



HAL
open science

Structure Versus Randomness in Computer Music and the Scientific Legacy of Jean-Claude Risset

Vincent Lostanlen

► **To cite this version:**

Vincent Lostanlen. Structure Versus Randomness in Computer Music and the Scientific Legacy of Jean-Claude Risset. Journées d'Informatique Musicale, May 2024, Marseille, France. hal-04571550v1

HAL Id: hal-04571550

<https://hal.science/hal-04571550v1>

Submitted on 8 May 2024 (v1), last revised 15 May 2024 (v2)

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.

STRUCTURE VERSUS RANDOMNESS IN COMPUTER MUSIC AND THE SCIENTIFIC LEGACY OF JEAN-CLAUDE RISSET

Vincent Lostanlen

Nantes Université, École Centrale Nantes, CNRS, LS2N, UMR 6004, F-44000 Nantes, France

RÉSUMÉ

Si le compositeur et chercheur Jean-Claude Risset (1938–2016) n’a pas érigé de système théorique unificateur, sa volonté de “composer le son lui-même”, affirmée dès 1969, a fait date. Mon article propose d’en tirer les leçons du point de vue du traitement du signal audionumérique. Dans ce but, j’emprunte au mathématicien Terence Tao son concept de “dichotomie structure–aléa” et en décline les manifestations sur deux moments emblématiques de la musique par ordinateur : la synthèse additive et les paradoxes de hauteur. Je montre qu’à chacun de ces moments, Risset a étendu les notions traditionnelles de structure et d’aléa en musique et les a articulées de manière originale. En guise d’épilogue, je reprends le même motif pour présenter un travail de recherche en cours : les réseaux de neurones multirésolution. Ainsi, là où des musicologues ont entrepris de situer l’héritage de Risset à partir de son répertoire de compositions, je propose une grille de lecture alternative, interne aux conditions d’objectivation des énoncés mathématiques.

1. INTRODUCTION

No one would dispute that the advent of the Information Age has profoundly disrupted the history of music [49, 7]. However, the converse assertion — i.e., that contemporary musicianship can inspire breakthroughs in the mathematical sciences — has few outspoken advocates. Even fewer yet are the individuals who put this assertion in practice. In this context, the figure of Jean-Claude Risset (1938–2016) epitomizes the infrequent, although by no means incongruous, tenet that “art and science bring about complementary kinds of knowledge” [33]. In 1969, he presented his piece *Mutations* as “[attempting] to explore [...] some of the possibilities offered by the computer to compose at the very level of sound—to compose sound itself, so to speak.” [19] Musicologists Vincent Tiffon [54] and Márta Grabócz [9] have shown that this aesthetic practice of “composing sound itself” recurs as a motto in Risset’s writings, up until his autobiographical note from 2015 [46]. In this article, I propose to take the same motto as a starting point, yet while adopting a mathematical and technological outlook, more so than a musicological one.

“My scientific and artistic activities have nourished each other”, Risset declared in 1999 upon being awarded the CNRS gold medal [35]. Yet, Risset also insisted that “computer music should be evaluated as music, not as experi-

ment” [41]. Thus, even so Risset’s “scientific research was driven by musical desires” [35], he stated that, in general, “research must strive to preserve its independence, to not let itself be altered by a demand that would be too pressing or too precise” [26]. And elsewhere : “I do not think it is artificial to separate the scientist from the composer” [43]. In other words, if the motto of “composing sound itself” is to have a scientific meaning, we must uphold it from within science ; i.e., from its internal criteria of validity and its specific methods of inquiry.

For this reason, the concept of “structure–randomness dichotomy”, as proposed by mathematician Terence Tao, will be particularly insightful. We will reuse it to discuss two moments in Risset’s scientific legacy and will conclude with an example taken from our ongoing research.

2. AGAINST REDUCTIONISM

Before delving into our inquiry, we must avoid a potential pitfall : that of interpreting “sound itself” as the postulate that musicianship is mathematizable through and through, and as accessible from audio recordings alone. In this regard, Risset had warned us as early as 1977 [23] :

We have found limitations to the powers of mathematics in music. The emergence of music (as that of life) imposes a critical mass of complexity which poses a practical problem of control. [...] We can no longer hope for the mathematics to be the *champ constitutif fondamental* of music ¹. [...] I was once endeared by the idea that there might be a profound mathematical foundation to the dazzling power of music—I no longer think so.

This position is in stark contrast with some antecedents, such as Rameau or Schönberg, as well as some contemporaries, such as Andreatta or Mazzola. As an example, Risset showed respect for the non-acoustical element in music, made evident by pitch spelling [39] :

Music is different from language : it does not convey a message that could be translated into another musical dialect. Yet it clearly has certain linguistic aspects. It is irritating to hear certain physicists unaware of tonal syntax claim that it does not make sense to differentiate between a B and a C flat in an

1 . This sentence refers to a concept by Jean-Toussaint Desanti [5].

equally-tempered tuning. In fact the same pair of sounds can imply different resolutions depending on the way you name them [...].

3. STRUCTURE BEYOND REPETITION AND RANDOMNESS BEYOND CHANCE

With this caveat in mind, we may outline a coherent thread in the scientific production of Risset around the concept of “composing sound itself”. Risset often described this concept by visual or tactile metaphors, namely, “sculpture” and “microsurgery” [40]. Let us propose a mathematical standpoint to complement these metaphors.

In 2006, Terence Tao pointed out profound connections between several well-known theorems across seemingly unrelated areas of mathematics : combinatorics, ergodic theory, harmonic analysis, and hypergraph theory [51]. For each of them, Tao revisited its proof by treating the mathematical object under study as a “superposition of a structured object and a pseudorandom object”. Having acknowledged that his notion of structure could take one of many forms depending on context, Tao offered some characteristic adjectives : *i.e.*, autocorrelated, low-entropy, smooth, predictable, concentrated, algebraic. Vice versa, pseudorandom objects are decorrelated, high-entropy, noisy, unpredictable, dispersed, or probabilistic.

Structured objects and pseudorandom objects, as depicted by Tao, are inherent to mathematics. As such, they do not necessarily align with “structure” and “randomness” as commonly understood in music theory. In interviews, Risset would occasionally mention Schenkerian analysis as a precursor of Chomsky’s generative grammars [39], as well as some musical examples of symmetry by Webern and Gershwin [29]. Yet, his music rarely involved exact repetition, and he viewed the proliferation of his synthetic sounds as closer to a rhizome than to a tree [31]. Therefore, theories of tonal or serialist music cannot apply [39] :

Structures can often be defined in a rigorous, objective way, for instance in terms of mathematical descriptions. However what counts in art is not the objective structure—*per se*—but the structure *in us*, as we experience it through our senses. (A distinction proposed by Jacques Mandelbrojt)

Towards the other side of the dichotomy, the concept of randomness should also be understood at the highest possible level of abstraction and generality ; *i.e.*, beyond the well-known case of stochastic music [25]. While Risset praised Xenakis as an “immense creator” , he was reluctant to apply probability theory to describe musical complexity [4] and saw limited value in random exploration [28] :

[...] by varying the numerous parameters of a complex model at random, we cannot reasonably hope to achieve a precise goal ; try to randomly turn the faces of a Rubik’s cube to arrive at the solution.

Most of the mathematical models which are surveyed in this article have no stochasticity in them : they do not rely

on a pseudorandom number generation software (PRNG) to compute digital audio samples. Yet, beyond the case of PRNG, the concept of randomness has other uses in science, particularly computer science : *e.g.*, arbitrariness, contingency, indeterminacy, irreducibility, and mutability. These words are not interchangeable : *e.g.*, the stochasticity of a “random walk” is not synonym with the arbitrariness of “random-access memory”. Yet, at the risk of a misunderstanding, we choose to maintain the term “randomness” as an umbrella term throughout our article, both for reasons of simplicity and as an homage to Tao’s original idea.

The existence of arbitrary long arithmetic progressions of prime numbers, as proven by Green and Tao in 2004, is noteworthy in this regard [10]. Indeed, at first glance, the ordered sequence of prime numbers is not structured : unlike the Fibonacci sequence, it does not obey a linear recurrent formula. It is also not random : the sieve of Eratosthenes is a deterministic algorithm which checks whether a given number is prime or not with bounded time complexity.

Yet, at a more refined level of mathematical description, the prime numbers show certain aspects of structured objects and, at the same time, aspects of (pseudo)random objects. On one hand, Tao lists some structural properties of prime numbers : *e.g.* they are all adjacent to a multiple of six, with two exceptions ; their last digit is 1, 3, 7, or 9, with two exceptions. On the other hand, a theorem by Hadamard and de la Vallée Poussin states that n^{th} prime number is approximately equal to $n \log n$. According to Tao [53] :

[This theorem shows that] the primes have some large-scale structure, even though they can behave quite randomly at small scales. [...] It is possible to use this large-scale structure, local structure, and small-scale randomness to prove some non-trivial results.

We contend that the structure–randomness dichotomy of prime numbers, as exemplified by the Green-Tao theorem, is “microsurgical” in the same metaphorical sense than the way Risset described his own practice of computer music. In both instances, neither structure nor randomness is explicitly apparent in the material ; and yet, both “deal with large objects of unspecified (or unusable) structure” (Tao) [52] via patterns “partially periodic and partially random” [34] (Risset), as Michael McNabb and John Chowning did in order to imitate the singing voice.

4. ADDITIVE SYNTHESIS

The need for structured objects in computer sound synthesis was already apparent in the early years of computer music. In 1963, Max Mathews justified the development of the *Music n* software series by pointing out the impracticality of specifying broadband analog signals in terms of isolated digital samples [14] : “the numbers-to-sound conversion is useless musically unless a suitable program (set of computer instructions) can be devised for computing the samples from a single set of parameters.” In 2014, Risset recollected his work with Mathews [45] :

Thanks to programming, one can synthesize sounds in many different ways. But when Max Mathews began to write programs for sound synthesis, he soon realized that he would have to spend his life writing different programs to implement different musical ideas. So, he undertook to write a really flexible program, as universal as possible. The main key to flexibility was the modular approach. Starting with *Music 3* (1959), the *Music n* programs—written by Max and others—would be compilers, that is, programs that could generate a multiplicity of different programs. [...] Any connection of modules corresponds to a particular synthesis model : it is called *instrument* by analogy.

In the same text, Risset insists on the innovative character of *Music 3* as a forerunner of modular synthesizers :

Contrary to a common belief, Max Mathews’s modular conception did not copy that of synthesizers : on the contrary, it inspired the analog devices built by Moog, Buchla and Ketoff using voltage control—these appeared after 1964, while *Music 3* was written in 1959.

Adopting the parlance of software engineering, the vision of Mathews (and Risset) for computer music could be described as a quest towards *loose coupling* and *high cohesion*. Here again, we find an instance of structure–randomness dichotomy, in the sense that the program combines mutable and immutable elements. On one hand, the synthesis models known as “instruments” remain unchanged at the scale of the musical fragment to be generated : as such, they provide a structural backbone to the program. On the other hand, control parameters for the instrument remain undetermined until the composition of the piece : these parameters may be chosen at will without technical knowledge about the implementation of encapsulated modules.

For reasons of usability and program conciseness, mutable elements should have a straightforward musical interpretability and remain relatively lightweight in terms of information content. A journal article by John Robinson Piece, Jean-Claude Risset, and Max Mathews summarizes these considerations (emphasis is ours) [16] :

Jean-Claude Risset has undertaken an analysis of real trumpet tones to get insight into the physical parameters which influence mostly the tone quality. It was found that from a *proper specification of a few features*, like the attack, a fast random frequency fluctuation, a frequency spectrum depending upon the loudness, one could synthesize brass-like sounds.

In a previous publication, researcher Anne Veitl has pointed out that this “proper specification” is akin to a new system of musical notation, which is conceptually closer to tablature than to *solfège* [55]. Veitl describes the programming paradigm of *Music 5* as a shift away from notating “certain theorized aspects of perceived phenomena” towards “writing down the causes” of sounds. She also

makes the case that Risset’s knowledge of physical acoustics and psychoacoustics has served as an underpinning for “digital tablatures” in *Music 5*. We may complement and nuance Veitl’s observation by pointing out that Risset had an unorthodox practice of audio content analysis. To him, the ultimate purpose of decomposing real-world sounds was to produce a flexible and evocative resynthesis, more so than to explain the formal cause of sound production at the vibratory level. In particular, Risset points out practical limitations behind modeling a musical event in terms of a Fourier series with P partials, with $P \rightarrow +\infty$:

$$\mathbf{x}(t) = \sum_{p=1}^P a_p \cos(2\pi p \xi t). \quad (1)$$

Even so the Fourier representation is mathematically universal for absolutely integrable functions, the fundamental frequency ξ may be lower than the perceived pitch and the number P may be computationally intractable [12]. This is particularly true for percussive inharmonic sounds such as bells and gongs, which Risset featured in his 1969 *Introductory Catalogue of Computer-Synthesized Sounds*. Risset’s solution was two-fold : first, to relax the constraint of having all partial frequencies be integer multiples of a fundamental ; and second, to introduce temporal amplitude modulation in each sinusoidal component $a_p \cos(2\pi \xi_p t)$. The additive synthesis model becomes :

$$\mathbf{x}(t) = \sum_{p=1}^P a_p(t) \cos(2\pi \xi_p t). \quad (2)$$

Such an increase in flexibility comes at the cost of greater complexity and runs the risk of making the adjustment of each $a_p(t)$ intractable. Yet Risset had learned, from his interactions with Max Mathews, that this complexification could be curtailed by introducing appropriate subroutines in the computer program. Specifically, Risset proposed to drive the temporal envelopes $a_p(t)$ by exponential functions whose decay rates are in inverse proportion to frequencies ξ_p : $a_p(t) = \alpha_p \exp(-\gamma \xi_p t)$, where the series α_p and the dimensionless parameter γ are specific to the instrument but independent from choice of pitch. This idea is reminiscent of time–frequency analysis as defined by Dennis Gabor in 1946 and made computationally efficient by Cooley and Tukey’s fast Fourier transform (FFT) of 1966. In 1969, Risset and Mathews insisted on the potential of these approaches, which are now ubiquitous in digital audio [20] :

The importance of time factors in tone quality, quite familiar to musicians using tape manipulation and computer synthesis of sound, do not appear to be well publicized among physicists. Clearly one must perform some kind of “running” analysis that follows the temporal evolution of the notes.

The availability of pitch-synchronous analysis algorithms to identify the modulation signals $a_p(t)$ automatically from data is at the heart of Risset’s study of trumpet tones, and

his “exploration of timbre by analysis and synthesis” in general [36]. Indeed, it provides the opportunity to introduce ductility in the design of musical notes via composition of oscillators, as compared to the excessive rigidity of Fourier series. As Risset puts it in his *Catalogue* [22] :

Here no attempt has been made towards economy of specification : schematized data from real trumpet tones have been used.

Our claim is that this original combination of domain-specific knowledge and data-driven modeling in Risset’s treatment of additive synthesis is another instance of structure–randomness dichotomy in computer music. Going back to the examples of bell and gong-like sounds, the structured object is made evident by the following system of differential equations

$$\forall p \in \{1, \dots, P\}, \quad \frac{da_p}{dt} = -\gamma \xi_p a_p(t), \quad (3)$$

for which γ is a spectrotemporal invariant. Yet, in his piece *Inharmonique* (1977), Risset calls this invariant into question, thus recombining randomness with structure [30] :

In most natural sounds, the higher-frequency components tend to decay more rapidly than the lower-frequency ones, which can be formulated as a principle : the higher, the shorter. Because with additive synthesis one has complete control of the decay rates of all components, one can add interest to the computer timbres by occasionally violating the principle. [...] This kind of transformation was used extensively in my piece *Inharmonique*.

Another phenomenon worth discussing is the correlation between high-frequency content and loudness, particularly in the case of brassy instruments such as the trumpet. Risset insisted on the importance of this correlation many times over five decades, from his 1964 visit to Bell Laboratories [17] and until a 2015 article for *Organised Sound* : “the louder, the brighter” [47]. The key observation is that the Fourier decomposition is linear : if the signal x is multiplied by some constant k , the amplitude coefficient at frequency $p\xi$ is equal to ka_p . Yet, most families of acoustical instruments, including brass, are nonlinear. Thus, controlling the amplitude at the source via a simple multiplicative factor k does not yield the intended effect of change in loudness. Instead, Risset proposed to depart from Fourier’s linear decomposition and replaced it with a nonlinear control of the form

$$a_p = a_{p,\text{ref}} + \lambda(\alpha - \alpha_{\text{ref}})p, \quad (4)$$

where α is the loudness control and the amplitude slope parameter λ is a spectrotemporal invariant. Note that all amplitude levels are equal to $a_{p,\text{ref}}$ for $\alpha = \alpha_{\text{ref}}$ and grow with partial-specific slopes λp for $\alpha > \alpha_{\text{ref}}$. This allowed Risset to present “some examples of brass-like sounds synthesized with more economy of specification” than what had been obtained from pitch-synchronous analysis [22].

Once again, we find a dichotomy of structure and randomness : *i.e.*, a dual effort towards drawing structural connections between seemingly independent attributes of synthesis whilst introducing free play inside pre-established structures whenever necessary.

5. PITCH PARADOXES

The age-old debate on the physiological foundations of harmony might appear to be purely a matter of structure. A cursory look to the history of Western musical tuning systems reveals a long sequence of normative constructs, from the Pythagoreans to the Xenharmonists. While each of these constructs hinges on a certain notion of structural regularity, they leave no room, if any, for randomness. On the contrary, the original treatment of pitch paradoxes by Risset demonstrates that structure and randomness coexist in our perception of intervals. Risset first presented this work to the Acoustical Society of America in 1969 [21] :

Using computer-generated sounds, R. N. Shepard has demonstrated circularity in judgments of relative pitch. Other unusual sounds will be presented, including endlessly descending glissandi [...].

Here, Risset is referring to the Shepard tones, which are built as a series of pure tones which are one octave apart. A Shepard tone of frequency ξ in Hertz and bandwidth $(2J + 1)$ in octaves is defined by

$$x(t) = \sum_{j=-J}^J 2^j \cos(2\pi 2^j \xi t). \quad (5)$$

Towards the limit of infinite bandwidth ($J \rightarrow +\infty$), the signal x satisfies the self-similarity equation $x(t) = 2x(2t)$. In practice, setting $J = 10$ is sufficient for any ξ in the hearing range of humans. As a result of this summation, the pitch of x is fundamentally ambiguous on a pitch height scale from low to high; meanwhile, x retains a recognizable pitch class on a log-frequency circle whose circumference is equal to one octave. Risset has modified Shepard’s definition by replacing partial tones of frequencies $(2^j \xi)$ by exponential chirps whose frequency $(2^{j+(t/T)})\xi$ double after a predefined time span T . The result $x(t)$ is known as a Shepard–Risset glissando :

$$x(t) = 2^{(t/T)} \sum_{j=-J}^J 2^j \cos\left(2\pi \frac{2^j \xi \log 2}{T} 2^{(t/T)}\right) \quad (6)$$

Shepard tones and Shepard–Risset glissandos are both fractal signals. As such, they are archetypes of mathematical structure. Yet, the case of the Shepard–Risset glissando also exhibits randomness in the sense that the choice of reference frequency ξ becomes irrelevant : a pitch interval $\delta_p = \log_2(\xi'/\xi)$ may be converted into a time interval $\delta_t = (t' - t)$ provided that $T = (\delta_t/\delta_p)$. By transforming the discrete sequence of Shepard tones into a continuous upward (or downward) frequency modulation, the Shepard–Risset glissando is indifferent to a choice of temperament

or tuning system. The auditory illusion it elicits, which Risset included in his *Computer Suite for Little Boy* [18], suggests that the disentanglement between pitch height and pitch chroma is more fundamental to music perception than interval relationships between fixed pitch classes.

Although the synthesis of the Shepard–Risset glissando does not rely on pseudorandom number generation, its construction does reveal significant aspects of indeterminacy and irreducibility in auditory perception—*i.e.*, randomness as understood in a broader sense. Risset strived to avoid inducing too much of his own knowledge about music theory into computer programs. On the contrary, he saw the computer as a “workshop” [44] able to craft “specially contrived stimuli” [21] that would shatter the dogmas we inherited from the tonal tradition. This willingness to maintain as much indeterminacy as possible in the design of fractal sounds is also apparent in the behavioral experiments on the circularity of pitch judgments, conducted with Gérard Charbonneau (emphasis is ours) [1] :

Since factor analysis provides a geometrical model which reflects the structure of the data, *without involving preconceived ideas about them*, we see that pitch judgments on the stimuli above are well described by a circular diagram. These results are compatible with the helical diagram since the spectral envelope of artificial stimuli has been kept invariant so that they should not significantly differ in terms of pitch height : in this case, the helix degenerates into a circle [...].

In the conclusion of the same paper, Charbonneau and Risset insist on the fact that the experimental design is done “without a priori”; *i.e.*, with no visual or textual reference. Another paper by the same two authors goes further along the structure–randomness dichotomy by showing that the orientation of the helix is itself indeterminate [2] :

Our results give a diagram akin to the helix of Drobisch, Revesz, and Pikler, but here the helix is leftward, which reflects the fact that with our stimuli, unlike what typically happens, sounds become lower (f_0 decreases) as we climb the whole-tone scale.

Esqueda *et al.* have followed in the footsteps of Risset by showing that the same “barberpole illusion” of endless ascent can be achieved via octave-spaced spectral notches upon a white noise background [8]. The replacement of narrowband spectral peaks by broadband noise raises the level of indeterminacy in the construction of the glissando.

A lecture from 2008 by Risset offers an interesting retrospective on the role of structure and randomness in auditory paradoxes (emphasis is ours) [42] :

Digital synthesis allows to build very curious sounds, yielding acoustical illusions [...]. Purkinje, talking about optical illusions, wrote : “illusions, errors of the senses, are truths of perception”. *The mechanisms of auditory perception are highly idiosyncratic, yet are no-*

thing arbitrary : their complexity is understood in the perspective of auditory scene analysis, a notion introduced by Albert Bregman and mentioned by Christine Petit.

Another element of indeterminacy in the structure of the helical diagram is found in the “tritone paradox”; that is, the response of a listeners to two Shepard tones of frequencies ξ and $\xi' = \sqrt{2}\xi$. Since $1/\sqrt{2} = \sqrt{2}/2$, the frequencies $\xi/\sqrt{2}$ and ξ' are spaced by one octave : thus, the problem of finding whichever tone ξ or ξ' is higher in pitch is ill-posed, and elicits randomness in listener responses. In 1979, Risset and David Wessel wrote [24] :

Psychologist Diana Deutsch [...] has used sounds comprising many components one octave apart—which in principle gives them the same pitch height—and asked listeners to make comparisons between pitch classes half an octave apart [...]. Curiously, listeners do not all agree : for the same sound pair, some hear the tritone going upward, others hear it downward.

Since then, Diana Deutsch has shown that, even so the response of a human listener to the tritone paradox cannot be predicted with perfect certainty, it correlates with certain objective factors such as geographical background and vocal pitch range [6]. More recently, Claire Pelofi has shown that the judgment of tritone directionality may be affected by probing listeners with a random melodic sequence of Shepard tones whose pitch class distributions is non-uniform [15]. Pelofi also found that this priming effect depends, in turn, on the level of musical expertise of the listener. Once more, we find that a highly structured model of pitch perception (the spiral diagram) harbors traits of randomness ; and vice versa, that seemingly random responses to a paradoxical stimulus (the Shepard tritone) may be partly explained by sociocultural determinations.

6. MULTIREOLUTION NEURAL NETWORKS

Structure–randomness dichotomies remain active today as a framework to design and interpret computer music systems. The recent surge of “hybrid deep learning” is arguably a redux of its founding principles, only at a larger scale of operation ². As one of many possible illustrations, we hereafter present an example from our own research : namely, multi-resolution neural networks (MuReNN).

MuReNN comprises two stages, multiresolution approximation (MRA) and 1-D convolutions, of which only the latter is learned from data. We implement the MRA with a dual-tree complex wavelet transform (DTCWT) [50]. The DTCWT relies on a multirate filterbank in which each wavelet ψ_j has a null average and a bandwidth of one octave. Denoting by ξ the sampling rate of x , the wavelet ψ_j has a bandwidth with cutoff frequencies $2^{-(j+1)}\pi$ and

². The concept of “hybrid deep learning” arose in physical sciences and was brought to music signal processing by Gaël Richard. See in particular Richard’s keynote at DAFX 2020 and ISMIR 2023 featuring George Fazekas, Changhong Wang, Zhiyao Duan, and Gus Xia.

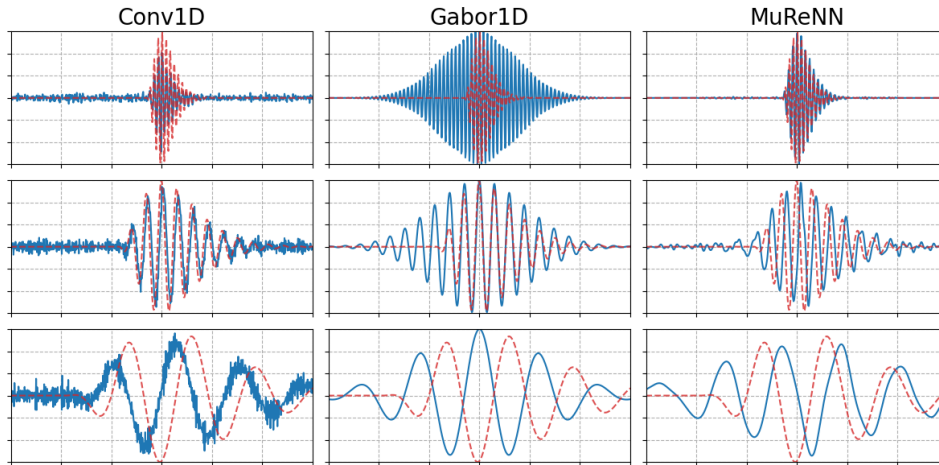


Figure 1. Compared impulse responses of Conv1D (left), Gabor1D (center), and MuReNN (right) after convergence, shown in the time domain. Rows corresponds to different Gammatone filters as targets : high-frequency (top), mid-frequency (center), and low-frequency (low). Solid blue (resp. dashed red) lines denote the real part of the impulse responses of the learned filters (resp. target). Figure reproduced from [13] with permission from coauthors.

$2^{-j}\pi$. Hence, we may subsample the result of the convolution ($\mathbf{x} * \psi_j$) by a factor of 2^j , yielding :

$$\forall j \in \{0, \dots, J-1\}, \mathbf{x}_j[t] = (\mathbf{x} * \psi_j)[2^j t], \quad (7)$$

where J is the number of multiresolution levels.

The second stage in MuReNN consists in defining learnable filters ϕ_f . Unlike in a one-dimensional convolutional neural network (Conv1D), those filters do not operate over the full-resolution input \mathbf{x} but over one of its MRA levels \mathbf{x}_j . More precisely, let us denote by $j[f]$ the decomposition level assigned to filter f , and by $2L_j$ the kernel size for that decomposition level. The number of such filters sharing the same level j is the number of filters per octave of MuReNN, and may vary depending on j .

We convolve $\mathbf{x}_{j[f]}$ with ϕ_f and apply a subsampling factor of $2^{J-j[f]}$, hence :

$$\begin{aligned} \Phi_{\mathbf{W}}\mathbf{x}[f, t] &= (\mathbf{x}_{j[f]} * \phi_f)[2^{J-j[f]}t] \\ &= \sum_{\tau=-L_j}^{L_j-1} \mathbf{x}_{j[f]}[2^{J-j[f]}t - \tau] \phi_f[\tau] \end{aligned} \quad (8)$$

The two stages of subsampling in Equations 7 and 8 result in a uniform downsampling factor of 2^J for $\Phi_{\mathbf{W}}\mathbf{x}$.

The trainable weights in MuReNN may be optimized via reverse-mode automatic differentiation, either in a supervised or self-supervised setting. In a recent publication [13], we have formulated a supervised task of knowledge distillation, also known as teacher–student training, in which the “teacher” model is an auditory filterbank. The neural network weights in $\psi_f[t]$ are initialized with independent random Gaussian and the cost function is the cosine distance between the magnitudes of the teacher model and those of the student model.

As a point of comparison with MuReNN, we also train a plain (single-resolution) Conv1D model and a parametric model known as Gabor1D, in which only the center

frequencies and bandwidths of Gabor filters are learnable [56]. We use 3190 spoken English vowels from the NT-VOW dataset as our training corpus.

Figure 1 illustrate our findings for three Gammatone teachers at various frequencies and three models as students : Conv1D, Gabor1D, and MuReNN. On one hand, the learned filters of the Conv1D model follow the temporal envelope of the Gammatone teacher but their time-domain impulse responses remain noisy, even after one hundred training epochs. On the other hand, the learned filters of the Gabor1D model are noiseless by design and have an excellent Heisenberg time–frequency localization, but their symmetric (Gaussian) temporal envelopes is a poor fit for Gamma functions, which have a sharp attack and a slow decay. To put it bluntly, learning with Conv1D has “too much randomness” whereas learning with Gabor1D has “too much structure”. In this context, MuReNN appears to strike the happy medium : it outperforms both Conv1D and Gabor1D in terms of goodness of fit by learning to replicate Gammatone filters with the appropriate temporal envelope (unlike Conv1D) and with few high-frequency artifacts (unlike Gabor1D); and so at any resolution.

The MuReNN project, which is currently ongoing, deliberately withdraws into the scientific legacy of Risset, and particularly his motto of “composing sound itself” via well-adapted dichotomies between structure and randomness. In Equations 7 and 8, the structured objects are the wavelets ψ_j while the random objects are the learnable filters ϕ_f . As of today, much remains to be understood about the abilities and limitations of MuReNN, both from theoretical and practical standpoints. For now, what we can say is that MuReNN offers an actionable solution for parameter-efficient deep learning in the raw waveform with limited domain-specific knowledge.

7. CONCLUSIONS

In this article, I have proposed an original interpretation for Risset’s well-known motto of “composing sound itself” with the computer. After borrowing the concept of structure–randomness dichotomy from mathematician Terence Tao, we have shown how this concept may also apply for computer music research. In doing so, we have interpreted these two concepts at the highest possible degree of generality : *i.e.*, structure beyond repetition and randomness beyond chance. A common theme between all the presented contributions is that Risset made a playful use of structure for sounds which did not conform to classical theories of musical acoustics (*e.g.*, inharmonic tones); and vice versa, he introduced elements of contingency and indeterminacy where they were the least expected : *e.g.*, inside geometrical models of pitch perception. This theme is eloquently described in the concluding remarks of one of Risset’s last published lectures [48] :

The composer has a direct grip on the physical nature of sound, but music comes into being in the auditory realm : natural and cultural models, schemes, scenarios, archetypes, universals help to craft musically the collision of the physical and the sensible.

I acknowledge that our survey is not comprehensive. Much could be said, for example, about structure and randomness in Risset’s use of the Gabor wavelet transform [27] as well as of digital audio effects [38]. It is my hope that, in spite of these limitations, our article may inspire other theoreticians and practitioners of computer music to examine how structure and randomness interact in their own work. The work in progress on multiresolution neural networks (MuReNN), as sketched in Section 6, serves as an example to show that the dichotomy between structure and randomness remains an effective paradigm in the age of stochastic optimization with deep neural networks.

8. ACKNOWLEDGMENT

I thank Mikhail Malt for his helpful advice. This work is supported by ANR project MuReNN (ANR-23-CE23-0007-01).

9. REFERENCES

- [1] Charbonneau G., & Risset J.-C. “Circularité de jugements de hauteur sonore”, *Comptes rendus de l’Académie des sciences de Paris*, Série B, 623–626, 1973.
- [2] Charbonneau, G., & Risset, J.-C. “Jugements relatifs de hauteur : schémas linéaires et hélicoïdaux”, *Comptes rendus de l’Académie des sciences de Paris*, Série B, 289–292, 1975.
- [3] Conlon, D., Fox, J., & Zhao, Y. “The Green-Tao Theorem : An Exposition”, *EMS Surveys in Mathematical Sciences*, 1(2), 249–282, 2014.
- [4] de Gérando, S. “À propos de l’oeuvre pour bande seule et de l’écriture musicale de Jean-Claude Risset”, *L’Éducation musicale*, 507, 2003.
- [5] Desanti, J.-T. *La Philosophie silencieuse, ou Critique des Philosophies de la science*, Éditions du Seuil, 1975.
- [6] Deutsch, D. “The Tritone Paradox : Some Further Geographical Correlates”, *Music Perception*, 12(1), 125–136, 1994.
- [7] Devine, K. *Decomposed : The Political Ecology of Music*, MIT Press, 2019.
- [8] Esqueda, F., Välimäki, V., & Parker, J. “Barberpole Phasing and Flanging Illusions”, *Proc. International Conference on Digital Audio Effects (DAFX)*, Trondheim, Norway, 87–94, 2015.
- [9] Grabócz, M. “Introduction”, *Colloque en hommage à Jean-Claude Risset. Pour un monde sonore virtuel*, Paris, France, 2018.
- [10] Green, B., & Tao, T. “The Primes Contain Arbitrarily Long Arithmetic Progressions”, *Annals of mathematics*, 481–547, 2008.
- [11] LOSTANLEN, V. *Convolutional Operators in the Time–Frequency Domain*, PhD dissertation, École normale supérieure, 2017.
- [12] LOSTANLEN, V., ANDÉN, J., & LAGRANGE, M. “Fourier at the Heart of Computer Music : From Harmonic Sounds to Texture”, *Comptes rendus Physique*, 20, 461–473, 2019.
- [13] LOSTANLEN, V., HAIDER, D., HAN, H., LAGRANGE, M., BALAZS, P., & EHLER, M. “Fitting Auditory Filterbanks with Multiresolution Neural Networks”, *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY, US, 2023.
- [14] Mathews, M. V. “The Digital Computer as a Musical Instrument”, *Science*, 142(3592), 553–557, 1963.
- [15] Pelofi, C., De Gardelle, V., Egré, P., & Pressnitzer, D. “Interindividual Variability in Auditory Scene Analysis Revealed by Confidence Judgements”, *Philosophical Transactions of the Royal Society B : Biological Sciences*, 372(1714), 20160107, 2017.
- [16] Pierce, J. R., Mathews, M. V., & Risset, J.-C. “Further Experiments on the Use of the Computer in Connection with Music”, *Gravesaner Blätter*, 27(8), 92–97, 1965.
- [17] Risset, J.-C. “Analyse des sons de trompette à l’aide d’un calculateur électronique”, *Comptes rendus de l’Académie des Sciences*, 163, série B, p. 111–114, 1966.
- [18] Risset, J.-C. “Computer Suite from Little Boy”, *Resources Ircam*, 1968.
- [19] Risset, J.-C. “Mutations”, *Ressources Ircam*, 1969.
- [20] Risset, J. C., & Mathews, M. V. “Analysis of Musical-Instrument Tones”, *Physics today*, 22(2), 23–30, 1969.

- [21] Risset, J.-C. “Pitch Control and Pitch Paradoxes Demonstrated with Computer-Synthesized Sounds”, *The Journal of the Acoustical Society of America*, 46, 1A, 88, 1969.
- [22] Risset, J.-C. “An Introductory Catalogue of Computer-Synthesized Sounds”, Bell Telephone Laboratories, 1969.
- [23] Risset, J.-C. “Musique, calcul secret?”, *Critique*, 359, 414-429, 1977.
- [24] Risset, J.-C. & Wessel, D. “Les illusions auditives”, *Universalialia (Encyclopaedia Universalis)*, rubrique “acoustique”, 167–171, 1979.
- [25] Risset, J.-C. “Stochastic Processes in Music and Art”. In *Proc. International Workshop on Stochastic Processes in Quantum Theory and Statistical Physics*, Marseilles, France, Springer Berlin Heidelberg, 1981.
- [26] Risset, J.-C. “Le compositeur et ses machines”, *Esprit*, mars, 59–76, 1985.
- [27] Risset, J. C. “The Computer, Music, and Sound Models”. In *Proc. the International Conference on Wavelets : Time-Frequency Methods and Phase Space*, Marseilles, France, Springer Berlin Heidelberg, 1987.
- [28] Risset, J.-C. “Perception, environnement, musiques”, *InHarmoniques* 3, 1988.
- [29] Risset, J.-C. “Symétrie et arts sonores”. In É. Noël (ed.), *La symétrie aujourd’hui*, Collection Points Sciences, Éditions du Seuil, 188–203, 1989.
- [30] Risset, J.-C. “Additive Synthesis of Inharmonic Tones”, in Mathews, M. V. & Pierce, J., *Current Directions in Computer Music Research*, MIT Press, 1989.
- [31] Risset, J.-C. “Composer le son : expériences avec l’ordinateur, 1964–1989”, *Contrechamps*, 11, 107–126, 1990.
- [32] Risset, J.-C. “Timbre Analysis by Synthesis. Representations, Imitations, and Variants for Musical Composition”, in Picciali, A., De Poli, G., & Roads, C. *Representation of Musical Signals*, MIT Press, 7–43, 1991.
- [33] Risset, J.-C. “Sculpting Sounds with Computers : Music, Science, Technology.” *Leonardo*, 27(3), 257–261, 1994.
- [34] Risset, J.-C. “Aujourd’hui, le son musical se calcule”, in Loi, M., *Mathématiques et arts*, Hermann, 211–233, 1995.
- [35] Risset, J.-C. “La portée d’une recherche”, *Alliage*, 40, 96–102, 1999.
- [36] Risset, J.-C. “Exploration of Timbre by Analysis and Synthesis”, in Deutsch, D., *The Psychology of Music*, 2nd Edition, Academic Press, p. 113–169, 1999.
- [37] Risset, J.-C. “Tri-IX”, Ressources Ircam, 2002.
- [38] Risset, J.-C. “Examples of the Musical Use of Digital Audio Effects”. *Journal of New Music Research*, 31(2), 93-97, 2002.
- [39] Risset, J.-C. “The Perception of Musical Sound”, *Seminar to the Acoustic Group of the University of Texas at Austin*, 2003.
- [40] Risset, J.-C. “Quelques points de vue sur l’invisible”, *Filigrane. Traces d’invisible*, 2, 89–108, 2005.
- [41] Risset, J.-C. “Fifty Years of Digital Sound for Music”, *Proc. the Sound and Music Computing Conference*, Lefkada, Grèce, 2007.
- [42] Risset, J.-C. “Musique et parole : de l’acoustique au numérique”, in Dehaene, S. & Petit, C., *Parole et musique*, Odile Jacob, 2008.
- [43] Risset, J.-C. “Du songe au son”. L’Harmattan, 2008.
- [44] Risset, J.-C. “Remarques de Jean-Claude Risset”, *Musica/Tecnologia*, 4(4), 33–38, 2010.
- [45] Risset, J.-C. “Sound and Music Computing Meets Philosophy”, *Proc. ICMC/SMC*, 2014.
- [46] Risset, J.-C. “Notice biographique”, Ircam–Centre Pompidou, 2015.
- [47] Risset, J. C. Recollections and Reflections on Organised Sound. *Organised Sound*, 20(1), 15-22, 2015.
- [48] Risset, J.-C. “Nature contre nature”, in Grabócz, M., *Modèles naturels et scénarios imaginaires*, Harmann, 2019.
- [49] Roads, C. *The Computer Music Tutorial*. MIT Press, 1996.
- [50] Selesnick, I. W., Baraniuk, R. G., & Kingsbury, N. C. “The Dual-Tree Complex Wavelet Transform”, *IEEE Signal Processing Magazine*, 22(6), 123-151, 2005.
- [51] Tao, T. “Long Arithmetic Progressions in the Primes”, *International Congress of Mathematicians*, 2006.
- [52] Tao, T. “Structure and Randomness in Combinatorics”, *Proc. IEEE Symposium on Foundations of Computer Science*, 3–15, 2007.
- [53] Tao, T., “Structure and Randomness in the Prime Numbers”, *Dean’s Seminar, University of Sydney*, 2008.
- [54] Tiffon, V. “Les musiques mixtes : entre pérennité et obsolescence.” *Musurgia*, 12(3), 23–45, 2005.
- [55] Veitl, A., “Le logiciel MUSIC V, technologie d’écriture musicale : rappels historiques et éléments d’analyse”, *Journées d’informatique musicale*, Grenoble, 2009.
- [56] Zeghidour, N., Teboul, O., Quitry, F. D. C., Tagliasacchi, M. “LEAF : A learnable frontend for audio classification”, *Proc. International Conference on Learning Representations*, 2021.