Audio, Visual, and Audio-Visual Egocentric Distance Perception by Moving Subjects in Virtual Environments

MARC RÉBILLAT, Université Paris-Sud, and École Polytechnique
XAVIER BOUTILLON, École Polytechnique
ÉTIENNE CORTEEL, *sonic emotion labs*
BRIAN F. G. KATZ, Université Paris-Sud

We present a study on *audio*, *visual*, and *audio-visual* egocentric distance perception by moving subjects in *virtual* environments. Audio-visual rendering is provided using tracked passive visual stereoscopy and acoustic wave field synthesis (WFS). Distances are estimated using indirect blind-walking (triangulation) under each rendering condition. Experimental results show that distances perceived in the virtual environment are systematically overestimated for rendered distances closer than the position of the audio-visual rendering system and underestimated for farther distances. Interestingly, subjects perceived each virtual object at a modality-independent distance when using the audio modality, the visual modality, or the combination of both. WFS was able to synthesise perceptually meaningful sound fields. Dynamic audio-visual cues were used by subjects when estimating the distances in the virtual world. Moving may have provided subjects with a better visual distance perception of close distances than if they were static. No correlation between the feeling of presence and the visual distance underestimation has been found. To explain the observed perceptual distance compression, it is proposed that, due to conflicting distance cues, the audio-visual rendering system physically *anchors* the virtual world to the real world. Virtual objects are thus attracted by the physical audio-visual rendering system.

Categories and Subject Descriptors: J.4 [**Computer Applications**]: Social and Behavioral Sciences—*Psychology*; I.3.7 [**Computer Graphics**]: Three-Dimensional Graphics and Realism—*Virtual reality*; H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems—*Audio input/output*; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*Auditory (non-speech) feedback*

General Terms: Experimentation, Human Factors

Additional Key Words and Phrases: Virtual environments, large-screen immersive displays, wave field synthesis, spatialized audio, distance estimation, spatial perception

ACM Reference Format:

Rébillat, M., Boutillon, X., Corteel, E., and Katz, B. F. G. 2012. Audio, visual, and audio-visual egocentric distance perception by moving subjects in virtual environments. ACM Trans. Appl. Percept. 9, 4, Article 19 (October 2012), 17 pages.
DOI = 10.1145/2355598.2355602 http://doi.acm.org/10.1145/2355598.2355602

Authors' addresses: M. Rébillat: Équipe Audition, Laboratoire de Psychologie de la Perception, École Normale Supérieure, Paris, France; email: marc.rebillat@polytechnique.edu; X. Boutillon: LMS, École Polytechnique, F91128 Palaiseau, France; email: boutillon@lms.polytechnique.fr; É. Corteel, *sonic emotion labs*, 42 bis rue de Lourmel, F75015 Paris, France; email: etienne.corteel@sonicemotion.com; B. F. G. Katz, LIMSI-CNRS, rue John von Neumann, Université Paris-Sud, F91403 ORSAY, France; email: brian.katz@limsi.fr.

DOI 10.1145/2355598.2355602 http://doi.acm.org/10.1145/2355598.2355602

1. INTRODUCTION

Virtual reality (VR) systems aim to provide subjects with a virtual world where they would behave and learn as if they were in the real world [Brooks 1999]. Audio-visual (AV) VR-systems that combine large immersive screens and many loudspeakers have been developed to provide subjects with a virtual space merging coherently the holophonic spatial audio and 3D visual renderings [Faria et al. 2005; Springer et al. 2006; Rébillat et al. 2008]. Here the term *holophonic spatial audio* stands for technologies such as wave-field synthesis (WFS) [Berkhout et al. 1993], Ambisonics [Gerzon 1985], and others [Komiyama et al. 1991], which attempt to physically recreate the same sound field that a real sound source would have radiated, and thus provide subjects with a natural spatialized sound rendering. Such AV VR-systems are very appealing because they are minimally intrusive (no headphones needed, only lightweight glasses) and allow subjects to move freely in the rendering area, while always holding a correct AV perspective. With the emergence of these multimodal systems, the question arises of the correct perception of the virtual space by moving subjects and, more specifically, of rendered distances within it [Loomis and Knapp 2003; Interrante et al. 2008].

1.1 Measurement Protocols for Estimating of Perceived Egocentric Distance

Because distance perception is a cognitive task, measurement protocols are needed to estimate perceived absolute egocentric distances. Existing measurement protocols can be divided into three main classes [Klein et al. 2009; Grechkin et al. 2010]: verbal estimations, perceptually directed actions, and imagined actions. In *verbal estimation* protocols, subjects assess the perceived distance in terms of familiar units, such as meters. In *perceptually directed action* protocols, an object is presented to the subject, who then has to perform an action, such as blind-walking, without perceiving the object. In *imagined action* protocols, the action is imagined instead of being performed, and response times are used to infer the results of the action. The advantage of perceptually directed actions is that they lead to distance estimations that are more accurate and less variable than distance estimations provided by verbal reports [Fukusima et al. 1997; Loomis et al. 1998; Russell and Schneider 2006; Andre and Rogers 2006]. Moreover, using *perceptually directed actions*, estimated distances can be directly inferred from actions, whereas a potential systematic bias exists in distances estimated using *imagined action* protocols, due to the conversion of a directly measured value of time into an indirect measure of estimated distance [Grechkin et al. 2010]. Hence, *perceptually directed actions* were preferred in the present study.

Among perceptually directed actions, *direct* blind-walking and *indirect* blind-walking (triangulation) are two possible alternatives both of which lead to accurate distance estimations [Fukusima et al. 1997; Loomis et al. 1998]. Due to physical spatial constraints imposed by the presence of large screens and many loudspeakers, only indirect blind-walking (triangulation) is possible in the kind of AV VR-systems under study here [Klein et al. 2009]. An advantage of the triangulation measurement protocol is that it is applicable to the measurement of audio, visual, and audio-visual perceived absolute egocentric distances, without any need to adapt the procedure to each different modality. One disadvantage is that small errors in pointing can lead to large differences in indicated distance for very distant targets. Furthermore, the error is not symmetric, since one degree of rotation in one direction can equate to a smaller change in linear distance than an equivalent rotation in the opposite direction.

1.2 Perceived Distance in the Visual and Auditory Modalities in Real or Virtual Environments

In classical *visual* VR-systems, such as head-mounted displays (HMD), perceived visual distances have been observed to be systematically underestimated [Loomis and Knapp 2003; Interrante et al. 2008]. This is not the case in the real world [Wiest and Bell 1985]. VR-systems based on large immersive screens were thought to offer a better distance perception [Plumert et al. 2005]. Studies focusing

on visual distance perception in virtual environments rendered by large immersive screens have found that visual distances were underestimated using these systems, exactly as in HMD systems [Armbruster et al. 2008; Naceri et al. 2009; Klein et al. 2009; Grechkin et al. 2010; Alexandrova et al. 2010].

In the *audio* real world, it is well established that near-auditory distances (<2m) are overestimated, whereas far-auditory distances (>2m) are underestimated (see Zahorik et al. [2005] for a review). Much less is known regarding auditory distance perception in virtual auditory systems based on holophonic spatial audio. In Corteel [2004] and Rébillat et al. [2008], it was shown that holophonic spatial sound renderings can be used effectively to render distances for static sources with moving subjects and that perceived distances are compressed with respect to rendered distances. When subjects are static, [Komiyama et al. 1991; Kearney et al. 2012] showed that performances in an holophonic audio virtual environment matched well with real-world performances in terms of distance perception.

In *audio-visual* virtual environments, perceived visual distances appear to be underestimated, near-auditory distances to be overestimated, and far-auditory distances to be underestimated. Audio and visual perceived distances are thus *a priori* inconsistent for a given rendered distance. Some effort has been made to study how audio and visual distance cues are merged in virtual environments [Côté et al. 2011]. Results suggest that static subjects perceived AV distances similarly to visual distances. However, subjects are rarely static when immersed in virtual worlds. Hence it is important to study how AV distances are perceived by subjects benefiting from static and dynamic AV distance cues in a virtual environment.

Furthermore, to provide subjects with a virtual world where they would behave as if they were in the real world, VR-systems should be fully "transparent" to subjects. Transparency is understood here as "the extent to which the computer displays are capable of delivering an inclusive, extensive, surrounding, and vivid illusion of reality to the senses of a human participant" [Slater and Wilbur 1997]. However, AV VR-systems are not perfect and suffer from some drawbacks that could potentially limit their transparency. So it is important to assess whether this limitation has an influence on the AV virtual space perceived by subjects, and whether there exists a spatial link created by AV VR-systems between the real world and the virtual world.

1.3 Objectives

In this article, we study *audio* (A), *visual* (V), and *audio-visual* (AV) egocentric distance perception in the action space (1.5m to 6m) by moving subjects in *virtual* environments. AV rendering is provided via the SMART-I^2 platform (Spatial Multi-user Audio-visual Real-Time Interactive Interface) [Rébillat et al. 2008, 2009] using tracked passive visual stereoscopy and acoustic wave field synthesis (WFS). This AV VR-system allows subjects to move freely in the rendering area, and everywhere in this area maintains stable AV perspective cues. Distances are estimated by means of perceptually directed action (indirect blind-walking, triangulation) under A, V, and AV conditions. This experiment aims at studying how A, V, and AV distances are perceived by subjects taking benefit of static and dynamic AV distances cues in a virtual environment. A second objective is to assess whether the fact that the AV rendering system is not fully transparent induces a spatial link between the real and virtual worlds.

2. METHOD

2.1 Experimental Design

Five virtual objects (denoted A, B, C, D, E) placed in the subject's *action space*, that is, the space where one "moves quickly, talks, and if needed can throw something to a compatriot or at an animal" [Cutting 1997], were rendered (see Figure 1). Two initial or starting positions for participants were

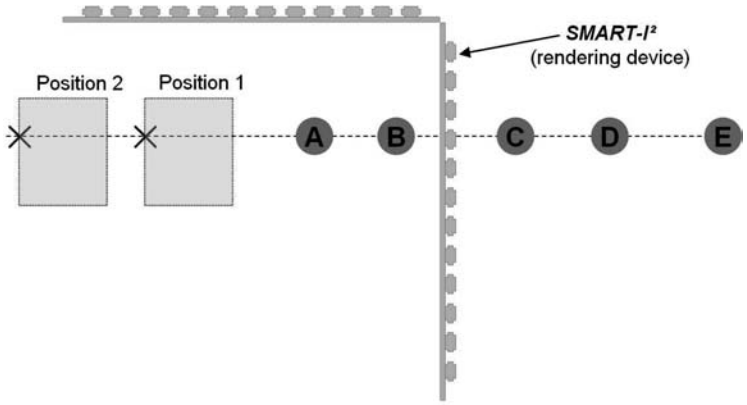


Fig. 1. Overview of the experimental setup. *Virtual objects*: grey disks labeled A, B, C, D, or E. *Start positions*: black « × ». *Exploration areas* are represented by the grey rectangles.

tested: Position 1, where subjects stood 2.3m in front of the right panel of the SMART-I^2, and Position 2 where they stood 3.3m from it (see Figure 1). Virtual objects are at the same locations with respect to the rendering system for both starting positions. Virtual objects were located at distances of 1.5m, 2m, 2.5m, 3.5m, and 5m from Position 1, equating to distances of 2.5m, 3m, 3.5m, 4.5m, and 6m from Position 2.

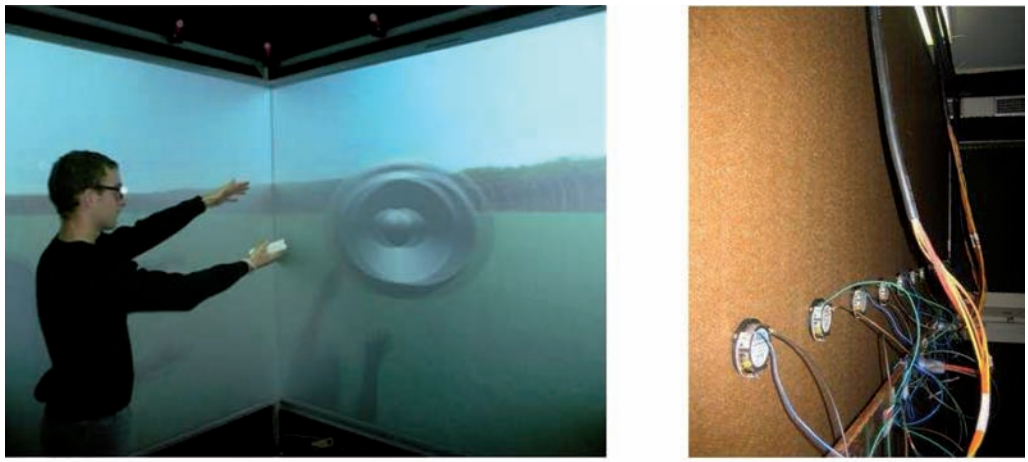
A total of 40 volunteers (30 men, 10 women) between 21 and 49 years of age participated in the experiment with half of the subjects starting from Position 1 and the other half starting from Position 2. All subjects had self-reported normal vision (possibly corrected) and normal hearing. Each subject had to estimate the distances of the five virtual objects four times under each rendering condition. They performed three sessions of 20 iterations each after a training phase of two iterations under each rendering condition. In the training phase, rendered distances were 3m and 7m for Position 1, and 4m and 8m for Position 2. Subjects took pauses between sessions, and the entire experiment lasted approximately one hour. The session order was balanced between the six possible permutations of the three rendering conditions.

Hence the chosen experimental design was a mixed design with three factors: rendered distance d_r (five levels, within-subjects); rendering condition (three levels, within-subjects); and starting position (two levels, between-subjects). The dependent variables are perceived distance d_p, time t_{XP} spent in the exploration phase (see Section 2.4), and exploration path length l_{XP}.

2.2 Experimental Setup

Experiments were conducted in the AV virtual environment produced by the SMART-I^2 platform [Rébillat et al. 2008, 2009]. In this system, front-projection screens and loudspeakers are integrated together to form large flat multi-channel loudspeakers, also called *Large Multi-Actuator Panels* (LaMAPs). The rendering screens consist of two LaMAPs (2m × 2.6m, each supporting 12 loudpeakers) forming a corner (see Figure 2). The reporting interface used in the present experiment was a *wiimote*.

Visual rendering was produced using tracked passive stereoscopy rendered at 80 frames per second with a resolution of 1280 × 960 pixels on each screen. Interocular distance for stereoscopic rendering was fixed at 6cm for all subject. At both starting positions (the black « × » in Figure 1, 3(a), and 3(b)), the horizontal field of view was approximately 150° and the vertical field of view approximately 70°. Since it has been shown that graphical resolution [Ryu et al. 2005; Grechkin et al. 2010] and field of



(a) Front view (b) Back view

Fig. 2. *Left*: a subject and an audio-visual object in the virtual world provided by the SMART-I^2. Visual rendering is projected on the front faces of the two LaMAPs which form a corner. *Right*: electro-dynamical exciters are glued on the back of each LaMAP.

view [Creem-Regehr et al. 2005] have no influence on *visual* distance perception, these experimental parameters should not influence the experimental results.

Spatial audio rendering was realized via acoustic wave field synthesis (WFS) [Berkhout et al. 1993]. This technology attempts to physically recreate the acoustic sound field corresponding to a virtual source at any given position in the horizontal plane, without the need for tracking. Realtime audio signal processing was achieved by a Wave 1 rendering engine provided by *sonic emotion*. The inter-loudspeaker distance of 21cm corresponds to an aliasing frequency $f_{al} \simeq 1.1$kHz, to which the sound field is correctly reconstructed [Corteel 2004]. It was demonstrated by Sanson et al. [2008] that sound fields reconstructed by WFS are sufficiently consistent to allow for accurate localization, even when frequencies above f_{al} are present. In Corteel et al. [2007], it was shown that even, if not *exact*, azimuthal cues above the aliasing frequency f_{al} are generally consistent with azimuthal cues below f_{al} when using MAPs.

Furthermore, fine temporal and spatial calibration has been performed to ensure that the audio and visual renderings are fully coherent.

2.3 Audio, Visual, and Audio-Visual Stimuli

The visual environment was an open, grassy field, with a forest at 50m (Figure 2(a), trees were $\simeq$ 7m tall). The associated audio environment consisted of the sound of wind in the trees accompanied by some distant bird song (overall background level of 36dBA). The audio environment was created by 12 plane waves equally distributed in the horizontal frontal field of rendering (that is, between -70° and 70°). Environmental sound levels were adjusted to be slightly above the background noise produced by the video-projectors (background noise level of 34dBA).

The chosen visual target object was a footless 3D loudspeaker, approximately spherical, with a diameter of $\simeq$ 30cm (Figure 2(a)). The stand was removed to avoid window violations when the object was displayed in front of the screen. The floating loudspeaker was positioned at a height of 1.6m, and shadows were displayed.

Table I. Available Audio-Visual Cues

Available cue	Modality	Class
Object size/level	A, V	Relative*
Motion parallax	A, V	Absolute
Time-to-impact	A, V	Absolute
Binocular/binaural cues	A, V	Absolute†
Height in the visual field	V	Absolute

*These cues are absolute if the subject is familiar with the object.
†Binocular cues convey more information for sources closer than $\simeq$ 2m.

The associated *audio target* object was a 4kHz low-pass filtered white noise with a 15Hz amplitude modulation. Low-pass filtered white noise was chosen in order to have a wide spectral content and to allow subjects to rely on numerous audio localization cues. The white noise was modulated in amplitude by a sine wave to produce attack transients, which are also useful in sound localization [Blauert 1999]. No simulated room-effect (i.e., ground reflection) was included. The sound level of the omnidirectional audio object corresponds to a monopole emitting 78dB(SPL) at 1m, well above the environmental sound level at each of the tested distances.

Audio and visual objects were always displayed coherently, that is, at the same spatial position. In addition, their visual size and audio level decreased naturally with distance. As the experimental design allowed subjects to move within the rendering area (see Section 2.4), they could rely on a large number of cues naturally available in the corresponding real environment for the estimation of distances, including dynamic cues. In particular, *motion parallax*, which denotes changes in the angular direction of a point source, occasioned by the subject's translation, was available. This cue has been shown to be useful for distance estimation using the visual [Beall et al. 1994; Nawrot and Stroyan 2009] or the auditory modality [Speigle and Loomis 1993; Porschmann and Storig 2009]. Another dynamic cue, the estimated time-to-impact for a constant velocity between the moving subject and the static source (also denoted acoustic or visual τ), can also be used [Ashmead et al. 1995; Porschmann and Storig 2009]. Available AV distance cues are summarized in Table I.

The AV background environment was kept active in all the rendering conditions. In the *audio* condition, the spatialized sound corresponding to the virtual object was played while no image of the virtual object was shown. The only visual image consisted of the open, grassy field with a forest in the background. In the *visual* condition, the 3D image of the virtual object was displayed with no corresponding sound. The only audio signal consisted of the sound of wind in the trees accompanied by some bird songs. In the *audio-visual* condition, the spatialized sound of the virtual object was rendered with its corresponding 3D image and the AV environment.

2.4 Experimental Task

Distance estimation was performed here in two phases: a *presentation* phase, see Figure 3(a), and a *reporting* phase, see Figure 3(b). Subjects began each iteration at one of the two possible *start positions*, indicated by a black « **x** » in Figures 1, 3(a), and 3(b).

Before starting the presentation phase, subjects had to indicate that they were ready to perform this phase by pressing a *wiimote* button. In the presentation phase, subjects move around in the *exploration* area which was a rectangle of $1 \times 0.8\text{m}^2$. Subjects were instructed to move in the exploration area in order to acquire "a good mental representation of the virtual object and its environment." A typical path followed by a subject during the *presentation* phase is depicted in Figure 3(a).

Once a "a good mental representation" was acquired, subjects were asked to press a button to indicate that they were ready for the reporting phase. At this point, the target stimuli was stopped, and the procedure for distance estimation by means of triangulated blind-walking began, as depicted in

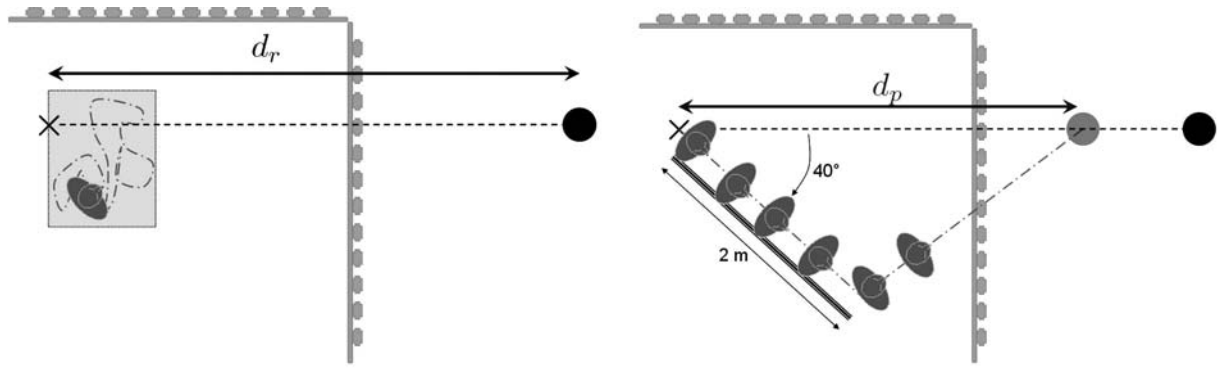


Fig. 3. Presentation and reporting phases. *Start position*: black « × ». *Virtual object*: black disk placed at a rendered distance d_r from the *start position*. In Figure 3(a), the *exploration area* is represented by the grey rectangle, and the dotted grey line indicates a hypothetical exploration trajectory performed by the subject. In Figure 3(b), the *guide* is shown as a thick plain black line. The dotted grey line indicates a hypothetical trajectory performed by the subject. *Perceived object*: grey disk placed at the estimated perceived distance d_p from the *start position*.

Table II. Post-Session Questionnaire.

Q1†:	I had the feeling of locating a real object.
Q2†:	I had the feeling of looking at a TV instead of really being in an outdoor environment.
Q3†:	The virtual environment became real for me and I forgot the real environment.
Q4†:	I remember the virtual environment more as a place where I have been than as a computer generated image I have seen.
Q5*:	I had the impression that I could touch the virtual objects.
Q6*:	I felt present in the virtual world.
Q7†:	I felt surrounded by the virtual world.

†Statements adapted from Bormann [2005].
*Statements adapted from Armbruster et al. [2008].

Figure 3(b). Subjects closed their eyes, made a 40° right-turn to a handrail guide which was included to help during blind-walking, and walked blindly for an imposed distance of $\simeq$ 2m, following the handrail to the end. Subjects stopped at the end of the guide, turned in the direction where they thought the object was, and took a step forward in the direction of the source position. Subjects had been instructed that the perceived distance was to be calculated according to this step. They then indicated that they had completed the reporting phase by again pressing a button. Afterwards, they could open their eyes and go back to their initial "start position" for the next trial. The experimental protocol was fully automated, with the subjects being observed remotely so as not to disturb the sense of presence.

2.5 Post-Session Questionnaire

In the present experiment, subjects were asked to complete a 7-item questionnaire at the end of each of the three experimental sessions (A, V, AV). The goal of this questionnaire was to evaluate the feeling of *presence* that subjects experienced during each session. This questionnaire was built by adapting statements taken from Bormann [2005] and Armbruster et al. [2008], translated into French. Statements were rated on a 7-point Likert scale ranging from −3 to 3 with two anchors. The statements are provided in Table II.

3. ANALYSIS OF RESULTS

As differences larger than twice the standard deviation were observed on the mean estimated distances for three subjects relative to the mean estimated distances of all subjects, data from these 3 subjects were removed from further analysis. Only results from the 37 remaining subjects (17 for Position 1 and 20 for Position 2) are shown in this section.

3.1 Extraction of the Dependent Variables

This section explains how the different dependent variables (d_r, t_{XP}, l_{XP}) were derived from the experimental data. The position of the head of the subject (central point between the eyes) is recorded for each iteration during both the *presentation* and *reporting* phases at 100Hz.

The duration of the *presentation* phase, t_{XP}, was obtained by measuring the time between the subject's button presses for "ready for a new trial" and for "ready for the reporting phase," as explained in Section 2.4. The exploration path length l_{XP} was calculated by using the head position recorded during t_{XP}.

The exploration path-length walked during the exploration phase l_{XP} can be separated into the component that is walked parallel to the direction of the source l^P_{XP} and the component walked in the orthogonal direction, l^O_{XP}. To be comparable, these two paths were normalized by the maximum physical path lengths in each direction, which here are the sides of the exploration area (that is, $s_O = 1$m and $s_P = 0.8$m). The dependent variable $P = l^P_{\mathrm{XP}}/s_P$ (respectively, $O = l^O_{\mathrm{XP}}/s_O$) denotes the number of times the subject walked the length of the exploration area in parallel (respectively, orthogonal) to the source direction.

Perceived distances d_p were estimated from the *triangulation* trajectory as follows: a line ($y = ax+b$) was fitted to the trajectory points during the forward step (118 ± 67 points have been used for the fit). The estimated perceived distance is given by Eq. (1):

$$d_p = -\frac{b}{a}. \tag{1}$$

The relative 95%-confidence intervals on the estimated distance are deduced from the 95%-confidence intervals of the linear fit regression coefficients a and b for each iteration (`regress` function in Matlab). Accross all iterations, subjects, and distances, the 95%-confidence intervals for the relative distances, that is, for d_p/d_r, estimated using the triangulation trajectory, is $\pm 8.5\%$. Thus the triangulation procedure and the associated data treatement provide a reliable estimation of the perceived distances.

3.2 Presentation Phase

In this section, the influence of the rendered distance d_r and of the condition (A, V, AV) on time t_{XP} and on the path length l_{XP}, respectively, spent and walked during the presentation phase are analyzed. Data collected for Position 1 and Position 2 are pooled together as a one-way ANOVA performed on the exploration time t_{XP} with factor starting position showing no significant difference ($F = 1.94$ and $p < 0.17$). Results of the analysis are shown in Figure 4.

A two-way repeated-measures analysis of variance (ANOVA) performed on the exploration time t_{XP} with condition (A, V, AV) and rendered distance d_r, as within-subject factors shows that condition is highly significant ($F(2, 64) = 6.44$ and $p < 0.003$); that rendered distance is significant ($F(4, 64) = 3.38$; and $p < 0.02$); and that there is no interaction effect between condition and rendered distance d_r ($F(8, 64) = 0.82$ and $p < 0.9$). *Post-hoc* tests, computed in terms of medians are shown for condition in Figure 4(a); they reveal that exploration times for each condition are significantly different. Post-hoc tests for rendered distances shown in Figure 4(b) reveal that exploration times for the virtual object B are slightly, but significantly, lower than those obtained for the virtual object E.

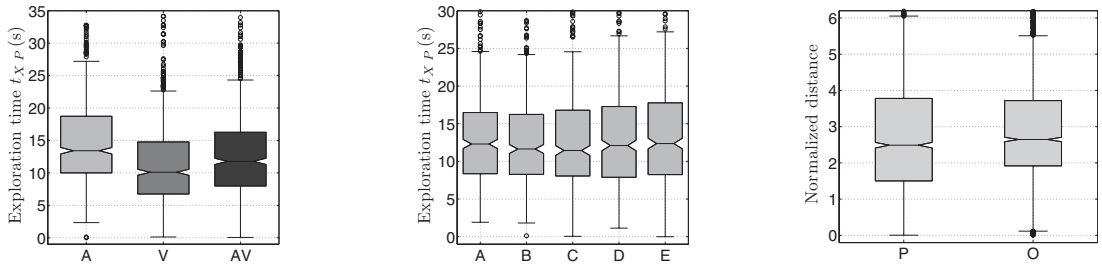


Fig. 4. Exploration time (t_{XP}) as a function of (a) the condition and (b) virtual objects, and (c) comparison of the normalized exploration path lengths in the direction of the virtual object (P) and in the perpendicular direction (O). On each box, the central mark is the median, the edges of the box are the 25^{th} and 75^{th} percentiles, the whiskers extend to the most extreme data points (outliers not considered). Points are drawn as outliers if they are greater than $q_3 + 1.5(q_3 - q_1)$ or less than $q_1 - 1.5(q_3 - q_1)$, where q_1 and q_3 are the 25^{th} and 75^{th} percentiles, respectively. Notches denote comparison intervals. Two medians are significantly different at the 5% significance level if their intervals do not overlap. Interval endpoints are the extremes of the notches.

Subjects spent more time in the exploration phase when estimating distances using the audio modality than when using the audio-visual modality. Furthermore, subjects spent more time in the exploration phase when estimating distances using the audio-visual modality than when using the visual modality only.

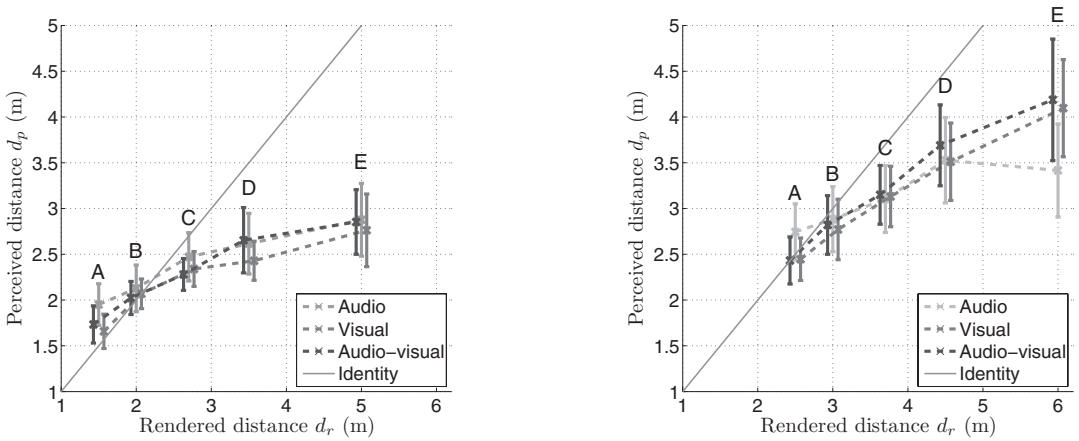
Normalized exploration path lengths in the direction of the virtual object (P) and in the orthogonal direction (O) are compared in Figure 4(c). The analysis shows that P is slightly, but significantly, longer than O. Given a certain exploration area, subjects walked 1.06 times longer in the direction parallel to the virtual object than in the direction perpendicular to the virtual object during the exploration phase.

3.3 Reporting Phase

In this section, the influence of the rendered distance d_r and of the rendering condition (A, V, AV) on the perceived distances d_p is analyzed for each starting position. Means and standard deviations of perceived distances for subjects starting from Position 1 are shown in Figure 5(a) and for subjects starting from Position 2 in Figure 5(b).

For Position 1, the perceived distances d_p were analyzed using a repeated-measures two-way ANOVA with rendered distance d_r and rendering condition (A, V, AV) as factors. Rendered distance d_r is significant at the 5% level with $F(4, 64) = 93.87$ and $p < 10^{-6}$. Rendering condition is not significant at the 5% level as $F(2, 64) = 1.96$ and $p < 0.09$. An "almost" significant effect could be seen at this level between rendered distance d_r and condition, since $F(8, 64) = 1.96$ and $p < 0.06$. The virtual object A is perceived as almost significantly farther in the A condition than in the V or AV conditions. As *post-hoc* tests, a series of Bonferroni corrected t-tests were performed and all the rendered distance pairs were found to be significantly different.

For Position 2, the perceived distances d_p were analyzed similarly. Rendered distance d_r is significant at the 5% level with $F(4, 64) = 48.79$ and $p < 10^{-6}$. The rendering condition is not significant at the 5% level, as $F(2, 64) = 1.84$ and $p < 0.17$. A significant interaction is found between rendered distance d_r and condition, as $F(8, 64) = 8.95$ and $p < 10^{-6}$. The virtual object A is perceived as significantly farther in the A condition than in the V or AV conditions. The virtual object E is perceived significantly closer in the A condition than in the V or AV conditions. As *post-hoc* tests, a series of Bonferroni corrected t-tests were performed and all the rendered distance combinations were found to be significantly different.



(a) Position 1 (b) Position 2

Fig. 5. Mean and standard deviation of perceived distances d_p as a function of rendered distance d_r for each rendering condition and each starting position. Vertical lines represent one standard deviation.

Thus, for both starting positions, the different distances d_r are correctly ordered and well recognized by subjects, independently of the rendering condition. Interestingly, subjects perceived each virtual object at a modality-independent distance when using the audio modality, the visual modality, or the combination of both. The audio-visual spatial rendering provided by the SMART-I^2 is, in this sense, fully coherent in distance. By comparing Figures 5(a) and 5(b), it can also be seen that the starting position has a direct impact on distance perception. This aspect of the results will be discussed in detail in Section 4.

A very small influence in the presentation order was observed: subjects presented with the audio condition in third position made slightly larger errors than subjects presented with the audio condition in the first position. However, this effect remained small. The possibility of any learning effect that could have occurred during the 60 trials of the experiment was checked by comparing groups of 10 successive trials. No significant differences between the relative errors made by the subjects among the different groups of trials were found. Thus, no learning effect appeared during the experiment.

3.4 Post-Session Questionnaires

At the end of each session (A, V, and AV), subjects rated 7 statements on a 7-point Likert scale with two anchors (see Table II). As differences between the various sessions are to be analyzed for each statement, any bias due to subjects has been removed using the following procedure: the rating $A_n^i(k)$ of the n^{th} subject for the k^{th} statements during session i (i = A, V, AV) has been transformed into $\underline{A}_n^i(k) = A_n^i(k) - M_n(k)$, with $M_n(k)$ the mean over the three sessions of the ratings of the n^{th} subject for the k^{th} statements. The *presence*-score has been built as the mean of the unbiased ratings $\underline{A}_n^i(k)$.

The results in Figure 6 show that the scores of the A condition are significantly lower than the scores of the others conditions (V and AV) and that the V and AV conditions are not significantly different. Thus, presence is rated significantly lower in the A condition than in the V and AV ones. Moreover, presence is rated statistically equivalently for the V and AV rendering conditions.

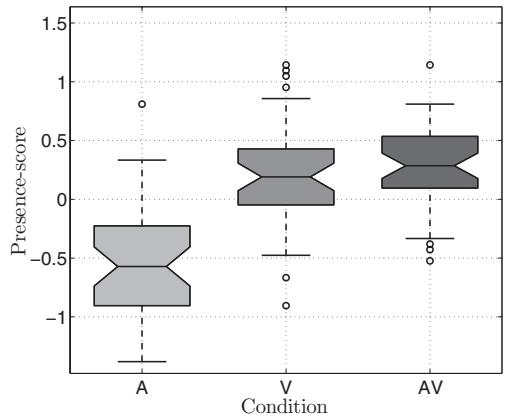


Fig. 6. *Presence* score for each of the rendering condition; for an explanation regarding boxplots, see the caption of Figure 4.

4. INFLUENCE OF THE PRESENCE OF THE AV VR-SYSTEM ON THE PERCEIVED VIRTUAL SPACE

4.1 Potential Conflictual Audio-Visual Spatial Cues

Like the vast majority of virtual and augmented reality systems, the SMART-I^2 system is not perfect, and potentially provides conflicting audio-visual spatial cues. As specified in Section 2.3, no room effect was synthesized in order to recreate acoustical conditions that were as close as possible to open free-field conditions. Nevertheless, even though the experimental room had been acoustically treated, there were still traces of a room effect, with a mid-frequency mean reverberation time T_{60}(500Hz to1kHz) of 0.45s. The ratio of the energies of the direct and reverberated components of the sound, which is an audio distance cue [Bronkhorst and Houtgast 1999], specifies to the subject a distance corresponding to the physical setup rather than the distance to the virtual object.

The technology used to provide the 3D visual rendering is not perfect either. To estimate the distance of the virtual visual object, subjects use two binocular cues. Focus cues (accommodation and blur in the retinal image) specify the distance at which the screen, instead of the virtual object, is seen. Vergence cues correspond to the distance at which the optical axes of the two eyes cross one another, that is, the virtual object. In a 3D visual rendering setup based on large immersive screens, focus cues are almost always in conflict with vergence cues [Howarth 2011] that can affect depth perception [Watt et al. 2005; Hoffman et al. 2008]. Finally, shadows projected on the virtual ground floor were visible only for the virtual objects D and E, but not for the nearer ones A, B, and C. For close distances, the lack of shadow is also in conflict with other spatial cues.

4.2 Anchor Hypothesis

The possible presence of conflictual audio-visual cues can potentially have an effect on distance perception. If subjects are experiencing audio-visual cues specifying two different distances, it is expected that the virtual object will be perceived somewhere between these two distances. Furthermore, the only distance at which all cues are in agreement corresponds to the physical location of the AV VR-system. Hence, distance perception is expected to be correct at that position. In the results presented in Figures 5(a) and 5(b), virtual objects rendered in front of the LaMAP (i.e., A and B) appear to be pushed toward the LaMAP, whereas virtual objects rendered behind the LaMAP (i.e., C, D, and E) seem to be pulled toward it. Furthermore, the distance at which the perceived distance equals the rendered distance corresponds roughly to the distance between the subjects and the physical location

of the AV VR-system (i.e., $D_s^{P1} = 2.3$m for Position 1 and $D_s^{P2} = 3.3$m for Position 2). Thus it is hypothesized that because some audio-visual cues specify the distance of the setup (here the screen), instead of the distance of the virtual object, the AV VR-system physically anchors the virtual world to the real world by attracting the virtual objects to it. At this point, it must be noted that two distances must be considered during the experiment: the distance that must be evaluated by the subject, which is defined explicitly as the distance between the object and the initial position ($P1$ or $P2$) of the subject, and the distance between the object and the subject, which varies during the exploration phase. The latter is used by the subject to evaluate the former. It is also hypothesized that anchoring, if such an effect exists, pertains to the average of this latter distance, which is experienced by the subject during the exploration phase.

Following Zahorik et al. [2005] for the audio modality and Wiest and Bell [1985] for the visual modality, it is assumed that a compressive model in the form $d_p = k \times (d_r)^a$ relates the perceived distance d_p to the rendered distance d_r. The coefficient a denotes the global perceptual compression, and is not expected to be influenced by the starting or exploring position of the subjects. However, the value of a may differ between the different rendering conditions. If subjects are located at a distance D_a from the rendering device, the anchor hypothesis predicts the value of k, and the relation between d_p and d_r should be

$$d_p = D_a \times \left(\frac{d_r}{D_a}\right)^a. \tag{2}$$

Thus, the anchor hypothesis predicts for each starting position that:

$$\text{Position 1} \quad \rightarrow \quad D_a^{P1} = D_s^{P1} + \left\langle \underline{l_{\text{XP}}^P} \right\rangle \quad \text{with} \quad D_s^{P1} = 2.3 \text{ m}, \quad \text{and} \quad a^{P1} = a \tag{3}$$

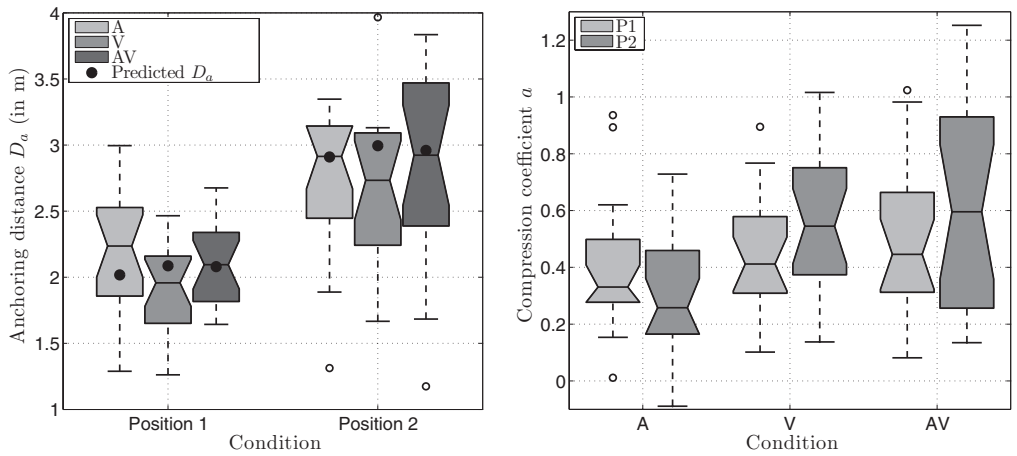
$$\text{Position 2} \quad \rightarrow \quad D_a^{P2} = D_s^{P1} + \left\langle \underline{l_{\text{XP}}^P} \right\rangle \quad \text{with} \quad D_s^{P2} = 3.3 \text{ m}, \quad \text{and} \quad a^{P2} = a \tag{4}$$

where $\langle \underline{l_{\text{XP}}^P} \rangle$ is the average of the algebraic walking displacement of the subject in the direction of the source during the exploration phase. If no correlation is observed between several of the crossing distances between curves in Figure 5 (for Position 1 and Position 2) and the physical distances to the screen, the anchor hypothesis does not stand.

4.3 Experimental Evidence of the Anchor Effect

For the two starting positions that have been tested, the values of D_a and a, for each subject and for each rendering condition, were estimated by fitting the compressive model of Eq. (2) to the collected data. Among all fits, a mean $R^2 = 75.3\%$ was obtained, highlighting the high quality of the model. The estimated and predicted anchoring distances D_a and the compression coefficients a are plotted versus the rendering condition (A, V, AV) and the starting position (Position 1, Position 2) in Figure 7.

From Figure 7(a), it can be observed that the anchoring distances D_a corresponding to each rendering conditions are not significantly different for a given starting position. Moreover, for all rendering conditions the anchoring distances D_a are significantly larger for Position 2 than for Position 1. For the audio condition, median values of $D_a^{P1} = 2.24$m and $D_a^{P2} = 2.91$m are obtained, whereas Eqs. (3) and(4) predict 2.02m and 2.90m, respectively; see Figure 7(a). For the visual condition, median values of $D_a^{P1} = 1.96$m and $D_a^{P2} = 2.73$m are obtained (vs. 2.09m and 2.99m, respectively). For the audio-visual condition, median values of $D_a^{P1} = 2.09$m and $D_a^{P2} = 2.92$m are obtained (vs. 2.08m and 2.96m, respectively). From Figure 7(b), it can be observed that the compression coefficient, a, corresponding to each rendering condition, is not significantly different for all rendering conditions between Position 1 and Position 2. A median value of $a = 0.31$ is obtained for the audio condition, $a = 0.48$ for the visual condition, and $a = 0.45$ for the audio-visual condition.



(a) Anchoring distance D_a (b) Compression coefficient a

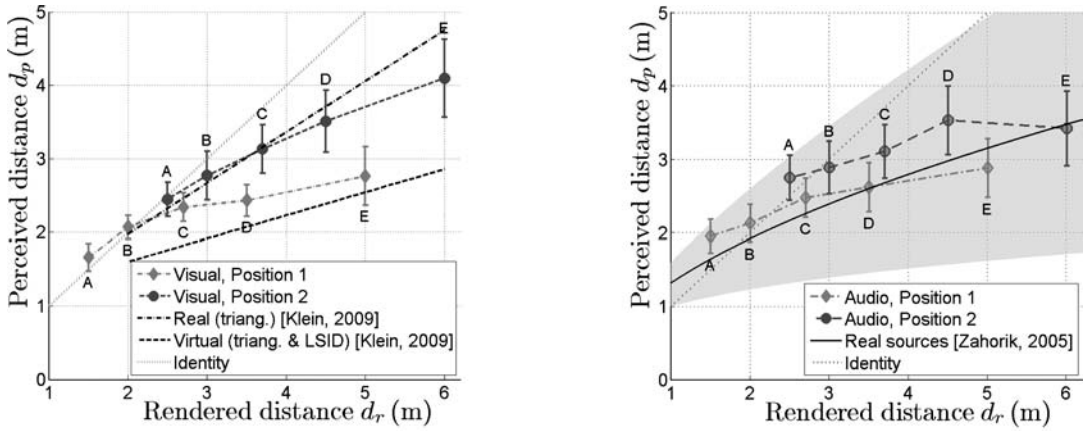
Fig. 7. Anchoring distance D_a and compression coefficient a versus rendering condition (A,V, AV) and starting position (Position 1, Position 2); for an explanation regarding boxplots, see the caption of Figure 4.

The anchor hypothesis predicts, according to Eqs. 3 and 4, that the anchoring distance D_a should be at the positions indicated by the black dots in Figure 7(a). The quantitative agreement between these predictions and the experimental anchoring distances is excellent, specifically for the AV condition. Furthermore, the anchor hypothesis predicts that the compression coefficient a should not differ between Position 1 and Position 2. This is effectively the case, as shown in Figure 7(b). This is experimental evidence arguing in favor of the anchor hypothesis proposed in Section 4.2.

5. GENERAL DISCUSSION

5.1 Perceived Distance in *Visual* Large Screen Immersive Displays (LSID)

Klein et al. [2009] have studied, using triangulation, visual egocentric distance perception in an open grassy field in the real world and in a virtual world rendered via LSID. Thus their results can be directly compared to the results obtained here for *visual* modality. The main differences in protocol between the two experiments is that, during the presentation phase, subjects were static and at 1.22m from the screen in [Klein et al. 2009], whereas they were allowed to move in the exploration area at 2.3m or 3.3m from the screen in the present experiment; see Figure 3(a). Results for the visual modality and the results of Klein et al. [2009] obtained in the real and virtual worlds are plotted in Figure 8(a). From this figure, it can be seen that for Position 1, the results of the present experiment closely follow the real world results of Klein et al. [2009] for $d_r < 3$m and tend toward those for the virtual world when $d_r > 3$m. For Position 2, the results of the present experiment closely follow real world results of Klein et al. [2009] up to $d_r = 4.5$m, before decreasing slightly. Thus we can conclude that moving during the presentation phase may have provided subjects with a better visual distance perception of close distances. We can also see that in the virtual world, results from Klein et al. [2009], the anchoring distance D_a (estimated here as the distance for which $d_r = d_p$) is around 1.4m, and is close to 1.22m, the distance between the subjects and the screen. This also constitutes experimental evidence arguing in favor of the anchor hypothesis proposed in Section 4.2.



(a) Visual modality (b) Audio modality

Fig. 8. Comparison of results obtained for *audio* and *visual* modalities with some results from the literature. For the *visual* modality, [Klein et al. 2009] have studied, using triangulation, egocentric distance perception in an open grassy field, in the real world and in a virtual world rendered by LSID. For the *audio* modality, Zahorik et al. [2005] proposed a psychophysical curve relating perceived distance to real distance from a review among 84 experiments. The shadowed zone denotes the standard deviation associated with the results from the 84 experiments.

5.2 Perceived Distance in *Audio* VR-Systems Based on Holophonic Sound Rendering

As discussed in Section 4.2, a compressive curve in the form of $d_p = k(d_r)^a$ has been shown to be a good model for the psychophysical function that relates estimates of perceived distance to physical source distance for the audio modality Zahorik et al. [2005]. A review among 84 experiments is presented in Zahorik et al. [2005] with a mean value of $a = 0.54$ obtained for the compression coefficient when fitting a compressive model to all the available data. It was also observed that experimental protocols [Loomis et al. 1998] (verbal report, perceptually directed action) and listening conditions [Speigle and Loomis 1993] (static or moving) have very little influence on the obtained values of a.

Results from Section 3.2 for the *audio* condition are compared to this compressive model in Figure 8(b). By fitting such a model to the perceived audio distances collected for Position 1, values of $a = 0.33 \pm 0.03$ and $k = 1.72 \pm 0.09$ are found, with $R^2 = 98\%$ of the variance observed in the experimental data explained by the compressive model. The fit for Position 2 gives values of $a = 0.29 \pm 0.07$ and $k = 2.13 \pm 0.31$, with $R^2 = 84\%$. The model $d_p = k(d_r)^a$ thus fits very well with the experimental data for both starting positions. The perception of auditory distance seems to be slightly more compressed in the virtual world than predicted in the real world using the average compressive model. However, a more rigorous experimental protocol, which compares directly real world and virtual world distance perception using the same distances and reporting method (as in Kearney et al. [2012] for example), is needed to assess this point. It can nevertheless be concluded that WFS is able to synthesize sound-fields which are perceptually meaningful in terms of distance for moving subjects and static virtual sources placed in the action space, but with, apparently, slightly more compression than in the real world.

5.3 Utility of Dynamic Distance Cues

It is important to see that, as shown in Section 3.2, all of the subjects spontaneously walked during the exploration phase and that the exploration durations, t_{XP}, were different among the different

modalities, with $t_{XP}(A) > t_{XP}(AV) > t_{XP}(V)$. Thus subjects attempted subjects to gain information from the AV dynamic cues and seemed to proceed differently depending on the available modality. Moreover, they walked slightly more in the direction parallel to the virtual object than in the direction perpendicular to the virtual object during the presentation phase. This highlights the importance of dynamic cues in virtual audio-visual environments and provides some information into how perceptual cues may be weighted.

5.4 Feeling of Presence and Visual Distance Underestimation

The major problem of the presence feeling is that it is subject-dependent. For some of the subjects, presence was higher in the AV condition than in the V condition. For others, the opposite was true. This is potentially a consequence of the chosen audio stimulus (low-pass filtered white noise, see Section 2.3) which has been reported as unpleasant by some subjects, and thus may have decreased their feeling of presence. This may also explain why no significant differences were found for the presence-score between the V condition and the AV conditions (see Figure 6).

Furthermore, it has been suggested that because AV VR-systems provide a higher degree of *presence* than visual-only VR-systems, they potentially lead to less visual distance underestimation [Interrante et al. 2008]. The correlation between the feeling of *presence* and visual distance underestimation is studied here. For each subject, the presence variation Δ_P, induced by the addition of the spatialized audio stimuli is calculated as the difference between the *presence*-score of that subject in the AV condition and the *presence*-score of that subject in the V condition (see Section 3.4). Similarly, the *linear visual underestimation factor variation* Δ_α, induced by the addition of the spatialized audio stimuli is calculated as the difference between the linear underestimation factor of that subject in the AV condition and the linear underestimation factor of that subject in the V condition. The *linear underestimation factor* is computed as the linear slope of the psychophysical curve relating d_p and d_r. As a result, Δ_α and Δ_P are not found to be correlated (correlation coefficient of $\Gamma = -0.12$ and $p < 0.45$). Hence this does not allow us to conclude that a higher degree of presence leads to less visual distance underestimation, and illustrates the limited efficiency of post-session questionnaires as a tool to measure fine variations of presence.

6. CONCLUSION

In this article, a study of audio, visual, and audio-visual egocentric distance perception by moving subjects in *virtual* environments is presented. Audio-visual rendering was provided by tracked passive visual stereoscopy and acoustic wave field synthesis (WFS). For each rendering condition, the estimation of perceived distances was based on a perceptually directed action using the method of indirect blind-walking. Distances perceived in the virtual environment were systematically overestimated for rendered distances closer than the audio-visual rendering system and underestimated for farther distances. Interestingly, subjects perceived each virtual object at a modality-independent distance when using the audio modality, the visual modality, or the combination of both. Regarding the *audio* modality, WFS was able to synthesize perceptually meaningful sound fields. Dynamic audio-visual cues are used by subjects when estimating the distance of virtual objects. Moving may have provided subjects with a better visual distance perception of close distances than if they were static. No correlation between the feeling of presence and visual distance underestimation has been found. Finally, to explain the observed perceptual distance compression, it is proposed that, due to conflicting distance cues, the audio-visual rendering system physically anchors the virtual world to the real world by attracting virtual objects to it.

ACKNOWLEDGMENTS

The authors thank all the volunteers who took part in the experiment and *sonic emotion* for providing the *Wave 1* WFS rendering engine. Thomas Chartier and Philippe Cuvillier, now former students from the École Polytechnique (France), are also thanked for their help in the design and preliminary tests, and Marc Fuzellier for his useful help during the second phase of the experiment. Special thanks are given to Matthieu Courgeon for time spent on the visual rendering. Finally, the authors would like to thank Antonio Trujillo-Ortiz for providing a reliable Matlab version of the repeated-measures two-way analysis of variance test.

REFERENCES

ALEXANDROVA, I. V., TENEVA, P. T., DE LA ROSA, S., KLOOS, U., BULTHOFF, H. H., AND MOHLER, B. J. 2010. Egocentric distance judgments in a large screen display immersive virtual environment. In *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization* (*APGV '10*). ACM, New York, 57–60.

ANDRE, J. AND ROGERS, S. 2006. Using verbal and blind-walking distance estimates to investigate the two visual systems hypothesis. *Percept. Psychophys. 68*, 3, 353–361.

ARMBRUSTER, C., WOLTER, M., KUHLEN, T., SPIJKERS, W., AND FIMM, B. 2008. Depth perception in virtual reality: Distance estimations in peri- and extrapersonal space. *Cyberpsychol. Behav.11*, 1, 9–15.

ASHMEAD, D. H., DAVIS, D. L., AND NORTHINGTON, A. 1995. Contribution of listeners approaching motion to auditory distance perception. *J. Experiment. Psychol.-Hum. Percept. Perform. 21*, 2, 239–256.

BEALL, A. C., LOOMIS, J. M., AND PHILBECK, J. W. 1994. Absolute motion parallax weakly determines visual scale. *Investigative Ophthalmology Visual Sci. 35*, 4, 2111–2111.

BERKHOUT, A. J., DE VRIES, D., AND VOGEL, P. 1993. Acoustic control by wave field synthesis. *J. Acoustical Soc. Amer. 93*, 5, 2764–2778.

BLAUERT, J. 1999. *Spatial Hearing, The Psychophysics of Human Sound Localization.* MIT Press, Cambridge, MA.

BORMANN, K. 2005. Presence and the utility of audio spatialization. *Presence-Teleoper. Virtual Environ. 14*, 3, 278–297.

BRONKHORST, A. W. AND HOUTGAST, T. 1999. Auditory distance perception in rooms. *Nature 397*, 6719, 517–520.

BROOKS, F. P. 1999. What's real about virtual reality? *IEEE Comput. Graph. Appl. 19*, 6, 16–27.

CORTEEL, E. 2004. Caractérisation et extensions de la wave field synthesis en conditions réelles. Ph.D. dissertation, Université de Paris 6.

CORTEEL, E., NGUYEN, K. V.,WARUSFEL, O., CAULKINS, T., AND PELLEGRINI, R. 2007. Objective and subjective comparison of electrodynamic and MAP loudspeakers for wave field synthesis. In *Proceedings of the 30th International Conference of the Audio Engineering Society.*

CREEM-REGEHR, S. H., WILLEMSEN, P., GOOCH, A. A., AND THOMPSON, W. B. 2005. The influence of restricted viewing conditions on egocentric distance perception: Implications for real and virtual indoor environments. *Perception 34*, 2, 191–204.

CÔTÉ, N., KOEHL, V., PAQUIER, M., AND DEVILLERS, F. 2011. Interaction between auditory and visual distance cues in virtual reality applications. In *Proceedings of the Forum Acusticum*.

CUTTING, J. E. 1997. How the eye measures reality and virtual reality. *Behav. Res. Meth. Instr. Comput. 29*, 1, 27–36.

FARIA, R. R. A., ZUFFO, M. K., AND ZUFFO, J. A. 2005. Improving spatial perception through sound field simulation in VR. In *Proceedings of the IEEE International Conference on Virtual Environments, Human-Computer Interfaces and Measurement Systems*. IEEE, Los Alamitos, CA, 103–108.

FUKUSIMA, S. S., LOOMIS, J. M., AND DASILVA, J. A. 1997. Visual perception of egocentric distance as assessed by triangulation. *J. Exper. Psychol. Hum. Percept. Perform. 23*, 1, 86–100.

GERZON, M. A. 1985. Ambisonics in multichannel broadcasting and video. *J. Audio Eng. Soc. 33*, 11, 859–871.

GRECHKIN, T. Y., NGUYEN, T. D., PLUMERT, J. M., CREMER, J. F., AND KEARNEY, J. K. 2010. How does presentation method and measurement protocol affect distance estimation in real and virtual environments? *ACM Trans. Appl. Percept. 7*, 4, 26.

HOFFMAN, D. M., GIRSHICK, A. R., AKELEY, K., AND BANKS, M. S. 2008. Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. *J. Vision 8*, 3, 33.

HOWARTH, P. A. 2011. Potential hazards of viewing 3-D stereoscopic television, cinema and computer games: a review. *Ophthalmic Physiol. Optics 31*, 2, 111–122.

INTERRANTE, V., RIES, B., LINDQUIST, J., KAEDING, M., AND ANDERSON, L. 2008. Elucidating factors that can facilitate veridical spatial perception in immersive virtual environments. *Presence-Teleoper. Virtual Environ.17*, 2, 176–198.

KEARNEY, G., GORZEL, M., RICE, H., AND BOLAND, F. 2012. Distance perception in interactive virtual acoustic environments using first and higher order ambisonic sound fields. *Acta Acustica* with *Acustica 98*, 1, 61–71.

KLEIN, E., SWAN, J. E., SCHMIDT, G. S., LIVINGSTON, M. A., AND STAADT, O. G. 2009. Measurement protocols for medium-field distance perception in large-screen immersive displays. In *Proceedings of the IEEE Conference on Virtual Reality*. IEEE, Los Alamitos, CA, 107–113.

KOMIYAMA, S., MORITA, A., KUROZUMI, K., AND NAKABAYASHI, K. 1991. Distance control system for a sound image. In *Proceedings of the 9th Audio Engineering Society International Conference on Television Sound Today and Tomorrow.*

LOOMIS, J. M., KLATZKY, R. L., PHILBECK, J. W., AND GOLLEDGE, R. G. 1998. Assessing auditory distance perception using perceptually directed action. *Percept. Psychophys. 60*, 6, 966–980.

LOOMIS, J. M. AND KNAPP, J. M. 2003. *Virtual and Adaptive Environments: Applications, Implications, and Human Performance t5rIssues*. Lawrence Erlbaum.

NACERI, A., CHELLALI, R., DIONNET, F., AND TOMA, S. 2009. Depth perception within virtual environments: A comparative study between wide screen stereoscopic displays and head mounted devices. In *Computation World. Future Computing, Service Computation, Cognitive, Adaptive, Content, Patterns*, 460–466.

NAWROT, M. AND STROYAN, K. 2009. The motion/pursuit law for visual depth perception from motion parallax. *Vision Res. 49*, 15, 1969–1978.

PLUMERT, J. M., KEARNEY, J. K., CREMER, J. F., AND RECKER, K. 2005. Distance perception in real and virtual environments. *ACM Trans. Appl. Percept. 2*, 216–233.

PORSCHMANN, C. AND STORIG, C. 2009. Investigations into the velocity and distance perception of moving sound sources. *Acta Acustica* with *Acustica 95, 4*, 696–706.

RÉBILLAT, M., CORTEEL, E., AND KATZ, B. F. 2008. SMART-I2: Spatial multi-user audio-visual real-time interactive interface. In *Proceedings of the 125th Convention of the Audio Engineering Society.*

RÉBILLAT, M., KATZ, B. F., AND CORTEEL, E. 2009. SMART-I2: Spatial multi-user audio-visual real-time interactive interface. A broadcast application context. In *Proceedings of the IEEE 3D-TV Conference*. IEEE, Los Alamitos, CA.

RUSSELL, M. K. AND SCHNEIDER, A. L. 2006. Sound source perception in a two-dimensional setting: Comparison of action and nonaction-based response tasks. *Ecological Psychol. 18*, 3, 223–237.

RYU, J., HASHIMOTO, N., AND SATO, M. 2005. Influence of resolution degradation on distance estimation in virtual space displaying static and dynamic image. In *Proceedings of the International Conference on Cyberworlds*. 43–50.

SANSON, J., CORTEEL, E., AND WARUSFEL, O. 2008. Objective and subjective analysis of localisation accuracy in wave field synthesis. In *Proceedings of the 124th Convention of the Audio Engineering Society*.

SLATER, M. AND WILBUR, S. 1997. A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence-Teleoper. Virtual Environ. 6, 6*, 603–616.

SPEIGLE, J. M. AND LOOMIS, J. M. 1993. Auditory distance perception by translating observers. In *Proceedings of IEEE Symposium on Research Frontiers in Virtual Reality.* IEEE, Los Alamitos, CA, 25–26.

SPRINGER, J. R., SLADECZEK, C., SCHEFFLER, M., HOCHSTRATE, J., MELCHIOR, F., AND FROHLICH, B. 2006. Combining wave field synthesis and multi-viewer stereo displays. In *Proceedings of the IEEE Virtual Reality Conference*. IEEE, Los Alamitos, CA, 237–+.

WATT, S. J., AKELEY, K., ERNST, M. O., AND BANKS, M. S. 2005. Focus cues affect perceived depth. *J. Vision 5*, 10, 834–862.

WIEST, W. M. AND BELL, B. 1985. Stevens exponent for psychophysical scaling of perceived, remembered, and inferred distance. *Psychol. Bull. 98*, 3, 457–470.

ZAHORIK, P., BRUNGART, D. S., AND BRONKHORST, A. W. 2005. Auditory distance perception in humans: A summary of past and present research. *Acta Acustica* with *Acustica 91*, 3, 409–420.

Received August 2011; revised June 2012; accepted July 2012