ABSTRACT
Virtual Environments (VEs) are being widely used to improve communications between collocated users. A development of a VE is mostly concerned in offering a high quality graphics visualization system to enhance presence. It is known that sound can reduce visual workload and improve awareness of information outside users’ field of view. Impressively, these values are usually forgotten, and the sound system is just used to reproduce some 2-D sound effects.

To encourage the use of a 3-D sound system when developing an Immersive Virtual Environment (IVE), we performed three different experiments. Our focus was the benefits of using 3-D sound to improve navigation and enhance presence. We also measured the subjective perception of sound when the loudspeakers were hidden or visible. 3-D sound proved to be essential not only to enhance presence (experiment 1) but also to improve navigation, orientation and way-finding (experiment 2). Experiment 3 proved that a simple, cheap and effortless technique enables greatest positioning errors between the image and the sound and improves users’ confidence and correctness.

Keywords
3-D sound, Sonification, Virtual Environments, Presence, Ventriloquism Effect.

1. INTRODUCTION
While designing VEs, researchers focus most of their efforts in providing the user with a high quality graphics visualization system. However, we believe that something is being left behind, namely sound. Sound is essentially used to reproduce the voice of a collocated user or some 2-Dimensional effects. For instance, if the user is facing a large-scale display, he/she can’t be aware of everything that is happening just with the eyes. Outside theirs field of view, might be some valuable information that if missed, will denigrate the sense of immersion. With the right amount of resources, 3-Dimensional sound systems can solve these kind of problems, because they are an essential complement for visualization techniques. If we are used to 3-Dimensional sounds in every day life, why not have 3-D sound systems to enhance presence in VEs?

To encourage the use of a 3-D sound system when developing an IVE, we performed three different experiments (Fig. 1). Our focus was the analyses of 3-D sound to improve navigation and enhance presence. We also measured the subjective perception of sound when the loudspeakers were hidden or visible.

To prove that it is worth hiding the loudspeakers to improve the subjective perception of sound in a VE, we performed a very simple task. The user needed to decide if a shifted sound from a talking-head, was or was not in the head’s mouth. Each user did the experiment with the loudspeakers hidden and visible. In the setup where the loudspeakers hidden, users tolerated greatest distances between the image and the sound and improves users’ confidence and correctness.

3-D sounds and their influence in enhancing presence, we framed our experiment in 6DSpaces Science Center. Our Outdoor Virtual Environment (OVE) was a simulation of a journey through the sea and the forest where vision, smell, haptics and hearing stimuli were activated. We used a conjunction of two methods to measure presence: the one proposed by Slater [19] and the Emotiv EPOC EEG headset. We found that the tactile perception of wind is the most important stimulus in OVEs and that a 3-D sound system is more important to enhance presence than a video-projection. We also discovered that Slater’s and EEG methods used together, help reducing some of the subjectivity inherent to Presence as an attribute.

The navigation experiment was framed in project CEDAR (Collaborative Engineering Design and Review), an hands-free Avatar-based Teleconference Environment with a large-scale display and a 3-D sound system, used to help collaborative engineering design and review which focus its attention in the oil industry (Fig. 2). The scenario used represented a real need for PetroBras. Through a comparison of three setups: video projection-only, video projection and 2-D sound, video projection and 3-D sound, we proved that audio cues make navigation effortless and improve sense of orientation. We found that with 3-D sound, users chose the shortest paths; without audio cues, users lost orientation more often and with 3-D sound users rotate more than with 2-D. We also found that 3-D sound approaches real life behaviours to the VE, therefore enhancing presence. The worst results were in the video projection-only, which led us to say that sounds, especially 3-D ones, are essential for tasks that need a rapid judgement and awareness of what’s surrounding.

To prove that it is worth hiding the loudspeakers to improve the subjective perception of sound in a VE, we performed a very simple task. The user needed to decide if a shifted sound from a talking-head, was or was not in the head’s mouth. Each user did the experiment with the loudspeakers hidden and visible. In the setup where the loudspeakers were hidden, users tolerated greatest distances between audio and visual stimuli (a greatest ventriloquism ef-

1http://www.6ds.com/ accessed on 11/08/2012
fect). This simple, cheap and effortless technique is very useful because it enables greatest positioning errors between the image and the sound: it doesn’t exist any sound system that can position virtual sound sources without errors.

2. RELATED WORK

In this section it is made a comparison between Visual-Only, Auditory-Only, Audio-Visual and other VEs that simulate more than the two main physical modalities (smell, touch or taste). After that, some works that used sonification techniques to convey information or perceptualize data are presented.

2.1 How to measure Presence in Virtual Environments

The feeling of presence in VEs has never benefited from a combined and generally established measurement approach [20]. Therefore, experiments are typically performed combining different measuring methods such as questionnaires, behavioural and physiological responses or comparison, and others. All of these approaches have their own set of problems, but when used together, they lead to clearer results. Depending on the type of VE, a correct mix of different methods seems to be the way to give us a more reliable definition of Presence [1].

2.2 Visual-Only Virtual Environments

The vision is known as the most important modality to navigate and localize objects in a VE, because it provides very detailed information of what we look at.

Head-Mounted Displays (HMDs) [21] and CAVE-like systems [8] are very concerned in the visualization of 3-D models and in the navigation on virtual spaces. Augmented Reality [2] is most concerned at bringing 3-D virtual images into the real world, or at substituting real images by virtual ones. Allosphere [15] is the most powerful high-resolution immersive environment to date and was considered essential to the next step in identifying quantum computers patterns. All these systems are very good in enhancing presence and awareness solely with photo-realistic images and visual cues, but having a 360 degree image reproduction is simply not enough.

Sonification: a form of auditory display, is the use of non-speech audio to convey information or perceptualize data.

The Allobrain [22] project is an interactive, immersive, multimodal art installation that uses functional magnetic resonance imaging (fMRI) brain data to construct a virtual world. It was installed in the Allosphere. As this large scale display produces a visual overload [6], Allobrain used artificial semi-autonomous agents to include normally unnoticed stimuli. These agents could enhance user’s presence and awareness but only if they were sonified by the third-order 3-D Ambisonics system. All the experience depended on sound spatialization because of information being missed due to eyes focusing elsewhere or the agent being positioned behind the user’s head.

In a comparison of visual-only and auditory-only information to navigate in a CAVE-like virtual world done by Gröhn et al. [13], it was also found that the user needed auditory cues, mostly outside its field of view, to navigate towards the objective faster.

2.3 Auditory-Only Virtual Environments

Even tough that vision can provide very detailed information, it is fully dependent of the user focus area. Hearing is an always-open channel and it can capture information independently from the area that the eyes are looking at. A 3-D sound spatialization is a way to “place” the sound sources anywhere in the virtual world and to enable a 3-Dimensional information retrieval.

Nevertheless, even with real sound sources, the user isn’t as accurate as in a visual-only VE, in identifying the exact position from where the information is coming from [18]. Therefore, the sense of immersion in auditory-only VEs is normally lower, because users can’t feel so physically present when they are disoriented.

Most of the research in auditory-only VEs is focused on people who cannot see or on dark spaces, where vision is reduced almost to zero. That “blindness” condition plus a 3-D spatialized sound reproduction, enhances in a better way, presence and awareness.

2.4 Audio-Visual Virtual Environments

Presence and awareness are fully correlated with the amount of physical modalities used in a Virtual Simulation [20]. The correct combination of audio and graphics in VEs will enhance presence and awareness and therefore improve collaborative or navigation tasks. If the user is capable of using both hearing and vision, a VE must use 3-Dimensional graphics and sounds.

In Boneel’s et al. [5] study it was found that a
Sonification emphasizes the role of sound as a mediator of meaningful information, making interactions more effective and pleasant. Interaction designers aim at designing such interactions to make them more effective and pleasant. Sonic Interaction Design designers aim at designing such interactions to make them more effective and pleasant. 

Gröhn et al. [13] concluded that the combination of auditory and visual cues are very dependent of each other in a navigation task. Users use auditory cues to define the approximate location of an objective and visual cues for the final approach.

2.5 Multi-Sensory Virtual Environments

A broader range of sensory cues promotes a greater sense of presence and therefore a better experience in a VE. A VE which uses the best graphics and sound-systems available, usually accuses latency issues, specially in interactive activities. The combination of all the physical modalities means that a large number of cross-modal effects can be exploited to improve the performance. In some cases it is possible to reduce graphics or sound-systems quality without perceptiveness, just by adding a new modality to the VE [5].

Chalmers et al. [7], performed an experiment to identify how the different modalities (noise, temperature, smell and combinations of these) affect the visual component. The users were asked to chose between two animations: one with High-Quality and the other with Low-Quality or Selective Quality, which one had the best graphics quality. In each comparison some multi-modal distractors were used: noise (pink noise) only; high temperature only; strong perfume, noise and high temperature; strong perfume and high temperature; mild perfume and noise; mild perfume, noise and high temperature; mild perfume and high temperature. They found that the influence of the other modalities had an impact, specially when using a combination of perfume and high temperature.

3. SONIFICATION TECHNIQUES

“Nowadays, human-object and human-human interactions are often facilitated by computational means. Interaction designers aim at designing such interactions to make them more effective and pleasant. Sonic Interaction Design emphasizes the role of sound as a mediator of meaningful interactions” [17].

3.1 Auditory Icons

Firstly introduced by Gaver [11], auditory icons are every-day sounds meant to convey information about computers events by analogy with everyday events. These sounds are intended to provide information about an event or object in the interface by representing the desired data using properties of the sound’s source, rather than properties of the sound itself. Auditory icons are easy to learn and remember as they are natural and related to our everyday lives.

3.2 Earcons

Earcons are abstract audio messages used in the user-computer interface to provide information and feedback to the user about computer entities [4]. In contrary to auditory icons, earcons are harder to remember and learn because they have no natural link or mapping to the objects or events they represent. On the other hand they are highly structured and can easily create families and hierarchies of objects and actions, with very simple audio messages.

3.3 Data Sonification

In order to correctly sonify an information display, it is important to take in account the nature of the data to be presented and the task of the user. Also, it is necessary to understand that some dimensions of sound are perceived as categorical (e.g., timbre) and other attributes of sound can only be perceived along a perceptual continuum (e.g., frequency, intensity) [14]. Not less important to have in mind, is that the interpretation and scale of a data-to-display relationship, varies with the conceptual dimension used (size, temperature, price, etc.) [23].

Data can be discrete (e.g., events or samples) or can be represented by a continuous flow of information. A correct sound mapping, depends on the data type: for example, categorical data types should be represented by categorically changing acoustic variables (e.g., timbre) while interval data is better represented by continuous acoustic variables (e.g., pitch or loudness) [14].

3.4 Soundscapes

Everything that is heard in a VE, is part of the soundscape. Nevertheless, if what we see doesn’t “fit” with what we hear, the sense of immersion is drastically affected. When we see images from a beach, we expect a soundscape with the seagulls screech and the waves hitting the sand.

When designing a VE, sounds must be carefully chosen. If users generate their own objects and sounds are produced in their interaction with the VE and other users, it is necessary to automatically characterize a given soundscape and search for sounds that best fit that characterization [16].

A work done by Finney at al. [10], showed that the Imperative properties of generated soundscapes from online databases, are preferred to the soundscapes done with real recordings and edition. This is very important, because every VEs can now focus their attention in characterizing soundscapes rather than in recording or synthesizing sounds.

4. THE STATE OF THE ART

According to the related work, we can assume that sound can enhance presence and awareness in VEs. We can also state that hearing and vision complement each other and in order to simulate realistic environments, 3-D sounds are important. Increasing the modalities of sensory input in a VE can also increase the sense of presence. This is not a
surprise, since vision, hearing, touch, smell and taste are what we experience in the real life.

Sonification techniques must be wisely chosen. Many experiments already stated that aspects like the learning curve and the relation with day life activities, must be considered before choosing earcons or auditory icons.

After our analysis of the related work we found out that the following topics are being left behind by the community:

1. Even though that there exist many experiments that clearly prove the importance of sounds to enhance presence and awareness, they lack on the analysis of the differences between using Stereo and 3-D sounds.
2. Another reality is that there aren’t experiments which weight the importance of each physical modality to enhance presence in VEs. These kind of analysis are important for example, to decide what to do after implementing 3-D graphics and sounds: introduce smell or haptics stimuli?
3. Least but not less important, there aren’t any standards to mount a 3-D sound system in a VE. Most of the times the sound system needs to adapt to the room: which can drastically influence sound localization (stated by many experiments). Also, this adaptation can contribute to a visual pollution (lots of cables and loudspeakers surrounding users). Never any experiment had evaluated the real impact of this pollution on sound localization or even in the sense of presence.

5. SIMULATION ROOM

Our simulation room had 3,10 x 5,70 x 3,20 m3. We placed our 8 loudspeakers distanced by 3,10m width, 3,80m length and 2,30m height. This disposition didn’t create a perfect cube, needed for a perfect Ambisonics\(^6\) sound reproduction. Nevertheless, and after some localization tests we found that a normal user could disambiguate sounds coming from the left, right, front, back, up or down positions.

All around the area covered by the sound system (the simulation area), there were black curtains that could hide the loudspeakers and create a separation from the technical area where were a secretary, the sound card and a desktop computer. This curtains could be moved so that the loudspeakers were unhidden (necessary for the third experiment). In the room there was also a projector that could produce a 2,50m x 2m image to the front-side (see Fig. 3).

6. EXPERIMENT 1 - MULTI-SENSORIAL STIMULATION AND PRESENCE IN OVES

A previous work performed by Slater [19], introduced a new method to measure presence in IVEs. In this experiment we followed Slater’s methodology to measure presence in an Outdoor Virtual Environment where four physical modalities were activated (vision, hearing, haptics and smell). We also used the Emotiv EPOC EEG headset in order to remove some of the ambiguity in data (see the VE used on Fig. 4).

\(^6\)Ambisonics is a recording and reproduction technique, invented by Gerzon with the help of his professor Fellgett [9, 12].

Ten participants alternately categorized the relative contribution of each modality to enhance presence: ones were instructed to pay attention to the illusion of being present in the place depicted by the VE (Place Illusion, PI), others to the illusion that events occurring in the VE were real (Plausibility Illusion, Psi). First they experienced the full environment (where all the physical modalities were activated), and then after deactivating some stimuli, they had to increment the VE with the modalities they thought more important to achieve the same illusion as the full environment.

The results showed that participants who were trying to achieve PI, gave a lot of importance to 3-D sounds and wind. The others who tried to achieve Psi gave more importance to wind and images than 3-D sounds. Nevertheless EEG data showed that this preference on images didn’t meant that they were more engaged in the VE when video-projection was enabled. Smell was considered the less important physical modality, but it was usually activated before feeling PI or Psi.

We believe that these results clearly showed the importance of the other physical modalities rather than vision, to enhance presence in OVEs. But more important, we found that in our research Slater’s method is not entirely related to the participant’s subconscious feelings, as measured by the EEG.

6.1 Properties

The property vector is \( P = [V, H, Ha, S] \), where V refers to vision (V(0) no visual stimulus, V(1) lamps, V(2) lamps and video projection), H to hearing (H(0) 2-Dimensional Audition or H(1) 3-Dimensional Audition), Ha to haptics (Ha(0) no haptics stimulus, Ha(1) wind, Ha(2) wind and heat) and S to smell (S(0) no smell, S(1) forest and sea smell).

It is important to understand that the smell of car fresheners in the room was constant. Nevertheless, when the smell...
variables were considered: The 4-tuple \([V,H,Ha,S]\) at which had one factor with two levels, PI or Psi. For those, three participants were given the PI or Psi instructions. Our experiment

6.3 Experimental Design

After getting familiarized with the VE components and capabilities, the users experienced the simulation with all stimuli in their maximum configuration (full environment) and the EEG data was recorded. Before experiencing the full environment the users were given one of the following two instructions:

- (PI) “What are the most important stimuli, to make me feel that I am no longer in this room, but in somewhere else?” Later (when some physical stimuli are disabled) we will ask you to get to that feeling, by enabling the stimulus or stimuli you think are essential for achieving it.
- (Psi) “What are the most important stimuli to make me feel that, what is happening is very close to reality?” Later (when some physical stimuli are disabled) we will ask you to get to that feeling, by enabling the stimulus or stimuli you think essential for achieving it.

After the full environment simulation, participants had one trial that, depending of the order of arrival, started from a different initial configuration. They were encouraged to make transitions and stop whenever they reached the same level of PI or Psi felt in the full environment. In each configuration the EEG was recorded, with markers like (strong wind on, smell off, image changed, user felt Psi, etc...) for an easier interpretation later on.

6.4 Results and Discussion

We followed three methods of analysis. First, we considered the configuration at which users declared a match. We used the Bayes’ Theorem to estimate the probability of a match being declared when the user is experiencing the configuration \([V,H,Ha,S]\). Second, we individually looked at every stimulus and estimated the probability of one being selected by the user on each transition and the probability the stimuli being active when the user reported feeling PI or Psi. Third, we made an analysis of the EEG data and used the Pearson correlation to find a relationship between different environments. We consider a different environment when a VE-A has less or more activated stimuli than a VE-B. The users experienced every stimuli in their maximum configuration (full environment) in the beginning and in the end of the experiment. We made the assumption that the EEG levels of engagement/boredom and the instantaneous excitement are equal at both the beginning and end of the experiment. Therefore we estimated the correlation \(r\) between each different configuration (initial, 1, 2, 3, 4) and the average between the first and last full environments.

3-D audio is important to feel PI and less important, although not significantly, to feel Psi.

We consider that a configuration with vision at its maximum and hearing at its minimum, means that 3-D audio is not so important. Only some of the Psi group had chosen a configuration like \([2,0,*,*]\) on a match. From those users, only one had just one more stimulus activated (Haptics). The rest had at least two more stimuli activated (Smell and Haptics).

Excitement never exhibited a significant increase due to the activation of the video-projection. For the PI group, engagement didn’t had it either, while in Psi group video-projection was responsible for a significant increase. Another interesting observation was that engagement in the Psi group had a larger correlation value than in PI group, due to the activation of 3-D audio. This might mean that, Psi users weren’t well aware of the differences conveyed by 3-D sounds and just realized it after saying that they had felt Psi. This was an observation that great part of the users made, when doing the experiment and after activating 3-D audio: “Now I realized the differences on a 3-D sound system”.  

First choices are Haptics and Hearing.

In the first configuration, PI users activated Hearing(1) or Haptics(1), while the Psi users, Haptics(1 or 2). Haptics(2) only for those who started from an initial configuration like \([0,0,1,0]\). Since the first choice is considered to be the most relevant when trying to feel PI/Psi, and considering that in the questionnaires, only 20% of the users had answered that if they repeated the experiment, they would change the order of choices, we conclude that wind (Haptics(1)), was the most important stimulus to enhance presence in this OVE. The second one was the 3-D Audio (Hearing(1)).
Wind increases the Engagement as measured by the EEG. Wind Haptics(1) was responsible for the highest values both for Psi and PI users. If we had to choose the stimulus that brings the greatest benefit for an OVE, we would say wind.

7. EXPERIMENT 2 - NAVIGATION TASK

In this experiment we carried out a navigation test in CEDAR. Six users needed to identify and mark colliding pipes in three different sets: video projection only, video projection and 2-D sound, video projection and 3-D sound. Depending on the set, the user had some visual or audio-visual cues of the collision nearest to him or her. If the set was the one with video projection only, audio cues were disabled, else audio-visual cues were enabled.

We focused our attention on four factors: number of found collisions; search time between collisions; normalized path length (travelled path length/distance between collisions); and rotation degrees between collisions. The results showed that audio cues make navigation effortless and improve sense of orientation. We found that with 3-D sound, users chose the shortest paths; without audio cues, users lost orientation more often and with 3-D sound users rotate more than with 2-D. We also found that 3-D sound approaches real life behaviours to the VE, therefore enhancing presence.

7.1 Virtual Environment Specification

The VE was a simulation of an oil extraction location in the middle of the ocean. Above the water were an oil tanker and an oil platform. Both the the platform and the tanker had some oil pipes connecting each other, and others connected to the oil wells (Fig. 5).

The pipes were previously positioned in a way that, some were colliding with each others, and others with the coral regions. The navigation control method was very simple to use. The user could move around by the means of a Wii remote control and when near the collision he could mark it by the means of Kinect.

During the experiment, the user had some visual or audio-visual cues of the collision nearest to her. If the set was the one with video projection only, audio cues were disabled, else audio-visual cues were enabled. It is important to note that, in the setup with 2-D sound and video projection, the user could only listen to the sounds coming from hers front.

7.2 Procedures

When the participants arrived at the laboratory they were given verbal information about the experiment and some instructions to read.

Before starting the experiment on each different set (video projection only, video projection and 2-D sound, video projection and 3-D sound), the user was encouraged to test and learn the navigation mechanism and to mark some collisions in a test environment using the different visual or audio-visual cues. This adaptation had no time-limit and it could be repeated until the user felt comfortable with the navigation system and the cues.

The user had 3 minutes to mark as much collisions as possible in each different set. The time left was always visible in the top right corner of the screen.

7.3 Experimental Design

Figure 5: Experiment 2 VE. Pictures 1, 2 and 3 shows the oil tanker and platform. Underwater we can see the coral regions with some oil pipes (4), the Christmas trees (5 and 6) and the pipes connecting the tanker with the platform (7).

The first set of the VE was dependent of the order of arrival of each Participant. In each set, the user had 3 minutes to find the most number of collisions as possible. From the measured data we focused our attention on four factors:

1. Number of found collisions, the user needed to be relatively close to the collision so that it could be marked and considered found;
2. Search time between collisions, the user wasn’t forced to mark the collision that the visual or audio-visual cues indicated;
3. Normalized path length (travelled path length/distance between collisions);
4. Rotation degrees between collisions, the distance between the user and the display screen wasn’t considered for this measure.

7.4 Results and Discussion

Each user was able to find and mark more collisions when audio cues were enabled. On average users found ±3 more collisions on the setups with 2-D and 3-D sound. Also the maximum number of collisions to mark was almost achieved when 3-D sound was enabled (21 out of 22).

The time spent between marking one collision and another was significantly lower when the audio cues were enabled. The setups with 2-D and 3-D sound improved the searching time between collisions in ±6 seconds. Video and 3-D setup established a maximum of 20 seconds between one collision and another.

The normalized path length on Video and 3-D sound setup was the lower.

The users rotated significantly more between collisions when audio cues were disabled. Users rotated more on 2-D than on 3-D.

To understand the users’ behaviour, we analysed the three longest paths between two marked collisions in the three minutes time. If we pick the total distance or the total
rotation and analyse it through the total time spent between collisions, we can have an overview of a typical behaviour on each setup.

Audio cues are essential to improve users' orientation, way-finding and ease of travel in a VE.

The results were clear, audio cues can make navigation in VEs, an effortless task. The objective in this experiment was to find and mark as much collisions as possible in three minutes time. Therefore the objective was best achieved for the video and 2-D setup, which had an higher average and total number of found collisions.

In what concerns to the performance on achieving the objective, we can also state that the two setups with audio cues, presented better results. Less time spent between collisions means that the user got lost less often. Lower normalized path lengths means that the user chose more often the shortest paths. More rotation means more disorientation or can also mean more awareness on what’s surrounding. Having this we can say that:

- With audio cues, users spent less time between collisions so they got lost less often.
- With 3-D sound, users chose shortest paths.
- Without audio cues, users lost orientation more often.
- With 3-D sound users rotated more than with 2-D. We believe that in this case, more rotation meant more awareness on surrounding objects.

3-D sound helps finding the shortest paths more rapidly.

Since the values of search time and normalized path length between collisions presented very similar results for 2-D and 3-D, it was worth looking at the normalized path length through time. Being closer to the next collision in a shortest period of time, means that the user decided where to go more rapidly. Also the first periods of time are when some subconscious decisions are made, if they aren’t correct, most likely a user changes his mind before being half way from the next collision. When comparing two very similar results, it matters if in the initial moments, the user was or was not closer to objective.

The results showed that in the 3-D set, users were closer to the objective in the initial moments. This lets us say that not only 3-D sound helps finding shortest paths more rapidly, but also approaches real life behaviours to the Virtual Environment. Our subconscious decisions are related to our day life experiences, they are effective on a VE if the user doesn’t need to change his mind after taking them.

3-D sound approaches real life behaviours to the Virtual Environment.

When people that can see are lost and need to decide where to go, the first thing that they normally do is look around and see what’s surrounding. To make the best decision possible, one most first “take a clear picture” of the world at 360 degrees.

Results showed that in the initial moments users rotated more with 3-D sound than with 2-D. This lets us do the following observation: most likely the users with 2-D sound, as soon as they had a visual match of a collision, they moved forward to it. It didn’t matter if it was the nearest collision or not, they avoided looking around because they weren’t sure if it worth the effort. On the other hand the users with 3-D sound felt more comfortable to look around and follow what they were hearing, because before rotating they already knew that a collision was there.

Like said before, with 3-D sound subconscious decisions are more effective. Also people seem to be more comfortable to look around and pay attention to what’s surrounding, because before rotating they already have an idea of what’s there.

8. EXPERIMENT 3 - MACHINE-LESS VENTRiloQUiSM

The ventriloquism effect is the maximum unperceived distance between audio and visual stimuli. It is also the association of something physical to a sound source. This means that, even tough that a virtual sound source can be perceived as coming from elsewhere rather than the loudspeakers, the human will always associate the sound to two sources: the image and the loudspeakers.

In this experiment we evaluated the ventriloquism effect with the loudspeakers hidden and visible. The visual source was a talking-head placed in the middle of the screen, the audio-source was increasingly shifted away from the image.

The results of the seven participants, proved that hiding the loudspeakers is an easy and effective way to improve the subjective perception of sound in a VE. This is because the maximum unperceived distance between audio and visual stimuli becomes greater, and therefore users tolerate greatest positioning errors (intrinsic to every sound systems). Results also proved that users’ confidence and correctness are greater when judging where the sound is coming from.

8.1 Virtual Environment Specification

The VE had two different sets: with the 8 loudspeakers hidden and with the 8 loudspeakers visible.

Since each user has its own particular perception of sounds position, in the beginning of the experiment a personal calibration with pink noise \(^7\) was made. It is important to

\(^7\)http://en.wikipedia.org/wiki/Pink_noise, accessed on
note that this calibration was performed only to identify the user perception of the front-left, front-right, front-top and front-bottom position, this is the front-face of the Ambisonics Cube.

There was a video-projection of a talking-head at the center of the screen. We chose solely a talking-head so that there weren’t any more distractors which could influence the perceptiveness of audio and visual stimuli and because, faces are what people are most audible and visually familiarized. The image was always in the same position during all the experiment, the correspondent sound was progressively shifted away from the image.

The user was seated in the middle of the simulation area: his head position was approximately at the same distance from each loudspeaker (the perfect sweet spot for the Ambisonics Sound System). He also had a remote control, to point where did he perceived the sound as coming from. In each shift, the user needed to say if he perceived the sound as coming from the talking-head’s mouth or from elsewhere. If the user perceived it as coming from elsewhere, he was asked to point out where did the sound was coming from (Fig. 6). The size of the pointing aura could be bigger or smaller, reflecting the amount of confidence that the user had when pointing that the sound was coming from a certain position (bigger = less confidence).

### 8.2 Experimental Design

In each set the user needed to answer if he perceived the sound as coming from the talking-head’s mouth, until he identified shifts in every positions (Left, Right, Top, Bottom). They were told that, in each 3 shifts they needed to listen and see all the loop at least one time (one, two, three, four, five, six, seven, eight, nine, ten, zero). This was necessary because, different numbers (different sound frequencies) might be perceived as coming from different positions (spectral cues [3]).

The user was told that he could rest 10 minutes between the experiments in each set. We believe that resting is essential, so that fatigue wouldn’t distort the results. This time was also used to hide/unhide the loudspeakers.

From the measured data we focused our attention on the following:

1. The maximum unperceived distance between the audio and visual stimuli (ventriloquism effect);
2. The correctness on identifying a shift, this is if the user identified a shift but he pointed it as coming from another position;
3. The number of times that a shift was perceived when the sound was coming from the same place as the talking-head’s mouth, useful to identify users’ fatigue;
4. The values of calibration, so that they could be compared with other users’ calibration.

### 8.3 Results and Discussion

When the loudspeakers were hidden, users unnoticed greatest distances between the audio and visual stimuli. The average when the loudspeakers were visible was 22 (STD 9.8 cm), when hidden 37 (STD 30 cm).

When the loudspeakers were hidden, users pointed where the sound was coming from, with more confidence and less error. The average when the loudspeakers were hidden was 61 (CI 12 cm) for distance error and 111 (CI 19 cm) for circle diameter. When the loudspeakers were visible 73 (CI 8 cm) and 119 (CI 15 cm).

**Hiding the loudspeakers can improve the subjective perception of sound in a VE.**

We believe that this simple technique can really improve the way people perceive sounds in a VE with a first order Ambisonics system. Not only helps reducing the imperfections intrinsic to every sound systems, but also helps creating a better illusion, i.e. a better ventriloquism effect.

After looking at the results we can say that, when the loudspeakers were hidden users unnoticed greatest shifts between the image and the sound. The average was 15 cm greater, this meaning that the user tolerated greatest distances between sound and visual stimuli. Also the confidence and correctness were improved, which lets us say that users are influenced by the loudspeakers when judging where the virtual sound source is coming from.

**Users’ calibration average is useful to identify room acoustics problems.**

The values from each user’ calibration are very useful so that we can calculate an average and therefore identify some problems that the room might have. There exist several tools and techniques to do this kind of acoustics corrections, but this method is cheap and effortless.

In our case, the average of users’ calibration told us that our room was probably influenced by some sound reflections (due to its short dimensions), resulting in a shift of 27 cm right and 21 cm up.

### 9. CONCLUSIONS

In an effort to contribute to the community we carried out three different experiments. They were independent from each others but all together tried to answer to our initial question: “Is it worth having 3-D sound systems to enhance presence in VEs?”. After analysing others’ experiments we decided to use a first order Ambisonics system with 8 loudspeakers. For all the three experiments we used the same room. During this time we also mounted a replica of our simulation room in the island of Madeira, which currently serves as a ludic place for tourists and others experiment a travel through the Laurissilva forest of Funchal, Portugal.

In the first two experiments, we answered our initial question: 3-D sound proved to be essential not only to enhance presence (experiment 1) but also to improve navigation, orientation and way-finding (experiment 2). In the third experiment we proved that a simple, cheap and effortless technique can improve the subjective perception of sound.

Amongst others, these were the most important conclusions from the three experiments:

- Wind and 3-D sound are important to feel Place Illusion (experiment 1).
- Wind increases the Engagement in Outdoor Virtual Environments (experiment 1).
- Users’ Excitement doesn’t increase due to the activation of visual stimuli in OVEs (experiment 1).
With audio cues, people get lost less often (experiment 2);
With 3-D sound people chose the shortest paths more often (experiment 2).
With 3-D sound people rotate more than with 2-D (experiment 2).
With 3-D sound subconscious decisions are more effective (experiment 2).
With 3-D sound people seem to be more comfortable to look around and pay attention to what’s surrounding (experiment 2).
When the loudspeakers are hidden, people tolerate greatest distances between audio and visual stimuli (experiment 3).
Hiding the loudspeakers, improves users’ confidence and correctness, when judging where the sound is coming from (experiment 3).

Proving that it is worth the effort to mount a 3-D sound system in a Virtual Environment, is rewarding. The results were enthusiastic and will hopefully be a reference to researchers on these subjects. In our point of view, these three experiments are very important to reinforce the fact that sound systems (2-D and especially 3-D) must not be a last minute add-in, but yes an integral part of the architecture of an Immersive Virtual Environment.

10. FUTURE WORK
Our work studies 3-D sound and its importance on Virtual Environments to enhance presence. The three experiments that we did were essentially to encourage the use of a 3-D sound system, and for that we focused our attention on comparing environments with and without 3-D sounds. These kind of comparisons are essential to reinforce that not only all IVEs should use 3-D sounds, but that their design should depend on the sound-system and sonifications used.
We believe that there are numerous experiments that can be reused and truly contribute to the 3-D sound ascension, just by comparing their previous results when using a 3-D or a 2-D sound system or none. Localization, navigation and orientation are three important variables to analyse, but there are several others: collaborative tasks, sense of presence, emotions, task correctness, interactive gestures, multimodal interfaces, and so on.
It is easy to introduce 3-D sounds in any Virtual Environment and it is important to think on the sound system in the firsts architectures. But for the purpose of research, it doesn’t matter if we mounted a 3-D sound system in the beginning or in the end, there are always possibilities to evaluate the benefit that 3-D sounds brings to a VE.
As told before, there aren’t many experiments that report the good and bad between using different audio/visual configurations to perform a specific task. Experiments that wish to contribute in this area can start by comparing 2-D and 3-D sounds, but can also go a little further and compare other physical modalities like smell, taste or touch. These kind of evaluations are a need, so that some standards to mount VEs can be made.

Finally and very important, we look forward to see others doing the same machine-less ventriloquism experiment we did. We are convinced that our results are true, independently of the sound system or the simulation room used, but this premise can only get strong with some more identical results.

11. REFERENCES


