Interactive Soundscapes: Developing a Physical Space Augmented through Dynamic Sound Rendering and Granular Synthesis

Matteo Lionello
Sound and Music Computing
Aalborg University Copenhagen (Denmark)
mlione16@student.aau.dk

Marcella Mandanici, Sergio Canazza, Edoardo Micheloni
Dept. of Information Engineering
University of Padova (Italy)
surname@dei.unipd.it

ABSTRACT

This project presents the development of an interactive environment inspired by soundscape composition, in which a user can explore a sound augmented reality related to a real soundscape. The user walks on the application's responsive floor controlling through her/his position the soundscape rendering. In the same time s/he can explore the sound structure entering central zone of the responsive floor, where some granular synthesis processes are activated. One main goal of the project is to sensitize the participants regarding the surrounding sonic environment looking also for pedagogical studies. A first evaluation of an audio excerpt produced by a user's soundwalk shows that the application can produce good quality and interesting soundscapes, which are fully consistent with the real environments from which they were inspired.

1. INTRODUCTION

Every day, in any place, we are all surrounded by an eternal and universal symphony. It is eternal because, in John Cage terms, as long as there is a person able to hear, there always will be a sound to be heard. The music is already around us, and we are immersed in it. What is the role of the artist in front of this already done music? Are we only passive listeners with no control over it or are we "*responsible for giving it form and beauty?*" [1]. This question has been written at the beginning of "*Tuning of the world"* by Murray Schafer, where he gave a definition of the term "*Soundscape*" indicating the set of sounds persisting into a certain context. This established the basis of the Soundscape Composition which will be afterwards carried on by Barry Truax, colleague and successor to Murray Schafer. Barry Truax deeply studied the granular synthesis and composed musical pieces implementing it in real time through electronic oscillators, sample sounds of small size, and finally using soundscape recordings. Soundscape composition is a form of electroacoustic music, characterized by recognizable contexts and environmental sounds. The original environmental context is preserved: all the memories, past experiences and associations are called to the listener's mind through artist's compositional strategies [2]. The artist is the medium between the soundscape and the listener. The aim of the artist is to invoke in the listener associations, memories and fancies referred to a particular soundscape.

1.1 Active and Deep Listening

Two different concepts have been strongly remarked by Francois Pachet and Pauline Oliveros. The first one from Pachet is held in the term "*Active Listening*" [3], which refers to the capability of a user to have an active and concrete control on what s/he listens to. The second concept from Oliveros is held in the expression "*Deep Listening*" [4], which outlines the concentration of the audience and the musicians themselves on the meaning beyond the sounds that are perceived and played. Both the concepts are basic and strongly present in the installation that we introduce here. The work presented in this paper is inspired by Barry Truax's "*Pacific Fanfare, Entrance to the Harbor*" [5] where the listener is accompanied by Truax in a sound-route in the harbor of Vancouver. In the 1994 revision, Truax implemented the recorded soundscape with some compositional tools bringing the listener between two parallel worlds which follow each other during the track, one related to the real Vancouver harbor soundscape, and the other one from the musician's interpretation of it. In this kind of soundscape composition the listener is led by the composer on a fixed flow of acoustic impressions captured by the artist and both of them are involved in a deep listening of the soundscape in Oliveros' terms. Music interactive systems allow to extend the user's freedom exploring the sound augmented environment with no external guides but her/his own auditory feedback. In this way the listener is no longer considered as a passive audience but s/he can use some degree of control on the music, proposing different perceptions of what is reproduced using memory associative routes. Moreover, her/his interaction depends on the sound context the system reproduces, promoting a deep listening of the surrounding sonic events.

1.2 Interactive Soundscapes

Interactive Soundscape is a 3x4 mt. floor under the range of a camera. When the user enters the installation s/he is projected in a sonic environment composed by sounds related to a real environment [1]. The aims of this project are both the spatial exploration of the reconstructed soundscape and the structural exploration of the sound elements peculiar to that soundscape. Using a granular synthesis process, brief time-windows are extracted from the sounds composing the soundscape. The user changing her/his position inside the installation modifies the parameters applied to the processing of the sound sources. Thus, moving around it is possible to explore the internal structure of the sounds projecting the user into a world as much real as the original one, emphasizing the acoustic properties of each sound. The role of the artist takes a step back leaving the tools for the musical composition to an audience not necessarily musically educated and not even aware of them. Besides the artistic purposes, the installation can be used in pedagogical field looking how children manage the spatial information to study cognitive processes [6] [7]. When the user gets the feedback from the different sounds distributed over the augmented space, cognitive maps are built by abstraction processes according to them. Thus, when a user enters the tracked area s/he can use what has been previously learned, enriching her/his cognitive map of the surrounding space.

2. RELATED WORK

Spatial cognition is widely used in interactive installations such as soundscape rendering and audio augmented environments. The design of composed soundscape for virtual environments requires a formal description and organization of the sound events which are used in it. The distinction between background audio and sonic events has been well defined by the Music Technology Group at Pompeu Fabra Universidad in Barcelona. They developed an online platform "*Soundscape Modeling*" that widely describes the state of the art about the possible applications of soundscape rendering and design, providing a powerful tool for authoring process. An online platform facilitates the process and generation of realistic soundscapes, focusing especially on computer games and mobile and web applications. Sample research from `Freesound.org` [2] is simplified by an automatic audio classification of sound features to compose graph model related to particular taxonomies. The soundscape is composed according to a two layers based architecture. On the first layer all the background sounds are propagated, whereas on a upper layer there are dynamic sounds events with which the users are able to interact. In a server/client architecture, the server gets real-time information from the clients coordinates related to a space providing a web stream of the soundscape. Combining SuperCollider [3], Freesound and GoogleMaps it is possible to provide soundscapes formed by collage of sounds according to specific position in a virtual city-visit in GoogleMap [4]. Mobile equipped with GPS and Internet can be used to enrich the experience of walking through a city listening to a particular soundscape that is streamed by the application and guided by the client's position. Compared to "*Soundscape Modeling*" where the interaction is interfaced through GPS or relative position in a virtual world, detecting a user's interaction in a augmented reality environment can be a challenging topic. Responsive floor interactive systems are one of the main system used to detect the user's interaction. They are mostly implemented through computer vision, audio capture and tracking techniques. For the Swiss Expo in 2002, in Neuchatel has been developed a complex artistic installation called Ada [8], which consisted of an artificial organism capable to interact actively with its visitor capturing their interaction and replying using visual and sound activities. The aim of the project was to provide Ada of the ability of expressing its internal status, which varied according to the input from people's interaction. "Her" core was composed by a tracking system recording weight, direction, speed and location of the visitors walking on its responsive floor; a vision system which used cameras controlled by the localization of the users for visual information; and an audio system to "*help identify salient individuals*". Beside computer vision and microphones, another technique used for tracking the user inside an augmented environment is triangularization through wireless devices. The LISTEN [9] project supplies to museum and art exhibition's visitors an augmented and personalized visit. Through motion-tracked wireless headphones, users exploring the physical space get immersed in a virtual environment which augments the space itself. A soundscape is composed dynamically differently for each visitor according to how s/he moves inside the space, to previous collected information of his path and his preferences and interests gathered during her/his walk. The experience provided by the installation varies from person to person, according to their behavior in the environment, allowing thus a higher level of visitors engagement.

3. SYSTEM DESCRIPTION

Interactive Soundscape uses a set of four loudspeakers, one camera and a computer running the two parts of the software and connected to both the audio system and the camera. It is presented as an indoor space under the range of a camera placed at the top of the room with its optical axis perpendicular to the floor. The installation is an interactive system composed by a tracking system capable of producing a couple of Cartesian coordinates of a user. Her/his relative position is transferred via OSC [5] [10] to a module written in PureData [6] [11] to process the audio data. The user entering the tracked zone is virtually transposed into a certain sonic environment built around her/him. Ac-

[1] A video of the installation is available at `https://vimeo.com/146137215`.

[2] Freesound.org is a collaborative on-line sound database, providing meta-data and geo-localization data for the uploaded sounds. `http://freesound.org`

[3] Supercollider is a developing environment and programming language used for real time audio synthesis. `http://supercollider.github.io/`

[4] Google Maps is a web mapping service developed by Google, including satellite images, topographic city and rural maps, and 360 degree street view perspective `https://maps.google.com/`

[5] Open Sound Control is a networking protocol developed at CNMAT by Adrian Freed and Matt Wright for musical instrument controls to comunicate in a network

[6] PureData is a visual programming language Open Source developed by Miller Puckette, more information can be found at `https://puredata.info/`

cording to the user's real location her/his virtual position is modified as well and, consequently, the processing of the audio data is differently shaped. The result is a soundwalk obtained by an acoustically augmented reality.

3.1 The Soundscape's Design

While the granularization technique is traditionally used on a physical fixed support, this installation projects is a real-time result of the interaction of a user with a physical environment. The human-machine interaction takes place through the spatial cognition of the user who decides to move according to her/his spatial perception combined with her/his auditory memory of the areas which s/he has already explored. As a tribute to the work "*Pacific Fanfare*" by Barry Truax, for testing the installation it has been decided to propose a sound environment similar to the one played by the artist: a harbor soundscape.

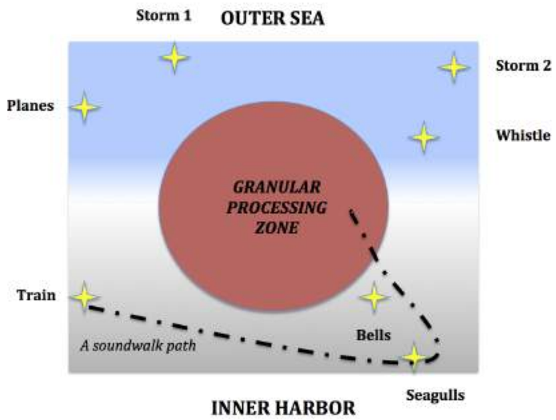

Figure 1. Graphic representation of the sound events positioning on the Interactive Soundscapes active area, with a user's soundwalk path.

In order to reproduce this soundscape, train and plane sounds were used to simulate the harbor industrial area, seagulls and bells sounds for the docks, weaves, boats whistles and far storms to simulate the farthest zone of the docks to the sea, and human activity sounds to simulate the inner land of the harbor. [7] The waves and the human activity sounds are used as the background of the soundscape and are processed with the granular synthesis in the central circular area shown in Figure 1, while the other files are used as events (some of them processed with reverberation and other with granular synthesis). Other configurations for different soundscapes are possible. Starting from a set of audio files for a particular context, a supervisor can load them in the PureData patch where the virtual location of the audio sources is decided setting a pair of coordinates. Those coordinates refer to the location where the user will hear the maximal amplification from that audio. Likewise it can be set the type of decreasing amplitude as the user goes further from an audio source. The granular zone is set choosing the circle radium and center over which the granular synthesis takes place. It's necessary to select the audio files to use for the granular synthesis, whereas files not selected files will be processed with a deep reverb. By walking inside and outside boundary distinguishing the two different areas, a 1400 ms crossfade between the two flows - one made from granular synthesis and the other one from sum of the amplitude of the audio sources - is applied.

3.2 The Design of the Parameters

The source location for the audio files can be chosen to be inside as well as outside the augmented space. In the case the coordinates are set outside, the user will never be able to reach that source and the sound will never gain the maximal intensity. The synthesis parameters depend on the distance between the user and the center of the granular zone. Along the edge the grains are 100 ms long, the density is maximized (100% of probability for two grains to succeed), the parameters for the band pass filter applied to the grains uses low cut off frequencies between 150 and 300 Hz, whereas the quality factor's values are ranging between 5 and 20, and the variance parameter describing the heterogeneity in the granular streams modeling is set to 0.1 in order to model all the streams around a narrow neighborhood of the values set by the user's position. As the user goes closer to the center of the granular zone, the soundscape results more and more chaotic: in the center the duration of the grains is set equal to 10 ms, the density is 25% and the cut off frequencies and the quality factors reach high values. The granularization happens in real time during the playback of the audio extracting every grain in each stream one after one from the original audio file. In Figure 1 it is possible to see a representation of the composed soundscape and the spatial organization of sounds inside the augmented space. The squared area represents the area covered by the camera, the red area is where the granular synthesis occurs, the blue and gray shaded backgrounds starting from the top and the bottom of the figure represent the amplitudes of the two different background sounds which are mixed together. The diamonds refer to the events composing the soundscape on the background layer.

4. SYSTEM ARCHITECTURE

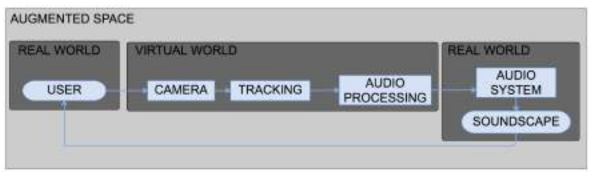

Figure 2. Flow chart of the system architecture of Interactive Soundscapes

The system architecture is made of two different software modules (Fig. 2). The first module is designed to elaborate the images captured by a camera, oriented perpendicularly respect to the floor. The camera covers a surface of variable dimension depending on the height at which it is

[7] All the audio files have been selected from `http://freesound.org`

positioned. A software named *Zone Tracker* [12] elaborates the visual information, subtracting the instantaneous images from a dry image of the ground. From the resulting image, calculated by averages of morphological transformation, a "*blob*" is obtained and its barycenter represents the user position in the area covered by the camera. The coordinates of the user's barycenter are sent, through OSC, to a second software module apt to the interaction sound processing. The second module is written in PureData, which processes the user's position with the loaded audio files, creating for each file an audio stream scaled in amplitude in function of the user position. A second stream is as well created for each file, performing a granular synthesis of the first one.

4.1 Granular synthesis

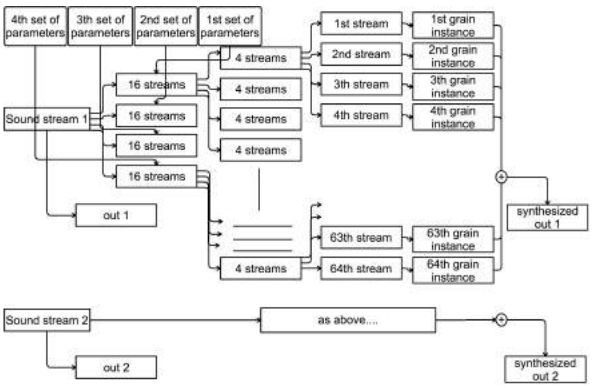

Figure 3. Granular streams architecture for two different audio sources. Each original audio input is split into 4 different classes of streams characterized by different granular parameters. The resulting asynchronous granular streams are summed together into a synthesized output stream.

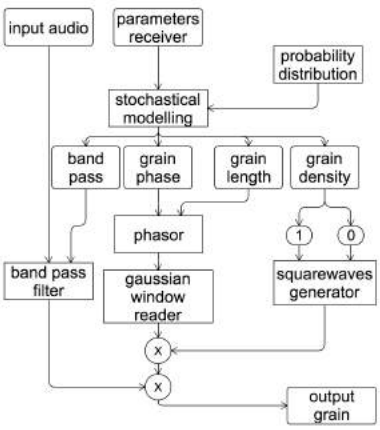

Figure 4. Grain structure. Each of the 64 audio streams for each audio source are filtered by a band pass filter depending on the user's position and enveloped according to the input parameters.

The streams of grains is obtained imposing to one audio a train of windows function of short length. Granular synthesis has been designed to operate on individual audio stream from each source. In order to obtain 64 granular sub-streams for each source, the main stream has been recursively divided in 4 sub-stream for 3 times as shown in Figure 3. To make the final stream richest and as various as possible, each of the first 4 sub-streams (see Figure 4) are defined with a set of parameters controlling duration, density, mid-frequency and Q factor of the granular stream. Each of these parameters are then stochastically modeled depending to a dynamic variance parameter. All the 64 sub-streams are then summed in one output resulting stream. Even more the resulting streams relating to each source are directed in one last output audio flow. To develop in a various and plentiful way each of the 4 sub-flows sharing the same set of control parameters, they have been modeled through a probability distribution such that there could be no two sub-streams identical in any aspect.

5. EVALUATION

To provide a first evaluation of the soundscape generated in the Interactive Soundscape environment we prepared an online listening test with the following aims:

1. to assess if the parameters of the audio engine which produce the computer generated soundscape respond to the perceptual expectations of the listeners
2. to evaluate how listeners perceive the computer generated soundscape
3. to see how listeners link the computer generated soundscape to a real environment.

Thirty-one invited listeners took part in the test, 25 males and 6 females. Subjects were aged between 20 and 57 years (mean = 31.5 years, *SD* = 8.8). Seventeen subjects have studied a musical instrument for more than 5 years (54.8%), nine have studied a musical instrument for less then 5 years (29%), and five have never studied a musical instrument (16.1%). For the test we selected an audio excerpt recorded during an exploratory session performed by a single user, whose soundwalk path is depicted in Figure 1. The excerpt lasts approximately 1 min and 30 s and comprehends both soundscape and granular synthesis produced sounds, with granular processing beginning at s16. Figure 5 reports a rough reconstruction of the event appearance and timing in the first 21s of the audio excerpt chosen for user's evaluation. In this representation at least five events are detected before the user enters the granular processing zone (waves, train, birds, bell and seagulls) [8]. The audio excerpt is not a simple soundscape recording but is a computer generated soundscape which refers to a real soundscape but which is not real. Beyond reverberation effects, it contains sounds obtained through a granular synthesis process which can significantly affect the perception of the user. Thus, to assess if the parameters of the audio

[8] This interpretation is not unique. Due to the complexity of the soundscape recordings employed and to the differences in the listening conditions (which in an online test cannot be easily controlled), different listeners may report different events or a different event order.

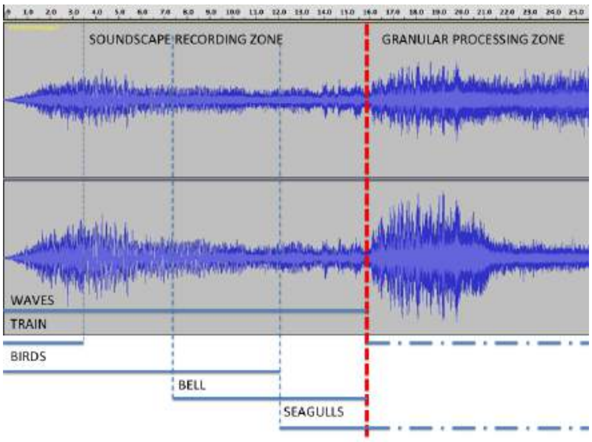

Figure 5. Graphic representation and event timing in the first 21s of the audio excerpt chosen for user's evaluation.

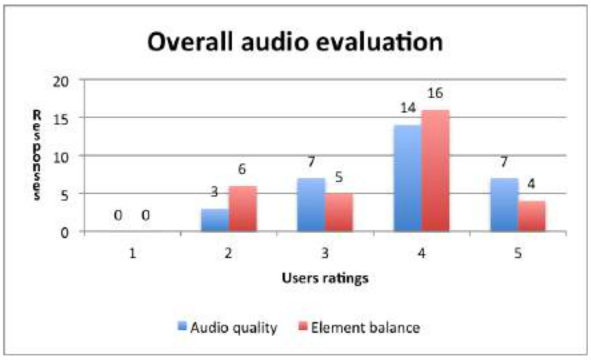

Figure 6. Histogram of users ratings for audio quality (in blue) and soundscape elements balance (in red), on a range from 1 (very bad) to 5 (very good).

engine respond to the perceptual expectations of the listeners, we asked our subjects to assign a rating ranging from 1 (very bad) to 5 (very good) to the audio quality and to the soundscape elements balance of the audio excerpt [9]. Results reported in Figure 6 show that the good rating (4) is the group which collected the majority of listeners responses (45.2% for audio quality and 51.6% for element's balance).

For the second part of our test we relied on the method described in [13], which identify *pleasantness* and *eventfulness* as the two main components for soundscape evaluation. Adapting the circumflex model of affect described by [14] to the soundscapes qualities, the authors define a circular two dimensional space which includes eight soundscape descriptors: eventful, exciting, pleasant, calm, uneventful, monotonous, unpleasant and chaotic. Thus, we asked our listeners to evaluate the *pleasantness* and *eventfulness* of the computer generated soundscape on a scale ranging from 1 to 5 in order to build a perceptual map of how listeners perceive the computer generated environment. The results are reported in Figure 7. The major number of positions are concentrated in the *eventful–exciting–pleasant* quarter of the perceptual map area and collect an

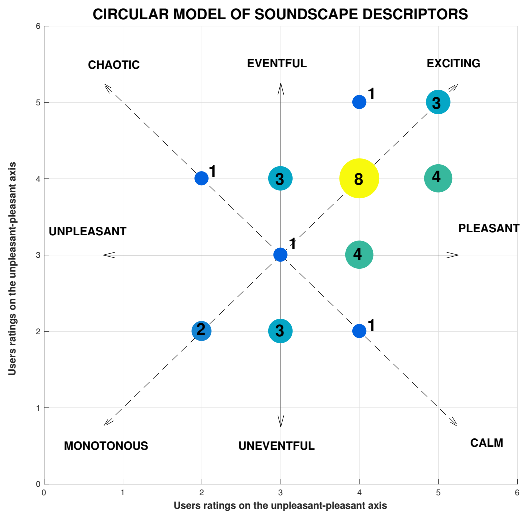

Figure 7. The perceptual map resulting from the ratings of the participants to the computer generated soundscape test. The numbers near the circles show the collected responses for each position.

overall of 24 responses. The greatest group (in yellow, with 8 responses) stands at point 4 on the *monotonous–exciting* diagonal axis.

In the version of Interactive Soundscapes employed for the test we reproduced a harbor soundscape through the use of many sound recordings. The result is a complex and multifaceted soundscape, which offers to listeners various aspects of the environment and which could also lead to different interpretations. To see how listeners linked the audio excerpt to a real environment, we asked our subjects to describe at least one physical place that could have created such a soundscape. A textual analysis of the answers [10] reports that the top phrase containing 2 words is "train station" with 6 occurrences. The single word with the major number of occurrences is "harbor" (11) followed by "station" (7), "train" (6), "port" (4) and dock (4). Words referred to natural places or sounds such as "beach", "rain", "foggy" and "sea" report only 1 occurrence. No occurrence is reported for seagull or bird sounds.

6. CONCLUSION

More than a half of the test subjects assigned a good or very good rating to the audio quality and element balance of the computer generated soundscape. This finding, together with the fact that more than 77% of the test subjects assigned an evaluation between 4 and 5 in the eventful-exiciting-pleasant zone of the circular map of soundscape descriptors, seem to indicate that this implementation of the Interactive Soundscape application can generate sonic experiences of good quality. Moreover, the soundscape is more exciting than monotonous and more pleasant than unpleasant, and this is a result which could not be given for

[9] The complete Likert scale is 1(very bad), 2 (bad), 3 (medium), 4 (good) and 5 (very good).

[10] The online tool used can be found at `https://www.online-utility.org/text/analyzer.jsp`

granted in advance considering that the audio excerpt is the product of a soundwalk in a virtual environment and not the reproduction of a real soundscape. Due to methodological and technical difficulties it was not possible to assess if subjects could perceive the sound changes obtained through the granular synthesis and how they evaluated them. This is an important task which requires a strict control on the listening conditions and a major control on the audio file's content. As a matter of facts, the presence of a great number of different sounds, both in the foreground and in the background, can confuse the listeners and does not guarantee a sufficient sound discrimination. No textual description of the physical place which could have generated the soundscape was delivered to the test subjects. They spontaneously identified it mainly as "train station" and "harbor", showing thus that the sound rendering is recognizable as a real soundscape and that it is consistent with what we wanted to reproduce.

6.1 Further work

The next steps for this work will be to collocate it in a 3D audio context using a wave-field synthesizer and providing the system to detect interactions from more people at the same time. The sound information will be local instead of being the same in the whole area, and the parameters collected from the participants will be used only for the manipulation of the dynamical events over the soundscape.

Acknowledgments

We want to thank Leonardo Amico for his support in programming the *Zone Tracker* application.

7. REFERENCES

[1] R. M. Schafer, *The soundscape: Our sonic environment and the tuning of the world.* Inner Traditions/Bear & Co, 1993.

[2] B. Truax, "Soundscape composition as global music: Electroacoustic music as soundscape," *Organised Sound*, vol. 13, no. 2, pp. 103–109, August 2008.

[3] F. Pachet, "Active listening: What is in the air?" Sony CSL, Tech. Rep., 1999.

[4] P. Oliveros, *Deep listening: a composer's sound practice.* IUniverse, 2005.

[5] (1999). [Online]. Available: http://www.sfu.ca/~truax/vanscape.html

[6] A. Camurri, S. Canazza, C. Canepa, G. L. Foresti, A. Rodà, G. Volpe, and S. Zanolla, "The stanza logomotoria: an interactive environment for learning and communication," in *Proceedings of SMC Conference 2010, Barcelona*, 2010.

[7] M. Mandanici, A. Rodà, and S. Canazza, "The harmonic walk: an interactive physical environment to learn tonal melody accompaniment," *Advances in Multimedia*, vol. 2016, p. 2, 2016.

[8] K. Eng, A. Babler, U. Bernardet, M. Blanchard, M. Costa, T. Delbruck, R. J. Douglas, K. Hepp, D. Klein, J. Manzolli *et al.*, "Ada-intelligent space: an artificial creature for the swissexpo. 02," in *Robotics and Automation, 2003. Proceedings. ICRA'03. IEEE International Conference on*, vol. 3. IEEE, 2003, pp. 4154–4159.

[9] G. Eckel, "Immersive audio-augmented environments," in *Proceedings Fifth International Conference on Information Visualisation*, 2001.

[10] M. Wright, "Open sound control-a new protocol for communicationg with sound synthesizers," in *Proceedings of the 1997 International Computer Music Conference*, 1997, pp. 101–104.

[11] M. Puckette, "Pure data: another integrated computer music environment," in *Proceedings of the second intercollege computer music concerts*, 1996, pp. 37–41. [Online]. Available: https://puredata.info/

[12] S. Zanolla, S. Canazza, A. Rodà, A. Camurri, and G. Volpe, "Entertaining listening by means of the Stanza Logo-Motoria: an Interactive Multimodal Environment," *Entertainment Computing*, vol. 2013, no. 4, pp. 213–220, 2013.

[13] Ö. Axelsson, M. E. Nilsson, and B. Berglund, "A principal components model of soundscape perception," *The Journal of the Acoustical Society of America*, vol. 128, no. 5, pp. 2836–2846, 2010.

[14] J. Posner, J. A. Russell, and B. S. Peterson, "The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology," *Development and psychopathology*, vol. 17, no. 03, pp. 715–734, 2005.