Studying the Perception of Sound in Space: Granular Sounds Spatialized in a High-Order Ambisonics System

Danilo Rossetti
Universidade Federal do Mato Grosso, Cuiabá, MT

Jônatas Manzolli
Universidade Estadual de Campinas, Campinas, SP

Abstract: Space is an important aspect of music composition that received little emphasis in certain periods of the history of Western music. However, in the twentieth century, electroacoustic music reintroduced this element by utilizing the different spatialization methods available. In this article, we discuss the methodology of analysis and the results of a study of the perception of sound in space involving sounds generated by granular synthesis that were spatialized in a high-order ambisonics system performed at CIRRMT's Immersive Presence Laboratory at McGill University. We investigated two hypotheses of sound perception in space: 1) the variation of the granular synthesis parameters produces differences in the morphology of sounds perceived in their time-varying spatial distribution; and 2) the presence of more morphological sound variations when higher orders of ambisonics are employed, in addition to the generation of sound fields with more depth. In order to verify these hypotheses, we employed a method of analysis based on graphical representations of data derived from audio descriptors, such as graphical curves, a graphic representation of volume and phase space graphics. As a result, we observed that when higher orders of ambisonics are employed, the listener perceives more variations in frequency and intensity in sound spatialization, in addition to a preponderance of the decorrelation effect. Such effect is related to the presence of more diffuse sound fields.

Keywords: sound perception in space. computer-aided musical analysis. high-order ambisonics. granular synthesis. electroacoustic music.

Estudando a percepção do som no espaço: sons granulares espacializados em um sistema de ambissonia de ordem superior

Resumo: *O espaço é um importante aspecto da composição musical que não obteve destaque em alguns períodos da história da música ocidental. No entanto, no século XX, esta categoria foi reintroduzida pela música eletroacústica através da utilização dos diferentes métodos de espacialização disponíveis. Neste artigo, discutimos a metodologia de análise e os resultados obtidos por um estudo da percepção do som no espaço envolvendo sons gerados por síntese granular e espacializados em um sistema de ambissonia de ordem superior, realizado no Laboratório de Presença Imersiva do CIRMMT,*

ROSSETTI, Danilo; MANZOLLI, Jônatas. Studying the Perception of Sound in Space: Granular Sounds Spatialized in a High-Order Ambisonics System. *Opus*, v. 26 n. 2, p. 1-26, maio/ago. 2020. http://dx.doi.org/10.20504/opus2020b2610

Submetido em 22/11/2019, aprovado em 13/7/2020.

na Universidade McGill. Nós investigamos duas hipóteses da percepção sonora no espaço: 1) a variação dos parâmetros da síntese granular produz diferenças morfológicas na distribuição espacial dos sons percebidos ao longo do tempo; e 2) há a presença de maiores variações morfológicas nos sons espacializados quando ordens maiores de ambissonia são utilizadas, além da geração de campos sonoros com mais profundidade. A fim de verificar essas hipóteses, empregamos um método de análise baseado em representações gráficas oriundas de dados extraídos por descritores de áudio, tais como gráficos de curvas, uma representação denominada volume e gráficos de espaço de fases. Como resultados, observamos que, quando maiores ordens de ambissonia são utilizadas, o ouvinte percebe maiores variações de frequência e intensidade na espacialização sonora, além de uma presença preponderante do efeito de descorrelação. Este efeito está relacionado com a presença de campos sonoros difusos.

Palavras-chave: *percepção do som no espaço. análise musical com suporte computacional. ambissonia de ordem superior. síntese granular. música eletroacústica.*

In this article, we discuss a case study aiming to test the interaction between granular synthesis and higher-order ambisonics. More specifically, the intention is to understand what happens when grains are diffused in space and the behavior of their movement. The first question that arose was whether a listener could perceive differences in this behavior if the ambisonics orders were changed. The decision to test granular synthesis (GS) specifically is because we have been working with this synthesis technique and analyzing works composed with this technique. In a recent article (ROSSETTI; MANZOLLI, 2019a), we observed that the GS technique produces a considerable sound flux in the sounds produced. The GS produces emergent sound organizations since the small grains and their interaction in microtime result in a macrostructure (perceived sound) that has properties and features that single grains do not share. Moreover, we believe that with the ambisonics spatialization those emergent features are emphasized. It displays the grains in movement in the acoustic space, generating considerable spatial cues for the listener.

Ambisonics is a surround-sound technology originally developed in the 1970s which aims at not making loudspeakers audible as separate sources of sound and is designed to give stable and uncolored acoustic images (FELLGETT, 1975: 20). In the 2000s, ambisonics technology was extended into high-order ambisonics (HOA), which presents the "holophony" feature: the acoustic reconstruction over a wide listening area provided by many loudspeakers (DANIEL, 2009: 3). It has already been observed that an increase in orders of ambisonics permits an increase in the spatial resolution of the ambisonics representation of the sound field (DANIEL, 2001: 11).

Furthermore, space is an important manipulable category in music composition in which auditory perception is a central element to be considered. In the past, works of music explored this principle especially during the Renaissance period in which we mention the famous motet by Thomas Tallis, *Spem in Alium* (1570), where eight groups of five voices each (totaling forty voices) were placed around the listeners. On the other hand, the adoption of the proscenium stage in the following centuries as a standard for concerts and opera made this possibility more difficult as musicians and singers began to perform music in front of the audience.

In the nineteenth century, a few attempts were made by composers working with musicians that were placed in different locations inside the theaters. One of them occurs in the *Tuba mirum* of Hector Berlioz's *Requiem* (1837), where four small brass ensembles must be placed separately at the four angles of the large choral and instrumental mass (SOLOMOS, 2013: 420-421).

At the beginning of the twentieth century, composer Charles Ives was particularly concerned about the question of space. In his work *The Unanswered Question* (1906), scored

for three instrumental groups (string quartet, flute or woodwind quartet, and solo trumpet), the trumpet player is placed offstage to create a sensation of distance. Composer Henry Brant, according to Maria Anna Harley (1997: 70), continued the line of development commenced by Ives. His spatial music was built with simultaneous sound layers contrasting in spatial location and other various aspects. For Brant, space was an essential aspect of music composition and, in his compositions, space had a structural and aesthetic function. In 1992, the catalog of his works listed seventy-six spatial pieces and fifty-seven non-spatial works. His work *Antiphony I* (1953), with the orchestra divided into five groups (woodwinds, horns, muted brass, pitched percussion and strings) placed in different parts of the concert hall was one of his spatial works[1].

With the advent of electroacoustic music in the 1950s, space was also thought of as an important compositional feature with sounds diffused through loudspeakers placed in different locations inside the concert halls. The fixed-media work *Gesang der Jünglinge* (1955-56) by Karlheinz Stockhausen, originally mixed for five channels and later reduced to four channels placed around the public, is a pioneering example of this music genre (STOCKHAUSEN, 1958). In the 1970s, the Acusmonium or the *GRM* orchestra of loudspeakers reinforced this tradition (BAYLE, 1986). More recently, new techniques of sound spatialization in digital domain have been developed exploring the idea of sound immersion such as high-order ambisonics and wave-field synthesis (DANIEL, 2003).

Composers who worked with electroacoustic music also made their spatial attempts at instrumental music. In that context, we mention works such as *Gruppen* (1955-57) by Stockhausen for three orchestral ensembles placed around the public, and *Terretektorh* (1965-66) by Iannis Xenakis for 88 musicians scattered among the audience. These musical explorations led us to ask new questions concerning sound perception. For instance, when the sound sources (be they musical instruments or loudspeakers) are distributed in the space of a concert hall or of an auditorium, and not all concentrated on stage, psychoacoustic issues such as consonance and dissonance or beats between two close frequencies acquire other perceptive qualities (HARLEY, 1997: 75).

In 1958, Iannis Xenakis and Edgar Varèse composed electroacoustic music works to be diffused inside the Philips Pavilion at Expo'58 in Brussels for the multimedia spectacle *Poème électronique.* The spectacle was composed of two musical works: the homonymous work of Varèse and *Concret PH* by Xenakis, which were combined with image projections. During the spectacle, the works were diffused inside the pavilion by about four hundred loudspeakers (ROSSETTI, 2012: 155-157). *Concret PH* is a *musique concrète* work that anticipates the granular synthesis technique in the sense that it can be seen as a granular work composed in analogic domain. Its granular sonority was generated by means of tape-editing processes: the recordings were cut into very short sections and isolated from their original contexts and then reassembled in order to create new textures and grain clouds (ROSSETTI, 2012: 166-170).

This paper explores the question of sound and timbre perception in space. It discusses a study performed at the Immersive Presence Laboratory of the Center for Interdisciplinary Research in Music Media and Technology (CIRMMT) of the McGill University in Montreal, Canada in July 2018 bringing together granular synthesis and ambisonics spatialization. In this study, hundreds of granular sounds were diffused in space in an eight-channel system under a high-order ambisonics (HOA) technique of spatialization. The granular synthesis sound samples generated were recorded in two formats: the eight-channel acoustic sound result (before the sound diffusion), and a stereo

[1] A performance of *Antiphony I* by the New York Philharmonic Orchestra in 1960 preceded by an explanation by Leonard Bernstein can be found at: https://www.youtube.com/watch?v=S-CXjr4Qtag.

version recorded after the sound diffusion by a dummy-head microphone placed in the center of the diffuse sound field. After the study, Rossetti composed an eight-channel acousmatic piece employing sound material generated in this study called *Substâncias moldáveis* (2018). The relation between the study and the musical piece is detailed in a work presented at the ICMC 2019 (International Computer Music Conference) (ROSSETTI; MANZOLLI, 2019b).

The research conducted recently at the Interdisciplinary Nucleus of Sound Communication at the University of Campinas may contribute to investigate and provide answers to the hypotheses described above. In order to understand the internal qualities of grain clouds scattered in space, it is necessary to understand their spectral content and how it varies in time. The GS produces a time phenomenon called sound flux in addition to the perception of volume or the spectral richness (TRUAX, 1992). For the analysis of these elements we implemented new analytical tools and sound representations such as volume and spectral liveness (ROSSETTI; MANZOLLI, 2019a). These graphical representations were created in Max/MSP with the MuBu Library (SCHNELL et al. 2009). We also applied the tools developed at NICS in the analysis of acousmatic and live-electronic music. Our research concerned the analysis of the performance of live electronic works (ROSSETTI; MANZOLLI, 2017) and the definition of the notion of emergent timbre (MCADAMS 2013. ROSSETTI; TEIXEIRA; MANZOLLI, 2018).

These developments started at NICS in the first decade of the 2000s aiming to understand specific features of human auditory perception and to perform a classification of the works analyzed and their musical styles. The methodology adopted in these research studies included the utilization of audio descriptors in order to extract quantitative data from digital audio recordings pertaining to the MIR research field.

In this context, Pereira (2009) proposed an automatic musical signal classification based on temporal and spectral parameters, expanding a previous taxonomy proposed by Barbedo and Lopes (2006) and presenting new categories and methods for this classification. Monteiro (2012) followed the line of research started by Pereira, focusing specifically on computer-aided composition, and created a real-time audio descriptor data extraction library in PureData, the *Pdescriptors* Library, addressing harmonic, perceptual, spectral, and temporal features of sounds.

In the following years, the research at NICS employed MIR tools in musicology and music analysis. Monteiro (2016) presented an analysis of spectro-temporal representations of sounds in order to perform an automatic segmentation of audio recordings. Simurra (2016) developed a computing environment dedicated to the orchestration of instrumental extended techniques describing their sonic characteristics. That environment was also implemented in PureData using the *Pdescriptors* Library. The result of Simurra's study was the development of a tool to create an instrumental palette in order to assist the composer in the orchestration of a musical piece (SIMURRA; MANZOLLI, 2016).

In the next sections, we introduce the guidelines of the study that are followed by a theoretical review of the perception of sound in space in regard to its acoustic and psychoacoustic features. Then, we approach and discuss technical questions and the study's methodology. Finally, we provide a discussion of results of the analysis of sound perception in space.

Guidelines of the Study

Our study intends to contribute to the multichannel sound field as a compositional tool, where a collection of sound design techniques is available (LYON, 2019). Among these techniques, as examples we mention the multichannel granular synthesis (GARAVAGLIA, 2016),

the timbre spatialization (NORMANDEAU, 2009), and the techniques for spatial distribution using frequency-domain processing (TORCHIA; LIPPE, 2004).

Two observations guide our study:

1. Variations on the GS parameters produce differences in morphologies of the sounds perceived in their time-varying spatial distribution;
2. There would be more morphological sound variations when higher orders of ambisonics are employed in addition to the generation of sound fields with more depth.

The first observation is related to sound perception in space considering the methodology and the procedures that were employed, i.e., it is directly connected with the process of the GS. The morphology of a sound has to do both with its macroform (the idea of a sound object), i.e., its dynamic envelope distribution in time (attack, decay, sustain and release features) (SCHAEFFER 1966: 389-403) and with its microform, i.e., its partials or the distribution of the grains in time and frequency domains in scales of milliseconds (ROADS, 2001: 87-88).

As Vaggione (1995: 2) states, in addition to the sound object definition as a unique entity perceived, as defined by Schaeffer (1966), it can also be considered as a collection of micro-samples, as the elements of its structure are defined in a discrete manner in digital domain. Thus, from this approach, the structure of the sound object can be manipulated not only in its globality but also from its computational code acting directly in its internal structure. The digital representations employed for sound manipulation in this study are computer programs or patches that operate the GS processes and ambisonics spatialization by the utilization of objects and numerical variables in a Max/MSP patch. The patches are process-oriented abstraction functions or algorithms of sound generation and transformation devices (PUCKETTE, 2006. VAGGIONE, 2010: 60-63).

The second observation concerns the enhancement that the ambisonics technique might provide when it is applied to the GS. It is important to highlight that even if the ambisonics technique is originally employed to codify and reproduce directions and amplitudes of sounds in environment (MALHAM, 2009: 169), the aim of associating GS and HOA is not exclusively to define the positions and directions of the grains, but to produce diffuse sound fields associated with the movement, speed, and other features of the displacement of the grains. The production of diffuse sound fields in spatialization is understood as the presence of uncorrelated parts coming from surround directions corresponding to the reverberation phenomenon. This is also a sound feature explorable in ambisonics technology (DANIEL, 2003: 12). Particularly, the decorrelation of audio signals, a by-product of reverberation, is an interesting feature to be explored since it produces significant sound effects in space. The GS process tends to generate decorrelated sounds, i.e., the existence of slight differences between the sounds synthesized for the output channels impacting their spatial imagery (KENDALL, 1995: 71-75). The microtemporal decorrelation between the loudspeakers of a diffusion system is an important procedure in electroacoustic compositions. This spatial effect can be produced, in addition to the GS, by promoting very short differences in the temporal alignment of identical mono sound files (in the order of milliseconds) routed into different loudspeakers in an audio editing and mixing software program resulting in a sensation of space and movement (KENDALL, 1995. VAGGIONE, 2002).

Audio descriptors are computational tools that extract information of digital audio recordings belonging to the research field of MIR – Music Information Retrieval (PEETERS, 2004). Depending on the statistical measures employed in the calculation of the descriptors, we can extract information from different features of the audio analyzed. The calculations can give us,

among others, temporal, energy, harmonic and spectral features of the analyzed recordings. One of the first results in the beginning of this field of research was to provide a classification of musical styles or genres. Nowadays, audio descriptors are increasingly being employed in the field of scientific musicology (PARNCUTT, 2007), such as in performance analysis or electroacoustic music analysis (acousmatic or live electronic).

Review on Perception of Sound in Space

Zahorik, Brungard and Bronkhorst (2005: 411-415) addressed a study of the auditory distance perception in humans and affirmed that it is mostly influenced by acoustic cues such as intensity, direct-to-reverberant energy ratio, spectrum, and binaural cues. Intensity has long been considered the primary acoustic cue for distance: it decreases insofar as the distance between the source and the listener is increased. Under ideal conditions, this relationship is related to an inverse-square law that when the source distance is doubled, a 6 dB loss in sound pressure is identified.

Regarding the reverberation effect, for environments with sound-reflecting surfaces, the ratio of energy reaching the listener directly to energy reaching the listener via reflecting surfaces is inversely related to the distance of the sound source. These authors consider that sounds with significant reverberation have uniform energy over its sound envelope. On the other hand, close sources (mostly with direct energy) have a more varying contour of the sound envelope. In this context, Bronkhorst and Houtgast (1999) formulated a computational model of auditory distance perception that is determined mainly by three variables: virtual source distance, number of reflections, and relative level of reflections. They demonstrated that the increase in the number of simulated reflections in a virtual auditorium result in the increase in the perceived apparent distance. The model suggested that the listener combines two sources of information when estimating the distance of a sound source in a room: the energies of direct and reverberant sound.

In relation to the spectrum, Zahorik, Brungard, and Bronkhorst (2005) affirm that for distances greater than 15 m, the sound-absorbing properties of the air modify the sound spectrum, attenuating high frequencies in the order of 3 to 4 dB loss per 100 meters at 4 kHz. Finally, binaural cues play an important role in the perception of sound sources near the head improving the localization for lateral sound sources, although some ambiguity is found in the perception of distant sources.

Malham (2009: 163), similarly to Zahorik, Brungard and Bronkhorst (2005), affirms that the auditory perception of distance depends on reverberation, early reflections, intensity and its variation, and loss of high frequencies. Regarding reverberation, he affirms that the energy in a reverberation field remains considerably constant for all combinations of source and listener locations, which means that when a given source intensity is constant, the intensity of reverberation remains the same. Moreover, considering perceptual issues, Malham highlights that auditory perception is a three-dimensional sense and the sounds we consciously hear are the source objects producing sounds. Instead, the majority of things we see are more or less passive reflectors. It means that the sound sources are active participants in the environment (MALHAM, 2009: 165-166).

Kendall (1995: 71) explores the phenomenon of decorrelation of audio signals. It refers to "a process where an audio source signal is transformed into multiple output signals with waveforms that appear different from each other, but which sound the same as the sound source" (KENDALL, 1995: 71). Normally, it is connected to reverberation or chorusing in acoustic

and digital domains. Decorrelation occurs in sound synthesis, especially in granular synthesis, when small temporal differences between the output channels occur.

Kendall affirms that signal decorrelation has a dramatic effect on the perception of sound imagery and the degree of decorrelation of sounds is a significant predictor of perceptual effects in natural environments or in audio reproductions. Decorrelation has at least five effects on the perception of spatial imagery in audio reproduction: 1) timbral coloration and combing effect; 2) production of diffuse sound fields such as the late field of reverberant concert halls; 3) externalization in headphone reproduction; 4) no image shift when the listener changes position relative to stereo loudspeakers; and 5) the precedence effect[2] is defeated, so one can perceive the same sound signal from multiple loudspeakers.

In terms of compositional features of space, Trochimczyk (2001: 39) adds that if the spatialization conception may belong to the outside time category, its realization is always temporal as the sound layers originated from different spatial locations are arranged in succession and simultaneity. Space in music is never empty, absolute, or homogeneous and is revealed through the attributes of the sound material. In that context, auditory perception is an essential element of compositional spatial arrangements.

The Methodology of the Study

We employed four different kinds of mono sound-sources in the process involving GS and HOA spatialization: two acoustic instrumental sounds with defined pitch (violin artificial harmonics with tremolo and a whistle sound of the flute) and two small excerpts of granular acousmatic works by Iannis Xenakis (*Concret PH*) (XENAKIS, 1997) and Horacio Vaggione (*Schall*) (VAGGIONE, 1995) having prominent textural features and no perceivable prominent pitch.

Programming and Recording

The digital sound processing was performed in Max/MSP employing objects of the High-Order Ambisonics Library (HOA) developed by the *Centre de recherche Informatique et Création Musicale* (CICM) of Paris 8 University (SÈDES; GUILLOT; PARIS, 2014). We highlight the use of the Max/MSP object *hoa.2d.process~* that combines in one unique operation electronic treatments such as GS with HOA sound spatialization. During the study, the ambisonics orders 3, 5, 7, and 9 were addressed in an eight-channel sound diffusion field.

Inside the patch, the object *hoa.2d.process~* is placed between the operations of ambisonics encoding and decoding and the granulation is performed upon a buffered delay line (WAKEFIELD, 2006: 123-124). In this sense, the type of sound transformation and spatialization process emphasizes the generation of diffuse sound fields based on the decorrelation of the audio signals over the loudspeakers. In terms of perception, a huge movement of grains over the space of diffusion is provided, provoking considerable aesthetical interest related to the perception of sound-depth (KEARNEY et al., 2010: 2).

[2] The precedence effect is a binaural psychoacoustical effect that happens when a sound is followed by another and separated by a short time delay below the listener's echo threshold resulting in the perception of a single auditory event (WALLACH; NEWMAN; ROSENZWEIG, 1949).

We recorded the generated sounds inside the Max/MSP patch in two different manners: 1) the digital eight-channel acoustic sound produced before the sound spatialization in the room; and 2) a stereo recording of the sound diffusion in the room by a Neumann dummy-head microphone placed in the center of the diffusion space. In Figure 1, we show a picture of the CIRMMT laboratory layout during the study.



Figure 1: Loudspeakers and dummy-head microphone layout inside the laboratory.

Figure 2 shows the Max/MSP patch for the seventh-order ambisonics. The mono sound source is played by the *sfplay~* object and is subsequently transposed by the *pitchshift~* object to an eight octaves range (from four octaves above to four octaves below). In this transposition, the random change in pitch occurs every 100 msec. This procedure affects directly the frequency range of the grains produced by the GS.

The GS and the ambisonics spatialization are performed inside the sub-patch. At first, the mono sound is encoded in the seventh order of ambisonics[3]. Then, the granular synthesis is performed according to the values of the four addressed parameters (grain size, grain delay, feedback, and rarefaction rate). In the decoding process, after choosing the decoding mode (in this case the ambisonics mode) the number of loudspeakers is defined (eight). Thus, eight tracks resulting from the GS and HOA processes are generated and, finally, sent to the eight loudspeakers placed in the room.

[3] In an ambisonics system, the harmonics are functions employed in the decomposition of space. They can be circular (2D) or spherical (3D) and have as variables the azimuth, in circular harmonics, and azimuth and elevation in spherical harmonics. The harmonics depend on a degree *l* and an azimuth *m*. In circular harmonics, each degree *l* comprehends two harmonics of orders $m = -l$ and $m = l$ (the degree 0 has only a single harmonic of order 0). For the spherical harmonics, each degree *l* has $2l + 1$ harmonics, whose orders stand between $-l$ to *l*. The decomposition of a sound field in the harmonics domain is performed by the order of decomposition *N*, comprehending all degrees from 0 to *N*. Thus, in 2D, the sound field is composed of $2N + 1$ circular harmonics and, in 3D, is composed of $(N + 1)^2$ spherical harmonics (COLAFRANCESCO; GUILLOT; PARIS, 2015). From this explanation, in 2D, for an order of decomposition $N = 7$, there are 15 circular harmonics, which are (degree, order): (0, 0), (1, -1), (1, 1), (2, -2), (2, 2), (3, -3), (3, 3), (4, -4), (4, 4), (5, -5), (5, 5), (6, -6), (6, 6), (7, - 7) and (7, 7).

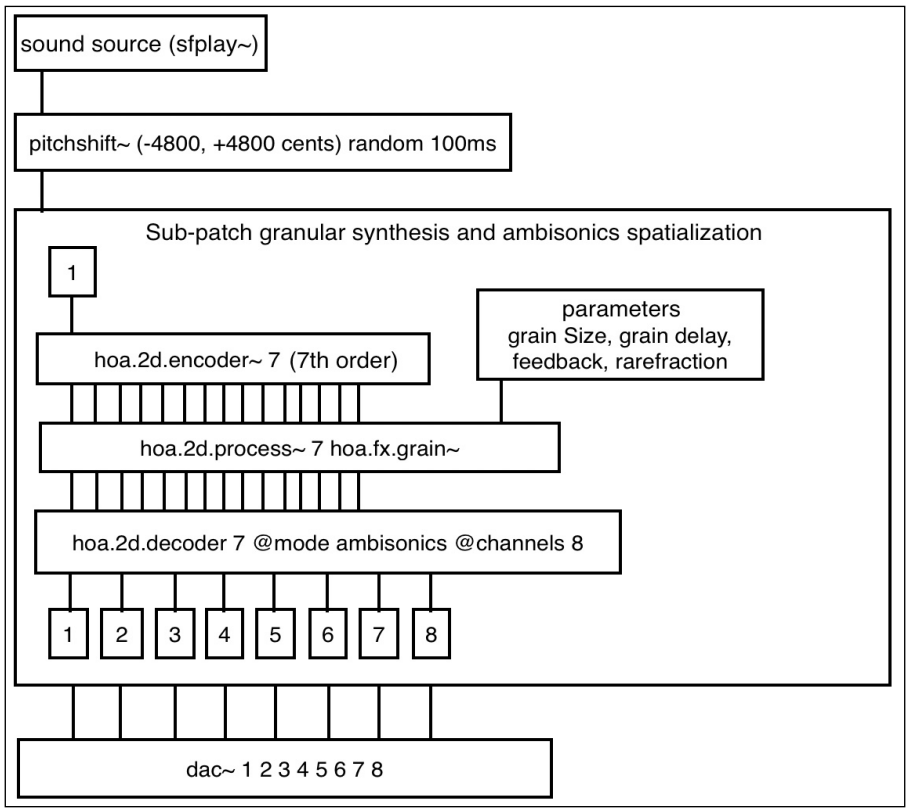


Figure 2: Diagram of the Max/MSP patch of the study.

Ambisonics and Dummy-Head Recording

The ambisonics system is a technological solution that combines the problem of codifying directions and amplitudes of sounds and reproduces them in a practical system of loudspeakers. The sound information is directionally codified in equations of spherical or circular harmonics, which are an efficient format for the manipulation of complex sound fields. The ambisonics system uses a set of signals known as *B-format* (MALHAM, 2009: 169). In this system, each loudspeaker (at least three) radiates a wave of amplitude and phase so as to reconstruct a simulacrum of the intended (recorded) surround-field approaching relevant psychoacoustic criteria. Gerzon (1975a: 24) affirms that the aim of this surround sound system is to reproduce a recording at the listener's ear accurately. In that sense, the process of encoding the sound field in ambisonics systems exists to fulfill this requirement. The first-order ambisonics encoding format consists of four spatial components associated to the following pick-up patterns: one omnidirectional (sound pressure *W*) and three bidirectional components (X, Y, Z), associated with the pressure gradient (DANIEL, 2009: 2). The decoding, in its turn, acts to accommodate the different shapes of the speaker layout to differently shaped listening rooms.

In the B-format of first order of ambisonics, the equations of codification assume that the sounds are located inside a spherical surface that surrounds the listener (with a theoretical radius of 1). If the sound moves out of the sphere, i.e., the radius exceeds the value of 1, the directional information will not be decodified correctly. Thus, the coordinates of the sound source should conform to the following rule (MALHAM, 2009: 170) as defined in Eq.1 the B-format sound source coordinates conformation:

$$x^2 + y^2 + z^2 \leq 1 \qquad \textbf{(1)}$$

where *x* is the distance within the X-axis (front-rear), *y* is the distance within the Y-axis (left-right) and *z* is the distance within the Z-axis (up-down)[4].

Further, when a mono signal is located in the surface of the sphere, the coordinates are considered according to Eq. 2 (MALHAM, 2009: 171) that describes mono signal location in the sphere surface.

$$\begin{aligned} x &= cos\theta cos\,\phi \\ y &= sin\theta cos\,\phi \\ z &= sin\,\phi \end{aligned} \qquad \textbf{(2)}$$

where θ is the anticlockwise angle of rotation from the center and ϕ is the angle of elevation upwards or downwards the horizontal plan. Figure 3 presents the conventions of ambisonics for the axes and angles.

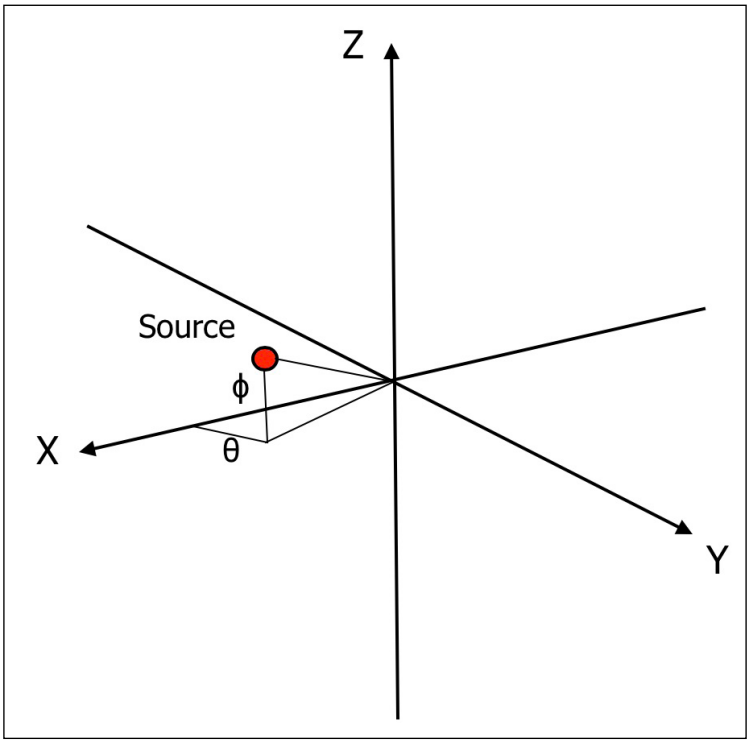


Figure 3: Ambisonics conventions for axes and angles (MALHAM, 2009: 171).

According to Malham (2009: 171), those coordinates can be directly used to produce B-format output signals (X, Y, Z and W), as described in Eq. 3 the B-format output signals for coordinates (w, x, y, z).

[4] The conventional names of the ambisonics axes are the same used in European standards for vibrational effects in the human body. They are different from the mathematical notation of Z (up-down), X (left-right) and Y (front-rear).

$$\begin{aligned} W &= (signal) * 0.707 \\ X &= (signal) * cos\theta cos\phi \\ Y &= (signal) * sin\theta cos\phi \\ Z &= (signal) * sin\phi \end{aligned} \qquad \textbf{(3)}$$

where the constant 0.707 in W, despite not being mathematically rigorous, is derived from engineering considerations, aiming to assure a relative distribution of the signal levels in each channel.

The decoding of ambisonics signals generally involves mathematical and psychoacoustic formulations. For large environments such as concert halls, Malham asserts that the better alternative is to apply the in-phase conception. In the in-phase decoding, each loudspeaker input receives a combination of B-format signals, corresponding to the position of the loudspeaker in relation to the center of the sound field. This conception is expressed by the following equation, in which the loudspeaker has a frontal anticlockwise angle α and an elevation angle β, above or below the horizontal plan (MALHAM, 2009: 181), as described in Eq. 4 the in-phase decoding related to coordinates (w, x, y, z).

$$P_{loudspeaker} = W + k(\mathrm{X}\cos\alpha\cos\beta + Y\sin\alpha\cos\beta + Z\sin\beta) \qquad \textbf{(4)}$$

where *k* is the constant that determines the directional source response and, taking into account the 0.707 factor from engineering for the channel W, employs the value of 0.707 for a directional cardioid source-response that is needed for the in-phase decoding.

Gerzon (1973) defined the spherical harmonics until the third order in employing cartesian coordinates (X, Y, Z). In a subsequent publication (Gerzon, 1977), these definitions were changed to polar coordinates r, θ, φ, which became the standard for the ambisonics system definitions. More recently, there have been new improvements in ambisonics technology, such as its extension to higher spatial resolution systems (DANIEL; NICOL; MOREAU, 2003). This extension arises by considering omni and bidirectional encoding patterns as a restricted subset of the spherical harmonics basis with a higher ability for angular discrimination. The spatial decoding basically consists of matrixing HOA signals to derive loudspeakers signals, and, for concentric and regular arrays, spatial decoding performs the discrete Fourier transform applied to the spherical harmonics domain. From these operations, HOA presents the possibility of "holophony", or the acoustic reconstruction over a wide listening area, provided by a system of many loudspeakers (DANIEL, 2009: 3).

The dummy-head recording technique (or binaural) emerged in the same period of the conception of the ambisonics system (GERZON, 1975b). It makes use of the ORTF (*Office de radiodiffusion télévision française*) technique with a pair of cardioid microphones spaced apart by 17 cm (the average diameter of a head) and angled at about 110º (the position of the ears in relation to one's head), shown in Figure 4. It is known that binaural recording gives sharper sound images than ordinary stereo. But only this miking technique is not enough to accurately reproduce the image of a recorded sound field.

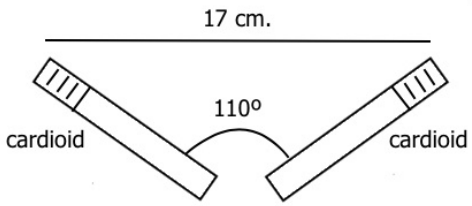


Figure 4: ORFT technique.

It is important to consider that the *pinnae* (the flaps on the ears) and their various ridges reflect and refract the sound waveform before it enters the ears. In fact, this feature plays an essential role in locating sounds (Figure 5). The effect of the ear ridges is for a sound impulse to arrive at the ear followed a few tens of milliseconds later by delayed impulses reflected off the ridges. When those delayed impulses are not detected by the ear, it assumes that the sound is behind the listener. Studies on reproducing recording sound fields have shown that when only binaural recording and no replicas of human *pinnae* are used, the subjects found it difficult to localize the sounds, mainly confusing between front and rear. This is the reason why the ORTF recording technique and *pinnae* replicas are combined in dummy-head recordings.

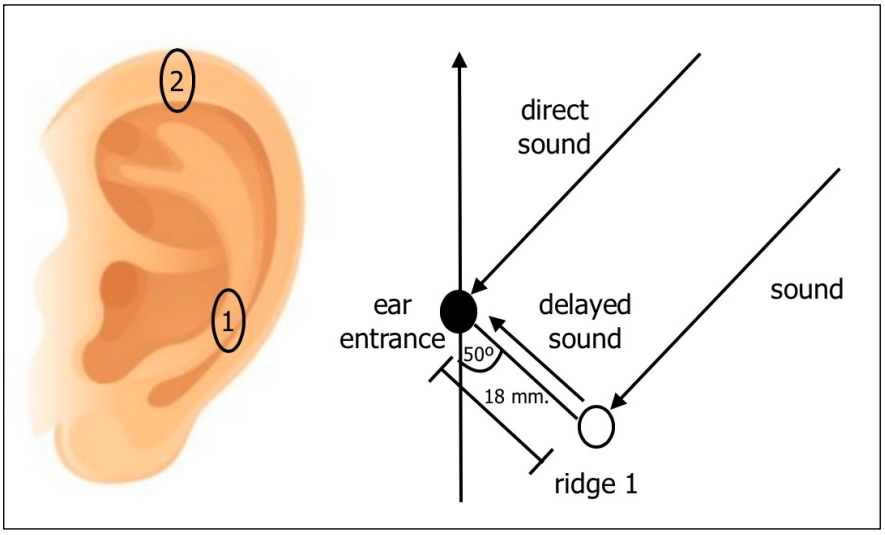


Figure 5: The pinna and reflecting ridges 1 and 2 that determine horizontal and vertical sound direction through the delay between direct and reflected sounds from the ridges (GERZON, 1975b: 44).

The Granular Synthesis Parameters and its Manipulation

Granular synthesis (GS) (ROADS, 2001: 85-118) is a kind of synthesis based on grains, a microacoustic event with a duration near the threshold of human auditory perception normally presenting a duration between 20 and 400 msec. The grains, by definition, capture two perceptual dimensions: the time-domain and the frequency-domain information, and each grain has a waveform shaped by an amplitude envelope. The GS is an automation process in which thousands of grains are combined over time, creating sound textures (clouds or masses) of different densities and frequency ranges. The textures generated are controlled by different variables and each grain has its own waveform and amplitude envelopes.

In a broader sense, the granular paradigm (that encompasses other granular techniques such as micromontage) is related to the corpuscular description of sound. The grains, which are samples of a very short duration, belong to the microtime scale. The GS consists of agglutinating a large number of grains and manipulating their global density and other morphological qualities (SOLOMOS, 2003: 396-397). We highlight the emergent feature of the GS process: the interaction of the grains in microtime results in a macrostructure that is perceived as timbre or texture with properties that isolated grains do not have. The macroform emerges from the internal activity of the grains in microtime and from the parametrical control applied to the numeric parameters (ROSSETTI; MANZOLLI, 2019a: 206).

The parameters of the GS for this study, which define the perceived timbre of the sound masses, were the grain size (msec.), grain delay (msec.), feedback and rarefaction rate[5] (from 0 to 1). The grain delay (3,000 msec.) and feedback (0.99) remained constant for all sound samples produced. On the other hand, the values of the grain size and the rarefaction rate varied according to the combination showed in Table 1 below.

Combination	Grain size (msec.)	Rarefaction
1	300	0.1
2	150	0.45
3	80	0.7
4	20	0.9

Table 1: Values of the granular synthesis parameters.

Two types of sounds with different morphologies were produced in the study. In the first type, its morphology consists of one attack, resonance, and release with a duration between 10 and 20 seconds. In this case, the parameters of the granular synthesis are fixed. In the second type of sounds, there is a gradual transition between the values of the synthesis parameters as shown in Table 1 (transition between the combinations 1 and 2, 2 and 3, and 3 and 4). Those sound samples have a duration of 30 seconds and are composed of five seconds of the initial parameters, twenty seconds of the linear gradual transition between both parameters, and five seconds of the final parameters. In Table 2 we detail the two types of sounds generated.

Type	Duration (sec.)	Characteristics (sound envelope morphology)
Sound 1	10-20	Prominent attack, resonance, release
Sound 2	30	Soft attack (fade-in), gradual transition, fade-out (release)

Table 2: Duration and sound envelope morphology of the sounds generated by the GS/HOA.

By the end of this part of the study, we produced 112 recorded sound samples from four different sound sources, four different orders of ambisonics, and two different sound morphologies. Considering the theoretical review approached, we performed a preliminary analysis with the following audio descriptors: loudness, spectral centroid, and spectral spread.

[5] 0 means a totally continuous sound, and 1, silence. The values between 0 and 1 express different levels of discontinuity among the grains.

The second part of this analysis consists of building curves, graphics, and other representations based on these descriptors.

Analysis of the Results

We present the results of the analysis of the sounds produced in the study. We will focus on the analysis of the generated sounds coming from the flute source played with the whistle technique in G#6 pitch. In the first analysis, we discuss the spectral content of the sounds in different conditions of the granular synthesis (different values of the variables) according to their development over time based on the volume representation. In the second analysis, we compare the amount of variation in terms of the perceived intensity and frequency region of brightness of sound 1, spatialized in the third and seventh orders of ambisonics. For this analysis, we employ the loudness and spectral centroid curves and phase space graphics, in addition to a comparison between their averages and standard deviations. In Figure 6, we present the waveform and sonogram of the flute sound sample, source of the GS, performed in the Sonic Visualiser software.

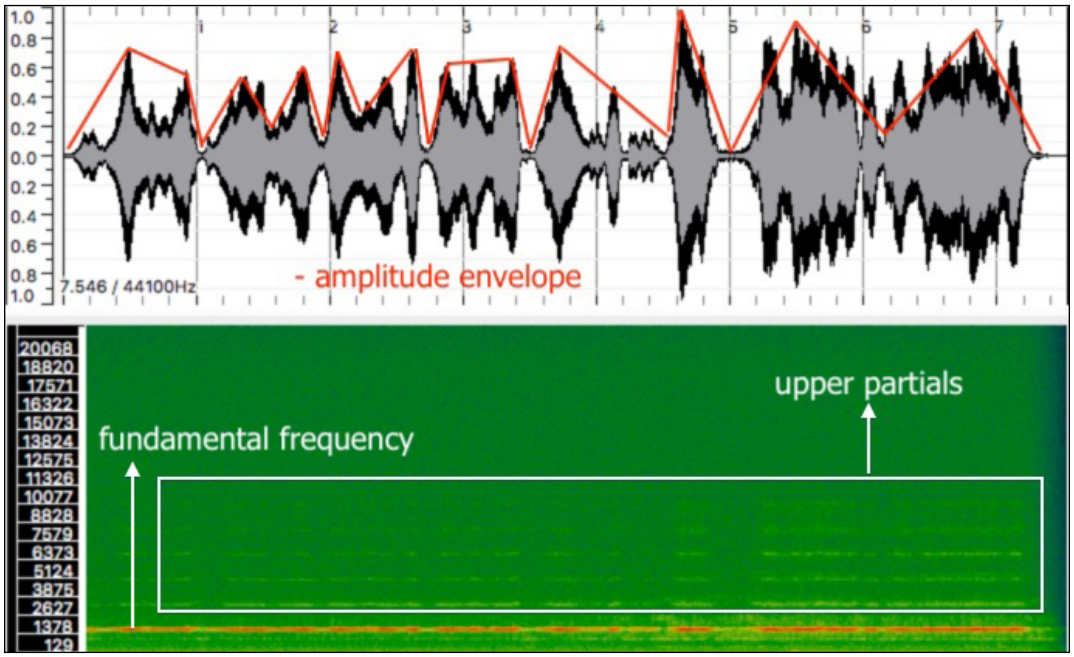


Figure 6: Waveform and sonogram of the sound sample of the flute played with whistle technique.

The sound sample of the flute has the pitch of G#6 situated in the high register of the flute with a duration of about 7.5 seconds. Although the flute is played in this region, the sonority achieved is considerably soft. In the sonogram (Figure 6, in green at the bottom), we can observe the partials of the flute distributed proportionally emphasizing the pitch perception. The fundamental frequency is around 1,660 Hz and we can observe five other upper partials in respect to the harmonic series. In the waveform representation (Figure 6, at the top), we observe that even if the whistle technique provides a stable pitch, the amplitude variation is quite considerable and noticeable. We believe that a sound with these characteristics can provide interesting sound clouds and textures when applied as a source for the granular

synthesis. Moreover, when associated with the HOA spatialization technique, we intend to analyze and discuss the spatial behavior of the grains if different orders are applied.

Analysis of the Sound Morphology Produced by the Granular Synthesis

For this analysis, we apply the volume sound representation based on the definition of BStD (brightness standard deviation) by Malt and Jourdan (2009), a representation of the sound texture evolution. We correlate this graphical representation with the concept of volume originally defined by Truax (1992: 32-35) as the perceived magnitude of a sound or the "psychological" space it creates and occupies. The volume is dependent on the parameters of spectral richness (dependent on frequency and intensity), temporal density (temporal spacing of independent spectral components such as multiple sound sources and phase-shifted or time-delayed events), and on the development of the sounds in time.

The graphical representation of the volume is constructed by the superimposition of three curves generated from data of the audio descriptors: the spectral centroid defined as the barycenter of the sound energy distribution of its spectral envelope and perceptually related to sound brightness (in Hz); the spectral spread, or the spread of the spectrum around its mean value; and loudness, being the psychoacoustic measure related to the perception of sound amplitude (in dB). We have introduced the graphical representation of volume from a musicological analysis in a previous article (ROSSETTI; MANZOLLI, 2019a: 219). Volume representation was built in Max/MSP with objects from the MuBu external Library.

In this article, we discuss two sounds of the second type which have a gradual transition between the GS values. The sounds addressed are sound A, which is a transition of combinations 1 and 2 (values informed in Table 1), as shown in Figure 7; and sound B, a transition between combinations 3 and 4, as shown in Figure 8. Both stereo versions were recorded by the dummy-head microphone. The window size of the FFT (Fast Fourier Transform) employed for the calculation of the values of the descriptors is 8,192 samples with a hop size of 1,024 samples. Both sounds are spatialized in the seventh order of ambisonics.

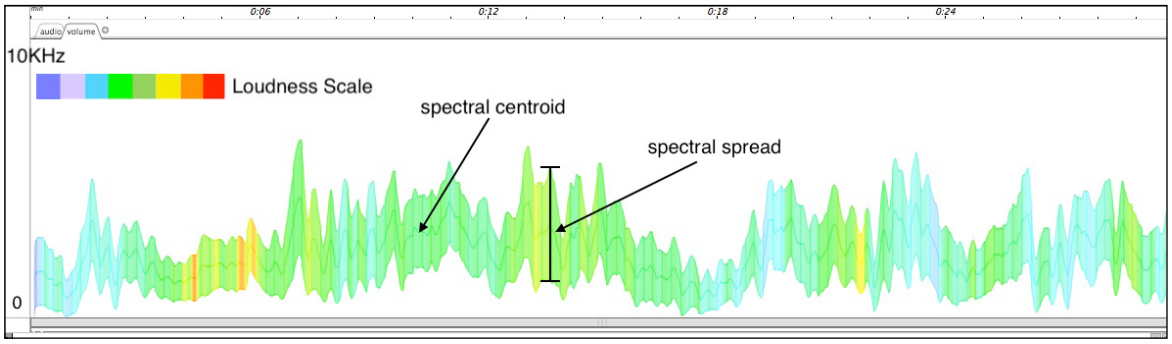


Figure 7: Volume of Sound A. Dummy-head recording of the GS sound: transition of combinations 1 and 2.

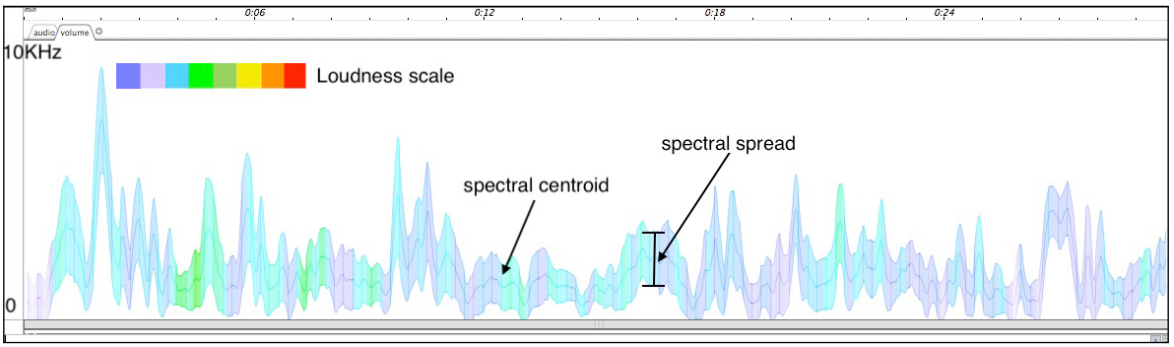


Figure 8: Volume of Sound B. Dummy-head recording of the GS sound: transition of combinations 3 and 4.

The volume of sound A, which was generated by a gradual transition in the GS process with values from 300 to 150 msec. of the grain size and 0.1 to 0.45 of the rarefaction rate, presents a considerably dense sound texture with medium sound intensity. The spectral centroid values, in Hz, indicate the frequency region that concentrates the brightness perception of the timbre, while the spectral spread values (also in Hz), summed and subtracted from the spectral centroid values, indicate the frequency range of the variation of the sound texture. Higher values of the spectral spread indicate a higher volume area and, consequently, a timbre with a denser and richer spectrum over time. The loudness level, shown by the color of the volume area, indicates a timbre with a medium intensity sensation (green color). At the beginning and at the end of this sound, a lower intensity sensation is perceived, shown by the light blue color. At around 5 sec., the intensity sensation reaches its highest level, evidenced by the yellow and orange colors.

The volume of sound B, which was generated by a gradual transition in the GS process with values from 80 to 20 msec. of the grain size and 0.7 to 0.9 of the rarefaction rate, presents a lighter sound texture with a medium-low intensity sensation. The brightness sensation of the timbre in terms of frequency, shown by the spectral centroid curve, reaches a high peak in its beginning (at around 3 sec. – close to 7 kHz) and continues to reach minor peaks during its development. In general, the values of the spectral spread are smaller than the values found in sound A, showing a timbre with a lower volume. This is shown by the narrower colored area in the volume representation (mainly with light blue and lilac colors), reflecting in a perceived light sound intensity. The lighter sound texture of sound B can be explained by the smaller grain size employed in the GS consisting mainly of the attack transients of the sound source and also by the higher rarefaction rates which produce sounds with more spaces between the grains inside the frequency range, in other words, the presence of less dense or emptier textures in terms of grains and partials.

Presence of Morphological Variations and Diffuse Sound Fields

Here we compare two stereo sound samples of the flute, both as transitions of combinations 1 and 2 (Table 1) of GS variables with the duration of 30 sec. (sound A) and recorded with the dummy-head microphone. One of the sound samples is spatialized in the third order and the other in the seventh order of ambisonics. The aim of this analysis is to verify the accuracy of the sound spatialization and the production of diffuse sound fields by different HOA orders. For this analysis, we apply the spectral centroid and loudness descriptors in order to have information on the variation of the region of the brightness perception in terms of frequency (spectral centroid)

and the variation of perception of intensity (loudness). The choice of spectral centroid and loudness is related to the features they describe. These features are relevant in the perception of sound in space, as discussed by Zahorik, Brungard and Bronkhorst (2005) and Malham (2009).

In Figures 9 and 10, we present the curves[6] of the descriptors applied to sound A spatialized in the third order of ambisonics (Figure 9) and the same sound spatialized in the seventh order of ambisonics (Figure 10) combined with an analysis of the peaks and valley values found. For the FFT calculation of the descriptors we defined the window size at 16,284 samples and the hop size at 2,048 samples.

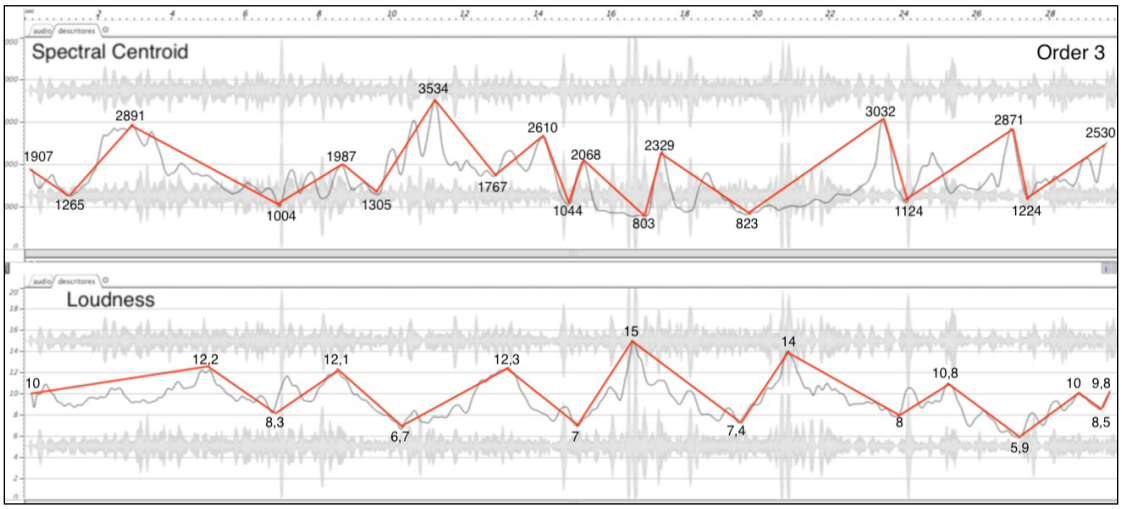


Figure 9: Spectral centroid and loudness curves of sound A spatialized in the third order of ambisonics.

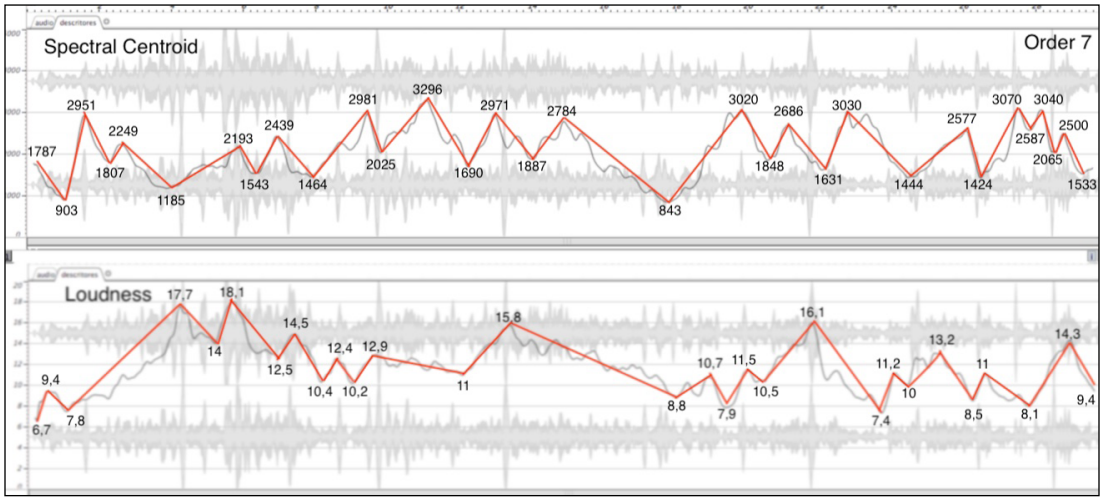


Figure 10: Spectral centroid and loudness curves of sound A spatialized in seventh order of ambisonics.

In Table 3, we summarize the maximum and minimum values found for the descriptors, their ambitus, and the number of peaks and valleys found in the curves of sound A spatialized in the third and seventh orders of ambisonics.

[6] The waveforms of the dummy-head recordings of different orders of ambisonics (Figures 9 and 10) may be slightly different. This can occur due to the random factor defined in the pitch-shift object of the GS process (see Figure 2).

		Maximum	Minimum	Ambitus	Valleys/Peaks
Spectral Centroid	3rd order	3,534 Hz	803 Hz	2,731 Hz	9
	7th order	3,296 Hz	843 Hz	2,453 Hz	15
Loudness	3rd order	10	5.9	4.1	8
	7th order	18.1	6.7	11.4	14

Table 3: Maximum, minimum, ambitus, and number of valleys and peaks of spectral centroid and loudness curves of sound A in the third and seventh orders of ambisonics spatialization.

By observing the data in Table 3 regarding the spectral centroid, we notice that although the ambitus of the third order of ambisonics is higher than the seventh order (2,731 Hz against 2,453 Hz), the variation of the spectral centroid values of this order is lower. This information is given by the number of valleys and peaks perceived in the thirty-second development of sound A. For this sound spatialized in the seventh order of ambisonics, we find fifteen valleys and peaks while there are nine valleys/peaks in the spatialization in the third order. This variation is also visible in Figures 5 and 6 by the line configuration which presents more variations in the spectral centroid curve of the seventh order.

Regarding the loudness curves of sound A spatialized in both third and seventh orders of ambisonics, we find a huge difference in their ambitus: the spatialization in the seventh order produced an ambitus of 11.4 whereas the spatialization in the third order produced an ambitus of 4.1. As in the spectral centroid curve, we find more variation in the loudness curve of the seventh order ambisonics spatialization. This is shown by the number of valleys and peaks found (14 against eight). Thus, in the sound spatialized in the seventh order, there is more variation in terms of the perception of sound intensity during the thirty seconds of sound A.

After the definition of the curves of the spectral centroid and loudness descriptors, we calculated the average and standard deviation of the values of sound A spatialized in the third and seventh orders of ambisonics. In Figure 11, we show the histograms that represent those values. The average is shown in black and the standard deviation in grey. The spectral centroid of sound A spatialized in the third order of ambisonics presents an average of 1,599 Hz and a standard deviation of 537 Hz. The same descriptor for sound A spatialized in the seventh order of ambisonics has an average of 2,106 Hz and a standard deviation of 533 Hz. The average of loudness of sound A spatialized in the third order of ambisonics is 9.34 and the standard deviation is 1.63, while the average of the same descriptor for sound A spatialized in the seventh order of ambisonics is 11.64 and the standard deviation is 1.91.

According to the information obtained, sound A spatialized in the seventh-order ambisonics has a brightness frequency region situated in a higher region than sound A spatialized in the third order of ambisonics. The standard deviation of the spectral centroid values of both sounds is very similar, although similar frequency deviation of values represents different musical intervals depending on the frequency region they are situated in (higher frequency regions need higher standard deviation values to produce the same musical intervals than lower frequency regions).

As to the loudness values, sound A spatialized in the seventh order of ambisonics has an average higher than the same sound spatialized in the third order of ambisonics, also with their standard deviation. This means that this sound when spatialized in the seventh order of ambisonics achieves higher intensity levels. Also, the intensity level is subject to more variations of sound pressure reflecting on more variations of sound intensity perception.

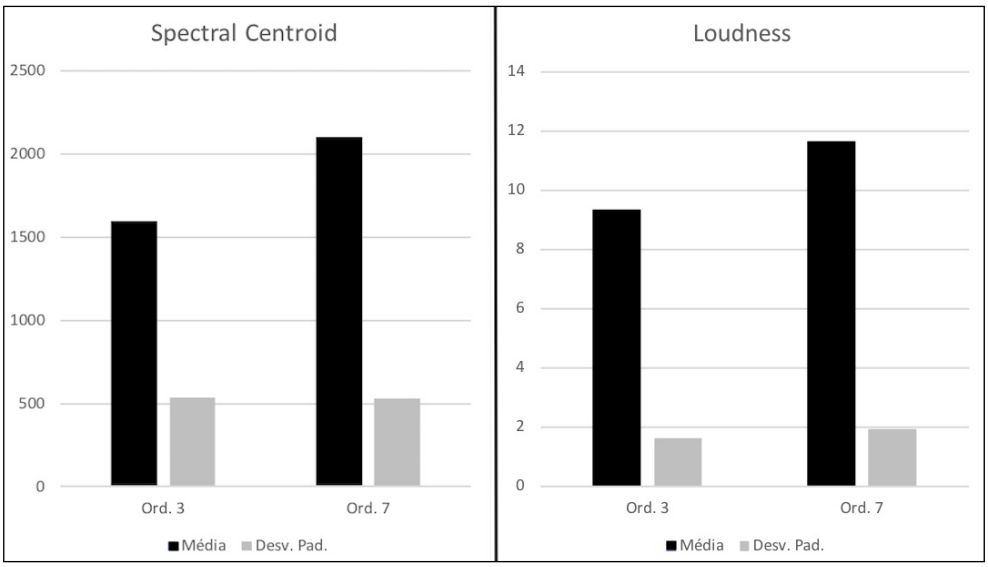


Figure 11: Average and standard deviation of spectral centroid and loudness values for sound A spatialized in third and seventh orders of ambisonics.

Finally, we present, in Figures 12 and 13, two phase space graphics calculated in Max/MSP with the MuBu library representing the values of spectral centroid (X-axis) and loudness (Y-axis). Figure 12 graphic addresses the configuration of the points of the third order of ambisonics and Figure 13 shows the configuration of the seventh order of ambisonics. The objective of this representation is to observe possible regions of accumulation or dispersion of points in order to understand the behavior of those descriptors in an out-of-time representation. The X-axis comprises values between 0 and 5,000 Hz, and the Y-axis comprises values between 0 and 20. The window size chosen for this calculation was 2,048 samples and the hop size was 1,024 samples.

We have already introduced the use of phase space graphics in music analysis in a recent work, discussing the spectral liveness tool that is calculated with spectral flux and spectral flatness descriptors (ROSSETTI; MANZOLLI: 2019a: 209). In another article, we constructed phase space graphics using the same descriptors as shown in Figure 12 (spectral centroid and loudness) (ROSSETTI; ANTUNES; MANZOLLI, 2020). As explained in this article, we consider music as a complex dynamic system in which time has emergent properties. According to Ilya Prigogine, in complex dynamic systems, time is an independent dimension that evolves in a one-way direction (PRIGOGINE, 1997: 26). Henri Poincaré, upon studying the celestial mechanics, pendulum movements, and systems with unstable equilibrium, discovered that we can represent a dynamical state by a point (x0, y0) in a space formed by the coordinates and momenta (PRIGOGINE, 1997: 31-32). The assemblage of all the points form a phase space graphic. In these graphics the notion of time is excluded. Instead, we can observe the emergent presence of regions of accumulation of points (and also the presence of dispersion), meaning that some phenomena are recurrent or not in time in the system analyzed.

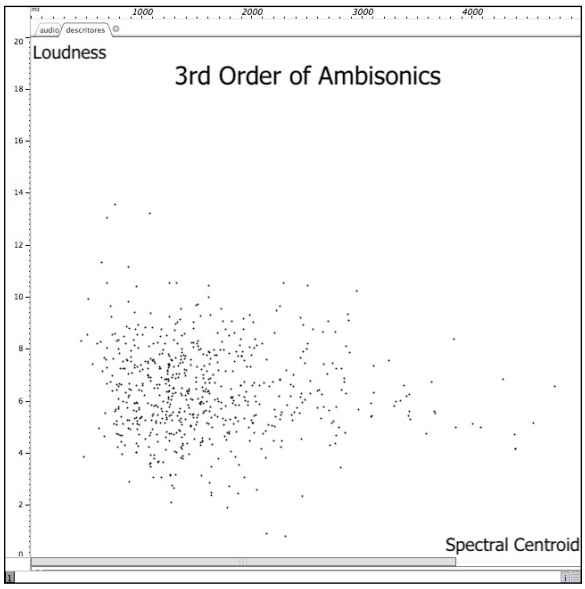


Figure 12: Phase space graphics of spectral centroid and loudness values of sound A spatialized in the third order of ambisonics.

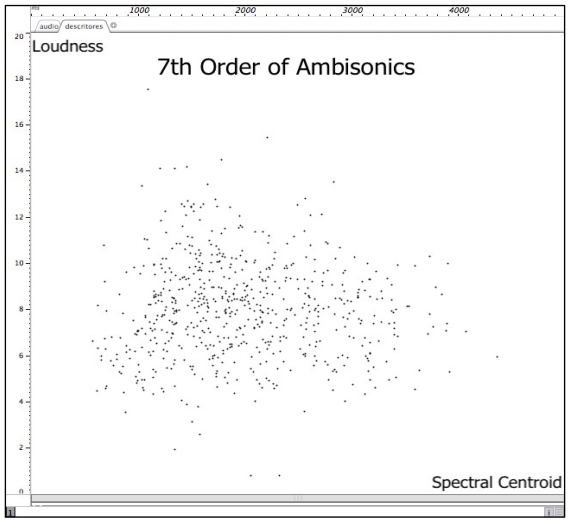


Figure 13: Phase space graphics of spectral centroid and loudness values of sound A spatialized in the seventh order of ambisonics.

By observing Figures 12 and 13, we notice that sound A spatialized in the seventh order of ambisonics produces a more dispersed phase-space representation considering spectral centroid and loudness values. In both figures, the spectral centroid values vary from 700 to 4,000 Hz, i.e., this sound is situated in a similar frequency region. In addition, it is possible to perceive that in the third order of ambisonics (Figure 12) there is a concentration of points between 1,000 and 2,000 Hz, while in the seventh order of ambisonics (Figure 13) the points are more equally dispersed in the ambitus of 700 and 4,000 Hz. On the other hand, regarding the loudness values, sound A spatialized in the seventh order of ambisonics (Figure 13) is distributed in a higher area, between 4 and 14, whereas sound A spatialized in the third order of ambisonics (Figure 12) covers an area between 4 and 10. The higher ambitus of the seventh order of spatialization in which the points are dispersed confirms the higher variation of values considering this order of ambisonics spatialization.

This representation also confirms the important role of amplitude variations (represented by the loudness descriptor) in the perception of sound distance and its variations.

Discussion of the Results and Conclusions

Considering the first hypothesis, which affirms that the variation of the parameters of the GS produces a variation in the time-varying spatial distribution of the sound morphology generated, the volume representation applied to sounds A and B has the property to show this variation. Sound A has a higher volume because it presents a higher spread between the spectral centroid values and a higher loudness level reflected by the colors of the representation. On the other hand, sound B has a lower volume that indicates a perceived timbre with less density. There is a lower spread between spectral centroid values and a lower level of loudness due to the sparser presence of the grains.

The second proposed hypothesis approaches the problem of sound accuracy in space and the production of diffuse sound fields (i.e., the presence of uncorrelated parts coming from surround directions) from HOA spatialization. It was considered that higher-order ambisonics would produce higher sound accuracy and more diffuse sound fields. Even though these are properties of HOA that are perceivable in listening experiments, they are not demonstrable with only one representation. We found it necessary to combine different representations and calculations based on the values of spectral centroid and loudness descriptors.

By articulating the theoretical approach proposed with the chosen descriptors, the loudness would represent the perception of the intensity variation while the spectral centroid would indicate the brightness frequency region variation regarding sound distance. A higher variation of both descriptors values would indicate a more diffuse sound field that is related to the decorrelation effect. And this effect, from a musical analysis or a musicological standpoint, can be related to a positive aesthetic feature. The analysis proposed for this hypothesis confirmed that the same sound spatialized in a higher order of ambisonics (in this case comparing the third and seventh orders) produced a more diffuse sound field. In Table 4 below we summarize the relevant data collected for this analysis.

	Ambisonics Order	Ambitus	Valleys/ Peaks	Average	Standard Deviation	Region of Point Accumulation (phase space)
Spectral Centroid	3rd	2,731 Hz	9	1,599 Hz	537 Hz	1,000 – 2,000 Hz
	7th	2,453 Hz	15	2,106 Hz	533 Hz	700 – 4,000 Hz
Loudness	3rd	4.1	8	9.34	1.63	4 – 10
	7th	11.4	14	11.64	1.91	4 – 14

Table 4: Information collected in the analysis of the sound field and accuracy in ambisonics spatialization.

As the proposed theoretical approach indicates, for sound perception in space, the production of diffuse sound fields can be associated with the late reverberation effect, even if both space-dependent effects have different natures. In sound studies and electroacoustic composition, the production of this effect particularly in GS processes is related to the microtemporal decorrelation effect, i.e., similar sounds diffused in different loudspeakers with very short time-offsets (in the order of few milliseconds). In our analysis, spectral centroid and loudness values were manipulated by different representations and calculations in order to

verify their behavior. As it can be observed in Table 4, the same sound spatialized in a higher order of ambisonics produces more variation in the perception of sound in space in terms of frequency and intensity. This information can be associated with the definition of a diffuse sound field, in which the perception information is vast because of the sounds coming from surround directions with no unique source localization.

The study described in this article was conceived from the intuition that immersive sound systems such as the HOA technology are important to be considered a tool in music composition. Our auditory system has the property of perceiving and identifying the position of sounds with accuracy coming from 360º in different height planes. Thus, if one has the means in composition to conceive the space as a manipulable feature besides time, rhythm, pitch, dynamics, orchestration, among others, he or she should not abdicate this possibility. We believe that this is a sound and musical field that has huge new possibilities to be further developed.

The forthcoming analyses of the audio samples that were recorded will deal with the eight-channel acoustic sound result, recorded inside the patch before the sound diffusion. We believe that it will be possible to compare the spectral properties of these audio samples with the recorded result of the sound diffusion by the dummy-head microphone.

Acknowledgements

São Paulo Research Foundation (FAPESP) supported Danilo Rossetti's post-doctoral research (process 2016/23433-8), and the National Council for Scientific and Technological Development (CNPq) supports Jônatas Manzolli's Pq Fellowships (processes 304431/2018-4 and 429620/2018-7).

We would also like to thank Professor Marcelo Wanderley and the CIRMMT for the opportunity to perform this study during a short-term residence at the Immersive Presence Laboratory.

References

BAYLE, François. À propos de l'Acousmonium. *La Revue Musicale.* Paris, v. 394-397, p. 144-146, 1986.

BARBEDO, Jayme; LOPES, Amauri. Automatic Genre Classification of Musical Signals. *EURASIP Journal of Advances in Signal Processing.* London, v. 2007, p. 1-12, 2007.

BRONKHORST, Adelbert; HOUTGAST, Tammo. Auditory Distance Perception in Rooms. *Nature*. London, v. 397, p. 517-520, 1999.

COLAFRANCESCO, Julien; GUILLOT, Pierre; PARIS, Eliott. Tutoriel de la bibliothèque HOA por Max/MSP. Paris [s.n.]: Centre de recherche en Informatique et Création Musicale/Université Paris 8, 2015.

DANIEL, Jérôme. Evolving Views on HOA: From Technological to Pragmatic Concerns. THE AMBISONICS SYMPOSIUM (1.), 2009, Graz. *Proceedings...* Graz: Institute of Electronic Music and Acoustics, 2009, p. 1-18.

DANIEL, Jérôme. Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia. Thesis (Ph.D. of Music). Université Paris 6, Paris, 2001.

DANIEL, Jérôme. Spatial Sound Encoding Including Near Field Effect: Introducing Distance Coding Filters and a Viable, New Ambisonic Format. *Audio Engineering Society International Conference* (23.), 2003, Copenhagen. *Proceedings...* New York: Audio Engineering Society Press, 2003.

DANIEL, Jérôme; NICOL, Rozenn; MOREAU, Sébastien. Further Investigations on High Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging. CONVENTION OF THE AUDIO ENGINEERING SOCIETY (114.), 2003, Amsterdam. *Proceedings...* New York: Audio Engineering Society Press, 2003.

FELGETT, Peter. Ambisonics. Part One: General System Description. *Studio Sound*. London, v. 17, n. 8, p. 20-22, 40, 1975.

GARAVAGLIA, Javier. Creating Multiple Spatial Settings with "Granular Spatialization" in the High-Density Loudspeaker Array of the Cube Concert Hall. *Computer Music Journal.* Cambridge, v. 40, n. 4, p. 79-90, 2016.

GERZON, Michael. Ambisonics. Part Two: Studio Techniques. *Studio Sound*. London, v. 17, n. 8, p. 24-26, 28, 30, 1975a.

GERZON, Michael. Design of Ambisonics Decoders of Multispeaker Surround Sound. AUDIO ENGINEERING SOCIETY CONVENTION (58.), 1977, New York. *Proceedings...* New York: Audio Engineering Society Press, 1977.

GERZON, Michael. Dummy Head Recording. *Studio Sound.* London, v. 17, n. 4, p. 42-44, 1975b.

GERZON, Michael. Periphony: With-Height Sound Reproduction. *Journal of the Audio Engineering Society*, v. 21, n. 1, p. 2-10, 1973.

HARLEY, Maria Anna. An American in Space: Henry Brant's "Spatial Music". *American Music*. Champaign, v. 15, n. 1, p. 70-92, 1997.

KEARNEY, Gavin. et al. Depth Perception in Interactive Virtual Acoustic Environments Using High Order Ambisonic Fields. INTERNATIONAL SYMPOSIUM ON AMBISONICS AND SPHERICAL ACOUSTICS (2.), 2010, Paris. *Proceedings...* Paris: Institut de Recherche et Coordination Acoustique/Musique, 2010.

KENDALL, Gary. The Decorrelation of Audio Signals and Its Impact on Spatial Imagery. *Computer Music Journal*. Cambridge, v. 19, n. 4, p. 71-87, 1995.

LYON, Eric. Multichannel Sound Design. INTERNATIONAL COMPUTER MUSIC CONFERENCE (45.), 2019, New York. *Proceedings...* San Francisco: International Computer Music Association, 2010.

MALHAM, Dave. El espacio acústico tridimensional y su simulación por medio de Ambisonics. In BASSO, Gustavo; DI LISCIA, Oscar Pablo (Org.). *Música y espacio: ciencia, tecnología y estética.* Quilmes: Universidad Nacional de Quilmes Editorial, 2009, p. 161-202.

MALT, Mikhail; JOURDAN, Emmanuel. Le BStD – Une représentation graphique de la brillance et de l'écart type spectral, comme possible représentation de l'évolution du timbre sonore. L'ANALYSE MUSICALE AUJOURD'HUI, CRISE OU (R)ÉVOLUTION?*,* 2009, Strasbourg. *Proceedings...* Paris: Société Française d'Analyse Musicale, 2009, p. 1-14.

MCADAMS, Stephen. Musical Timbre Perception. In: DEUTSCH, Diana (Org.). *The Psychology of Music.* 3rd Ed. San Diego: Academic Press, 2013, p. 35-67.

MONTEIRO, Adriano. *Criação e performance musical no contexto dos instrumentos musicais digitais*. Dissertation (Master of Music). Institute of Arts, University of Campinas, Campinas, 2012.

MONTEIRO, Adriano. *Medidas de áudio com fundamentação neurofisiológica aplicadas para a segmentação automática de músicas feitas de sons não redutíveis ao conceito de nota musical*. Thesis (Ph.D. of Music). Institute of Arts, University of Campinas, Campinas, 2016.

MUSIQUES POUR PIANO ET ELECTROACOUSTIQUE. Horacio Vaggione (composer). Philip Mead (interpreter, piano). Bourges: Chrisophée Électronique, 1995 (Compact Disc, track 5).

NORMANDEAU, Robert. Timbre Spatialization: The Medium is the Space. *Organised Sound*. Cambridge, v. 14, n. 3, p. 277-285. 2009.

PARNCUTT, Richard. Systematic Musicology and the History and Future of Western Musical Scholarship. *Journal of Interdisciplinary Music Studies.* Graz, v. 1, n. 1, p. 1-32, 2007.

PEETERS, Geoffroy. *A Large Set of Audio Features for Sound Description (Similarity and Classification) in the CUIDADO Project* (Cuidado Project Report). Paris: [s.n.], 2004. Available online at: http:// recherche.ircam.fr/anasyn/peeters/ARTICLES/Peeters_2003_cuidadoaudiofeatures.pdf. Accessed July 2019.

PEREIRA, Erica. *Estudos sobre uma ferramenta de classificação musical*. Dissertation (Master of Music). School of Electrical and Computer Engineering, University of Campinas, Campinas, 2009.

PRIGOGINE, Ilya. *The End of Certainty: Time, Chaos, and the New Laws of Nature*. New York: The Free Press, 1997.

PUCKETTE, Miller. *The Theory and Technique of Electronic Music*. Singapore: World Scientific Publishing, 2006. Draft available online at: http://msp.ucsd.edu/techniques/latest/book-html/. Accessed July 2019.

ROADS, Curtis. *Microsound*. Cambridge: The MIT Press, 2001.

ROSSETTI, Danilo. Modelos de composição e percepção de Xenakis: *Concret PH* e o Pavilhão Philips. *Opus* (Online), v. 18, n. 1, p. 153-178, 2012.

ROSSETTI, Danilo; ANTUNES, Micael; MANZOLLI, Jônatas. Compositional Procedures in Electronic Music and the Emergence of Time Continuum. *Organised Sound*. Cambridge: University Press, v. 25, n. 2, 2020, p. 156-167.

ROSSETTI, Danilo; MANZOLLI, Jônatas. Analysis of Granular Acousmatic Music: Representation of sound flux and emergence. *Organised Sound*. Cambridge: University Press, v. 24, n. 2, p. 205-215, 2019a.

ROSSETTI, Danilo; MANZOLLI, Jônatas. Creating Timbre in Space: Granular Synthesis and Ambisonics Spatialization Study and Composition. INTERNATIONAL COMPUTER MUSIC CONFERENCE (45.), New York. *Proceedings...* San Francisco: International Computer Music Association, 2019b.

ROSSETTI, Danilo; MANZOLLI, Jônatas. De Montserrat às ressonâncias do piano: uma análise com descritores de áudio. *Opus* (Online)*,* v. 23, n. 3, p. 193-221, 2017.

ROSSETTI, Danilo; TEIXEIRA, William; MANZOLLI, Jônatas. Emergent Timbre and Extended Techniques in Live-Electronic Music: An Analysis of *Desdobramentos do Contínuo* Performed by Audio Descriptors. *Música Hodie*. Goiânia, v. 18, n. 1, p. 16-30, 2018.

SCHAEFFER, Pierre. *Traité des objets musicaux:* essai interdisciplines. Paris: Seuil, 1966.

SCHNELL, Norbert. et al. Mubu & Friends – Assembling Tools for Content-Based Real-Time Interactive Audio Processing in Max/MSP. INTERNATIONAL COMPUTER MUSIC CONFERENCE (35.), 2009, Montreal. *Proceedings...* San Francisco: International Computer Music Association, 2009.

SÈDES, Anne; GUILLOT, Pierre; PARIS, Elliot. The HOA Library, Review and Prospects. INTERNATIONAL COMPUTER MUSIC CONFERENCE (40.), 2014, Athens. *Proceedings...* San Francisco: International Computer Music Association, 2014, p. 855-860.

SIMURRA, Ivan. *Contribuição ao problema da orquestração assistida por computador com suporte de descritores de áudio*. Thesis (Ph.D of Music). Institute of Arts, University of Campinas, Campinas, 2016.

SIMURRA, Ivan; MANZOLLI, Jônatas. *Sound Shizuku Composition:* a Computer-Aided System for Extended Music Techniques. *MusMat: Brazilian Journal of Music and Mathematics*, v. 1, n. 1, p. 86-101, 2016.

SOLOMOS, Makis. *De la musique au son: l'émergence du son dans la musique des XXe-XXIe siècles*. Rennes: Presses Universitaires de Rennes, 2013.

STOCKHAUSEN, Karlheinz. Musique dans l'espace [1958]. *Contrechamps.* Paris, v. 9, p. 78-100, 1988.

TORCHIA, Rayan; LIPPE, Cort. Techniques for Multi-Channel Real-Time Spatial Distribution Using Frequency-Domain Processing. CONFERENCE ON NEW INTERFACES FOR MUSICAL EXPRESSION (4.), 2004, Hamamatsu. *Proceedings...* Hamamatsu: Shizuoka University of Art and Culture, 2004, p. 116-119.

TROCHIMCZYK, Maja. From Circles to Nets: On the Signification of Spatial Sound Imagery in New Music. *Computer Music Journal.* Cambridge, v. 25, n. 4, p. 39-56, 2001.

TRUAX, Barry. Musical Creativity and Complexity at the Threshold of the 21st Century. *Interface.* London, v. 21, p. 29-42, 1992.

VAGGIONE, Horacio. Décorrélation microtemporelle, morphologies et figurations spatiales. JOURNÉES D'INFORMATIQUE MUSICALE (9.), 2002, Marseille. *Proceedings...* Marseille: GMEM Centre National de Création Musicale, 2002.

VAGGIONE, Horacio. Objets, représentations, opérations. *Ars Sonora.* Paris, v. 2, 1995. Available online at: http://www.ars-sonora.org/html/numeros/numero02/02e.htm. Accessed July 2019.

VAGGIONE, Horacio. Représentations musicales numériques: temporalités, objets, contextes. In: SOULEZ, Antonia; VAGGIONE, Horacio (Org.). *Manières de faire des sons*. Paris: L'Harmattan, 2010, p. 45-82.

WAKEFIELD, Garry. Third-Order Ambisonics Extensions for Max/MSP with Musical Applications. INTERNATIONAL COMPUTER MUSIC CONFERENCE (32.) 2006, New Orleans. *Proceedings...* San Francisco: International Computer Music Association, 2006, p. 123-126.

WALLACH, Hans; NEWMAN, Edwin; ROSENZWEIG, Mark. The Precedence Effect in Sound Localization. *The American Journal of Psychology*. Champaign, v. 62, p. 315-336, 1949.

XENAKIS – ELECTRONIC MUSIC. Iannis Xenakis (composer). Albany: Electronic Music Foundation, 1997. (Compact Disc, track 2).

ZAHORIK, Pavel; BRUNGART, Douglas; BRONKHORST, Adelbert. Auditory Distance Perception in Humans: A Summary of Past and Present Research. *Acta Acustica United with Acustica.* Stuttgart, v. 91, p. 409-420, 2005.

Danilo Rossetti does research in computer-aided composition and musical analysis. His is a assistant professor (full-time faculty) of the Department of Arts of the Federal University of Mato Grosso, collaborator professor of the Graduate Music Studies at the Institute of Arts of UNICAMP and head of the CNPq research team Computer-Aided Music Composition, Analysis and Performance. He holds a PhD in Music Composition from the University of Campinas (UNICAMP) conducting research stage at the Center de recherche Informatique et Création Musicale of the Université Paris 8. He accomplished post-doctoral research at the Interdisciplinary Nucleus of Sound Communication at UNICAMP, under the supervision of Prof. Dr. Jônatas Manzolli with the support of FAPESP. He was one of the winners of the Funarte Award for Classical Music in 2016 in the solo, acousmatic, and mixed-music category. His compositions were performed in conferences and festivals such as ICMC, CMMR, NYCEMF, CICTeM, NIME, BIMESP, SBCM, ANPPOM and CDMC 30 anos. d.a.a.rossetti@gmail.com

Jônatas Manzolli combines contemporary musical creation and cognitive sciences focusing on the dialogue between music and science. The interdisciplinary study results in electroacoustic, instrumental, and multimodal works. A composer and mathematician and full professor of the Art Institute, Unicamp, Brazil, he is a pioneer in Computer Music research in Brazil. He has been a guest researcher at the Institute for Neuroinformatics, Switzerland, the SPECS Group at the Universitat Pompeu Fabra, Barcelona and the CIRMMT, McGill University, Montreal.
Jônatas Manzolli's most notorious achievements have emphasized the delicate relationship between man and machine, including the use of artificial intelligence and digital interfaces such as *Ada: Intelligent Space* (2002) and the *Multimodal Brain Orchestra* (2009). His compositions also include large orchestral settings such as the multimodal opera *Descobertas* (2016). He has received numerous grants and awards including the recent Rockefeller Foundation "*Arts & Literary Arts*" Award to be an artist in residence at the Bellagio Centre (2018). jotamanzo@gmail.com