ABSTRACT

The main goal of this paper is to describe certain aspects of the construction of a prototype, while introducing generally some of the key issues resulting of the design of its interface. The word LeinKlang (from the German) means sound (Klang) screen (Leinwand). This project consists in the construction of a digital music instrument. Physically, it appears as a screen, against which users interact, by shadow or video projection, and behind it, intentionally out of view, a set of 24 small loudspeakers.

Keywords – Multimedia and AudioVisual Creation, Interface Design, Perception, Sound and Visualization.

1. INTRODUCTION

The 24 loudspeakers that form the instrument’s “sound projection screen”, are homogeneously distributed in the area of the rectangle defined by the perimeter of the screen (3.5 x 2.5 meters). This instrument, besides the usual characteristics of a digital music instrument [3], allow one to “visualize the sound”, its movement speed and its spatial positioning (in the first stage through Cartesian coordinates – x,y). The speeds at which the user’s actions are performed, as well as the distance between it and the screen introduce variables both in the frequency and sound dynamic scopes. The different combinations of these variables simultaneously allow spectral sound manipulation, such as for example, the association of certain coordinates to region-specific filters or the convolution of different signals. Since this project is an effort to better understand the different modalities of digital music instruments control and manipulation, the inclusion of some thoughts on the matter of sound visualization [9] is in order as well as game modes associated to the audiovisual binomial relationship. To this end, we propose two versions of the instrument. The first one with visual support (see fig.1) and the second version only with sound transformations.

2. CONCEPT AND PROJECT DESCRIPTION

In a global context, we can place this proposal in the widely ranging scope of games. In an initial approach, the contact and interaction with the user is expected to induce an attractive and intuitive entertaining state [7]. In a deeper approach and regarding a higher musical knowledge, the user may be able to effectively compose sound/music consciously and with full mastery of the wide range of possibilities the instrument offers. Finally, in an even more complex approach, the instrument is thought to allow one past the game aspect and the structural composing aspect into the aforementioned sound visualization. The consequence of interactions between the approaches explained above (game, composition and visualization) allow us to establish a universe that can be characterized by a fictitious consciousness. It can be said that this universe characterized by a fictitious consciousness, introduces a pioneering and innovating character, regarding the global landscape when it comes to the building of digital music instruments.

3. INTERACTION WITH THE SONIC MATERIAL

Physically, the instrument presents itself with a grid of loudspeakers, arranged in a matrix of eight columns (vertical) by three rows (horizontal). This grid may be hidden behind a screen, over which the user’s silhouette is projected in real time, or, alternatively, it can appear
“as is”, allowing some of the hardware components to be visible: the loudspeakers socketed into small plaques of transparent Plexiglas, flanked by the individual audio amplifiers and all the cabling that it entails. By allowing the presentation of the installation to be twofold, the opportunity to study two different kinds of interaction with the musical instrument arises [3].

Therefore, on one hand we have an instrument that tries to direct the interaction through the projection of the intervenient body on the screen. On the other hand, the second variant proposes to induce the user into the domain of acoustic perception, where the musical parameters require a much more accurate and focused listening. (see fig.2).

Two web cams, laid in opposing axes (on the horizontal plane) track the movement of the user. They are hidden inside two of the five columns that dictate the interaction space. The computer that processes the motion tracking algorithm and the sound generation is also hidden in the central column.

In the screen version of the installation, the user has at his disposal in front of the amplifier grid a surface that mirrors (with more or less distortion) his own movement and spatial location (see fig.2). These two aspects will unleash the interaction process. Thus, the user’s movement limits the number of sound sources that are active at a given time, and, consequently, determines the sound projection’s dynamics in a bipolar way [3]: the farthest away the user is from the grid, the more the instrument attempts to compensate the dynamic, increasing the volume adequately. However, this aspect also indexes an analogy to the dimension of the sound sources’ acoustic space, where the architectural volumetry grows in proportion to the sound volume. In this version, the sound sources that are bound by the shape of the user’s body emit generated sounds, or a number of small sound samples. Globally, the interaction can be characterized by a metamorphosis of a particular sound universe to a global universe [2].

In the version without the screen, bearing in mind that one intends to induce the user into a much more detailed and accurate sonic universe, all the sound are active at all times, emitting generated sounds or samples up to a maximum number of twenty four different instances (number of loudspeakers available). However, considering the particular to global approach [2], the coordinates of the user’s head perimeter allow that the corresponding sources (bound by an almost spherical shape) excite and modify themselves in a much more rich and varied fashion, in comparison with the rest. Thus, it is possible to experience a kind of background sonic landscape, disturbed and modulated by precise elements and subject to a specific composition according to the user’s manipulation.

The sound generation and treatment was programmed in Max/MSP, and the multi channel distribution algorithm based on the SPAT library of the IRCAM.

4. MOTION TRACKING MODULE

This module’s objective consists in elaborating the algorithm that can consistently track the user’s location in relation to the screen or the Plexiglas plaques. Initially, we wanted to track seven different points, distributed as key points of the human silhouette (head, shoulders, hands and knees), however, it became obvious the much of the loudspeakers’ sonic information would become redundant to the user; being close to the loudspeakers, the user would have little to no perception of the sonic matter dictated by the position of his knees, for example. We then decided to modify the algorithm in such a way to locate and track only the head’s position (also calculating its shape).

The chosen tracking method was the optic flux, namely the Lucas Kanade method. Optic flux defines the analysis of the quantity and direction of movement, usually translated into vectors between pixels [6], The Lucas Kanade method is characterized by a great robustness when tracking the perimeter of relatively large objects (the variation of its perimeter in the video matrix), and its weakness in detecting rapid movements.

As regards the different programming steps, the chosen items were the programming kit Max/MSP Jitter [10] and the Jitter library, CV.jit library by Jean-Marc Pelletier [11], that features a number of objects based on the Lucas Kanade method. Methods of background subtraction and blob tracking were used in this project. Background subtraction refers to the elimination of all visual information in the video matrix that is not relevant. In this case, all information that is not related to the user, and therefore not subject to motion tracking.
(the grid or the screen). Later, the resulting video matrix is binarized, so that all the surviving information (the remainder of the background subtraction) is presented on the matrix as a white blob (user) against a black background (subtracted) (see fig3.)

![Figure 3](image3.png)

From the moment a blob is detected, the CV.jit library presents various pieces of information related to it, namely:

- Centre: coordinates of the centre of the area covered by the blob;
- Orientation: angle (in radians) of the approximate disposition of the blob’s area;
- Involving rectangle: rectangle resulting from the area described by the most extreme pixels in the blob’s perimeter, on both the x and y axis (4 values: \(x_1, y_1, x_2, y_2\)).

With this information, we were able to calculate the coordinates of the head in the blob that represents the user in the video matrix. The y coordinate is simply the highest y value of the blob perimeter. The calculation of the x coordinate is not so direct, being obtained through the orientation of the resulting blob. Therefore, it was necessary to convert the angle from radians to Cartesian coordinates, where x and y are the latter, \(r\) is the radium, and \(\alpha\) is the former (see form. 1 and form. 2).

\[
y = r \cos \alpha \\
x = r \sin \alpha
\]

Bearing in mind that the cv.jit.orientation object calculates the angle, we need only to find the radium, which is the value of the distance (in pixels) between the blob’s centre and the y coordinate (highest y values in the blob’s perimeter). After a scaling process to fit the resulting values into the captured video matrix interval (0 to 119), the coordinate set corresponded, with an insignificant margin of error, to the positioning of the user’s head (see fig4).

![Figure 4](image4.png)

5. REFERENCES


6. ACKNOWLEDGMENTS

This project was carried out with the support of CITAR – Centro de Investigação em Ciências e Tecnologias das Artes da Universidade Católica Portuguesa – the Fundação para a Ciência e Tecnologia, Portugal and ZKM| Zentrum für Kunst und Medientechnologie – Institut für Musik und Akustik of Karlsruhe. Special thanks to Daniela Coimbra, Álvaro Barbosa and António de Sousa Dias for reviewing this work.