Developing an Synthetic Binaural Interactive Soundscape Based on User 3D Space Displacement Using OpenCV and Pure Data

Isaac Batista(✉) and Francisco de Paula Barretto(✉)

Laboratory of Research, Creation and Innovation, Federal University of Bahia, Salvador, Bahia, Brazil
isaac.dneves@gmail.com, francisco.barretto@ufba.br

Abstract. This poster describes a transdisciplinary research concerning the development of an interactive binaural soundscape that responds to user 3D displacement. Soundscapes can be described, according to R. Murray Schaffer (1977) in his book "The Tuning of The World" as an immersive environment which is defined by the combination of all acoustic resources, natural and artificial within a given area as modified by environment. These soundscapes might be generated through the recording of an "natural" specific environment using ambient microphones or it might be completely synthesized using pre-recorded acoustic resources in order to simulate an specific soundscape. The development of artificial soundscapes opens up a whole new range of possibilities such as the development of virtual environments that are able to create an immersive context that allows the user to feel inside the virtual environment. When we create a soundscape that is able to respond to user displacement in a 3D environment like a room, for example, can navigate in a virtual 3D sound environment. Trough the use of a system of cameras we are able to track the user position in a 3D environment and adjust accordingly the previous created soundscape using binaural synthesis. The present project uses OpenCV and Pure Data in order to develop this interactive environment that might be used to develop games for the deaf and to create soundscapes to Virtual Reality environments.

Keywords: Binaural · Soundscapes · Interactive · OpenCV
Pure data · Virtual reality

1 Introduction

Acoustic environments have several types of sound sources. As sound comes from vibrations, anything that is not completely static emits sound, at different intensities, tones and timbres. With so many sound sources generating so much information, we can't pay attention to almost everything we hear because we start to consider it as noise. We can take as an example a person who lives

C. Stephanidis (Ed.): HCII Posters 2018, CCIS 851, pp. 231–236, 2018.
https://doi.org/10.1007/978-3-319-92279-9_31

near an airport and is disturbed by the noise of neighbors upstairs while not even noticing the noise coming from the planes that take off and land near their residence. The more we are exposed to sounds the more we become used to the point of not noticing them anymore.

According to the ISO 12913-1 [1] soundscapes are constructed from a relation between physical and perceptive phenomena, and can be seen as the human perception of the acoustic environment. The soundscape can be analyzed differently by each person and in different contexts. This research aims to propose an interactive binaural soundscape in order to better understand the relationship between individual, place and context that influences the experience of different people in a particular acoustic environment.

We must use reproduction techniques, from a technological perspective, focusing on the perception of the soundscape. The physical characteristics of the sound that allow the location of the sound source in space and the perception of all interaction of this source with the environment are interpreted when there is a difference of information between the two ears. Binaural reproduction allows us to perceive this difference of information so that we can simulate the soundscape with fidelity.

2 Theoretical Background

2.1 Soundscapes

We can call soundscapes the sound or combinations of sounds that form an immersive environment. Characteristic bells, noises, musical compositions, the specific sonority of each environment, all together creates a soundscape. The soundscape can be defined by three elements: figure, background and field [2]. The figure is where the focus of interest of the listener is concentrated, the background are other elements that make up the scenario where the figure is and the field is the place where everything occurs and that creates this figure-background relationship.

Focusing on sound perception also means focusing on geographic studies that emphasize the importance of a place within a process of meaning-making composed of elements such as geographic coordinates, technological means, social arrangements that develop and many other information. In historical research, hearing has a major disadvantage because the lack of records does not make it possible to have some information about the loudness and intensity of certain objects and environments. Unlike the visual, all the historical contribution we have about the sound landscape is through auditory testimony.

The great problem of auditory testimony is precisely how the relation of sound sources between figure and background is determined. As in Schafer's own example, anyone looking into the clear water of a lake can perceive the reflection itself or the lake bottom, but not both at the same time [2]. From the moment that attention is focused on one sound source, the other becomes an unperceived background and this information is lost in the testimony.

No matter how much the physical dimension contributes, the consideration of sound as a figure or background is intrinsically linked to the habits trained through social and cultural construction and the individual's interest and state of mind. Sound is classified by its physical characteristics, the way it is perceived, its function and the emotional and affective states that it stimulates. All these subjective and objective parameters cannot be analyzed separately when it comes to sound landscape, it is necessary an interdisciplinarity that creates a correlation between the hi-fidelity records and the listener's perception of that sound source.

We relate to the sound imagining visual representations of its reproduction, we perceive its positioning between the speakers, whether they are loudspeakers or headphones [3]. When we imagine or perceive a sound as if it is in the center between two speakers it is not there in fact. What happens is that it is coming out with the same intensity and at the same time of all the speaker and giving us that feeling of symmetry. If we make a panning of this recording, we can move its image left or right, inside the limits defined by the position of the speakers. In addition, we also create a notion of spatiality through effects like delay and reverb, that adds to the sound record a sense of depth. But even so, it is necessary to use extra tools and methods to create a high-fidelity auditory immersion.

2.2 Binaural Sound Representation

Let's separate two audio playback modes between spatialization and auralization. The first one searches for surrounding sound fields, with ambience, through signal processing based on psychoacoustics that give a sense of spatiality. On the other hand, auralization aims to simulate legitimate sound fields with the highest fidelity possible based on wavefront propagation, generating a perception of a sound field that integrates the listener, the primary source (figure) and the secondary sources (background) that composes the sound field [4]. Auralization can be achieved by multichannel through a speaker-matrix or binaural, using a two-channel playback based on the auditory spatial cues of psychoacoustics. The latter is much more relevant to this research and therefore will engage our discussions.

Binaural reproduction is based on how our brain processes the acoustic information that reaches our ears. It is necessary to use a headphone for this technique, because this way ears will receive a direct sound wave, without influence of the external field after the reproduction of the speakers. Our auditory system can distinguish directivity and distance from sound sources from Head Transfer Functions (HRTF). The main HRTF indicators are the Interaural Time Difference (ITD) and the Interaural Level Difference (ILD), which analyze, respectively, the time and intensity difference that the same sound is perceived between one ear and the other [5].

In high-frequency sounds, that have a small wavelength, an acoustic shadow forms in the farthest ear from the sound source, which leads to a decrease in intensity in comparison to the closest ear and makes the ILD quite noticeable. At low frequencies, which have longer wavelengths, it does not generate an acoustic

shadow but a delay occurs in the wavefront, which makes it possible to perceive the direction through the ITD.

ILD and ITD are auditory locator indicators for the horizontal plane. For the vertical plane, the indicators are monaural, which generates a spectral alteration in the sound due to reflections of the chest, head and external ear. These changes occur due to the interference of the reflected waves mentioned above and by direct waves, with a lag that increases some frequencies and degrades others.

Finally, the distance can be perceived mainly by the variation of intensity. In open environments, the sound pressure level (SPL) decreases by six dB when the distance between the listener and the source is doubled, when the distance is decreased by half the inverse happens. While in a closed field this variation of SPL happens due to the relation between direct field and the field of reverberation (where there is predominance of the direct field) the law of attenuation for open environments is still valid. From the point where the reverberant field predominates we evaluate the time difference between the direct sound, which travels a minor path and therefore arrives first, and the reverberant sound, which comes later by traveling a greater path and giving us the impression of origin of the walls.

Throughout life we have memorized this great variety of transfer functions corresponding to the directions and distances of the sound sources and with this we weigh all these functions and distinguish the exact location of the sound. Even though the HRTF of each person has variations due to existing physical differences, the result is still satisfactory and effective, allowing to create emulations with optimum fidelity.

3 Binaural Interactive Soundscape

When we hear someone testify about a sound landscape when we are already familiar with it, we easily recognize and visualize what is being said, because we already have that landscape formed in our mind. When the sound landscape is unknown, the person creates an image quite different from the one the witness wanted to pass on to it, as in most of the information that carries a subjectivity. Taking Salvador as an example, which is a tourist city and with a strong sound peculiarity, it is remarkable the lack of understanding of the people of other cities when someone tries to comment on the sound of the city.

The proposition of the binaural interactive sound landscape is to get the listener to have a first-person experience in the field of sound rather than an external presentation. This will make it possible for him to walk through characteristic places with a faithful sound record and make his own analysis of the sound of the place, feeling inserted in it. In addition, this experience also allows you to recognize sound sources and map their position. As described in Fig. 1, this research is based on computer vision system that will recognize user displacement and through TUIO protocol will send real-time information to a sound synthesis patch written in Pure Data.

The sounds used will be sounds that refer to specific places, precisely so that we can insert people who are not accustomed to that place in a sound field

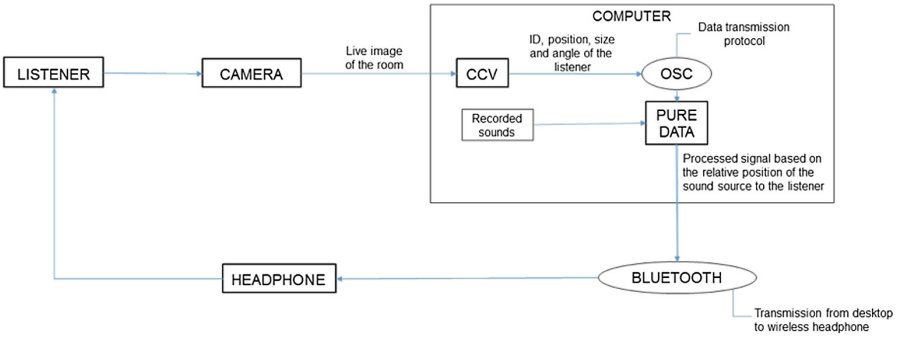

Fig. 1. High level diagram describing the system software and hardware involved

with elements other than what they usually hear, and thus have with them the experience of having to contemplate the and create their own relationship of figure-background in that sound field. We intend, initially, to use the sounds of the historic center and/or the fairs of Salvador, due to identity and also logistics, since it is the city where the University and the laboratory are located. But the intention is to create sound landscapes from other cities as well, to have a larger catalog and allow a better evaluation of reaction and interactivity with unfamiliar soundscapes.

The sound objects will be samples of recordings with an acceptable resolution and made at strategic points to facilitate the mapping and the creation of the dynamics of movement in relation to the listener.

The positioning and displacement mapping of the user will be done by an infrared camera, which makes the color difference not a problem and allows the mapping even in the dark environment, and thus incite a more intense hearing immersion. The camera will be used as an image source for Community Core Vision (CCV). CCV reads a video stream, identifies the object contained in the video, and decrypts data from the object (e.g. coordinates, size, rotation angle).

Since we will have only one person at a time participating in the experiment, Session ID, which is what identifies the video object, will only import to be used as a transfer of the other data sent. Other parameters that are also not relevant are the speed and rotation speed, after all, it is sent after the movement of the object and if the audio synthesizer is to process and execute the speed after the listener is already in the next position there will be a latency large and that will increase with the chain reaction. Because the change of direction and intensity of sound is not procedural between one position and another, it is necessary that this change occur in the shortest time possible, because in addition to making the differences are in small levels, our brain understands small time differences only as distance variation processing. Thinking about this, the ideal is to use a camera with 60 frames per second, so that it would work with 16.6 ms between the samplings, entering the concept in the Hass effect [6].

The data will be sent to the audio synthesizer, which in this case will be Pure Data, via Open Sound Control (OSC)/TUIO described in [7]. After the Pure Data patch receives the angle and user position in x and y coordinates, we calculate the azimuth with the difference between the angle of the position of the sound source in relation to the user position. After this information is calculated we can update the parameters of the synthesized audio samples.

4 Expected Results

Since we are going to do the experiments in a flat room, and at the beginning the parameter of the sound source elevation in Pure Data was not very efficient, the location will be much more noticeable in the 2D plane. But even so we expect a much greater efficiency than the stereo system with only panning and volume as parameters of 2D position, because, as said before, in stereo we have a sound image in 3rd person, while the use of azimuth with Earplug allows us to create an immersive sound field.

We will try to experiment with people who already know the soundscapes that will be reproduced so that they can speak about the similarities and differences of the natural sound fields for the synthesized recordings for binaural. We hope to see how the latency of all processing will work, and even if it occurs with considerable value, how much it will influence the perception being that the hearing will be the only reference in a dark room.

References

1. Axelsson, S.: The ISO 12913 series on soundscape: an update, May 2012. J Acoust. Soc. Am. **131**(4), 3381 (2012)
2. Schafer, R.: The Tuning of the World. Knopf, Borzoi book (1977)
3. Gibson, D., Petersen, G.: The Art of Mixing: A Visual Guide to Recording, Engineering, and Production. MixBooks, Mix pro audio series (1997)
4. Faria, R.R.A.: Auralização em ambientes audiovisuais imersivos. Ph.D. thesis, University of São Paulo (2005)
5. Zhang, W., Samarasinghe, P.N., Chen, H., Abhayapala, T.D.: Surround by sound: a review of spatial audio recording and reproduction. Appl. Sci. **7**(5), 532 (2017)
6. Haas, H.: The influence of a single echo on the audibility of speech. J. Audio Eng. Soc **20**(2), 146–159 (1972)
7. Kaltenbrunner, M., Echtler, F.: Tuio hackathon. In: Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces ITS 2014, pp. 503–505. ACM, New York, USA (2014)