Case of Study: Complexity Based AI Sound Synthesis and Musical Score Interpolator

An Introduction to the Computational Analysis of Music

Early intellectual work around the fundamental link between aesthetics and numbers can be traced back to ancient times were some spatial proportions, such as the golden ratio or the Mayan Ahau-Can rattlesnake cross pattern, were correlated with meaning and beauty. Within the temporal domains, Pythagoreans quantified harmonious musical intervals in terms of ratios, establishing a concept which over time evolved into the well-tempered and equal-tempered scales (1).

Transposing this ancient line of inquiry into the mid 20th century, Zipf based on previous work from Pareto, Lotka and others refined a statistical technique for capturing scaling properties of human and natural phenomena, which he further applied to musical pieces by studying melodic intervals and distance between repetitions of notes (2).

Later on, Voss and Clarke by conducting a large-scale study of music excerpts they discovered that pitch and loudness fluctuations followed a Zipf's distribution, which in general terms it states that the probability of an event P(f) of rank f scales as P(f)~f^n, being n=-1 the exponent corresponding to Zipf's ideal, also referred as 1/f noise or pink noise (3).

The observed ubiquity of this kind of 1/f distributions over a vast diversity of natural phenomena, from earthquake magnitudes to extinction of biological species, suggested that music could also be understood as a complex system guided by rules that impose constrains on how structures are organized.

Moreover, Voss and Clarke by the use of computer generated sounds they conducted a set of musical listening experiments with white noise (n=0), brown noise (n=-2) and pink noise (n=-1), finding that most listeners judged the latter being far more interesting than either white (which was "too random") or brown (which was "to correlated") (4).

Those findings around the perception of some music structures, conceived as musical elements organized through a temporal cohesion, suggested that 1/f noises may have an essential cognitive role in the creative process and that they can possibly be linked to an evolutionary biological base (5).

By departing from these framework of study but also incorporating into the play a set of AI driven algorithms, we describe here the operation and use of a musical system oriented both towards the sound synthesis and composition of novel musical pieces concurrent to some explicit complexity-based mathematical parameter.

Mathematical Background

Within the mathematical field of temporal series (TS) we can study the evolution of a chosen physical observable, such as musical pitch, as the chronological sequence of points measured in successive uniform intervals in time. In this manner we can classify the typology of TS as: periodic, random or non-periodic with correlations. One way to distinguish in between these three types of TS is according to its corresponding power spectrum (PS), mainly defined as the squared of its Fourier transform (FT).

In the particular situation were a fragment of the TS has a similar PS with respect to the entire TS it is said that the TS is self-similar. In that case we have that the TS must be a scale-invariant which behaves as a power law according to PS(f)~f^beta, were f represents the frequency and beta the PS exponent. In the case were beta=0 we have white noise, for beta=-1 we have pink noise, for beta=-2 brown noise and for beta=infinity we have a a periodic TS, which in a log-log representation of PS vs. f they exhibit a power law that is translated into a line with slope equals to beta.

Additionally we can make use of another statistical measure called auto-correlation function (AC) which describes the correlation of a TS with itself as a function of time differences and can be defined according to its inverse relation to the FT of the PS. In a broader sense the AC can be though to encode the systems long-term memory that spans and affects its current behavior.

Particularly, AC it is found to decay more slowly for pink noise spectrums in comparison to white, brown noises and periodic TS. In that sense we can use the AC as a sort of complexity based parameter that describes the amount of information required to produce a TS, being the periodic TS the simplest one to produce and 1/f noises the more complex one (6).

System Description

The overall scope of our musical system is to be able to produce and explore a new set of sounds and musical scores according to a predefined complexity parameter via two kinds of AI hidden and latent space interpolation algorithms.

In that sense, our first algorithm is used to synthesize a 3 second music sample by generating interpolations between a pure 1 kHz pure sine tone (beta=-infinity), corresponding to a periodic TS with minimal complexity, and on the other side a 1/f noise (beta=-1), generated via the Timmer & Koenig algorithm and corresponding to the maximal complexity sound as stated before.

The AI algorithm used to generate this interpolation corresponds to a WaveNet-style autoencoder model that conditions an autoregressive decoder on temporal codes learned from the raw audio from both sine and 1/f waveforms (NSynth)(7). This algorithm allows for morphing in between both sounds, interpolating in timbre to create a novel sound that inherits its original temporal codes but also provides access to a timbre hidden space in between.

Once the new sample is generated (waveforms and temporal encodings shown in the top left figure), we analyze its PS and calculate its beta exponent (plot in the bottom left figure). The results for the new generated sound show two spectral peaks around the main single 1 kHz frequency and a beta=-2.11, allocated in between beta=-1 and beta=-infinity.

In a similar fashion, once we have generated our sound sample to be used by our system we move to the generation of a MIDI sequence. In analogy to the first algorithm this new MIDI sequence results from the interpolation of a simple and a complex one on each extremes. The first one is constructed as a monotonic sequence composed of the same MIDI note = 48 (C3) repeated in 25 temporal steps (figure right top), the second one which corresponds to the maximal complexity sequence is generated by the use of discrete 1/f noise Voss algorithm around MIDI note = 48 (figure right top).

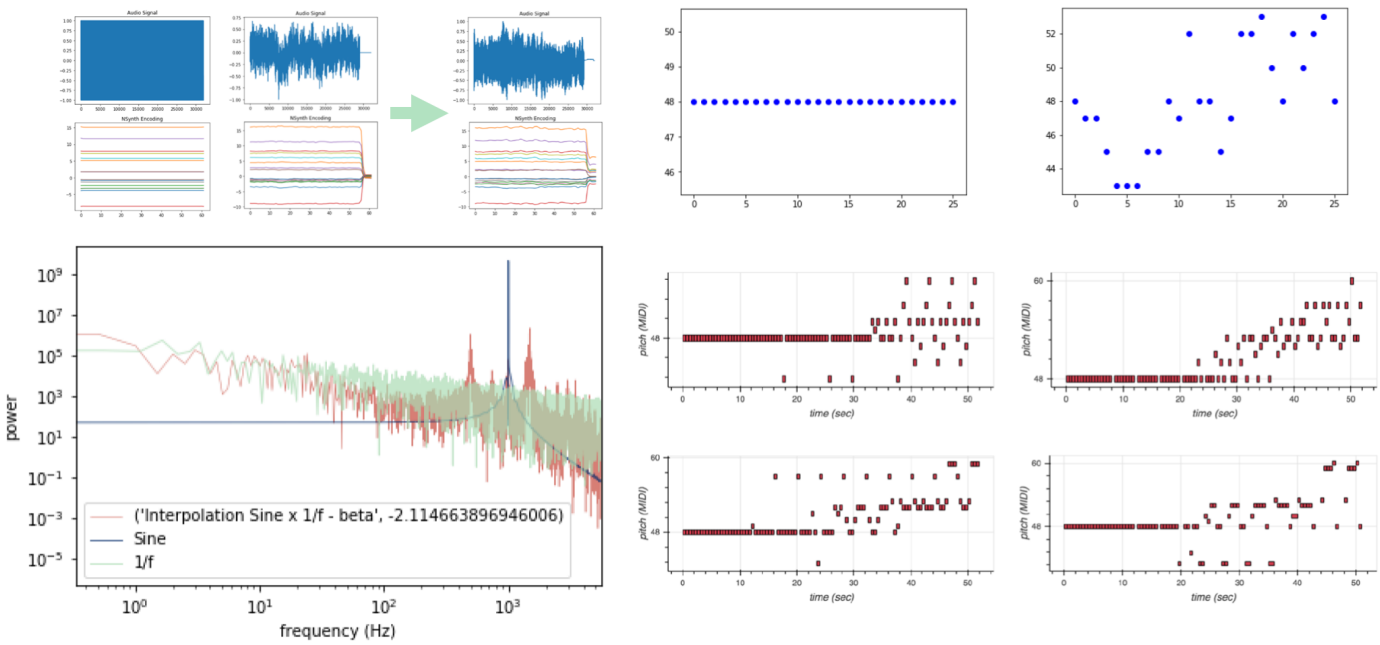
The latter AI algorithm used to generate the MIDI sequences is based on the Musical Variational AutoEncoder (MVAE) which allows to create a palette of musical scores through a Recurrent Neural Network (RNN) (8). This algorithm learns about the temporal structure of a MIDI sequence via

a hierarchical decoder and it can be used to fuse two MIDI sequences to explore latent-space interpolations of sequence extensions.

The result of the interpolation produces 4 possibles music scores with a structure that resembles the monotonic series at the beginning and later on morphs progressively into the 1/f discrete noise structure (figure right bottom).

Once the interpolated sound and score is established one can set up the overall system to reproduce its combination in a digital audio workstation. In that way, the final musical results inhabit in an hybrid simple-to-complex latent space sound territory that can be further explored with the same procedure by generating different realizations within the system.

The overall architecture of this system can be mainly explored as a new tool for music expression from one side, but moreover a subset of the generated sounds within the latent space could be further analyzed from a listener point of view, such as in the mentioned Voss and Clarke experiments, with the scope to investigate some of the intrinsic cognitive mechanisms beyond music perception.



References

(1) Manaris, Bill, Juan Romero, Penousal Machado, Dwight Krehbiel, Timothy Hirzel, Walter Pharr, and Robert B. Davis. "Zipf's law, music classification, and aesthetics." Computer Music Journal 29, no. 1 (2005): 55-69.

(2) Manaris, Bill, Dallas Vaughan, Christopher Wagner, Juan Romero, and Robert B. Davis. "Evolutionary music and the Zipf-Mandelbrot law: Developing fitness functions for pleasant music." In Workshops on Applications of Evolutionary Computation, pp. 522-534. Springer, Berlin, Heidelberg, 2003.

(3) Voss, R. F., and J. Clarke. "1/f noise in speech and music." Nature 258 (1975): 317-318.

(4) Voss, Richard F., and John Clarke. "”1/f noise”in music: Music from 1/f noise." The Journal of the Acoustical Society of America 63, no. 1 (1978): 258-263.

(5) Hsü, Kenneth Jinghwa, and Andrew Hsü. "Self-similarity of the" 1/f noise" called music." Proceedings of the National Academy of Sciences 88, no. 8 (1991): 3507-3509.

(6) Fossion, R., E. Landa, P. Stránský, V. Velazquez, JC Lopez Vieyra, I. Garduno, D. Garcia, and A. Frank. "Scale invariance as a symmetry in physical and biological systems: listening to photons, bubbles and heartbeats." In AIP Conference Proceedings, vol. 1323, no. 1, pp. 74-90. American Institute of Physics, 2010.

(7) Engel, Jesse, Cinjon Resnick, Adam Roberts, Sander Dieleman, Mohammad Norouzi, Douglas Eck, and Karen Simonyan. "Neural audio synthesis of musical notes with wavenet autoencoders." In International Conference on Machine Learning, pp. 1068-1077. PMLR, 2017.

(8) Roberts, Adam, Jesse Engel, Colin Raffel, Curtis Hawthorne, and Douglas Eck. "A hierarchical latent vector model for learning long-term structure in music." In International conference on machine learning, pp. 4364-4373. PMLR, 2018.

Short Bio

Carles Tardío Pi (http://carlestapi.hotglue.me). I am currently a Postdoctoral Research CONACYT Fellow at Synthetic and Systems Biology Laboratory from the Center for Genomic Sciences at UNAM, where I study plasmid dynamics, genetic variability and microbial evolution in spatially structured environments. I hold a PhD in Music Technology from UNAM-México, a MsC in Cognitive Systems and Interactive Media from UPF-Barcelona and a Degree in Physics from UAB-Barcelona. I have been involved in several art-science and sound projects, residencies and exhibitions and I am currently collaborating with Interspecifics collective based in Mexico City.