dots: an Audio Entertainment Installation using Visual and Spatial-based Interaction

Nikolaos Grigoriou, Nikolaos Moustakas, Andreas Floros and Nikolaos Kanellopoulos

Dept. of Audiovisual Arts, Ionian University, 49 100 Corfu, Greece
{av200432, av200406, floros, kane}@ionio.gr

Abstract. “dots” is an interactive sound installation that takes into account the spatial position of an arbitrary number of participants in order to algorithmically synthesize an audio stream in real-time. The installation core is a software application developed during this work, which employs advanced video and audio processing techniques in order to detect the exact participants’ positions and to weight-mix short audio granules. Audio mixing is performed using a virtual spatial gridding of the installation space in two dimensions. The synthesized audio stream reproduction is combined with a number of appropriately designed visual effects, which aim to enhance the participants’ comprehension and render the “dots” installation a high-quality interactive audiovisual platform.

1. Introduction

The continuous growth and evolution of digital audiovisual technologies is providing the scientific framework to design and develop/demonstrate new ways of artistic expression. The recently established high-definition video and multichannel audio formats and standards are becoming widely accepted by both the media-producers’ and the end-users’ market. Additionally, new terms and ideas originating from the general concept of interaction are nowadays frequently used to provide novel means of audio and visual production, where the audience is actively participating in the production process. As a consequence, a continuously growing number of new media artists have now the option to express their thoughts and feelings by developing and demonstrating their installations, using high quality audio/visual synthesis and playback techniques and equipment [1].

The above interactive installations can be also considered as advanced tools for creating and synthesizing in real-time high-quality audio/visual content. For example, in a typical case, the visual information obtained or processed from an interactive installation can provide feedback to a sound synthesis system for producing novel sound streams in real-time. The opposite approach can be alternatively employed for synthesizing complex visual content and environments using sound recorded signals as input. In any case, the user interaction is usually realized using an interaction algorithm that defines the rules under which selected synthesis parameters are varied. The variation amount is usually obtained from the installation environment, typically using a wide range of wired or wireless sensors, image capturing devices and video and/or audio recording equipment.

In this article we present the “dots” interactive platform, which aims to produce complex audio streams taking into account the movement and the instantaneous spatial positions of the participants. In order to provide a motion motivation to the participants, an additional visual component is produced and concurrently projected, which also depends on the participants positions and motion. Hence, a complex spatial, audio and visual interaction effect is achieved which apart from dynamically creating audio content can be also considered as an interactive game with colours and sounds. The visual content consists of a controlled number of dots with varying colours, providing the name “dots” to the overall installation platform.

The “dots” interactive platform presentation is here mainly performed in technological terms, typically focusing on the platform architecture and some elementary design / algorithmic issues and concepts. Additionally, within the framework of this work, we investigated and evaluated the interaction functionality and the overall installation performance under real-world conditions during an audiovisual festival / exhibition. This evaluation process also allowed the collection and interpretation of some observations related to the behaviour and the means interaction of the participating audience.

The rest of the paper is organized as following: In Section 2, the “dots’ installation architecture is described in detail, focusing on both the core application and the installation room requirements and design. Next, a demonstration of the installation interactive features is presented in Section 3, followed by a brief analysis of the functional and behavioural observations made during an installation exhibition. Finally, Section 4 concludes this work and accents further interaction and audio/visual enhancements that may be integrated in the “dots” platform in the future.

2. Analytic installation description

As mentioned previously, the “dots” platform developed during the present work is an interactive audio/video installation, an interactive game, combining space, time and audio/visual content. From the design point of view, it consists of a) the core application, which is responsible for handling the basic input / output video and audio signals and for realizing all the signal processing necessary for recognizing the spatial placement of the participants as well as for appropriately mixing and finally producing the audio signal and b) the installation space, where the participants are moving and interacting with the core application, producing the final, complex audiovisual output. Both core application and installation space subsystems are described in detail in the following two Sections.

2.1. The “dots” core application

The application was programmed with the open source tool Processing [2], a powerful software sketchbook and professional production tool used in many fields of signal processing science and arts. A number of particular libraries were additionally used for implementing the “dots” installation application (see Figure 1). For example, advanced image and signal processing techniques were employed to identify blobs from video real-time captures and extract all the spatial information needed. The libraries that were used are the following: the video capture library (included in the basic Processing installation package), the minim [3] and the “blob detection” library [4]. With the order that they were listed above, the first code library is responsible for the input of the video image into the core application, the second for the reproduction/playback of the synthesized audio component and the third for the recognition of the outlines from the footage of the video signal.

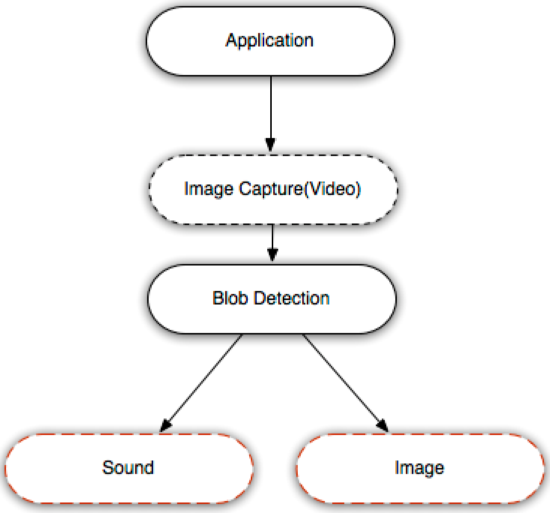


Figure 1: Core application architecture

For the accurate position recognition of the people moving into the installation space and in order to further process the incoming spatial information and to convert the extracted spatial information to an audio stream, a virtual grid was created. This grid was designed based on the coordinates of the specific installation space (and more particularly after testing where a participating person could stand or move in the room, in order to avoid non-active grid areas). Finally, for the purposes of this work, provided that the selected installation room has simple geometry with parallel surfaces only, the grid was designed to have nine discrete rectangle areas. It is also significant to mention that apart from the above discrete rectangle areas, some audio-fading grid territories were also defined within a specific distance of the grid borders. This distance was selected to be equal to 10% of a grid area width/length.
For synthesizing the visual component of the installation, the initially defined goal was to design very simple geometrical and coloured shapes that do not require detailed visual analysis by the participants. Hence, after an extensive sequence of design experiments, the final decision was to employ three circles (dots) with a set of colours of blue, yellow and red. As it is shown in Figure 2, these three dot types are also included into the “dots” installation logo. The appearance of these dots on the projection area of the installation is algorithmically defined by the instantaneous positions and the total number of the participants within a specific grid area. More specifically, a dot appears when one of the grid areas noted above gets triggered by the presence of a participant. On the other hand, the dot colour is defined by the number of the detected participants within the specific grid area: the blue dot appears when only one person is traced, the yellow when more than two people are traced and the red when there are more than three people in the specific active grid area.

dots

Figure 2: The dots colour shapes designed and the “dots” logo

Focusing on the audio output synthesis module, a number of sound granules were statically attached to each grid area. These granules are original, exclusively produced for the purposes of this work by the authors, with the use of an external synthesizer and a sound processing software tool. Alternatively, some of the audio granules are produced using a real-time sound synthesis engine (typically based on additive or FM synthesis) with predefined parameters.
The audio synthesis engine is activated similarly to the visual component, that is, when a participant is detected to be within a rectangle grid area. In this case, an audio granule attached to the specific grid area is randomly selected and mixed to form the final audio output. When the playback of this granule stops, another one is selected using the same selection algorithm, provided that the participant is still within the specific grid area borders. The random granule selection process is always performed using a normal probability distribution. When more than one participants are detected in the same grid area, no audio granules are selected. Instead, the playback volume for the corresponding audio granule is increased by 1dB for each additional participant. Hence, when all the audio granules are linearly mixed, the number of participants within a grid area defines the audio mixing-weights.
The playback of an audio granule immediately stops when no participants are detected within a grid area. Additionally, as mentioned previously, some audio-fading territories are defined closely to the grid border. When a participant is moving towards a grid border line, an audio fade-out process is activated, by linearly re-adapting the overall audio granule gain as a function of the distance to the grid border. A fade-in gain is correspondingly applied when a participant is moving away from the grid border (but he is still within the fade territory).

2.2. The “dots” installation space

The technical equipment that is needed for the realizing the complete installation is a personal computer (typically with a core2duo processor, 2GB of basic memory and ideally a high performance video processing expansion card), a High Definition video camera for creating the input video signal, a video projector with a minimum analysis of 1024x768 pixels and two monitor active speakers As it shown in Figure 3, the camera is located on the ceiling so that it can track and trace the movement of people that exist and move into the installation room. In the same Figure, the virtual grid mentioned in the previous Section is also illustrated.
The computer required for executing the core application is located inside the room, and preferably towards the rear of it. Based on the minimum requirements that the computer system must have, it would be good to take into consideration that the technical equipment must be silent so that there will not be any audience noise disturbance. Hence the computer was placed inside a special sound absorptive construction. The “dots” core application is running on the computer receiving the data from the camera and projecting the final result on the opposite wall with the help of the projector. As it is understandable the projector must be installed at the back of the room projecting the final visual effect on the opposite front wall.

The stereo loudspeakers employed are located next to the front two corners of the room and give feedback to the user with sounds that are produced each time with a different combination. In a future enhanced version of the installation, there will be reproduction surround sound from 4 or more loudspeakers in the room.
During this work, a number of important aspects related to the installation effective realization where considered, such as the existence of light (and more preferably the presence of ambient light) which is shed equally in the whole room, and the selection of a light coloured floor or a white carpet so that the blob detection accuracy is significantly increased. Moreover the colour used for painting the walls is an important installation parameter, as it should be the same for all room surfaces and preferably light (for example light blue or white).

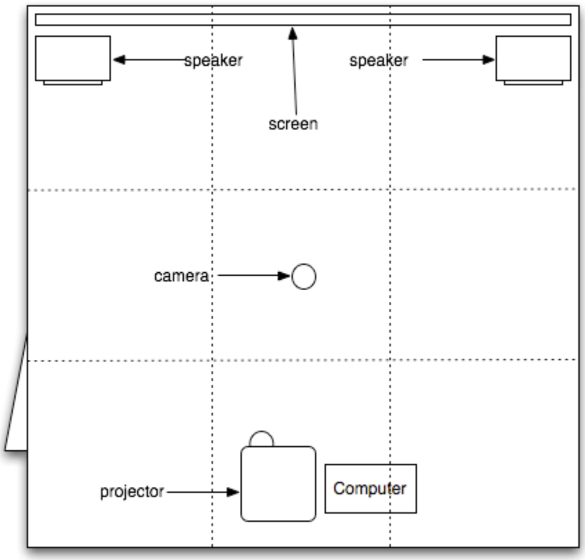


Figure 3: Typical installation space layout

Finally, a small CRT (or LCD) monitor is installed outside the installation room, in which the same reproduction of the projector inside the room is made. This means, that one visitor outside the installation room can see the exact position of each person together with the number of them that are in a specific part of the grid inside the room.

3. "dots" installation demonstration

The complete "dots" platform was realized for the purposes of the annual festival of the department of audiovisual arts, held in Corfu, Greece on May 2008. The installation room dimensions were 4 m (width) x 4m (length) x 3.4m (height). As mentioned previously, the camera was placed on the ceiling. Under these dimensions, the effective detection surface covered a satisfactory part of the installation room.
The colour of the room was plain and bright so that the contrast between the human and the room would be high. This condition is very significant for the blob detection algorithm, as it clearly defines the human edges/borders. Accordingly, a white carpet was placed on the floor in order to additionally increase the final video signal contrast. Additionally, the illumination inside the room had to be uniform and ambient; otherwise it would cause problems to the blob detection process, while the sensitivity of the detection was adjusted according to the specific illumination during the initial installation setup.
The equipment described in Section 2.2 had to be hidden in order to be transparent to the people participating into the installation and not to attract their attention. The camera that used was connected via a firewire interface, allowing fast and robust digital video signal transmission to the core application.
Finally, the gain of the different sound sources was appropriately and independently adjusted, depending on the acoustic properties of the room and the disturbance it would cause to neighbour installations.
An additional, important point of interest was to design an appropriate audiovisual sub-system to attract the attention of the people passing outside of the installation, as it was difficult to notice the installation functionality from outside. An ambient sound together with the small monitor that was installed outside of the installation room (as mentioned previously) showing the dots' patterns created by the participants' interaction finally (and efficiently) solved this problem (see Figure 4). The ambient sound excited the visitors' curiosity, so that most of them finally decided to enter into the installation room.

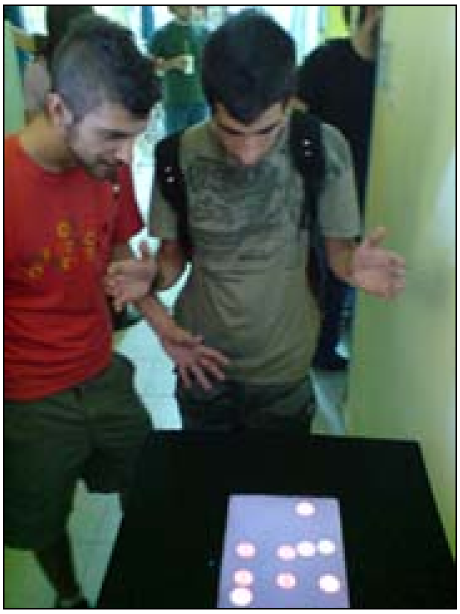
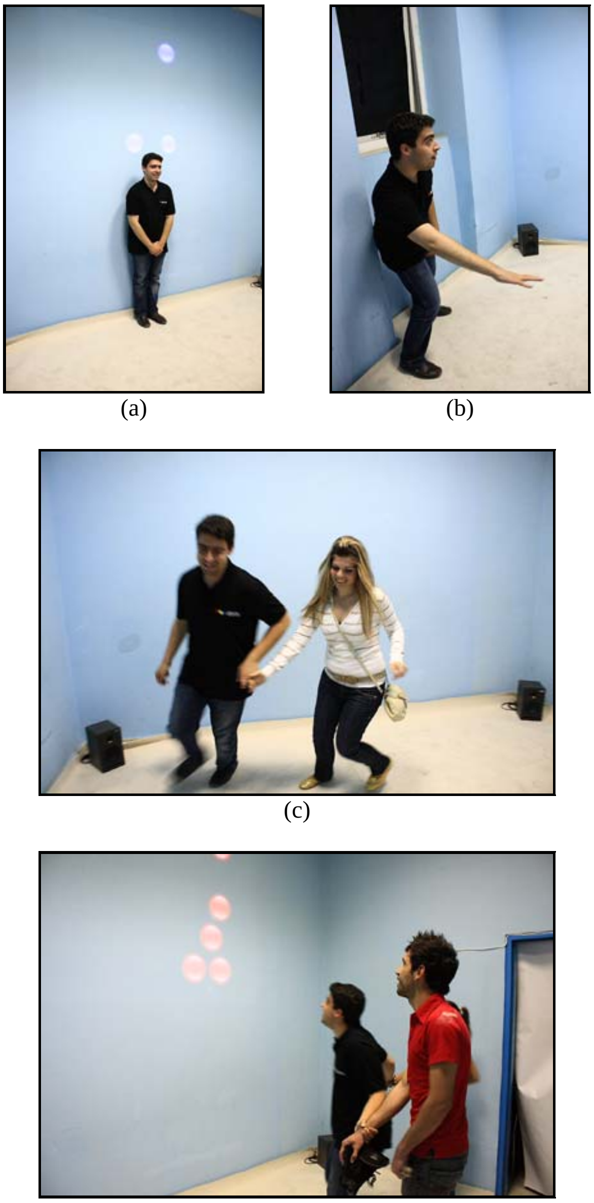


Figure 4: The external monitor subsystem

During a series of observations organized under real usage and interactions conditions with participants that didn't know anything about the experiments (see Figures 5(a) – (d)) it was found that the combination of both sound/audio and visual effects were the basic installation characteristics which gave "energy" to the audience movement. The visual content creates a strange feeling of following up, which in a small number of cases makes somebody moving and watching the projector without necessarily thinking about the existence and production of sound. On the other hand, some other participants were mainly interested in audio signal composition without paying much attention on the visual component. It was interesting that these users were making strange movements following the rhythm of the sound composition. However, the majority of the participants focused on discovering the relation between the audio and visual components by creating complex motion traces and using various spatial body movements. The discovery approach was audio-based (that is the participants were trying to relate the sound produced with the dots visual effect) or visual-based (the visitors were associating the visual effect with the produced audio stream). In most cases, the interaction algorithm was described as a complex audio/visual interaction, without a specific algorithmic structure. Specific movements towards the room borders were also frequently observed, aiming to determine the spatial interaction limits.



(a) (b) (c) (d)

Figure 5: Demonstrating the “dots” installation

Finally, it should be noted that during the interaction sessions, the complete installation was considered by many participants to be a complex game of sound and colours. Hence, in many cases, two or more participants decided to comprehensively play, producing interesting audio and visual effects.

4. Conclusions and ongoing work

In this work, an interactive sound installation prototype called “dots” is presented which aims to synthesize and produce high-quality audio streams in real-time, using a novel interaction algorithm that employs spatial and motion characteristics of the installation participants. More specifically within the installation room, the participants are motivated to move using a dynamically controlled visual content (i.e. a number of coloured dots, the number and colour of which depends on the number and the relative position of the participants within a virtually designed, two-dimensional grid). Accordingly, the human movement dynamically controls the sound synthesis engine, which performs weight-based mixing of short synthesized granules, taking into account the instantaneous participants’ positions within the virtual grid. Hence, a complex spatial, audio and visual interaction effect is achieved, which produces continuous audio variations and non-repetitive and time-variant sound patterns.

Apart from the analytic technical description of the proposed interactive sound synthesis platform and the discussion of a number design issues and concepts, during this work we additionally demonstrated the “dots” installation in the context of an audiovisual festival recently organized. A number of tests sessions and audience observations organized during the real-world installation demonstration have shown that the proposed interactive platform can be considered as both a visual-based or audio-based synthesis platform, as well as an audio-visual game with complex synthesis rules. Taking advantage of this approach, future enhancements of the proposed platform will be considered, mainly in terms of spatial audio reproduction through modern multichannel audio systems (such as 5.1, wavefield synthesis or ambisonics), panoramic, three-dimensional visual content projection using multiple video projectors and complicated, dynamically adjusted motion and interaction scenarios.

5. Acknowledgments

The authors wish to thank Mr. Loukas Ziaras for providing the photographic material presented in this work.

References

[1] N. A. Streitz, C. Rocker, Th. Prante, R. Stenzel, & D. Van Alphen, “*Situated Interaction with Ambient Information: Facilitating Awareness and Communication in Ubiquitous Work Environments*”. In Proc. HCI International, Crete, Greece (2003)
[2] http://www.processing.org (last visited May 29th, 2008)
[3] http://code.compartmental.net/tools/minim/ (last visited May 29th, 2008)
[4] (http://www.v3ga.net/processing/BlobDetection (last visited May 29th, 2008)