hcross]* (2017), also a texture synthesis algorithm developed by Axel Röbel and Hugo Caracalla [4].

3.4. Segment-wise summarization

Let us denote by $c \in \{1, \ldots, 12\}$ the channel variable in *Synopsis Input* and $\widetilde{\mathbf{S}}\boldsymbol{x}[c, t, k]$ the renormalized scattering coefficients at channel index c, time index t, and path index k. We apply the algorithm of sequential changepoint detection, as described in the previous subsection, independently on each channel c. After a process of trial and error, we adjust the activation entropy $\Delta H^{\ddagger}$ equal to $11\,\mathrm{bits}$ for all 12 channels.

For every channel c, we store the list of timestamps t_0, t_1, etc. as a vector $\mathbf{T}(c)[n]$ where n denotes the timestamp index. Note that the number of timestamps N_c, i.e., the dimension of $\mathbf{T}(c)$, varies from one c to another. Then, we average coefficients $\widetilde{\mathbf{S}}\boldsymbol{x}[c, n, k]$ on a per-channel and per-segment basis, thus yielding the three-way tensor

$$\mathbf{Y}(c, n)[k] = \frac{1}{\mathbf{T}(c)[n+1] - \mathbf{T}(c)[n] + 1} \times \sum_{\tau=\mathbf{T}(c)[n]}^{\mathbf{T}(c)[n+1]-1} \widetilde{\mathbf{S}}\boldsymbol{x}[c, \tau, k]^2, \quad (13)$$

in which each entry contains the right-sided Bregman centroid of the segment $(\mathbf{T}(c)[n], \mathbf{T}(c)[n+1])$ for channel c.

4. EVOLUTIONARY ALGORITHM

4.1. Jensen-Shannon divergence

The Kullback-Leibler divergence between two discrete probability distributions with K categories is defined as:

$$\mathrm{KL}\big(\mathbf{P} \| \mathbf{Q}\big) = \sum_{k=1}^{K} \mathbf{P}[k] \log\left(\frac{\mathbf{P}[k]}{\mathbf{Q}[k]}\right) \quad (14)$$

Note that the KL divergence is asymmetric: in the general case, $\mathrm{KL}\big(\mathbf{P} \| \mathbf{Q}\big) \neq \mathrm{KL}\big(\mathbf{Q} \| \mathbf{P}\big)$. To circumvent this problem, we adopt a symmetrized version of the KL divergence, known as the Jensen-Shannon (JS) divergence. The definition of the JS divergence is as follows:

$$\mathrm{JS}\big(\mathbf{P} \| \mathbf{Q}\big) = \sum_{k=1}^{K} \frac{\mathbf{P}[k]}{2} \log\left(\frac{2\mathbf{P}[k]}{\mathbf{P}[k] + \mathbf{Q}[k]}\right) + \sum_{k=1}^{K} \frac{\mathbf{P}[k] + \mathbf{Q}[k]}{4} \log\left(\frac{\mathbf{P}[k] + \mathbf{Q}[k]}{2\mathbf{Q}[k]}\right) \quad (15)$$

Adopting $\mathbf{M}$ as a shorthand for the midpoint distribution $\frac{1}{2}(\mathbf{P} + \mathbf{Q})$, the JS divergence rewrites as:

$$\mathrm{JS}\big(\mathbf{P} \| \mathbf{Q}\big) = \frac{1}{2}\mathrm{KL}\big(\mathbf{P} \| \mathbf{M}\big) + \frac{1}{2}\big(\mathbf{M} \| \mathbf{Q}\big), \quad (16)$$

which confirms that the JS divergence is symmetric. We use a MATLAB implementation of the JS divergence by Boris Schauerte [3]. We refer the reader to [21] for more details on JS and KL divergences.

4.2. Temporal consistency loss

Following the multichannel assignment of Table 1, we map channels 1 through 6 (resp. 7 through 12) of *Synopsis Input* to the left (resp. right) channel of *Synopsis Seriation*. In doing so, we rearrange segments to follow a joint objective of temporal consistency and binaural (left-right) similarity.

We encode the seriation of segments via four sequences of indices: γ_{L}, η_{L}, γ_{R}, and η_{R}. The sequence γ_{L} (resp. γ_{L}) contains the channel indices in *Synopsis Input* of the segments of the left (resp. right) channel in *Synopsis Seriation*. Furthermore, the sequence η_{L} (resp. γ_{R}) contains the segment indices in *Synopsis Input* of the segments of the left (resp. right) channel in *Synopsis Seriation*.

Recalling Equation 13, the Bregman centroid that is associated to the i^{th} segment in the left channel of *Synopsis Seriation* is $\mathbf{Y}\big(\gamma_{\mathrm{L}}(i), \eta_{\mathrm{L}}(i)\big)$. Let us denote by I_{L} the number of such indices i, i.e., the total number of segments in channels 1 through 6 of *Synopsis Input*. We formulate the loss of temporal consistency of the seriation in the left channel as the cumulated Jensen-Shannon divergence between adjacent segments:

$$\mathcal{L}_{\mathrm{temporal,L}}(\gamma_{\mathrm{L}}, \eta_{\mathrm{L}}) = \sum_{i=1}^{I_{\mathrm{L}-1}} \mathrm{JS}\Big(\mathbf{Y}\big(\gamma_{\mathrm{L}}(i), \eta_{\mathrm{L}}(i)\big) \Big\| \mathbf{Y}\big(\gamma_{\mathrm{L}}(i+1), \eta_{\mathrm{L}}(i+1)\big) \Big). \quad (17)$$

Likewise, we define a loss function $\mathcal{L}_{\mathrm{seq,L}}(\gamma_{\mathrm{L}}, \eta_{\mathrm{R}})$ for temporal consistency in the right channel, expressed in terms of index sequences γ_{R} and η_{R}:

$$\mathcal{L}_{\mathrm{temporal,R}}(\gamma_{\mathrm{R}}, \eta_{\mathrm{R}}) = \sum_{i=1}^{I_{\mathrm{R}-1}} \mathrm{JS}\Big(\mathbf{Y}\big(\gamma_{\mathrm{R}}(i), \eta_{\mathrm{R}}(i)\big) \Big\| \mathbf{Y}\big(\gamma_{\mathrm{R}}(i+1), \eta_{\mathrm{R}}(i+1)\big) \Big), \quad (18)$$

[2] Source code of the MATLAB Sound Texture Synthesis Toolbox: https://mcdermottlab.mit.edu/downloads.html

[3] Source code of the MATLAB "Histogram Distances" Toolbox: http://schauerte.me/code.html

where I_{R} s the total number of segments in channels 7 through 12 of *Synopsis Input*.

4.3. Binaural consistency loss

The terms $\mathcal{L}_{\mathrm{temporal,L}}$ and $\mathcal{L}_{\mathrm{temporal,R}}$ ensure that the stereo channels (left and right) of *Synopsis Seriation* are temporally coherent, in the sense that they formulate a traveling salesperson problem (TSP) for the Jensen-Shannon divergence. Conversely, we also encourage *Synopsis Seriation* to have binaural consistency, in the sense that, at every time t, the simultaneous audio contents of the left and right channel should have a low mutual Jensen-Shannon divergence. To formulate the binaural consistency loss, we consider the concatenation of renormalized scattering coefficients in the left channel according to the channel sequence γ_{L} and to the segment sequence η_{L}. Recalling the notations of Section 3.3, we obtain the following sequence of K-dimensional vectors:

$$\mathbf{Z_L}(t)[k] = \Bigg(\bigotimes_{i=1}^{I_{\mathrm{L}}} \widetilde{\mathbf{S}}\boldsymbol{x}\Big[\gamma(i), \mathbf{T}(\gamma_{\mathrm{L}}(i))[\eta_{\mathrm{L}}(i)] : \mathbf{T}(\gamma_{\mathrm{L}}(i))[\eta_{\mathrm{L}}(i)+1], k\Big]\Bigg)(t)[k], \quad (19)$$

where the symbol $\otimes$ here denotes sequence concatenation and the colon notation $\mathbf{X}[a:b]$ represents range indexing inside the array $\mathbf{X}$ with a included and b excluded.

Like in the equation above, the sequences γ_{R} and η_{R} define the list of renormalized scattering coefficients that correspond to the right channel $\mathbf{Z}_{\mathrm{R}}$. To alleviate notation, we leave the dependency of $\mathbf{Z_L}$ in γ_{R} and η_{R} as implicit, as well as the dependency of $\mathbf{Z_R}$ in γ_{R} and η_{R}.

We define the binaural consistency loss as tbe cumulated simultaneous Jensen-Shannon divergence between the left and right channels, in the feature space of renormalized joint time–frequency scattering coefficients:

$$\mathcal{L}_{\mathrm{binaural}}(\gamma_{\mathrm{L}}, \eta_{\mathrm{L}}, \gamma_{\mathrm{R}}, \eta_{\mathrm{R}}) = \sum_{t=1}^{T} \mathrm{JS}(\mathbf{Z_L}(t), \mathbf{Z_R}(t)) \quad (20)$$

where the constant T here denotes the number of the piece, expressed in scattering transform frames.

4.4. Parallelized evolutionary optimization

We formulate the arrangement of segments in *Synopsis Seriation* as the minimization of the following loss function:

$$\mathcal{L} = \mathcal{L}_{\mathrm{temporal,L}} + \mathcal{L}_{\mathrm{temporal,R}} + \nu\mathcal{L}_{\mathrm{binaural}}. \quad (21)$$

In practice, we set the hyperparameter ν equal to 100 after a process of trial and error. Note that $\mathcal{L}$ is not a differentiable function of the parameter $\Theta = \big(\gamma_{\mathrm{L}}, \eta_{\mathrm{L}}, \gamma_{\mathrm{R}}, \eta_{\mathrm{R}}\big)$. Therefore, the resort to deep learning techniques on top of the scattering transform, as was recently proposed by [8] for drum sound synthesis, is inapplicable in this case.

Instead, we seek a quasi-optimal arrangement of segments via an evolutionary algorithm. We use a MATLAB implementation of the genetic algorithm by Joe Kirk [4] . We initialize the parameter Θ at random, while guaranteeing that the tuples $(\gamma_{\mathrm{L}}, \eta_{\mathrm{L}})$ and $(\gamma_{\mathrm{R}}, \eta_{\mathrm{R}})$ form two bijections. Then, we perform random mutations of the seriation sequences by swapping segments, either within the left channel or within the right channel. We retain mutations if and only if they improve the joint objective $\mathcal{L}$.

To speed up convergence, we parallelize computation over 100 CPU cores with different random seeds. Every 10^5 permutations, we mutualize results across cores and re-launch seriation with the best of the 100 seriations as the new initialization. We repeat this process for a duration of 96 hours, i.e., 400 CPU-days in total.

5. GRAPHICAL DESIGN

5.1. Printed booklet and extended digital booklet

Editions Mego have published *Synopsis Seriation* as a double CD as well as a digital download. The compact disc set includes a 16-page printed booklet and the digital download a 29-page booklet in PDF format. In academic computer music, the booklet is often used as a space featuring liner notes and commentary on the production. Yet, such functions tend to shift in fields of experimental, underground, non-academic computer music significantly. Indeed, the content of the booklet may extend visual concepts developed for the cover of a release or offering space for related paraphernalia that does not necessarily comment on the music as explicitly as written liner notes.

Over the past decades, we have experimented with different types of content accompanying CD releases. The CD release *Sun Pandämonium* [9] (2003) includes a 12-page booklet with monochrome coloured paper without any additional textual elements. The booklet accompanying the CD release *Speculative Solution* [10] (2011) contains essays by the philosophers Elie Ayache, Robin Mackay and Quentin Meillassoux. That booklet is distributed over 160 pages, taking the notion of accompanying text to an extreme.

The booklet of *Synopsis Seriation* suggests yet another detour, acknowledging the underlying scientific methodology as much as a design concept flirting with cyphers, symbols and signs. In this section, we discuss the contents of the extended digital version of the *Synopsis Seriation* booklet.

5.2. Computer-generated visualizations

After the first page, displaying the record's cover image, the booklet contains a collection of computer-generated visualisations of sample segments of *Seriation Input*. As reproduced in Figure 5 (left), this collection begins with a scattering representation ("scattergram") of a segment covering the duration from 1 minute 28 seconds to 1 minute 33 seconds from channel 1 of the input piece *Formulation*.

[4] Source code of MATLAB Traveling Salesman Problem Genetic Algorithm Toolbox: https://github.com/rubikscubeguy/matlab-tsp-ga

Figure 5. *Synopsis Seriation*, digital booklet, pages 2-3. On the left, page 2 of the digital booklet: scattergram covering the duration from 1 minute 28 seconds to 1 minute 33 seconds from channel 1 of the input piece *Formulation*. On the right, page 3 of the digital booklet: correlation matrix covering the same segment.

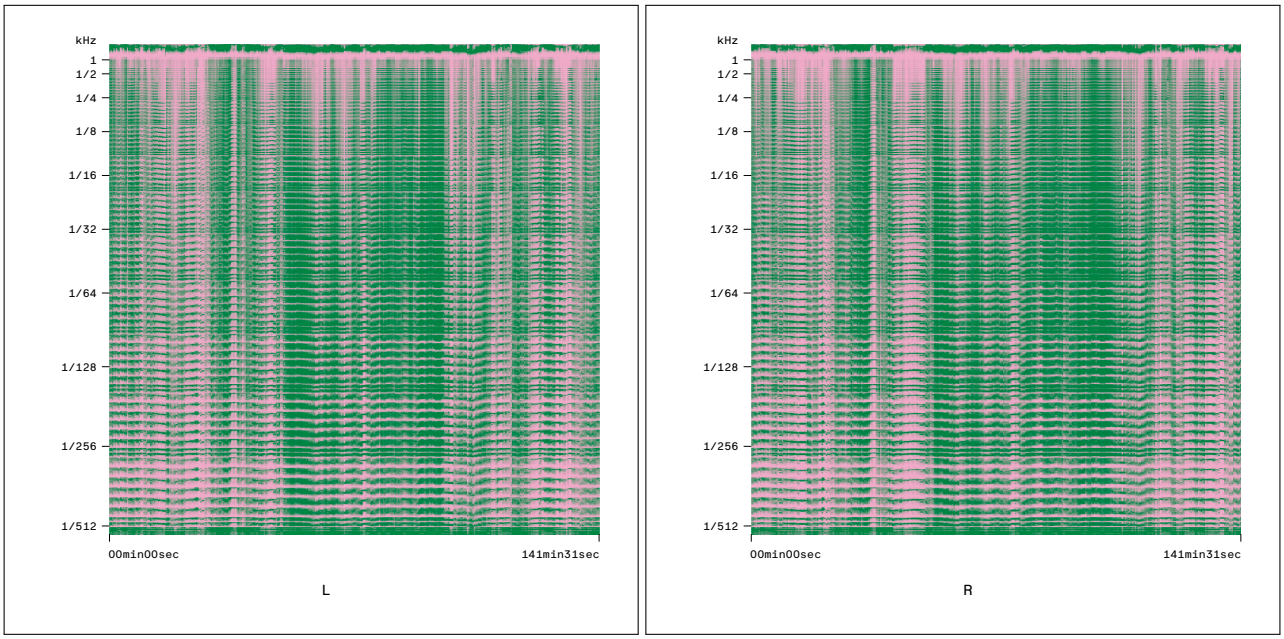


Figure 6. *Synopsis Seriation*, digital booklet, pages 14-15. On the left, page 14 of the digital booklet: scattergram covering the entire left channel of *Synopsis Seriation*. On the right, page 15 of the digital booklet: scattergram covering the entire right channel of *Synopsis Seriation*.

In doing so, we have ordered coefficients in $\mathbf{S_2}\boldsymbol{x}$ by decreasing λ, then decreasing temporal rate α, and finally decreasing frequential scale β. For reasons of legibility, we only display ticks that correspond to values of λ on the y-axis, and omit the display of α as well as β.

To retrieve the physical quantities that are associated to each scattering path $p = (\lambda, \alpha.\beta)$, we build upont implementation that was presented in the project "Scattering to Text" for the book *Florian Hecker: Halluzination, Perspektive, Synthese* (2019) [5] [15].

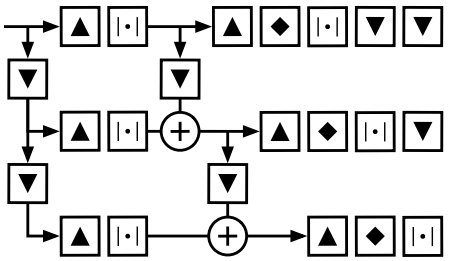
Alongside the scattergram of the audio segment at hand, we visualize the spectral correlation matrix that corresponds to the same segment (Figure 5, right). This visualisation stems from the texture synthesis algorithm [4] that has been employed in the input pieces *Formulation As Texture [hcross]* and *Formulation Chim 111 [hcross]*. This alternating mode of depicting visualisations of short segments continues throughout the booklet with the exceptions of pages 14 and 15 – the equivalent to the centrefold of the printed booklet. These show scattergrams of the entirely seriated output that makes up the sonic content *Synopsis Seriation*. Pages 14 and 15 display the entire left and right channel respectively (see Figure 6).

The scattergrams and correlation matrices are reproduced in pink and green, colours rich in contrast, selected for aesthetic reasons and to print the visualizations with two complementary colours, valid in all three print channels (C, M and Y) of the CMYK colour model. This ensures the most reliable efficiency of colour representation in the additive and overlapping process applied in print. The visualizations were generated using the Matplotlib colourmaps 'cool' and 'inferno' initially. Whereas the 'cool' colourmap is defined via two colour channels (R and G) of the RGB colour model, and is, therefore, easier to extract these into separate layers which can be assigned with a colour such as pink or green in the print process. Yet, a direct transformation from the RGB to CMYK tends to be more challenging without significant colour changes. Colourmap 'inferno' poses the opposite scenario: it is defined via all three colour channels of the RGB colour model; however, a direct transformation to CMYK is more straightforward. To circumnavigate this difference, the designers worked with the 'cool' colourmap, extracting its colour channels, representing pink via the M channel and green via the C and Y channels of CMYK. The resulting combination of pink and green ultimately resembles in contrast and intensity that of the 'inferno' colourmap, yet employs the more elegant approach of two colours only, as featured in the 'cool' colourmap.

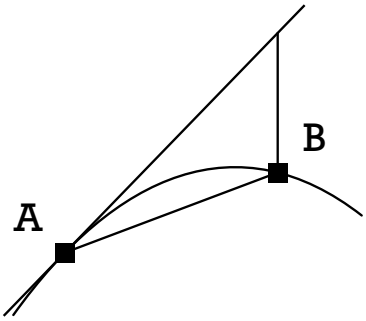
5.3. Diagrams

Page 28 corresponds to the last page of the printed booklet and contains four diagrams featuring an abstraction of the computational processes applied in *Synopsis Seriation* (see Figure 7). Diagrams, that "[...] are in a degree the

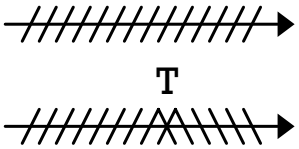
[5] Samples of the "Scattering to Text" project are available at: www.sternberg-press.com/product/halluzination-perspektive-synthese/



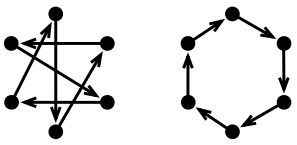
(1) Time-frequency scattering extracts spectrotemporal modulations according to a multiresolution pyramid scheme. Symbols ▲, ▼, and ♦ denote high-pass, low-pass, and band-pass filters respectively.



(2) The relative entropy (dashed line) between models A and B is equal to the difference between the log-likelihood of observing B under prior A and the entropy of B.



(3) Top: null hypothesis. Bottom: alternative hypothesis, denoting a changepoint at time T.



(4) Evolutionary algorithms update a random Hamiltonian cycle (left) iteratively to find the shortest route (right).

Figure 7. *Synopsis Seriation*, digital booklet, page 28.

accomplices of poetic metaphor" [5] inhabit a central role at the intersection of mathematics and philosophy [5] and equally interface music and mathematics [18].

The top diagram of Figure 7 represents the computational graph of time–frequency scattering. It is a redux of [1, Figure 7] while being devoid of any textual denotation of mathematical objects. For example, the low-pass filter, usually denoted by $\mathbf{h}$ in the wavelet theory literature, is here rendered as a downward triangle: ▼. Likewise, we render high-pass and band-pass filtering by means of the upward triangle (▲) and lozenge (♦) symbols respectively. Lastly, we adopt the notations $|\cdot|$ for complex modulus and $\bigoplus$ for concatenation.

The second diagram in Figure 7 is a pedagogical explanation of Bregman divergences, adapted from [11]. The explanation goes as follows: given two people A and B standing on a concave hill H, how high should B jump upward to appear in the line of sight of A? In the case of Synopsis Seriation, A and B are musical segments (left and right, for example), and the hill H corresponds to the Shannon entropy in dimension K, with K the number of time–frequency scattering coefficients.

The third diagram of Figure 7 represents the generalized likelihood ratio associated to the decision rule of sequential changepoint detection. Slash ($/$) and backslash ($\backslash$) symbols correspond to past ($\mathbf{X}_\mathrm{P}$) and future ($\mathbf{X}_\mathrm{F}$) observations respectively. In the null hypothesis, $\mathbf{X}_\mathrm{P}$ and $\mathbf{X}_\mathrm{F}$ are independent and identically distributed. In the alternative hypothesis, $\mathbf{X}_\mathrm{P}$ and $\mathbf{X}_\mathrm{F}$ are independent but not identically distributed.

Lastly, the fourth diagram of Figure 7 presents a randomly chosen Hamiltonian cycle on an hexagon (left) next to the shortest Hamiltonian cycle (right). This presentation illustrates the traveling salesperson problem, i.e., to seek the shortest route while traversing every node once.

In the CD booklet of *Synopsis Seriation*, we choose the diagrammatic form for conceptual and aesthetic reasons: how to display featured processes in a universal yet stylistically reduced way? By representing these diagrammatically — with one accompanying caption per diagram — such spores of information might trigger the listener's curiosity to investigate deeper despite their enigmatic appearance. What are these terminologies used, what are these techniques mentioned here? To initiate an inquisitiveness in the audience ultimately is also the function of booklets and liner notes.

6. RELATED WORK

While we are not familiar with any similar endeavors transforming the entire audio content of an installation artwork into the 2-channel format of a music publication employing information geometry, the formatting of compositions at the core of immersive installations into a music release format has a range of precursors.

Iannis Xenakis's composition *La Légende d'Eer* [13] (1978), originally conceived to be experienced as an automated diffusion in the *Diatope* pavilion structure erected in front of the Centre Georges Pompidou between 28 June and 21 December 1978, subsequently has been published in CD format [28]. The process of this CD production and other succeeding publications of *La Légende d'Eer* have been described in great detail [7]. The original 7-channel source material of *La Légende d'Eer* has been mixed to a stereo format by blending and superimposing the multi-channel material into the 2-channel requirement of the audio CD. This approach keeps the overall experience and the timing structure of the original piece, as experienced in the *Diatope*, intact; however, very details of the multi-channel source material might get obscured and veiled through such down-mixing. Xenakis himself worked on producing several stereo versions — of which one is featured on the Auvidis Montaigne CD release — at WDR Cologne's Hörspielstudio in 1981. Here he "[...] only applied slight stereo panning to the mix and did not try to approximate a translation of the eight-channel spatialisations into stereo" [7]. More recent publications of *La Légende d'Eer* acknowledge sound spatialisation patterns derived from the spatial layout in the *Diatope* [29, 30].

Such historic examples pose a set of questions about how an immersive piece can circulate as a standard music release and via this route, reaching new audiences which have not experienced the original presentation. Nevertheless, *Synopsis Seriation* differs from such representational endeavors, insinuating a medium (the CD format) and media (the sound material) specific strategy with an *internal* logic, where decisions stem from the interaction of the virtual listening agent with the analyzed sound material. *Synopsis Seriation* transforms an immersive spatial experience into a new arrangement and ultimately into a different spatial experience, where the entire sonic content of the *Seriation Input* is revealed to the listener in synoptical clarity and directness.

7. CONCLUSION

Joint time–frequency scattering finds a growing number of musical applications, from timbre similarity retrieval [16] to the classification of playing techniques [27]. In this paper, we have shown that it may also serve as an acoustic frontend for music segmentation and structuration. We have introduced a new problem in computer music named multichannel seriation and have addressed it by means of combinatorial optimization techniques. This research has resulted in a computer music piece named "Synopsis Seriation", which has recently been released in CD format. The accompanying booklet presents computer-generated visualizations of the piece as well as scientific diagrams describing its mathematical underpinnings.

In future works, we hope to make new creations with time–frequency scattering by bringing insights psychoacoustics and artificial intelligence closer together. We note that the recent release of the Kymatio library for scattering transforms in PyTorch and TensorFlow paves the way towards such a goal [2].

8. REFERENCES

[1] Andén, J., Lostanlen, V. and Mallat, S. "Joint time–frequency scattering". *IEEE Transactions in Signal Processing*, 67(14), 3704–3718, 2019.

[2] Andreux, M., Angles, T., Exarchakis, G., Leonarduzzi, R., Rochette, G., Thiry, L., Zarka, J., Mallat, S., Andén, J., Belilovsky, E., Bruna, J., Lostanlen, V., Chaudhary, M., Hirn, M., Oyallon, E., Zhang, S., Cella, C., Eickenberg, E. "Kymatio: Scattering transforms in Python". *Journal of Machine Learning Research*, 21(60), pp. 1–6, 2020.

[3] Banerjee, A., Merugu, S., Dhillon, I. S. and Ghosh, J. and Lafferty, J. "Clustering with Bregman divergences". *Journal of Machine Learning Research*, 6(10), 2015.

[4] Caracalla, H. and Roebel, A. "Gradient Conversion Between Time and Frequency Domains Using Wirtinger Calculus". In *Proceedings of the International Conference on Digital Audio Effects (DAFX)*, 2019.

[5] Châtelet, G. *Figuring space philosophy, mathematics, and physics*. Kluwer Academic Publishers, 2000.

[6] Dessein, A. and Cont, A. "An information-geometric approach to real-time audio segmentation". *IEEE Signal Processing Letters*, 20(4), 331–334, 2013.

[7] Friedl, R. "Towards a Critical Edition of Electroacoustic Music: Xenakis, La Légende d'Eer". In M. Solomos, ed. *Xenakis. La musique électroacoustique*. Paris: Harmattan, 109-–122, 2015.

[8] Han, H. and Lostanlen, V. "wav2shape: Hearing the Shape of a Drum Machine". *Proceedings of Forum Acusticum*, 2020.

[9] Hecker, F. *Sun Pandämonium*. Mego, MEGO 044, Vienna, compact disc, 2003.

[10] Hecker, F. *Speculative Solution*. Editions Mego, EMEGO 118, Vienna and Urbanomic, Falmouth, UK, compact disc, 2011.

[11] Huszár, F. "Scoring rules, divergences and information in Bayesian machine learning". PhD dissertation, University of Cambridge, 2013.

[12] Kendall, M. G. and Smith, B. B. "On the method of paired comparisons". *Biometrika*, 31(3), 324–325, 1940.

[13] Kiourtsoglou, E. "An Architect Draws Sound and Light: New Perspectives on Iannis Xenakis's Diatope and La Legende d'Eer (1978)". *Computer Music Journal*, 41(4), 8–31, 2018.

[14] Lostanlen, V. "Découverte automatique de structures musicales en temps réel par la géométrie de l'information". Master's thesis, Ircam/UPMC, 2013.

[15] Lostanlen, V. "On Time-frequency Scattering and Computer Music". In V. Müller, Ed. *Florian Hecker: Halluzination, Perspektive, Synthese*, Sternberg Press, 2018.

[16] Lostanlen, V., El-Hajj, C., Rossignol, M., Lafay, G., Andén, J., Lagrange, M. "Time–frequency scattering accurately models auditory similarities between instrumental playing techniques". *EURASIP Journal on Acoustics, Speech, and Music Processing*, 1, pp. 1–21, 2021.

[17] Mallat, S. "Group invariant scattering". *Communications on Pure and Applied Mathematics*, 65(10), 1331–1398, 2012.

[18] Mazzola, G. *The topos of music: Geometric logic of concepts, theory, and performance*. Birkhäuser, 2012.

[19] Fowler, L. "The Technical Sound—On Florian Hecker's *Synopsis* and the apparatus of Electronic Music". In *Florian Hecker – Synopsis*, Tramway Glasgow, 2017. https://tinyurl.com/37js92mz

[20] McDermott, J.H. and Simoncelli, E.P., "Sound texture perception via statistics of the auditory periphery: evidence from sound synthesis." *Neuron*, 71(5), pp.926-940, 2011.

[21] Nielsen, F. and Nock, R. "Sided and symmetrized Bregman centroids". *IEEE Transactions on Information Theory*, 55(6), 2882–2904, 2009.

[22] Smith, Z.M., Delgutte, B. and Oxenham, A.J., "Chimaeric sounds reveal dichotomies in auditory perception." *Nature*, 416(6876), pp. 870–90, 2002.

[23] Schwarz, Diemo and Beller, Grégory and Verbrugghe, Bruno and Britton, Sam. "Real-time corpus-based concatenative synthesis with CataRT". *Proceedings of the International Conference on Digital Audio Effects (DAFx)*, pp. 279–282, 2006.

[24] Schwarz, Diemo. "Corpus-based concatenative synthesis." *IEEE Signal Processing Magazine*, 24(2), pp. 92–104, 2007.

[25] Collins, Nick, and Bob L. Sturm. "Sound cross-synthesis and morphing using dictionary-based methods." In *Proceedings of the International Computer Music Conference. ICMA*, 2011.

[26] Waldspurger, I. "Exponential decay of scattering coefficients". In *Proceedings of the IEEE International Conference on Sampling Theory and Applications (SampTA)*, 143–146, 2017.

[27] Wang, C., Lostanlen, V., Benetos, E., Chew, E. "Playing Technique Recognition by Joint Time–Frequency Scattering". In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 881–885, 2020.

[28] Xenakis, I. *La Légende d'Eer*. Auvidis Montaigne, MO 782058, Paris, compact disc, 1995.

[29] Xenakis, I. *La Légende d'Eer*. Mode Records, mode148, New York, compact disc, 2005.

[30] Xenakis, I. *La Légende d'Eer*. Karlrecords, KR024, Berlin, vinyl and digital download, 2016.