Interactive Spatialization and Sound Design using an Evolutionary System

Jose Fornari
Interdisciplinary Nucleus for Sound Studies, NICS, Unicamp
Rua da Reitoria, 165. C.P. 6166
Campinas, 13091-970, BRAZIL
+55(19) 3521 4966
fornari@nics.unicamp.br

Adolfo Maia Jr.
Applied Mathematics, IMECC
Interdisciplinary Nucleus for Sound Studies, NICS, Unicamp
Rua da Reitoria, 165. C.P. 6166
Campinas, 13091-970, BRAZIL
+55(19) 3521 4966
adolfo@nics.unicamp.br

Jônatas Manzolli
Music Department
Interdisciplinary Nucleus for Sound Studies, NICS, Unicamp
Rua da Reitoria, 165. C.P. 6166
Campinas, 13091-970, BRAZIL
+55(19) 3521 4966
jonatas@nics.unicamp.br

ABSTRACT
We present an interactive sound spatialization and synthesis system based on Interaural Time Difference (ITD) model and Evolutionary Computation. We define *a Sonic Localization Field* using sound attenuation and ITD azimuth angle parameters and, in order to control an adaptive algorithm, we used pairs of these parameters as *Spatial Sound Genotypes (SSG)*. They are extracted from waveforms which are considered individuals of a *Population Set*. A user-interface receives input from a generic gesture interface (such as a NIME device) and interprets them as ITD cues. Trajectories provided by these signals are used as *Target Sets* of an evolutionary algorithm. A *Fitness* procedure optimizes locally the distance between the *Target Set* and the SSG pairs. Through a parametric score the user controls dynamic changes in the sound output.

Keywords
interactive, sound, spatialization, evolutionary, adaptation.

1. INTRODUCTION
Sound Spatialization has been studied by for decades [1-3]. Recently, Interactive Sound Spatialization (ISS) has been applied in hypermedia environments to develop an interactive control and integrated control of several sonic features for multi-user application [4]. ISS has been applied in the context of helping people with special needs where it can be successfully used in education and rehabilitation for cognitively disabled persons [5].

Psychoacoustic factors may be seeing as sonic cues that allow any listener to perceive and recognize a particular sound [6]. In experiments involving sound perception, it is common to take into consideration only the classical psychoacoustic factors, such as: loudness (perception of sound intensity), pitch (perception of sound frequency) and spectrum (perception of partials that compose the sonic frequency spectrum), where it is often relegated the importance of sound spatial positioning perception (SSPP).

However, SSPP turns to be very meaningful when we are in a space with several sound sources located in different places (ex: in a concert hall, watching a symphonic orchestra). SSPP can deliver important information which not just of aesthetically meaning but may be also concern our own safety (ex: driving a car or crossing a traffic road). SSPP is given by three types of hearing cues: interaural time differences (ITDs) [6], interaural level differences (ILDs) [7] and head-related transfer functions (HRTFs) [8]. ITDs refer to the difference in time for a sound to reach both ears. ILDs describe the amplitude differences in the frequency spectrum of sound as heard in both ears. HRTFs are a collection of spatial cues for a particular listener, including ITDs, ILDs and also taking into account the effects of the listener's head, outer ears and torso, to perceive sound localization. The interaural time difference was used for robotic sound source localization and cross-correlation [9].

Here we present an implementation of an interactive system that takes ISS from the perspective of adaptive evolution. Since the use of adaptive evolution produces emergent and macro structure properties [15 16], we understand the sonic result of our system as an adaptive evolution and our goal is to show how effective is this model to generate interesting sonic results, we further discussed this concept in [18].

In the following sections we present the theoretical model based on integration between interactive sound design and adaptive evolution. After, we describe the system features, mathematical model, system implementation and sound results.

2. SPATIALIZATION AND SOUND DESIGN AS ADAPTIVE EVOLUTION
Since 2001, we studied interactive, genetically generated music as a technique suited for composing highly textured music [10]. We developed also an Evolutionary Sound Synthesis methodology [11 12] and, recently, we incorporated spatial information in the sound genotype of our system [13 14]. Our study is based in the application of concepts from the theory of complex adaptive systems (CAS) [15] to sound design [18] and in this section we present the key theoretical concepts of our research and their relation with the implemented system.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
Nime'05, May 26-28, , 2005, Vancouver, BC, Canada.
Copyright remains with the author(s).

2.1 Adaptive Evolution

Our starting point is that spatialization sound systems can be seen as complex adaptive systems, (CAS). As described in [15] a CAS involves a large number of interconnected parameters that altogether exhibits a coherent emergent pattern in time. Spatialization can involve a large amount of loudspeakers, waveforms, acoustic cues, damping, and reflections, among others. A good spatialization system has to be able to integrate all these features and handle them with a simple user-interface that allows the user to use few parameters to control the whole system. CASs generate emergent and macroscopic properties [16], arisen from competition and cooperation. The large scale system behavior results from a large number of interactions made by many individuals. This is a self-organized process in which a complex system is driven to different organization states [17] and it could be applied to sound design, as we discussed in [18].

Starting from these ideas, we developed our spatialization model to control sonic populations' sets. Our goal is to apply evolutionary strategies to generate emergent patterns in sonic domain. Despite our model deals with few basic information and simple interactive role, it is able to produce complex spatial behavior. In this way, it is necessary to correlate interactive control, evolutionary strategy with gesture input and sound output control.

Finally, we connect these concepts in the following assumptions: a) control gesture inputs are used to generate Target Sets which will drive an evolutionary computation process based on Interaural Time Differences (ITD), b) spatialization is represented by a Sound Localization Field and c) spatial similarities are measured by a Fitness procedure within an Iterative Evolutionary Cycle presented below and d) adaptation between gesture inputs (i.e. Target Sets) and the sonic population is built as an evolutionary process controlled by genetic operators.

2.2 Sound-Marks Cues

Schafer [19] describes soundscapes as natural, self-organized processes usually resultants of an immense quantity of sound sources, correlated or not, but that conveys a unique audible experience that is at the same time recognizable and yet always original (as it never repeats itself). It is not difficult to relate these features of those belonging to complex adaptive system (CAS), previously mentioned. Soundscape composition might aim to computationally emulate self-organized biological or natural acoustic environments, as mentioned in [20].

Starting upon three main characteristics of a soundscape defined in [19]: *key-sounds, key-signals and sound-marks.* In our study we attempted to develop a system to interactively create sound-marks using two immersive concepts: *Sonic Localization Field and Spatial Similarity,* defined below.

The idea is to use ITD cues to interactively generate trajectories of evolutionary sound-marks. This evolution is controlled by spatial sound genotypes (SSG) related to two parameters: sound intensity and ITD azimuth angle. So we use genetic operators to reproduce new individuals in the population based on their spatial localization. Having the overall process running in real-time, it is possible to obtain as output a continuous audio stream.

3. INTERACTIVE SPATIALIZATION

3.1 ITD Cues

As previously described, the ITD, is a mechanism with which the human brain can resolve the delay time or phase difference between ears to determine the location of low frequency sound sources. For example, if both ears simultaneously hear a low frequency sound, then the source is either directly in front or behind the listener. If there is a delay of perception between ears, then the source will be perceived as being closer to the ear that receives its sound first. Time delays are significant in the localization of sound in the full range of frequency localization.

3.2 Sound-Marks in Sonic Localization Field

From the mathematical point of view our model consist of a space of triplets **G**={(W, I, L)}, named Genotype Space, where $0\leq I \leq 1$ is the waveform intensity factor and $-1\leq L\leq 1$ is the waveform ITD localization factor, given by the azimuth angle θ, where $L = (90^o - \theta)/ 90^o$ and $0^o \leq \theta \leq 180^o$. For more details, see [14]. The set of all possible values of the pair (I,L) is named *Sonic Localization Field* (SLF) (see Figure 1). In our model it is a semicircle and the pair (1,0) is associated to the sound with the greatest attenuation and it is located in front of the listener. Spatial dispersion in the SLF is characterized by the distribution of the finite set of pairs S = (I, L) as showed in Figure 1.

We define a *population* as any finite subset of elements of **G**. In our model we start from a initial population $P^{(0)}$ and a target population **T** and, iteratively, we construct a sequence of, say a number **R** of populations $G^{(1)}$, $G^{(2)}$,…, $G^{(r)}$, where the k-th population is a subset of **G** with N individuals (elements) $\boldsymbol{G}^{(k)} = \{G_1^{(k)}, G_2^{(k)}, \ldots, G_N^{(k)}\}$ and the individuals are given by triplets $G_i^{(k)} = (W_i^{(k)}, I_i^{(k)}, L_i^{(k)})$. The Target population has M individuals $\mathbf{T}= \{t_1, t_2, \ldots, t_M\}$ with the j-th individual given by $t_j = (W_j^{(T)}, I_j^{(T)}, L_j^{(T)})$. Spatial dispersion in the SLF is characterized by the distribution of the set of pairs $S_i = (I_i, L_i)$ as showed in Figure 1. These pairs on which are applied the Genetic Operators are named as Spatial Sound Genotypes (SSG). The Target Set $\mathbf{T} = \{T_k = (I_k, L_k)$, for k=1,…,M} can, in principle, be generated by a several gestural controllers associated to the position and motion of the user/musician in the space. This allows perceptual impressions guide interactively the evolutionary process of spatial distribution of the sound.

Since $\boldsymbol{G}^{(k)}$ and **T** are subsets of **G** we define the distance between these two sets as follows: consider the numbers

$$d_{ij}(G^{(k)},T) = \frac{1}{2}\frac{\left|I_i^{(k)} - I_j^{(k)}\right|}{A} + \frac{1}{4}\frac{\left|L_i^{(k)} - L_j^{(k)}\right|}{B} \qquad \textbf{(1)}$$

where the constants **A** and **B** are taken as the maximum of the intensity and localization factors, respectively, and the distance is normalized in the interval [0,1].

The distance d_k between $\mathbf{G}^{(k)}$ and **T** is defined by

$$\mathbf{d_k} \equiv \mathrm{d}(\boldsymbol{G}^{(k)},\mathbf{T}) = \min_{i,j}\, d_{ij}(G^{(k)},T) \qquad \textbf{(2)}$$

for i=1,…,N and j=1,2,…, M. Observe that this distance function take into account only the two parameters of the SSG.

The best individual in the k-th population $G^{(k)}$, $G_{i*}^{(k)} = (W_{i*}^{(k)}, I_{i*}^{(k)}, L_{i*}^{(k)})$ is that one that realizes the distance $\mathbf{d_k} \equiv \mathbf{d}(\boldsymbol{G}^{(k)}, \mathbf{T})$.

This new individual (the optimal one) now can be used by the Iterative Evolutionary Cycle as it is presented in Section 4. In order to control the sonic output we use the distance function above to define *Spatial Similarity* as follows:

Given two individuals of a population $\mathbf{G}_i^{(k)}$, $\mathbf{G}_j^{(k)}$, they are similar if the $\mathbf{d}(\mathbf{G}_i^{(k)}, \mathbf{G}_j^{(k)}) \leq \varepsilon$, where ε is an arbitrary small number and the distance **d** is defined in Eq. (1).

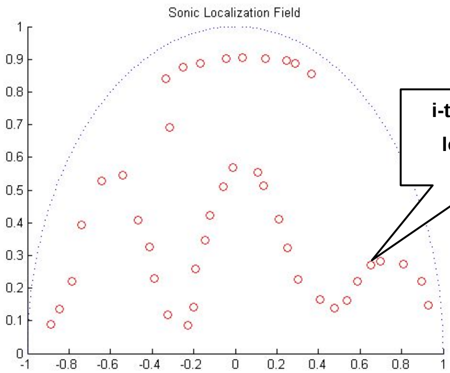

Figure 1. Sonic Localization Field (SLF).

3.3 Genetic Operators for Sound Spatilization

To be able to control the spatialization process as an evolutionary one, we defined two basic operations:

3.3.1 *Crossover*

Given the best individual of the k-th generation $\mathbf{G}_{i^*}^{(k)} = (W_{i^*}^{(k)}, I_{i^*}^{(k)}, L_{i^*}^{(k)})$ and the crossover rate α, with $0 \leq \alpha \leq 1$, individuals in the population will be renewed changing the values of their parameters as follows:

$$I_i^{(k+1)} = \alpha\, I_{i^*}^{(k)} + (1-\alpha).\, I_i^{(k)}, \text{ and}$$
$$L_i^{(k+1)} = \alpha.\, L_{i^*}^{(k)} + (1-\alpha).\, L_i^{(k)} \qquad \textbf{(3)}$$
$$\text{for } 1 \leq i \leq N, \text{ and } k = 0,1,\ldots, \mathbf{R}.$$

where **R** is the number of iterations.

3.3.2 *Mutation*

Similarly, given the best individual of the k-th generation $G_{i^*}^{(k)} = (W_{i^*}^{(k)}, I_{i^*}^{(k)}, L_{i^*}^{(k)})$ and the rate ß, with $0 \leq ß \leq 1$, the *mutation operation* is defined as:

$$I_i^{(k+1)} = \beta_1 \,.\, (\text{rand}) + (1-\beta_1).\, I_i^{(k)} \text{ and}$$
$$L_i^{(k+1)} = \beta_2 \,.\, (\text{rand}) + (1-\beta_2).\, L_i^{(k)} \qquad \textbf{(4)}$$
$$\text{for } 1 \leq i \leq N, \text{ and } k = 0,1,\ldots, r.$$

where "rand" is a random parameter in the interval [0,1] and the rates β_1 and β_2 controls the degree of randomness of the mutation operation. In our implementation below we have taken $\beta_1 = \beta_2$ for simplicity.

4. IMPLEMENTATION

We report here the first implementation of the system using MATLAB in which we have simulated the iterative evolutionary cycle (IEC) and the system interactive dataflow. The IEC consists of two main processes: a) an evolutionary sound synthesis module, which performs genetic operators that modify the waveform shape (see details in [11]) and b) an evolutionary spatial engine module which applies crossover and mutation over the population as described in section 3.2. For the interactive dataflow we simulate a user-interface that resembles the sonic localization field and we implemented also a parametric score with which the user controls de dynamic changes of the parameters.

4.1 Creating Populations

First of all, we generated waveform populations in which pre-recorded samples were cut using fixed *Time Slice*. In that case, we developed two procedures: a) automatic segmentation of a stored sample and b) random generation of a large population of sine waves with frequency and amplitude varying in pre-defined ranges. The time slices can vary from 50 milliseconds to 2 seconds and we our system can be a tool to be applied in micro and macro interactive sound design. Depends upon the size of the time slices the number of waveforms varied from dozens to thousands.

4.2 Evolutionary Engines

The evolutionary engines are used to control the evolutionary sound trajectory of the system. The Evolutionary Sound Synthesis (ESSynth) Engine has been presented in previous works [10 11] and we have modified it in order to use Eq. 2 for fitness evaluation. The Evolutionary Spatial Engine has also been described in [13 14] and it uses Eq. 3 and Eq. 4 as genetic operators.

4.3 User-Interface & Parametric Score

The general idea of the system is to allow the user to interact with it in two ways: a) real-time interaction using any a gestural device that is able to produce input for the Target Set that controls the fitness evaluation of the two evolutionary engines described in 4.2, and b) off-line interaction using a parametrical score in which the user controls cross-over and mutation rates, the region of the population that will be modified in real time, the upgrade time rate for each generation of the population and the delay in which a new waveforms is sent to a circular buffer.

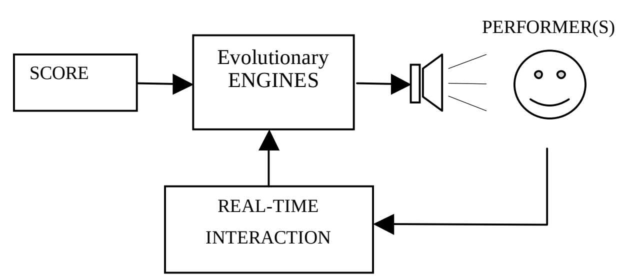

Figure 2. General diagram of the system dataflow.

The parametrical score is an ASCII file with a sequence of lines described in Table 1. The idea is to let the user to project a general evolutionary behavior and use the gestural control in real time to produce novelty. Since the population of individuals can be very large, we implemented a parameter to control a selection of sub-sets in the population. We use two integer numbers n_1 and n_2 and we define a *Population Segmentation Window (PSW)* as the subinterval $[n_1, n_2]$ where $0 \leq n_1 < n_2 \leq N$. Only the individuals belonging to the chosen PSW will be used in the IEC. In this way the parametrical score is used to give to a composer/musician the flexibility to explore regions or waveform neighborhood.

Table 1. Description of the Parametric Score

parameter	description	application
$0 \leq \alpha \leq 1$	crossover rate	*increases Correlation*
$0 \leq ß \leq 1$	mutation rate	*increases Randomness*
$0 \leq n_1 < n_2 \leq N$	population segmentation window	*defines a sub-set in the population*
T_b	upgrade time rate for each generation (in millisecs).	*defines the duration of each iterative evolutionary cycle*
T_m	delay new waveforms is sent to a circular buffer (in millisecs).	*defines the rhythm of the sound transformation*
Flags = 1, 2, 3	status selector	*indicates population segmentation (0), synthesis (1), spatial engine (2) or (3) end.*

4.4 Iterative Evolutionary Cycle

Finally, it is possible to describe the whole iterative process in which we implemented an adaptive evolution applied to spatialization and sound design. Notice that there are to main circuits (see Figure 3): a) off-line: it is controlled by the Parametrical Score (see Figure 3, left) and b) on-line: it is controlled interactively by the user using a gestural controller to produce a Target Set (see Figure 3, right). Both circuits are applied to the evolutionary engines and the sound output is given by the best individual of each generation.

5. SOUND OUTPUT

We presented here results of the MATLAB implementation. The tested parameters and the parametric scores are presented in Table 2. The first parameter on the score is always the time in milliseconds. Basically, we evaluated how the ITD cues work as part of the sound genotype and how the evolutionary synthesis method modifies the generated sound.

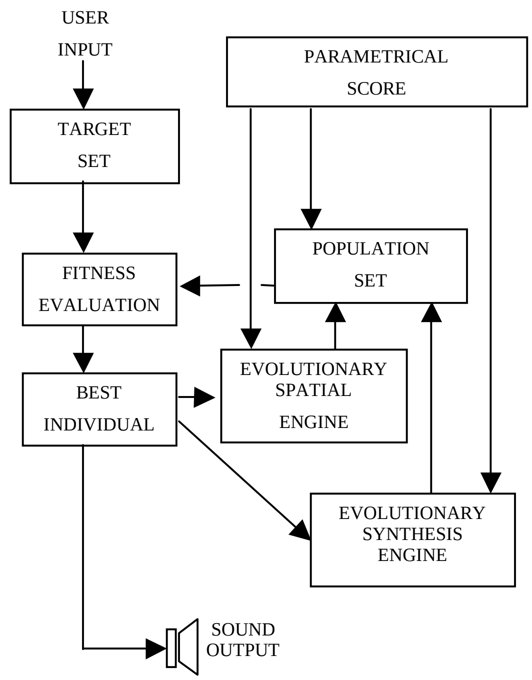

Figure 3. Iterative Evolutionary Cycle

Table 2. Parametric Score of the Sound Example

Parameter Line	Comment
0, 0, 25, 30	*Time Population Segmentation**
0, 1, .5, .0, 50,100	*Synth - alfa=0.5 and beta=0*
0, 2, .2, .3, 50,100	*Spatial – alfa=0.2 and beta=0*
50, 2, .2, .3, 50,100	*Spatial – alfa=0.2 and beta=0.1*
100, 2, .2, .3,50,100	*Spatial – alfa=0.2 and beta=0.2*
110, 2, .0, .0,50,100	*Spatial – alfa=0.2 and beta=0.3*
130, 3, .0, .0,50,100	*End of the process*

The example presented here is the result of a population generated using a sample of a speech material. Below in Figure 4 we presented the sound example generated with a population with time slices of 0.2 seconds. It is possible to verify the dynamic of the genetic operators in Table 2. When the process starts EESynth and Spatial are applied with mutation rates equal to zero. The SLF presented in Figure 4 (top) allocates all the spatialization information close to its maximum values. Parameters $\mathbf{T_b}$ and $\mathbf{T_m}$ described in Table 1 were not tested yet for we are working in a Pd implementation.

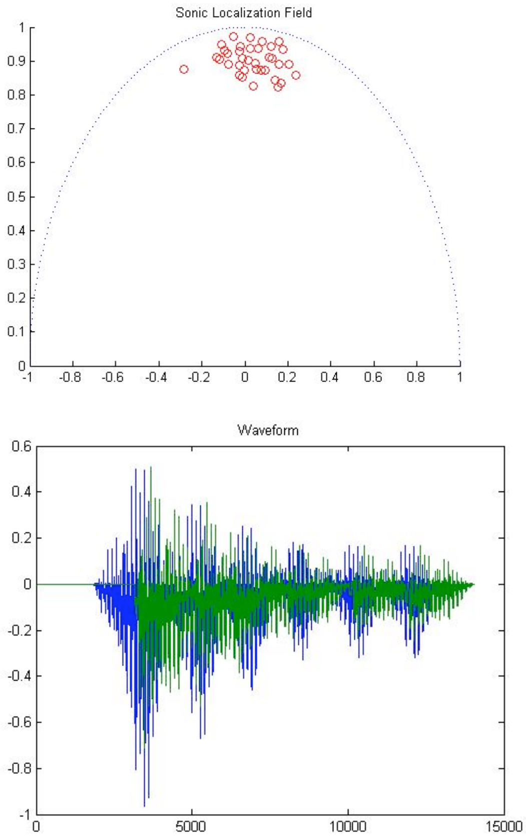

Figure 4. Sound Localization Field used to generate the sound example (top); resultant waveform (bottom).

6. CONCLUSION

We presented here a mathematical model and a MATLAB implementation of a system for interactive spatialization control and evolutionary sound synthesis. We reviewed some applications of interactive sound spatialization and introduced the concept of adaptive evolution based in the theory of complex adaptive systems. The system we presented is connected to recent studies in which we incorporated spatial information as sound genotype description. Here we linked these new achievements to the idea of using a gestural controller as Target Set. We think it is a promising model in which a gesture produced by NIME device can be used as real time controller of an evolutionary process. In this way, the user is providing novelty and the system will adapt according to this interactive rule. We think the fitness procedure is simple enough to let the user to be able to correlate the sound evolution with its spatialization. Using the MATLAB implementation we tested the control parameters using parametric scores and a simulation of the sound localization field using a graphic interface.

We also implemented a new version of the Evolutionary Sound Synthesis (ESSynth) Method as it was presented in [11]. The new ESSynth modifies the population according to the fitness evaluation described in Eq. (2). Differently, from previous work, we used the notion of proximity in the SLF to change also the shape of the waveforms in the population. This approach has been proved to be very efficient for interactive applications, because the user will have the best modified sound as it is located and perceived by ITD cues. It was also simple to implement Eq. (1) and the calculations involved are not expensive. Thus we project that the interactive real-time operation will be satisfactory. Further studies will include:

a) An application in Pd (Pure Data language) to test the usage of several sensor-devices to capture the position of performers into the stage, such as cameras and infrared sensors, to produce geometric trajectories to build curves in the Target Set.

b) It is also possible to explore the Inter-aural Level Differences (ILD), in this case the use of ITD and ILD will provide a near and far field effect. Also the use of reverb will provide extra psychoacoustics cues to interact with the user.

7. ACKNOWLEDGMENTS

We thank to the Brazilian Agencies CNPq that financially supports Manzolli and FAPESP that financially supports Fornari.

8. REFERENCES

[1] J. Blauert. *Spatial hearing: the psychophysics of human sound localization*. Cambridge: MIT Press, 1997.

[2] V. Pulkki, "*Virtual sound source positioning using vector base amplitude panning*". Journal of the Audio Engineering Society, vol. 45, pp. 456-66, 1997.

[3] J. M. Chowning, "The simulation of moving sound sources," presented at *Audio Engineering Society 39th Convention*, New York, NY, USA, 1970.

[4] Michael Cohen, Jens Herder and William L. Martens., *Cyberspatial Audio Technology*. J. Acous. Soc. Jap., vol.20, No.6, pp.389--395, Nov. 1999.

[5] Barbieri, T., Bianchi, A. and Sbattella, L. Minus-Two: Multimedia, Sound Spatialization and 3D Representation for Cognitively Impaired Children. In *Computer Helping people with Special Needs*, Lecture Notes in Computer Science, Springer Berlin / Heidelberg, 2004.

[6] Jack B. Kelly and Dennis P. Phillips. "*Coding of interaural time differences of transients in auditory cortex of Rattus norvegicus: Implications for the evolution of mammalian sound localization*". Hearing Research, Vol 55(1), pages 39-44. 1991.

[7] Birchfield, S.T. & Gangishetty, R. "Acoustic Localization by Interaural Level Difference". *IEEE International Conference on Acoustics, Speech, and Signal Processing* (ICASSP), Philadelphia, Pennsylvania, March 2005.

[8] Douglas S. Brungart and William M. Rabinowitz. "Auditory localization of nearby sources. Head-related transfer functions". The Journal of the Acoustical Society of America -- September 1999 -- Volume 106, Issue 3, pp. 1465-1479.

[9] Murray, J. C., Erwin, H. R., and S. Wermter, "Robotic sound source localization using interaural time difference and

cross-correlation". *In proceedings of the KI-2004*, September 2004.

[10] Fels, S. S. and Manzolli, J. “Interactive, Evolutionary Textured Sound Composition”. 6th Eurographics Workshop on Multimedia, pp. 153-164. Sept. 2001.

[11] Manzolli, J., Fornari, J., Maia Jr., A., (2001) Damiani F. The Evolutionary Sound Syn-thesis Method. Short-paper. ACM multimedia, ISBN:1-58113-394-4. USA.

[12] Fornari, J., Manzolli, J., Maia Jr., A., Damiani F., “The Evolutionary Sound Synthesis Method”. SCI conference. Orlando, USA. 2001.

[13] Fornari, J.; Maia Jr. A.; Manzolli, J.. “A Síntese Evolutiva Guiada pela Espacialização Sonora”. XVI Congresso da Associação Nacional de Pesquisa e Pós-graduação (ANPPOM). Brasília. 2006.

[14] Fornari, J.; Maia Jr. A.; Manzolli, J..”Creating Soundscapes using Evolutionary Spatial Control”. *In proceedings of the EvoMusarts*, Spring-Verlag,Valencia. 2007.

[15] Holland J. H. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology*, Control and Artificial Intelligence, MIT Press, Cambridge MA, 1992.

[16] **Holland** J.H. *Hidden Order: How Adaptation Builds Complexity* , Addison-Wesley 1996.

[17] Von Foerster, H., “*On Self-Organizing Systems and Their Environments.*” In: Self-Organizing Systems, M. C. Yovits und S. Cameron (Hg.), Pergamon Press, London, pp. 31–50, 1960.

[18] Caetano, M. Jônatas Manzolli, J. Fernando Von Zuben, F. Self-Organizing Bio-Inspired Sound Transformation. In *proceedings of the EvoMusarts*, Spring-Verlag,Valencia. 2007.

[19] R. Murray Schafer, M. (1977). “*The Soundscape*”. ISBN 0-89281-455-1.

[20] Truax, B.. (1978) “*Handbook for Acoustic Ecology*”. ISBN 0-88985-011-9.