A prospective report on the research developed at the Laboratory of Audio and Music Technology at USP

Regis Rossi A. Faria, Ricardo Thomasi, João Monnazzi, Eduardo Bonachela, André Giolito, Gabriel Lemos [1]

[1] Laboratory of Audio and Music Technology (LATM) – School of Arts, Sciences and Humanities of the University of São Paulo (USP)
Rua Arlindo Bettio, 1000 – 03828-000 São Paulo, SP

regis@usp.br, ricardothomasi@usp.br, joao.monnazzi@usp.br, eduardobonachela@usp.br, andregiolito@usp.br, gabriel.lemos@usp.br

Abstract

This paper presents a concise report on the research developed at the Laboratory of Audio and Music Technology (LATM) at the School of Arts, Sciences and Humanities of the University of São Paulo (EACH-USP). The laboratory was founded in 2011 targeting the areas of music technology, musical acoustics and bioacoustics, strengthening its scope in 2019 to the areas of sound and music computing and audio engineering. Six projects are presented herein, describing their application areas, goals, achievements and perspectives.

1. The laboratory history and research orientation

The Laboratory of Audio and Music Technology (LATM) started its trajectory in 2011 in the Music Department at USP Ribeirão Preto campus. Counting on an infrastructure combining a technical room adjacent to a performance and recording room in a building designed for the department, the initial focus was to support under-graduation education and to promote research lines in musical acoustics and music technology at the campus.

Figure 1: A view of an event at LATM studio

A research emphasis on auralization, surround sound and musical spatialization was present from the beginning, counting with available sets of loudspeaker arrays and multichannel interfaces. In eight years the laboratory successfully implemented a series of educational activities, such as the LATM seminars on musical acoustics and audio technologies, and had several scholarship students in scientific initiation programs. Along the years it held several scientific events, including an AES (Audio Engineering Society) regional meeting on new frontiers in musical instrumentation (2014), and introduced innovative research lines, resonating with both internationally relevant and active research topics and local scientific influences, for instance on chemistry and bioacoustics.

Supporting the investigation lines and activities of its main researchers, the laboratory became home of the AUDIENCE spatial audio project and started the Reactive Music project, an innovative long-term research exploiting new sound processing concepts. Starting in 2016, a prospective collaboration with the Laboratory of Ethology and Bioacoustics (EBAC) at FFCLRP-USP opened new frontiers for research in the analysis of the vocalization of terrestrial mammals and in the design of monitoring devices and proposals for bioacoustic inference.

In 2019 the laboratory underwent a shift to another campus in São Paulo, strengthening the research orientation to sound and music computing and audio technology. The laboratory is currently working on the design of a new infrastructure to accommodate its activities, while articulates with local communities demands and collaborates with leading research laboratories at USP, such as LabArteMídia (ECA-USP), CITI-USP (EP-USP), EBAC (FFCLRP-USP), NuSom (ECA-USP), and at other universities, such as UNICAMP.

2. Lab projects

This section presents the main projects currently under development at the laboratory. Some of them are active since the foundation of the laboratory, characterizing long term research lines. Most projects have been investigating specific fronts along the years, connected to graduation and post-graduation research goals.

2.1 Acoustic-digital feedback control strategies for electroacoustic performance

This research investigates music structural possibilities in emergence theories aiming at electroacoustic performance. A fertile investigation territory has been uncovered in paradigms that includes the acoustic environment as part of the musical system. In this sense, the *Audible Ecosystem* paradigm [1] and the concept of *Performance Ecosystem* [2] were taken as analytical scope, because they drive a peculiar perspective of acoustic and digital feedback instrumentalization as a principle to sound emergent structures generation and interface of interaction as well.

Four musical studies under development, called *Ecos 1-4,* had contributed as an experimentation territory for audible ecosystems observation, for testing instrumentalization feedback techniques and musical structuring. An acoustic-digital feedback control Max/Msp patch was developed based on an adaptive multiband gain control engine (AMGC), enabling an intuitive and responsive interaction with the resonant frequencies, favoring improvisation and multiple performance situations, such as laptop music, acoustic and electric instruments, body performances and interactive sound installations (Figure 2).

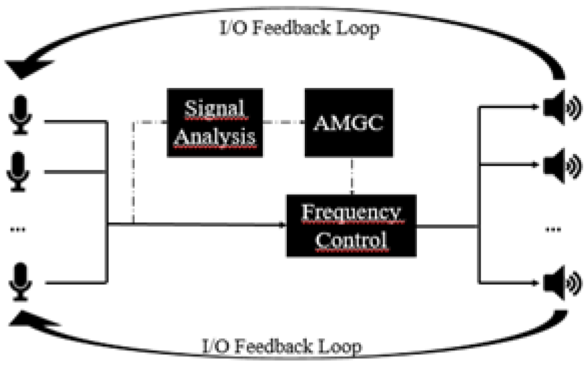

Figure 2: Digital-acoustic feedback loop control [9].

In this project we have been following uncovered strategies so far, such as (a) a new approach to multichannel sound spatialization, which in turns starts to have an impact on the emergent spectra morphology [3]; (b) other perspectives of digital filtering for sound synthesis; (c) strategies for instrument incorporation as ecosystem's part, under development in *Study for Ecosystemic Guitar* [4]*;* (d) a theoretical-analytical framework that relates the audible ecosystem's behavior and the sound patterns with a feedback and causal correlation-based taxonomy of emergencies.

From this, one hopes to draw a novel music structural thinking that may be helpful in situations where sounds are emergencies from performance interactions, and they are not known in advance [5].

2.2 Reactive Music - sound processors for reactive musical applications

A sound processor that uses stoichiometry chemical concepts metaphorically in order to produce new sounds out of the reaction between two reagent sounds is the core rationale in this project. This research is within the realm of the attractive Intelligent Musical Production research branch, towards interesting techniques and automations for sound mixing, mastering and effects.

To explore and to apply the chemistry knowledge to process real-time musical performance we have been designing, implementing and testing a suite of tools in the form of Pure Data (Pd) [6] patches, abstractions and externals.

The current package contains the following main sound processors: *timepitch* (an independent time and frequency transformer), *funew* (a band-pass filter for harmonic spectra) and *reator* (an environment to host musical reactions). Along the project phases, a number of auxiliary functions required to build the main sound processors were not available elsewhere, and were then developed, such as: *antideadlock* (a tool for consistent linkage between numbers and sliders); *faixa* (a numerical range converter), *instr* (a simple additive synth for testing processor behaviors) and *patchbay* (an output converter to several multichannel/surround formats).

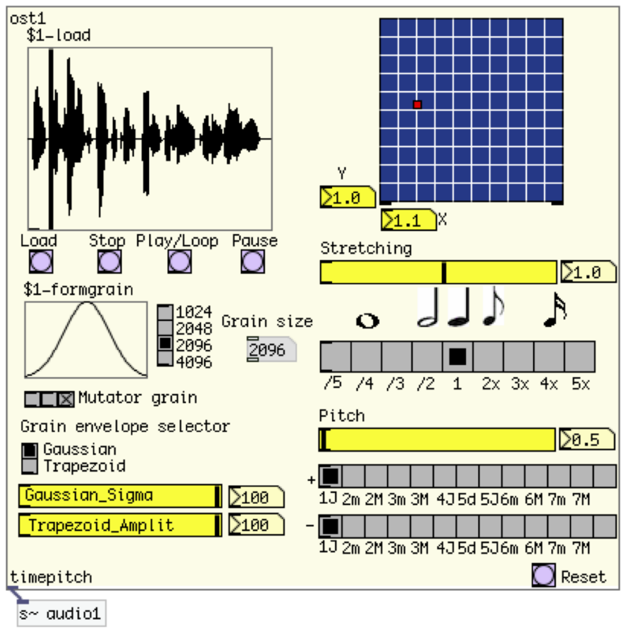

Figure 3: TimePitch sound processor

Figure 3 shows, for instance, the timepitch interface. This processor is built upon granular synthesis and can perform independent time stretching and pitch transposition with time steps in standard music note durations and frequency steps in standard music intervals.

Taking into account the progress achieved on interfaces and sound processing schemes [7], we have proposed a *musical reactor equation* that defines: (1) the quantities of each sound reagent that will react to each other, and (2) the functional reaction type.

The first type of processing explored was the mixing functional. The reactive mixing can be thought of as a mixing method in which the flow of reagent sounds is controlled according to a reaction equation, weighted by reagent parameters. We expect to use this material to advance in developing other types of reaction using ring modulation, convolution and morphing functionals. We are also aiming at building computationally effective plug-ins, since most sound processing tools require demanding computer resources.

A set of control interfaces, spectral/time processors, auxiliary objects and versions of the reactor were developed already. A prototype of the first plug-in was recently brought up using a compiled version of the Camomile VST3 effect [8] along with specific externals used in the reactor encoded in C++. It behaves as a host so as to enable the patch to become a plug-in itself.

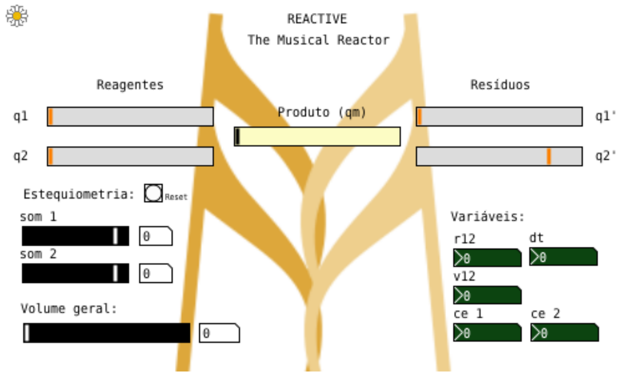

Figure 4: Reactive-camomile VST3 prototype interface.

The group is currently setting up a Github repository to provide the library distribution with free availability to benefit a community of musical artists, sound producers and anyone keen on musical technology. A next step in the project shall be a Pd-independent plug-in version, using for instance the JUCE framework and C++ encoding.

2.3 Musical Stereotypy: analysis of museme recurrency in film scoring

A reiteration of musical elements or structures that we call musemes [9] is identified in cinematographic musical tracks, which, linked to certain archetypes or moods, and reiterated over decades in similar connotative senses, gained extra-musical meanings. The characteristics of these recurrences in sound cinema, however, are not well known. We propose that these musemes are identifiable and extractable, and that this search can be optimized with the help of computational tools, such as musical and symbolic descriptors.

This project proposes an investigation on musemes reiterated in specific cinematographic genres, raising the quantity and the way in which these musical materials were used and the way in which they were linked to the images. By producing a mapping of examples and analysis of the recurrence of these musical stereotypes, it is intended to provide support material for audiovisual professionals and researchers for the study and creation of sound narrative in Western cinema.

It is our hypothesis that Photoplay Collections recurrently make use of themes, topics and musical structures associated with moods, archetypes and scenes, which we call musemes, and can thus be considered the first musical materials created specifically to accompany cinematographic presentations and with a defined intention of contribute to the filmic narrative. Following this perspective, we assume that the musemes in the cinematographic film scoring originated in musical collections, with the focus of the project being an investigation of the occurrence and reiteration of these structures in a defined cut of original musical tracks, created by several renowned composers of Hollywood cinema.

We propose an unprecedented tracking of cinema musemes using the method of interobjective comparison [10] aided by symbolic and musical descriptors, where we relate musical structures, their similarity and uses in the scene. We aim to track, qualify and quantify the recurrence of certain musemes, associated with specific archetypes and moods, within sound movie years (1930 to 2020), from a selection of films proposed for analysis.

Our specific objectives are defined strategies for museme description; toolkit development and validation for museme identification and characterization through music and symbolic descriptors, evaluate the currency of musemes appearance on film scorings.

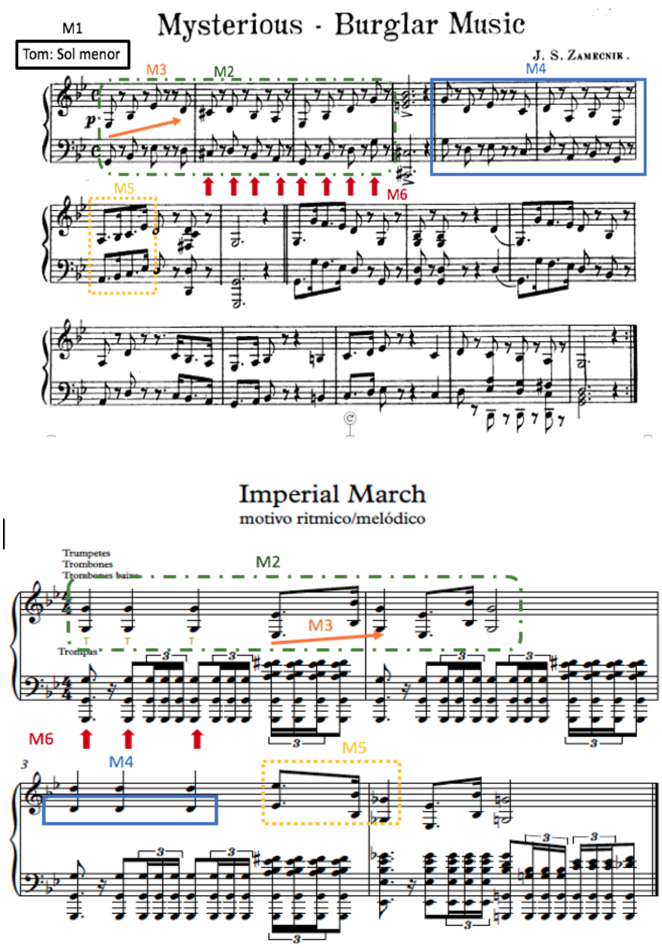

Figure 5: Interobjective comparison between the themes Mysterious (1914) (top) and Imperial March (1970) (bottom).

So far, we analysed four photoplays collections and tracked six musemes that can be linked to the shadow archetype [11]. Through a manual analytic method we found that musemes in the *Imperial March*, music theme of Darh Vader from Star Wars. This preliminary result shows that our hypotheses make sense and now we are gonna enlarge our number of analyses by using the symbolic descriptors Music21 [12] and JSymbolic [13]. By using these tools to improve and accelerate our analytic method, we expect to cover a number of film scores that allows us to have data that can answer us if there is a reiteration of the musemes found in the photoplays and, if so, what is the frequency in which they appear on movies.

2.4 Bioacoustic inference - sound analysis for animal identification and respiratory-based audio diagnosis

In the context of COVID-19 pandemic, fast and at-home identification of symptoms that require medical care by non-invasive diagnosis and without social exposure can be a very useful tool to ensure quick action in potentially severe cases of respiratory problems.

In this project, we seek to research descriptors and schemes for inference in bioacoustics to analyze recordings of respiration of healthy individuals infected by the new coronavirus made by cell phones at different stages of infection, and to verify the emergence of respiratory patterns indicative of attention. One step is to validate the effectiveness of analysis and classification techniques for differential audio-diagnostics and then, at a later stage, propose a tool such as a website or application for widespread use in society.

Original recordings exhibit different qualities, noise floor, velocities and number of cycles, and it is required a normalization procedure prior to proceed to content analysis. We have been working on such a pre-processing for a number of respiratory recordings.

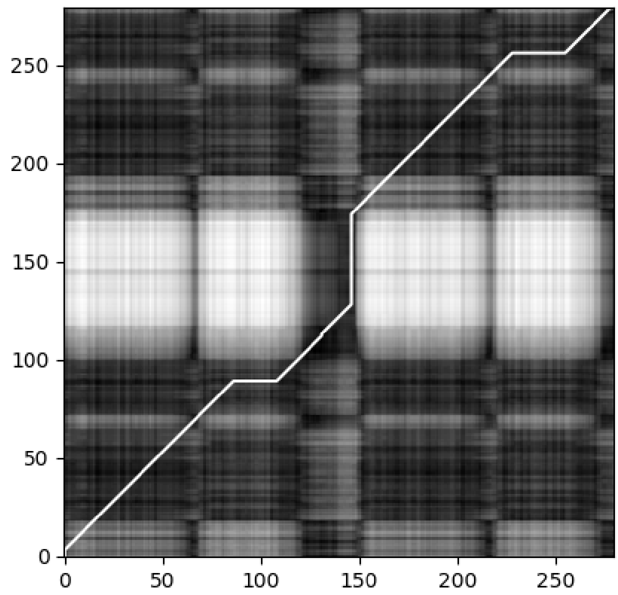

Figure 6: Comparison using MFCC + DTW between a normal respiratory sound (vertical axis) and one with crackle noise (horizontal axis)

After this step, we have been investigating the use of Mel-frequency cepstrum coefficients (MFCC) [14] and Dynamic time warping (DTW) [15] to analyze the sounds and evaluate its efficacy for comparison and classification. A library of reference sounds include normal respirations, and disease examples, such as Littmann Library [16] and ThinkLabs [17]. Figure 6 shows the maximum cost matrix between two respiratory sounds, and shows a white line indicating the shortest path between the two sequences, which can be an element that would allow classifying the data.

2.5 AUDIENCE - technologies for immersive audio (2D/3D sound)

AUDIENCE and OpenAUDIENCE are versions of a spatial audio software library being developed since 2005. The initial project and the current distribution of the programs is hosted at www.lsi.usp.br/audience. Since then, its unique functional architecture has been inspiring subsequent research projects, gathering a number of collaborators in what can be seen as a systematic effort to investigate and develop technologies for the production and reproduction of 2D/3D audio and immersive sound for applications in music, acoustics, audio engineering and bioacoustics [18] [19] [20].

Its development has been done in "periodic waves", with each phase addressing one spatial audio technology or a problem to solve. First wave addressed the Ambisonics in CAVEs and open-spaces, issues on source localization and directivity, then exploring the codecs integration such as MPEG SAOC and MPEG Surround, then sound scene auralization for musical and architectural scenes, multichannel and microphone array-based recordings, then wave-field synthesis and lately addressing 360/3D sound scene production and rendering using HOA and object-based strategies.

The system operation is based on concentrating all operations of the spatial sound production and reproduction chain in 4 functional layers, and all processes and functions follow this design architecture.

Resulting hardware products for spatial audio include: our own-developed Wave Field Synthesis (WFS) auralization rig with 18 linear loudspeakers, hosted at the partner-lab NEAC (Audio Engineering and Coding Center at EP-USP); a bi-hemispheric 360º sound scene capture board to which 2 portable recorders are attached; and a 3D microphone prototype shown at Figure 7.

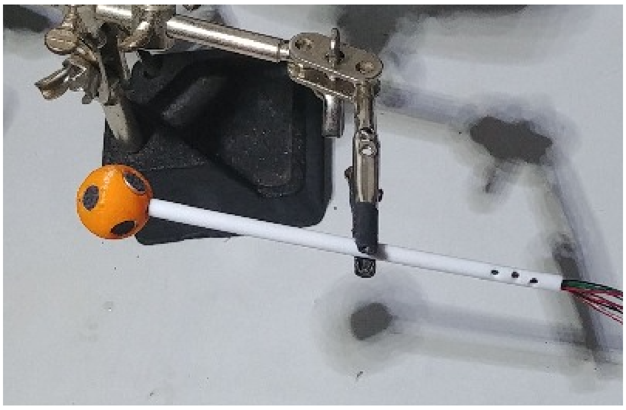

Figure 7: A 3D microphone array prototype with 6 orthogonal capsule

Software developed for spatial audio include the OpenAUDIENCE library: a whole distribution of patches and objects for sound immersion, auralization, and surround sound in Pure Data (Pd) [21]; solutions for the production and reproduction of spatial sound scenes (SSS); a WFS auralization engine; Figure 8 shows the main interface of the Spatial Sound Scene producer module.

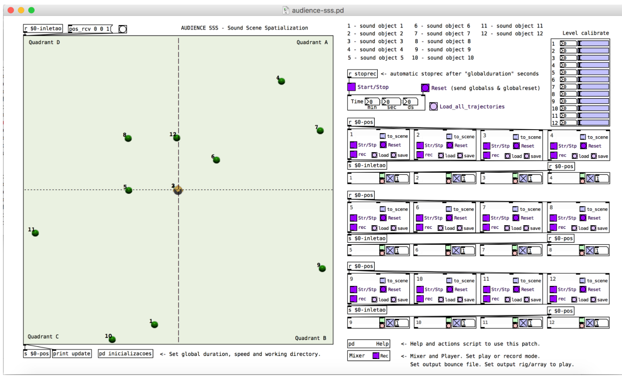

Figure 8: The Spatial Sound Scene renderer module

The community of spatial audio research is used to change its approach target from time to time, sometimes revisiting prior topics under a different convention, adding more knowledge and new products to them. Recently developments on higher order Ambisonics (HOA) and object-oriented audio scene rendering have been incorporated in 360 degree audiovisual media and adopted by the market, as in Facebook and YouTube.

Together with LabArteMidia (ECA-USP) we have been exploring the making-of for such spatial narratives and developing techniques for sound scene production and for rendering such immersive media formats. This scenario has been strongly influenced by extended reality (XR) and pulled by MPEG standards, such as the MPEG-H (3D Audio) and the forthcoming MPEG-I, targeting immersive and augmented reality. This has shown to be a trend for the new generation of digital TV worldwide, including in Brazil [22].

2.6 The spectrum in the machine: use of neural networks for sound synthesis in contemporary compositional contexts

Considering the growing use of machine learning in digital artistic practices in the last decade, especially its improvements in complex sonic style transfer techniques and sound synthesis, this research proposes a comparative study between different topological structures of Neural Networks applied to music generation processes in the waveform domain.

We propose an evaluation of the state of the art of these technologies in the field of contemporary sound creation, and we aim to identify the technical limitations and aesthetic possibilities for the application of neural networks in multimedia composition contexts. The method considers the elaboration of authorial artistic processes guided by successive generative experiments - especially in the implementation of Recursive Neural Networks (RNN) and the Generative Adversarial Networks (GAN) models [23].

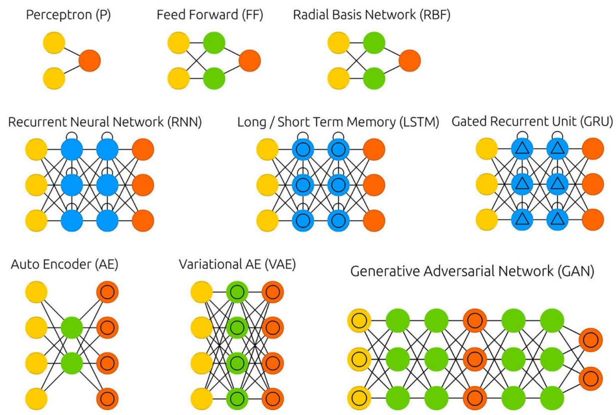

Figure 9: Graphic representation of Perceptrons, RNN and GAN topological structures

An expected result is to document in detail the creative process and the functioning of this technology so that, at the end of the research, we can make available programs dedicated to music synthesis using open-source software and code libraries that could then be accessed online and used by active developer communities.

The relevance of the research in the implementation of these models in the field of sound creation for the Brazilian context focuses on the study of the adequacy of machine learning techniques in the music synthesis process and its broader political and economic implications for the use of this technology.

The project is currently in its initial state, mainly focused in bibliographic revision for case studies and study of different end-to-end architectures based on RNN and GAN generative models implemented on Python. The implementation experiments will also be supported with additional computer resources by the Center for Artificial Intelligence (C4AI/Inova-USP) [24], in partnership with the activities held by the GAIA - Artificial Intelligence and Art Group (Inova-Usp) [25].

For the near future, we will be implementing specific networks aiming for practical experimentations in compositional contexts.

3. Reflections and future trends

The diversity of themes and convergence in strategic areas of cultural, artistic, scientific, ecological and health interest has been a relevant characteristic of the group. The projects are aligned to the lab's main research topics legacy, which act as attractors, but are naturally connected to individual inspirations and interests brought by our researchers in post-graduation (masters and doctorate) or under-graduate (scientific initialization) programs. We shall acknowledge the importance of collaborating with other laboratories and centers, which is the actual key to real multidisciplinary research, inspiring applications, joining complementary competence and infrastructure.

Techniques for context-oriented sound analysis. The approach named 'bioacoustics inference' is a systematic pursuit of descriptive and classification metrics using multivariate analysis aiming to support ecologic demands about monitoring biomes and acquiring natural soundscapes information. Applications to support medical diagnosis for health. A new cicle has begun for explorations on music synthesis, beyond previous generations works on modular analog combinations and algorithmical deterministic digital synthesis. Frontiers for sound scene individualization, immersive audiovisual production, and 3D immersion with microphone and loudspeaker arrays, extended reality.

The current laboratory website is available at sites.usp.br/latm/ and it also has a Github repository at github.com/latm-lab/, which is to be one of the portals to disseminate the lab products. The knowledge sharing is also supported by the NEAC website at www.lsi.usp.br/neac.

References

[1] A. Di Scipio. Sound is the interface': from interactive to ecosystemic signal processing. *Organised Sound , 8(3)*, 269–277. Cambridge University Press, 2003.
[2] S. Waters. Performance ecosystems: ecological approaches to musical interaction. *Electroacoustic Music Studies Network.* De Montford/Leicester: EMS, 2007.
[3] R. Thomasi and R. R. A. Faria. Moving along sound spectra: an experiment with feedback loop topologies and audible ecosystems. *International Computer Music Conference*. Santiago: ICMC, 2021.
[4] R. Thomasi and F. Kozu. Study for Ecosystemic Guitars: The electroacoustic improvisation in the sound emergence minefield. *The 21st Century Guitar: Unconventional Approaches to Performance, Composition and Research.* Lisboa: NOVA, 2021.
[5] R. Meric and M. Solomos, Agostino Di Scipio's music: emergent sound structures and audible ecosystems. *Journal of Interdisciplinary Music Studies, 3 (1,2),* 57-76, 2009.
[6] PUCKETTE, M. "Pd Documentation". http://crca.ucsd.edu/~msp/Pd_documentation/. Access: 23/07/2021.
[7] FARIA, R. R. A.; CUNHA JUNIOR, R. B. ; AFONSO, E. S. Reactive music: designing interfaces and sound processors for real-time music processing. In: Proceedings of the 11th International Symposium on Computer Music Multidisciplinary Research (CMMR 2015), Plymouth, 2015. p. 626-633.
[8] P. Guillot. Camomile: Creating audio plugins with Pure Data. In: Linux Audio Conference, Berlin, Jun. 2018. Available at: https://hal.archives-ouvertes.fr/hal-01816603. Access: July 2021.
[9] P. Tagg and B. Clarida. Music's Meaning: a modern musicology for non-musos. Nova York: MMMSP, 2013.
[10] P. Tagg and B. Clarida. Ten Little Title Tunes: Towards a musicology of the mass media. The Mass Media Music Scholar's Press. Nova York and Montreal, 2003.
[11] C. Vogler. The Writer's Journey: Mythic Structure for Writers. New Frontier. Rio de Janeiro, 1998.
[12] M. Cuthbert and C. Ariza. Music21: A toolkit for Computer-Aided Musicology and Symbolic Music Data. 2010. ISMIR, *International Society for Music Information Retrieval*, 2010. pp. 637-642.
[13] C. Mckay and J. Cumming. JSYMBOLIC 2.2: Extracting features from symbolic music for use in musicological and MIR research. 2018. *Proceedings of the 19th ISMIR COnference*, Paris, 2018. pp. 348-354.
[14] GORDILLO, Christian Dayan Arcos. Continuous speech recognition by combining MFCC and PNCC attributes with SS, WD, MAP and FRN methods of robustness. Thesis (Masters in Electrical Engineering), Pontifícia Universidade Católica, Rio de Janeiro, 2013.
[15] W. Fu, X. Yang and Y. Wang, "Heart sound diagnosis based on DTW and MFCC," In: 2010 3rd International Congress on Image and Signal Processing, 2010, pp. 2920-2923, doi: 10.1109/CISP.2010.5646678.
[16] Littmann Library. Lung Sounds, 2020, available at: http://www.3m.com/healthcare/littmann/lung.html. Accessed on July 5 2021.
[17] THINKLABS. Digital Stethoscope, 2020, available at: https://www.thinklabs.com/, accessed on July 5 2021.
[18] Faria, R.R.A. et al. AUDIENCE - Audio Immersion Experiences in the CAVERNA Digital. In: Proceedings of the 10th Brazilian Symposium on Computer Music - Current Frameworks for Music Information Representation, Belo Horizonte, 2005. p. 106-117.
[19] Faria, R.R.A et al, "Improving spatial perception through sound field simulation in VR," *IEEE Symposium on Virtual Environments, Human-Computer Interfaces and Measurement Systems*, 2005, pp. 103-108, doi: 10.1109/VECIMS.2005.1567573.
[20] Faria, R. R. A. AUDIENCE for Pd, a scene-oriented library for spatial audio. In: Proceedings of the 4th international Pure Data Convention, Weimar e Berlin, 2011.
[21] OpenAUDIENCE library for sound immersion and auralization, v. 1.0.3 (2012), available in: http://lsi.usp.br/neac/en/openaudience
[22] Forum SBTVD, Call for Proposals: TV 3.0 Project, (17/07/2020), available at: https://forumsbtvd.org.br/

wp-content/uploads/2020/07/SBTVDTV-3-0-CfP.pdf
[23]VEEN, Fjodor Van. *The Neural Network Zoo.* The Asimov Institute, 2016. Published online at: https://www.asimovinstitute.org/neural-network-zoo/. Accessed on July 25 2021.
[24] Center for Artificial Intelligence, 2020, accessed on July 23 2021 at: http://c4ai.inova.usp.br/pt/home-2/.
[25] GAIA - Grupo de Arte e Inteligência Artificial, 2020, https://sites.usp.br/gaia/, accessed on July 23 2021.