USING 3D SOUND FOR PROVIDING 3D INTERACTION IN VIRTUAL ENVIRONMENT

Vladimir Ortega-González
Arts et Metiers ParisTech, CNRS, Le2i
Institut Image, 2 rue T. Dumorey,
71100 Chalon-sur-Saone France
erikvladimir@gmail.com

Samir Garbaya
Arts et Metiers ParisTech, CNRS, Le2i
Institut Image, 2 rue T. Dumorey,
71100 Chalon-sur-Saone France
samir.garbaya@cluny.ensam.fr

Frédéric Merienne
Arts et Metiers ParisTech, CNRS, Le2i
Institut Image, 2 rue T. Dumorey,
71100 Chalon-sur-Saone France
merienne@cluny.ensam.fr

ABSTRACT

In this paper we describe a proposal based on the use of 3D sound metaphors for providing precise spatial cueing in virtual environment. A 3D sound metaphor is a combination of the audio spatialization and audio cueing techniques. The 3D sound metaphors are supposed to improve the user performance and perception. The interest of this kind of stimulation mechanism is that it could allow providing efficient 3D interaction for interactive tasks such as selection, manipulation and navigation among others. We describe the main related concepts, the most relevant related work, the current theoretical and technical problems, the description of our approach, our scientific objectives, our methodology and our research perspectives.

Keywords

3D sound, 3D Interaction, Virtual Reality, Human Computer Interfaces.

INTRODUCTION

Multimodal stimulation and multimodal interaction are two relatively recent research focuses. They aim at the study and development of new techniques for using two or more senses in interactive interfaces and virtual environments. The raise of the interest in these areas is explained by two main factors. The first is the mature current level of development of the visual stimulation and interaction technologies. The second is that, in spite of the maturity of the visual technologies, there are still some persistent needs for improving different aspects of the user experience such as perception (i.e. in terms of realism and intuitiveness among other criteria) and performance (i.e. in terms of task precision and execution time among other criteria).

The use of multimodal stimulation is gaining interest particularly in Virtual Reality and in Human Computer Interfaces. Virtual Reality (VR) is a domain focused in the artificial recreation of real scenarios and conditions for solving many different kinds of needs. Human Computer Interfaces (HCI) are concerned with the development and study of interaction mechanisms between users and devices. VR and HCI are nearby scientific domains which are both interested in interactive systems. They also share different challenges such as the improvement of the interaction and the integration of new multimodal stimulation techniques.

The study of touch and hearing is of a current special interest in research on multimodal systems. 3D interaction takes place into a tridimensional virtual environment. It could be very complex to provide this kind of interaction in a realistic manner. Authors have identified a set of basic 3D interactive tasks [1]. Some interesting issues in this field are the contribution of each sense (coupling and decoupling studies) and the sensory substitution.

3D audio is a mature technology concerned with reproducing and capturing the natural listening spatial phenomena. The contribution and pertinence of 3D audio stimulation in interactive systems is an interesting scientific problem. 3D sound has been traditionally studied as a mean of improving realism and immersion, but not for interaction.

In this paper we are focused on analyzing the pertinence of providing and improving interaction in 3D space (either in HCI or virtual environment) by using 3D audio stimulation. Our concern is to provide precise interaction.

3D INTERACTION AND AUDIO STIMULATION

3D Interaction in interactive systems

3D interaction is concerned with the interaction of human users and devices in a tridimensional space. 3D interaction is related to both HCIs and VR. There exist different techniques for 3D interaction and they were classified by Bowman (1998) [1] as the canonical tasks: selection, manipulation, navigation and system control. Most of the existing techniques are based mainly on visual stimulation and feedback.

Foley and Van Dam (1982) [2] published a reference book about interaction with computers. This early work is limited to 2D interfaces. This work compiles the basis of most of the later efforts for the design of interactive interfaces.

Bowman et al. (2005) [3] claimed that manipulation is one of the most fundamental tasks in physical and virtual environments. For the authors, many of the techniques for the navigation and system control are based on the existing techniques concerning manipulation. The existing techniques of manipulation are commonly restricted to rigid objects that do not change of shape. Manipulation and selection are often treated together. Moreover, selection could be seen as a subtask of manipulation. Also in Bowman et al. (2005) [3], authors established that according to the literature, the principal subtasks of manipulation are: selection, positioning, and rotation. The field of application of such techniques is wide because they attempt to be abstract. The same authors enlisted some relevant manipulation techniques and they suggested creating new ones only in the case that a large benefit could be derived.

We believed that 3D sound stimulation could be used for implementing 3D interaction for the different canonical tasks (selection, manipulation, navigation and system application). This integration could be done either for completing or substituting existing stimulation mechanisms such as vision of haptics.

Audio stimulation for interaction

Kramer (1992) [4] introduced the term of auditory display, which refers to the use of non-speech sound to convey meaningful information within the context of HCI applications.

Gaver (1986) [5] introduced the term of auditory icon referring to the use of natural sounds for conveying information to the user. Auditory icons are created as a perceptual mapping for creating an analogy between the sound and the information provided to the user. In Gaver (1989) [6], the author presented the SonicFinder which is an auditory interface based on auditory icons.

Blattner et al. (1989) [7] introduced, for their part, the concept of earcons referring to parameterized musical patterns used to represent different things such as entities, properties or events. They proposed a classification of earcons and presented principles for their applications in computer interfaces.

Rocchessso et el. (2003) [8] presented an approach for conveying information by audio in interactive applications based on dynamic sound models. The proposal considers the physical modeling of sound, the use of parameterized control mechanisms and the application of design and validation mechanisms based on auditory perception. In Roccheso and Fontana (2003) [9], divers authors discussed different issues about non-spatial auditory information in interactive systems.

Brewster (2003) [10] presented an interesting review of the theory and practice of the different techniques for representing information in HCIs by means of non-speech audio stimulation.

These works are important to understand the current state-of-art of auditory interfaces. We notice that there exist divers techniques for conveying efficiently information and feedback by means of audio stimulation. However, this flown is limited to certain kinds of information such as messages, one-dimensional data, alerting and the occurrence of events. Hence, these techniques does not provide enough information to allow 2D or 3D interaction likely to be necessary in Virtual Reality applications as well as in certain types of HCIs.

TECHNOLOGIES OF 3D SOUND

3D sound is a technology concerned with reproducing the natural listening spatial phenomena. It allows displaying sound sources with spatial properties such as directivity, depth and reverberation [11] (see figure 1). Directivity refers to the perception of the spatial direction from the position where the audio source is located to the listener. Depth refers to the sensation of the distance to the sound sources as well as the intensity of the sound effect. Reverberation refers to the simulation of the natural sound field produced by the acoustic interaction of the sound wave and the environmental elements. Particularly, it is interested on the simulation of the phenomena known as early reflections and reverbering. Considering these

features, the 3D Sound techniques provide at artificial manner to position spatial sound sources in a tridimensional space.

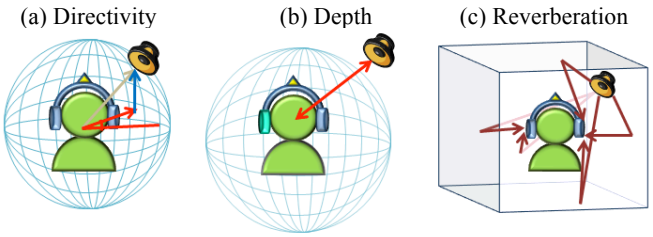

Figure 1. Spatial properties of 3D Sound.

The integration of 3D sound in virtual environment has been proposed as a manner to extend the interaction capabilities of current interactive systems. However, the spatial audio is far for being a common technique in such systems.

Sound Spatialization

There are different techniques for modeling different aspects of the human hearing phenomenon. Among them, we can mention the followings: the subtle differences on perceived sound for both ears (i.e. Inter-Aural Intensity and Time Differences -IID and ITD-), the effect of human body and human listening system (i.e. Head-Related Transfer Functions -HRTF- filtering) and the environmental acoustics (i.e. reverberating and early reflections). The HRTF is the most used technique for realistic 3D spatialization.

The objective of an HRTF is to recreate the natural altering effect on listened sounds caused by the morphology of each pinnae as well as by the diffraction and reflection effects due to the head and shoulders respectively. Begault [12] stated that HRTF represents the spectral filtering which occurs before the arrival of the sound to the internal ear drum. These effects vary from individual to individual forcing systems designers to choose between the use of individualized or generalized filters. For implementing the approach presented in this paper we employed the generalized HRTF database compiled by Bill Gardner et. al. [13].

The use of HRTF in VR applications normally requires a series of online processes. In typical 3D sound systems when a sound source is positioned, the HRTF filter corresponding to the orientation of the virtual source is loaded and applied continuously to the sound wave.

Kyriakakis et al. (1999) [14] presented an interesting and extensive review of the different common techniques for acquisition, rendering and displaying of spatial audio. Kapralos et al. (2008) [15] published a very complete and more recent review of the theory and practice of 3D sound in virtual environments. For the authors there is still a lot of work to do for generating convincing spatial sound in interactive real-time virtual environments. In this work the idea of 3D sound for interaction is not present. Lentz et al. [32] presented a rendering system architechture based on multichannel configurations which allows precise near-to-head spatialization for immersive facilities. This approach appears to be a good solution for applications with freely moving listeners without headphones.

There exist another group of techniques that reduce the problem of audio spatializing to the panning of the audio signal. This term is related to the positioning of the sound source by varying the intensity levels of the audio output channels. One common panning technique is Constant Power. This method uses sinusoidal curves to control the amplitude of the signal in different channels in order to maintain the same level of the total power of all channels [16].

However, there are other methods of panning such as Ambisonics and Vector Based Amplitud Panning (VBAP). Ambisonics is a microphoning technique that can be simulated to perform a synthesis of spatial audio [17]. In this technique the sound is reproduced by all speakers with different gain factors determined by sinusoidal functions. VBAP is a reformulation of the amplitude panning method based on vectors and vector bases [18]. Another related technique is the wave field synthesis (WFS) that implies the use of a large number of near loudspeakers for reproducing high fidelity sound fields. Theodoropoulos et al. [33] presented an interesting review of different architecture models for the WFS method.

There exist some other ways to obtain or model the HRTF and some of them are not based in real measures. Kistler et al. (1992) [19] proposed and evaluated a method based on Principal Component Analysis (PCA) for modeling the HRTF. The method uses the measures of several individuals. Their results show that the judgment of the users is similar for the original and the synthesized HRTF.

Duda (1993) [20] described the minimal requirements for an effective model of HRTF. In the same paper they reviewed different relevant proposals and classified as follows: spherical models, pinnae-echo models, structural models and models based on series expansions. The author proposed the usage of a hybrid model that could alleviate the drawbacks of the other separated approaches.

Brown et al. (1997) [21] proposed a method which includes head, pinna and room approximation by means of mathematical models based on structural models. They conducted an experiment for validating their approach obtaining an absolute mean angular localization error of 12 degrees. They concluded that the need of customization is not critical for azimuth but it is for elevation.

Another interesting issue is the choice between customized and generalized HRTFs. Customizing refers to the spatialization process using HRTFs measured from or adapted to each user. A generalized HRTF is obtained by processing the measures of one or many users. The generalization implies that the same HRTF is used for spatializing sounds to different users or for different groups of users. The use of customized

HRTF is not a very practical choice for interactive applications because of the technical needs for measuring the HRTF for each user. In this way, most designers of interactive systems with spatial sound capabilities prefer to use generalized HRTFs.

SOUND STIMULATION AND INTERACTION

Existing approaches of Audio Stimulation for Assisting Interaction

There are very few published works about the use of 3D sound in interaction applications for improving performance and perception of the user. Most of the available papers are also very recent. For this reasons, we believed that 3D sound in interactive applications is an emerging research field. In the following lines we included a selection of some of the most relevant works about this issue.

Cohen and Wenzel (1995) [22] published an interesting paper about the use and design of spatial sound in interactive interfaces. They discussed the current and the potential role of 3D sound in the context of multidimensional interaction and made an interesting review of existing and possible applications of 3D sound. They selected some cases of study for developing a discussion about different design issues. Authors recognize the difficulty of auditory localization and depth but proposed that combination of visual and auditory information is a good solution.

Cohen (1993) [23] introduced the use of gestural interaction for modifying the position and other characteristics of 3D sound sources. The author explored two systems with 3D audio interaction capabilities: Handy Sound and MAW. Authors proposed the use of 3D sound for extending the auditory display space. This work is relevant to our proposal because they are similar in the sense that it associates certain spatial and non-spatial characteristics of spatial audio (such as position, intensity and pitch) to certain interactive events (such as throwing, catching and manipulation of sound quality). In this way, the information of spatial audio is used for facilitating an interactive task and it is also enriched for facilitating others. One important difference with our proposal, which will be detailed later on, is that we provided spatial audio and enriched cue. The author made an interesting discussion of several issues about the design and implementation of these interactive mechanisms. The main limitation of this study is the lack of a methodology of evaluation which could have helped to scientifically validate their proposals.

Zahorik et al. (2001) [24] studied the effect of using visual feedback for helping users to improve their localization capabilities while using generalized HRTF. Authors conducted an experiment divided in three stages: before, during and after the training. They founded that the training with visual feedback contributes positively to localization accuracy.

King (2009) [25] published a review of the influence of visual information in the relearning process of auditory localization. According to the authors, vision is key for the learning and relearning stages. They discussed different issues such as the localization and the influence of vision, visual and audio neural interaction, and the effect of deprivation. Vision allows improving dramatically the spatial localization accuracy. In case of uncertainty or incoherence between vision and audio, the brain preferred the information of the most reliable source, which is vision in this case.

Bowman et al. (2005) [3] suggested that the use of 3D sound in 3D interfaces could contribute to the localization, to the realism, to sensory substitution and to the sonification of information in general. They provide some interesting guidelines about the integration of interaction techniques like spatial audio on such interfaces.

Kan et al. (2004) [26] presented a mobile application for giving directivity to sounds for communication proposes. This approach is based in 3D sound and the Global Positioning System (GPS).

Marentakis and Brewster (2004) [27] investigated on different gestures for interacting with 3D audio interfaces for localizing sound sources. They considered gestures such as pointing with the hand, pointing with the head and localizing in touch tablet. Authors determined the typical error for each gesture and found that the three of them are suitable for being used in future audio interfaces.

In another work, Marentakis and Brewster (2005) [28] conducted an experiment for comparing different interaction audio cues for improving the efficiency of localization of sources. This work is relevant to our study because they evaluated for the first time the introduction of different feedback cues of spatial sound. They tested different sources all of them positioned only in the horizontal plane obtaining mean deviation angles between 4 and 10 degrees approximately.

Ho and Spence (2005) [29] investigated on the benefits of spatial alerting in potentially dangerous situations for driving. According to the their results, the use of spatial audio allows an effective way of capturing the attention of the driver better than non-spatial alarms.

Vázquez-Álvarez and Brewster (2009) [30] presented an evaluation of the spatial audio capabilities of a mobile device for introducing multiple sound sources. The idea is to determine if users are capable of distinguishing different sources within a 3D audio interface working in current mobile technology. The effective display of multiple sources by spatial audio could facilitate the use of and switching between multiple information channels by allowing the differentiation of them.

Ménelas et al. (2009) [31] outlined briefly one of the first approaches where the idea of enriching the information of the HRTF is proposed. Authors are focused in multimodal interaction for target selection and they introduced the term of

modality metaphors. The stimulation modes are haptic and audio with total visual occlusion.

The need of sound stimulation for interaction

Sound can be used to provide interaction in Virtual Environment (VR), particularly, for assisting users for selection and manipulation of objects in the tridimensional space of a VE.

We have identified that among the most common applications of interactive systems there are some recurrent needs where the use of 3D sound stimulation techniques could be valuable. These needs are grouped in the following generic functions.

I. Localization. This is the most studied function of spatialized audio and it refers to the localization and identification of elements in a virtual scenario. It is can be related to the interaction task of selection which is common in virtual environment.

II. Manipulation assistance. It refers to the use of spatialized audio stimulus for assisting the user in tasks related to the manipulation of virtual objects such as translating and rotating.

III. Movement assistance. It refers to the use of 3D audio stimulation for assisting the user to follow targets, predefined trajectories or to execute certain movements.

IV. Spatial Alerting. This is an important need in different kinds of interactive applications. By using 3D audio stimulus, these functions could be improved in terms of intelligibility and efficiency on the problem of localization.

V. Spatial Tracking. 3D audio could be useful for facilitating and reinforcing the necessary information for processes such as object tracking inside and beyond the visual field.

These functions are not specific applications but generic functions applicable to different kinds of context. We attempt to measure and to analyze such hypothetical contribution. We pretended this classification to be as simple as possible in such way that the most of applications of 3D audio in interactive systems could be reduced to one or more of these functions. In Figure 2, we show the typical application scenario for each of these identified functions.

Providing interaction efficiently with 3D sound cues for functions such as the described above requires offering mainly a good level of performance and interactiveness. The level of performance refers to a group of objective parameters such the angular resolution (precision) and the execution time. The interactiveness refers to the perceptual (and subjective) aspects such as intuitiveness, ease of use and realism among others. Performance and perception are two key elements of our research approach and they will be discussed later on.

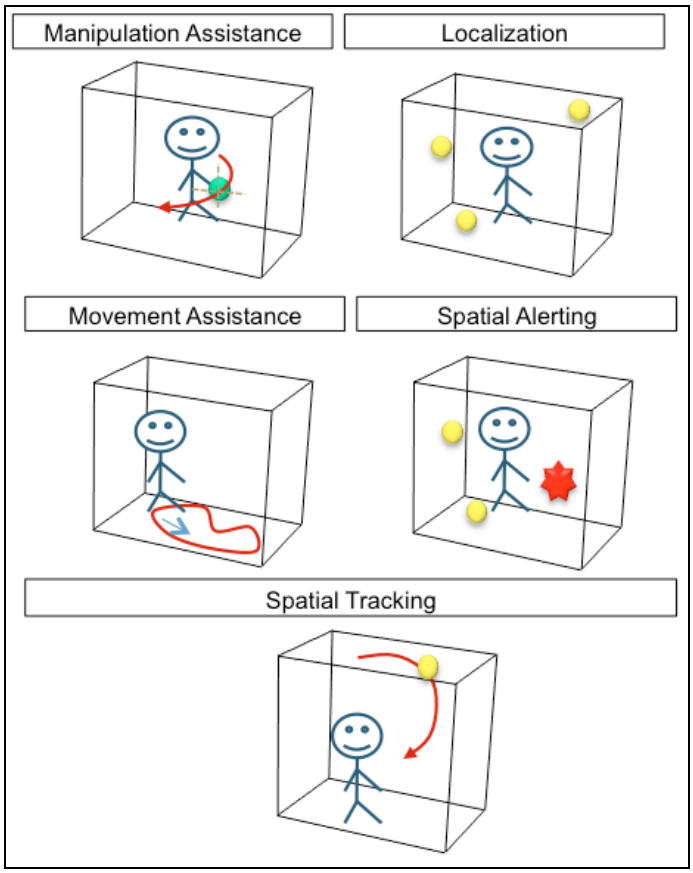

Figure 2. Potential functions of 3D sound spatial stimulation in virtual environment.

Metaphoric stimulation and 3D Sound

The use of Metaphors is a common strategy in VR and HCI. A metaphor is an association of two distinct elements. In VR and HCI they are commonly used for enriching the interaction and the flown of information between the user and the system.

The use of metaphors of 3D Sound is a novel approach in VR aiming to provide cues and to extend interaction capabilities. These metaphors could combine perceptual cues with models of human hearing with or without modifications.

Traditional approaches of 3D sound attempts commonly to prioritize realism. In contrast, metaphoric stimulation emphasizes other aspects such as perception or performance. In this way, 3D sound metaphors might not necessary correspond to real acoustics and hearing phenomena.

A 3D sound Metaphor is defined for providing interactive 3D sound stimulation to the user. This is carried out by dynamically adjusting the spatial properties of an audio stimulus depending on the user activity. The idea is to give particular cues to the user by applying certain frequency, intensity and panning effects to the audio stimulus. The

metaphors considered for our research are described in the next section.

Figure 3 summarizes the main techniques of sound stimulation for interactive systems. These techniques were already discussed in the section of related work. They are classified in three main groups: the audio cueing approaches, the conventional 3D sound techniques and finally the spatialiazed audio cueing. The last one is an original proposal of the authors. The purpose of this figure is to show the main antecedents of our approach. Audio cueing refers mainly to the techniques designed for conveying information to the user by audio stimulation in interactive systems. The 3D sound techniques refer to the classical techniques for artificial audio spatialization in virtual environment. The spatialized audio cueing combines the idea of meaningful audio stimulation (audio cueing) and the traditional spatial audio techniques. The integration of audio cueing and spatial sound represents one of our main contributions.

SPATIALIZED AUDIO CUEING

The spatialized audio cueing consist, as described before, in combining the audio cueing approach and the 3D sound techniques. The objective is to provide precise interactive cueing in 3D environment. The main applications are the described functions of 3D sound and particularly the 3D interactive tasks of selection, manipulation and navigation in virtual environment.

Problems and limitations of current approaches

The main limitations of the former approaches for audio cueing and 3D sound for providing 3D audio cueing are discussed in this subsection. The current 3D sound techniques provide 3D audio stimulation with an acceptable response time but a poor level of precision.

The existing audio cueing techniques are useful for providing spatial information to the users of interactive systems but it does not provide 3D audio stimulation. Nevertheless, the field of application of these techniques is large and the number and quality of the research in this domain is important.

3D sound metaphors are proposed as an approach for spatial audio cueing. It is conceived as an attempt to tackle the problem of lack of precision of the current 3D sound techniques by enriching this stimulation with audio cueing. The idea is to facilitate the task of 3D sound localization by modifying the audio stimulus depending on the user activity.

Theoretical and technical issues

In this subsection we discussed the main technical aspects to be considered for the design of 3D metaphors. These issues should be treated considering the current bibliography and research focus. They are related to three classes: the spatial hearing, the head movement and pointing gestures and finally, the implementation issues.

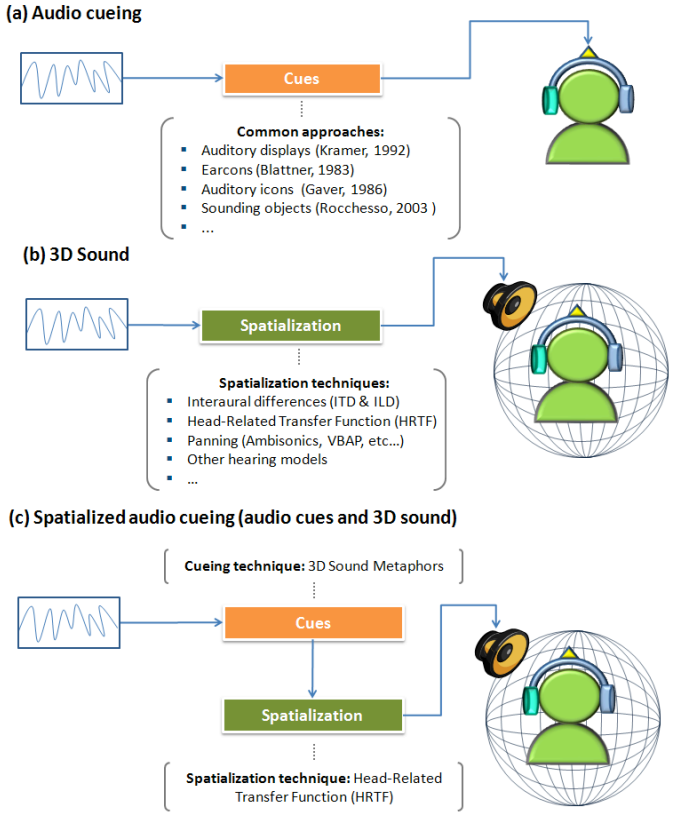

Figure 3. Audio stimulation techniques in virtual environment.

The spatial hearing phenomenon implies a series of limitations. Some of the most important are the lack of precision, the presence of reversal errors and that the hearing capabilities depend, among other factors, on the morphology of the hearing system of each individual. Humans are not very accurate in spatial sound localization which precision varies for the horizontal and the vertical plane. The absolute localization error is about 30 degrees, which is notoriously very high for providing precise cueing. The metaphors may allow reducing the error for allowing accurate localization.

The reversal errors refer mainly to the problem of ambiguity of spatial hearing. The front/back reversal error refers when the user is not capable to distinguishing whether the sound is coming from the front or from the back of the user. There also exist the up/down reversal error. The reversal errors represent also an obstacle for providing precise spatial cueing. The 3D sound metaphors attempt to eliminate or to reduce significantly the presence of such errors.

The dependence of spatial hearing on the individual characteristics of users represents an important technical problem. The classical spatial audio approaches offer different solutions based on the use of generalized spatialization

techniques. The approach of metaphors is based on generalized HRTF for avoiding the process of customization. Nevertheless, the use of generalized HRTF usually represents a moderate loss of precision compared to customized HRTFs. The enriched stimulation provided by the 3D sound metaphors should provide sufficient information for obtaining more accurate results than the customized HRTF. In this case, the use of customization is prevented.

The implementation of 3D sound implies the choice of the use or not of head tracking. Head movement is used by humans in many different situations for improving their spatial hearing capabilities. For example, it has been determined that head tracking helps to reduce the number of reversal errors. There are many different solutions for implementing head tracking. It is necessary to evaluate the different options and to select the most adapted one to the metaphors. Even if the usefulness of using head tracking in 3D sound systems has been shown, it is also necessary to study the contribution and to determine the need of using head tracking with the metaphor. It is important to find a good solution in terms of price and technical complexity of the head tracking as well as of the other resources because one of our objectives is to propose a low cost solution compared with other simulation mechanisms like immersive environments and haptic interfaces.

Another interesting problem is the selection of the pointing gestures. It refers to the manner on which the user will indicate the system the localization of the sound source. There exist different research works about this issue. We are particularly interested in two gestures: pointing with the head and pointing with a hand-manipulated instrument. It is necessary to determine if both of these gestures are well adapted to the 3D sound metaphor approach.

It is also important to determine whether the HRTF model represents a significant advantage when used with the metaphor compared with other panning techniques such as constant power panning.

Another interesting issue is the improvement of the perception of sound depth. As described in a previous section, there exist different techniques for simulating the perception of depth. Some of them are based in perceptual calibration. One simplified way to simulate depth is to vary the intensity of the sound depending on the distance to the source. We are interested in determining whether the integration to the intensity-distance model with the Interaural Time Differences (ITDs) could contribute to the sensation of depth. This is in order to improve the sensation of depth for the 3D sound metaphor.

The design of an efficient software and hardware architecture for the implementation is a key element. The main desired features are the following:

- 3D multimodal stimulation and interaction (vision, haptic and audio feedback)
- 3D tracking for head rotation and hand position.
- Time real audio spatial processing
- Compatibility with different audio output mechanisms such as headphones and speakers
- Different audio spatialization techniques
- Reverbering and depth simulation.

In the other hand, there are also the aspects about interactiveness and the perception of the interaction technique designed with the 3D sound metaphor. There exist different mechanisms for evaluating the impact on perception of audio stimulation techniques. There are also mechanisms for evaluating the pertinence of an interactive technique on virtual environment. For this work, we decided to design an evaluation mechanism based on the most relevant concepts to our focus of the existing perception and interaction mechanisms. We are interested in the following subjective criteria: ease of use, intuitiveness preference and realism.

The metaphor of 3D sound

In this subsection, we describe the defined characteristics of our 3D sound metaphor at this point. The objective of this 3D sound metaphor is to provide the user with cues about the horizontal and vertical orientation of a fixed reference point associated to the spatial sound source. The metaphor was created for providing interactive 3D sound stimulation to the user. This is carried out by dynamically adjusting the spatial properties of an audio stimulus depending on the user activity. The idea is to recreate the directivity properties of a spatial audio stimulus by applying certain frequency, intensity and panning effects to the audio stimulus. In our proposal we privileged the intelligibility of cues over the realism.

The spatial cues associated to the effects that compose the metaphor are detailed in Table 1. For each cue we specified the associated audio effect as well as the corresponding associated spatial feature of the sound source. The relationship of each effect and the corresponding spatial feature is specified by a behavior curve.

Table 1. Perceptual cues of the metaphor

Cue	Associated sound effect	Associated spatial feature
Verticality	Reverb	Elevation
Horizontality	Attenuation	Azimuth
Frontality	Occlusion	Azimuth
Angular proximity	Sharpening	Elevation and azimuth
Depth	Attenuation	Distance

The term verticality refers to whenever the sound source is located above or below the reference plane associated to the head of the user. Horizontality refers to the horizontal

angular deviation perceived by the user. The term of frontality refers to the ability to distinguish whether the sound source is located in the front or at the back of the user. Angular proximity refers to the capacity for accurately determining the provenance of a sound source when the absolute angular deviation is small (less than ten degrees).

The sound metaphor has two main objectives: to facilitate the first initial recognition of a spatial sound stimulus, and to allow an accurate recognition. The metaphor is implemented using the FMOD Firelight Technologies [34] sound library and specifically its following components: FMOD EX API and FMOD Designer.

One key element of the interaction technique working with the 3D sound metaphor is the association between the spatial features of the sound source and the degrees of freedom of a specific case of study. The idea is that the audio cueing depends on how the user manipulates or select an object and on how he navigates into an environment. In this way, the sound source should be positioned in order to provide the user with meaningful information that is supposed to be useful for executing the task more efficiently. Thus, the parameters of the degrees of freedom are somehow associated to the spatial features of the metaphoric sound source (azimuth, elevation and depth mainly). This association has to be done very carefully and it is very important for the success of the proposed interactive technique. It is possible to test different ways of making this association as an independent variable in the experiment design, for determining which is the best way to be done. However, we preferred to determine the nature of this association during the pre-tests and not to use it as a controlled variable of the experimental setup.

The experimental platform

In this section we briefly describe the experimental platform and its hardware and software architectures. The main relevant specifications are presented and the most important technical choices are explained. The main characteristics of the hardware components are detailed in the Table 2. This setup is low-cost and it can be easily implemented with a desktop station. The XSens MTi is a 2.5 DoF (360º in Roll/Heading, 180º in Pitch) orientation tracker based on accelerometers. The tracker is fixed on the headset and it is used to monitor the head orientation in the 3 axes: heading, roll and pitch. The Wiimote is used for provide the user with an ergonomic manner to indicate different events like, for example, when a sound source is localized.

The scheme of the developed software modular architecture is shown in Figure 4. In this scheme, we specify the main software components and how they were integrated. FMOD is a library for online audio processing which is commonly used in video game development. This library offers a great range of audio processing techniques. The core of the application was produced with OpenGL, Glut and Microsoft Visual C++. These resources are common development utilities that are characterized by its flexibility, robustness and for the abundance of references and additional component resources.

Table 2. Hardware architecture

Function	Device	Type	Inter-face	Speed	Spec.
Audio Processing	Creative Sound Blaster X-Fi Pro	Sound card	PCI	192 kHz	24 bit processor, EAX support
Audio Output	Sennheiser HD 201	Headset	Analo-gic	-	Dynamic closed
Interaction	Nintendo Wiimote	Wireless HID	Blue-tooth	100 Hz	10 buttons; vibration feedback
Tracking	XSens MTi	Accelo-rometer tracker	USB	24 Hz	2.5 DOF, error < 1.0 deg
Graphics	NVIDIA Quadro FX	Video Card	PCI	24 Hz	3840x2400 px; 128 Mb
Worksta-tion	HP xw8400	PC	-	2x1. 6GH z	2 GB RAM

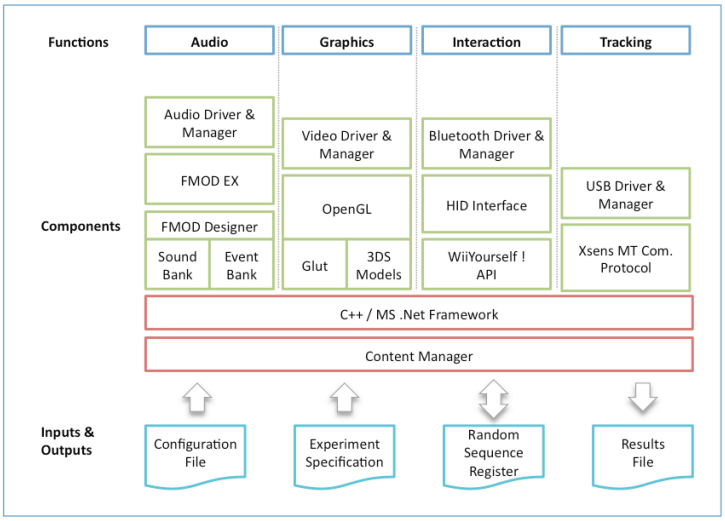

Figure 4. Software architecture.

METODOLOGY

In this section we describe the main characteristics of our research methodology for validating our approach of 3D sound metaphors. We expect to use this methodology in a series of different experiments that will be briefly described later on. In all cases, our objective is to determine the contribution of 3D sound metaphor.

Model of the experiment design

The experiment design is based on the within groups repeated measures modality. In this way, all users test the different

experimental conditions and make several repetitions. There are two experimental basic conditions: 3D sound and no sound. This could change depending on the independent variables for each case. Each experimental condition consists on different repetitions. The idea of having various repetitions is to get a representative sample in terms of complexity of the task depending of the characteristics of each repetition.

The hypothesis should be defined for each particular case. Nevertheless, we expect that in general the 3D sound metaphor must contribute to improve the user performance (either in precision or in execution time or in both of them) and must have a positive impact on user perception.

The criteria for selecting the group of users to participate in the experiments are the following:

- Aged from 20 to 30 years (males or females indistinctly).
- 15 subjects approximately for each experiment.
- With no audio sickness. Visual sickness is tolerated when they are overcome by using glasses.
- Minimal experience in interactive technologies. But it is preferable they not to be experts neither in the specific case of study nor in spatial hearing.

The reason for discarding experts is that it has been traditionally accepted that the effect of different stimulation techniques is easier to observe in non-expert groups of users.

One of the main inconvenient we faced in our experiment design is the "carry over" (learning effect) when the user repeats the same task in different conditions. It is natural to expect that the user will acquire a skill at the final part of the experiment compared to the beginning When using the "within groups" design with multiple repetitions it is very difficult to eliminate the learning effect. One alternative is to try to distribute its effect more or less equally by making each user to execute randomly the order of both elements: repetitions and experimental conditions.

Evaluation

We propose to implement an objective evaluation of user performance as well as a subjective evaluation of the perception of the user. This double evaluation will allow us to observe more completely the effects on both aspects: performance and perception respectively.

The performance evaluation is based on the measures of precision and execution time. It is important to mention that the precision measure depends on the characteristics of each case of study. In the other hand, it is important to carefully identify the period of time when the subjects are actually using the audio stimulation in order to record the correct elapsed time. The evaluation of performance is useful to characterize the technique and to determine its viability. We selected to analyze mean values, medians, dispersion measures and to use the Repeated Measures Generalized Linear Model analysis for determining and comparing the effects of the different experimental conditions to test as well as their statistic significance.

The evaluation of user perception is based on the criteria of ease of use, intuitiveness, realism and preference. These data will be collected by applying questionnaires after the accomplishment of the experiments. These questionnaires are designed by adopting common terms and standardized models. These results will be analyzed by comparing the percentages of responses.

RESEARCH PERSPECTIVES

In this section we described the main scientific issues we are interested in and we outlined the general research strategy to be followed. This project of research has three main axes: the study of the contribution and pertinence of the approach of 3D sound metaphors, the sensorial coupling and decoupling of the technique with the common stimulation mechanism of vision and haptics and finally the validation of the integration of the technique in different cases of study.

Stages of the research project

Figure 5 shows a tree of the main stages of the research strategy. As it is shown, the first stage consists in the scientific validation of the metaphoric audio stimulation. Firstly it is necessary to characterize the performance and the effect on perception of the 3D sound metaphor. Then, we proposed to make a comparison between the 3D sound metaphor and the classical technique of 3D sound of HRTF for sound localization. We plane to compare them and to determine the potential benefits of combining them.

The second stage refers to the integration of the 3D sound metaphor into a virtual environment for the interactive task of object manipulation. The objective is to show that the spatial audio cueing provides efficient interaction for 3D manipulation of an object with our without vision feedback. This coupling and decoupling study will help us to determine the contribution of each stimulation mechanism.

The third stage refers to the coupling and decoupling of the stimulation mechanisms of 3D audio, vision and haptics. The idea is to determine whether the spatial audio cueing could contribute to improve performance and perception in multimodal systems. Another important objective is to determine if the spatial audio cueing could be use to substitute other stimulation mechanisms such as haptics.

The last stage refers to the application of the interaction technique of 3D sound metaphors in real applied cases. The audio precise cueing could be useful when the

manipulation of objects and the assisted navigation in 3D environment is required.

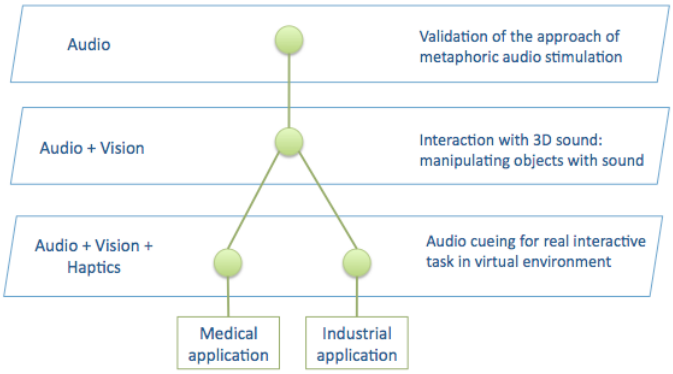

Figure 5. The integration of the different kinds of stimulation mechanisms and the interaction technique.

Cases of study

We have identified different cases of study that are of special interest for our proposed approach. The first task is the sound source localization. This task will be useful to validate our approach and to characterize the user performance. The idea is to determine the effect of our approach on the precision and the execution time. After this, we expect to compare this effect with other existing techniques for sound spatialization (e.g. HRTF).

We also identified a group of applied tasks that would be useful for the study of the pertinence of spatial audio cueing for assisting the user. One of these cases of study is the training of specialized gestures in the context of surgery simulation. Another is the assisted manipulation in virtual mechanical assembly. The simulation of crane operation in collaborative context is also a case of interest.

As one of the results of this scientific process, we also expect to determine guidelines for facilitating the integration of the proposed spatial audio interaction technique in different scenarios.

CONCLUSIONS

In this paper, we outlined the main elements of our research proposal for integrating the 3D sound in virtual environment for providing effective spatial audio cueing. Our approach consists in the integration of 3D sound metaphors. They are spatial sources generated by the HRTF and enriched with cues for improving the user performance and perception.

The main problem discussed in this paper is the need of providing precise position cueing in 3D environment by means of audio stimulation. We concluded that the current techniques for audio cueing and for sound spatialization are not suitable for providing this kind of interaction in an effective manner. The 3D sound metaphor is a combination of the traditional 3D sound techniques and a group of audio cues.

We discussed the main technical and theoretical issues for being considered for the development of this approach. We also gave a brief description of the 3D sound metaphor and the information (cues) it attempts to provide to the user. Our experimental procedures attempt to study the contribution of 3D sound metaphors and to compare this stimulation mechanism with typical vision and haptics feedback mechanisms.

The main characteristic of our methodology are described and explained. Finally, we describe our general research strategy. The precise audio cueing could be useful in different application fields such as chirurgical and industrial simulation of manipulation tasks as well as in assisted navigation. The hypothetical contribution of 3D sound metaphors in such fields could represent and important contribution for the use of audio stimulation for improving performance and for substituting other in virtual environment.

REFERENCES

[1] D.A.Bowman, Interaction techniques for immersive virtual environments: Design, evaluation, and application, Journal of Visual Languages and Computing 10 (1998) 37–53.

[2] J.D. Foley, A. Van Dam, Fundamentals of interactive computer graphics, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA,

1982.

[3] D.A. Bowman, E. Kruiff, J. LaViola, I. Pouppyrev, 3D User Interfaces Theory and Practice, Addison Wesley, USA, 2005.

[4] Kramer, G., 1992. An introduction to auditory display. SFI studies in the sciences of complexity. Addison Wesley Longman. URL Proceedings/1992/Kramer1992a.pdf.

[5] W.W. Gaver, Auditory icons: using sound in computer interfaces, Human Computer Interaction, 2(2) (1986).,167–177.

[6] W.W. Gaver, The Sonic Finder: An interface that uses auditory icons, Human Computer Interaction, 4:67-94, 1989.

[7] M.M. Blattner, D.A. Sumikawa, R.M. Greenberg, Earcons andi cons: their structure and common design principles, Hum.-Comput.Interact. 4 (1) (1989) 11–44.

[8] D. Rocchesso, R. Bresin, M. Fernstrm, Sounding objects, IEEE Multi-Media 10 (2) (2003) 42–52. doi:http://doi.

ieeecomputersociety.org/10.1109/MMUL.2003.1195160.

[9] F.F. Davide Rocchesso, The sounding object, Mondo Estremo, Italy.

[10] S.Brewster, Non speech auditory output (2003), Chap. 12: The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications, edited by Jacko, J.A. and A. Sears, Erlbaum Associates Inc., 220–23, USA.

[11] Vladimir Ortega-González, Samir Garbaya, Frédéric Merienne, An approach for studying the effect of high-level spatial

properties of 3d audio in interactive systems, Word Conference on Innovative Virtual Reality WINVR'09, ASME, France, 2009.

[12] Begault, D., 2005. 3D-Sound for Virtual Reality and Multimedia. AP Professional, USA.

[13] Gardner, B., Martin, K., 1994. HRTF Measurements of a KEMAR Dummy-Head Microphone. MIT Media Lab Perceptual Computing, USA.

[14] Kyriakakis, C., Tsakalides, P., Holman, T., 1999. Surrounded by sound. IEEE Signal Processing Magazine (16), 55–66.

[15] Kapralos, B., Mekuz, N., 2007. Application of dimensionality reduction techniques to hrtfs for interactive virtual environments. In: ACE '07: Proceedings of the international conference on Advances in computer entertainment technology. ACM, New York, NY, USA, pp. 256–257.

[16] Roads, C., 1996. The Computer Music Tutorial. MIT Press, Cambridge, MA, USA

[17] Gerzon, M. A., January-February 1973. Periphony: With-height sound reproduction. Journal of the Audio Engineering Society 21 (1), 2–10.

[18] Pulkki, V., 2001. Spatial sound generation and perception by amplitude panning techniques. PhD in Technology, Department of Electrical and Communications Engineering, Helsinki University of Technology, Helsinki, Finland, ISBN 9512255316.

[19] Kistler, D. J., Wightman, F. L., 1992. A model of head-related transfer functions based on principal components analysis and minimum-phase econstruction. The Journal of the Acoustical Society of America 91 (3), 1637–1647.

[20] Duda, R. O., 1993. Modeling head related transfer functions. In: Proc. 27th Asilomar conf. on Signal, Systems and Computers, Asilomar, CA. pp. 457–461.

[21] Brown, C. P., Duda, R. O., 1997. An efficient hrtf model for 3-d sound. In: IEEE ASSP Workshop on Applications of Signal Processing to Audio

and Acoustics.

[22] M. Cohen, Throwing, pitching and catching sound: audio windowing models and modes, Int. J. Man-Mach. Stud. 39 (2) (1993) 269–304.

[23] M. Cohen, E.M. Wenzel, The design of multidimensional sound interfaces (1995) 291–346.

[24] P. Zahorik, C. Tam, K. Wang, P. Bangayan, V. Sundareswaran, Localization Accuracy In 3-D Sound Displays: The Role Of Visual-Feedback Training, Proceedings of the Advanced Displays and Interactive Displays Federal Laboratory Consortium, 2001.

[25] A.J. King, Visual influences on auditory spatial learning, Philosophical Transactions of the Royal Society B: Biological Sciences, 364 (1515) (2009).

[26]A. Kan, G. Pope, A. van Schaik, C.T. Jin, Mobile spatial audio communication system, in: ICAD, 2004.

[27] G. Marentakis, S.A. Brewster, As tudy on gestural interaction witha 3d audio display, Lecture Notes in Computer Science (2004) 180–191.

[28] G. Marentakis, S. A. Brewster, A comparison of feedback cues for enhancing pointing efficiency in interaction with spatial audio displays, in: MobileHCI '05: Proceedings of the 7th international conference on Human computer interaction with mobile devices & services, ACM, New York, NY, USA, 2005, pp. 55–62. doi:http://doi.acm.org/10.1145/1085777.1085787. 7

[29] C. Ho, C. Spence, Assessing the effectiveness of various auditory cues in capturing a drivers visual attention., Journal of Experimental

Psychology: Applied 11 (3) (September 2005) 157–174.

[30] Y. Vázquez Álvarez, S. Brewster, Investigating background & foreground interactions using spatial audio cues, in: CHIEA'09: Proceedings of the 27th international conference extended abstracts on Human factors in computing systems, ACM, New York, NY, USA, 2009, pp. 3823–3828.

[31] B. Ménelas, Lorenzo Pincinali, Brian F.G. Katz. Patrick Bourdot, Mehdi Ammi, Haptic Audio Guidance for Target Selection in a Virtual Environment, HAID '09: Ancillary Proceedings of the 4th International Haptic and Auditory Interaction Design Workshop, Dresden, Germany, 2009.

[32] Tobias Lentz, Ingo Assenmacher, Michael Vorländer, Torsten Kuhlen, Precise Near-to-Head Acoustics with Binaural Synthesis, Journal of Virtual Reality and Broadcasting, Volume 3(2006), no. 2.

[33] Dimitris Theodoropoulos, Catalin Bogdan Ciobanu, Georgi Kuzmanov, Wave Field Synthesis for 3D Audio: Architectural Prospectives, CF'09, May 18–20, 2009, Ischia, Italy.

[34] Firelight Technologies, FMOD Official Documentation, 2005, www.fmod.org.