Depth Perception Within Virtual Environments: Comparison Between two Display Technologies

Abdeldjallil Naceri*†, Ryad Chellali†, Fabien Dionnet† and Simone Toma*†
*Dipartimento di Informatica Sistemistica e Telematica, Università degli Studi di Genova, Via Balbi 5, 16126 Genoa, Italy
†Telerobotics and Applications dept, Istituto Italiano di Tecnologia, via Morego 30, 16163 Genoa, Italy
{abdeldjallil.naceri, ryad.chellali, fabien.dionnet, simone.toma}@iit.it

***Abstract*—Depth perception is one of the key issues in virtual reality. Many questions within this area are still under investigation including the egocentric distance misestimation. In this paper we describe an experiment confirming distance underestimation from another point of view. The approach we developed is based on a very simple task: subjects had to compare relative depths of two virtual objects. The experiment compared performance using head mounted display and stereoscopic widescreen display to evaluate which visual cues subjects use to estimate depth of virtual objects. To minimize motoric effects, subjects were seated and their estimations were only verbal. Likewise, to avoid the well known effects of apparent size, namely the size-distance invariance, the experiment was also performed with conflict sequences: the presented objects had the same apparent sizes with different depths or the same depth but different physical sizes. The obtained results show significant differences between the two devices and confirm the *distance misestimation* phenomenon for head mounted display. Moreover, changing the background color or the shape of the presented objects also had an influence on subjects' performance.**

***Keywords*-Virtual reality, Human-machine interaction, Cognition, Depth perception, Head mounted display, Widescreen display.**

I. Introduction

Immersive viewing devices are key elements for virtual reality [1]. In the paper entitled *The Ultimate Display*, Sutherland [2] depicted a futuristic vision of synthetic worlds and the ways that humans experience virtual realities within these worlds. Nowadays, Sutherland's prophecies are widely spread. Virtual reality (VR), specifically virtual environments (VEs), are used in several areas to support sensitive and complex topics such as psychology, robotics, education, medical therapy and diagnosis, archeology, geography, neuroscience, etc. These research and application fields take advantage of VE capabilities. VE technologies offer flexibility and support innovations by allowing users to explore artificial environments with unconventional rules and interact with virtual objects.

Historically, this field started with computer graphics and 3D representations; VR was mainly concerned with visual channel. With the fast technological advances of the '80s, VR and VE systems began to address other senses. Haptic, tactile, vestibular and auditory sensory channels were introduced to mimic the human sensory system. Since that time the targeted goal has been being improved realism and increased immersion and sense of presence.

In other words, the goal is for users to experience realistic artificial worlds by providing plausible and coherent stimulation and allowing them to interact with these worlds and its objects. This goal has not yet achieved, for a variety of reasons, despite great advances in VE technologies.

Regarding the visual channel, to generating a realistic representation of the real world is very complex and requires a lot of computing resources. The image formation on human retina and its interpretation by human brain is not fully understood. A significant amount literature has been published on visual realism, however the contribution of the visual channel in immersion feeling is not well quantified and few studies have tackled this issue. Slater [3] showed that visual perception is affected by other sensory data streams including kinesthetic proprioception, motoric actions, sounds, etc. For example, a visual flow generated for walking motion increases immersion and presence within VE.

This visual flow phenomenon illustrates the unstable equilibrium of the perception process. This latter is based on a cross-modalities scheme where realism and consequently behavior are affected in an unpredictable way if any of the modalities is itself affected. Because VE systems that are capable of full, accurate simulation are not yet a reality, one can infer that creating such systems is not a trivial task. Consequently VEs rely on actions, interactions and feelings that are distorted, biased and/or incomplete causing malfunctions (fatigue and stress) and biases (physical misestimations). For visual realism the same observation is true: it depends on parameters which are not well identified nor well understood (for a review about the visual cortex and the binocular depth perception see [4]) and any defect or distortion of the visual features can have unexpected effects.

For example, Zago and colleagues [5] tested the validity of the internal model of gravity. They simulated a virtual falling ball displayed on a desktop computer display and a stereoscopic widescreen display (SWD), and subjects adapted to the desktop computer display but not to the SWD. Moreover, in another work to check the same hypothesis, Senot and colleagues [6] investigated the relative role of visual and non-visual information on motor-response timing in interception tasks. Subjects were asked to hit a virtual ball with a virtual racket using a keyboard while the scene was displayed through a head mounted display (HMD). Authors reported that the task was difficult and recorded a success rate lower than 50%. Therefore, these experiments raise the following two questions:

- Why do subjects adapt to desktop display and not to the SWD?
- Why was the task more difficult when using the HMD?

These questions suggest two additional questions about the nature of the obtained results in the previous experiments.

- Does the VR setup introduce a bias?
- Does the VR system distort data?

These questions are generic ones and the previous example is a showcase. Indeed, technological factors modify the behavior of users and consequently affect the understanding of a phenomenon (gravity internal model) not fundamentally concerned with perceptual schemes. More generally, display technologies and interfaces may introduce distortions and biases which are hard to identify and manage. Both users and developers of VR technologies must be less naive and be better informed about current VR limits.

In our work, we aimed to study the effects that visual display technologies can have on a very simple perceptual task. Namely, we wanted to compare SWD and HMD technologies in order to determine how they affect users' behavior.

This research is a part of a lager project to build a VR-based telerobotics system. For these systems, perceiving accurately and correctly remote environment is crucial. A fortiori, the visual perception must provide an exact replica of the remote reality with all corresponding and needed visual cues to ensure natural and coherent sensory motor-based tasks like grasping, reaching or obstacles avoidance and navigation. The replica goal is a theoretical one and we know that it is not reachable with the current technology. Nevertheless, by knowing the limitations of both SWD and HMD, we will be able to prevent inappropriate uses of these display technologies.

The paper is organized as follows. In Section II we give some entry points to research works concerned with visual perception and display technologies. In Section III we describe the designed protocol, the environment conditions, and detail the hardware and software we used. We present in Sections IV and V the obtained results and statistical analysis. In Section VI we discuss our findings concerning depth misestimation issues within VEs and way to exploit them.

II. Background and related work

Even after a decade of research, *visual space perception* remains an active topic with a lot of ongoing work and research efforts. This indicates the importance of the topic and the challenging problems it raises. In this section, we review fundamental notions about human depth perception through vision. We describe the display technologies used in the present work. Additionally, we introduce some techniques used to assess the effectiveness of perceiving distances in VEs. Finally, we discuss the issue of depth misestimation in VR.

A. Visual cues and depth perception

Depth perception results from the integration of several visual cues. Research in this area identify the visual features, the individual visual cues and the ways these cues are processed by the human visual system and combined within human brain to create a vivid three-dimensional perceptual world [7].

From a functional point of view, researchers are still working to understand some fundamental issues including the mapping between real space and its mental representation (visually perceived space), the connection between visual space and motor actions, and the contribution of each visual cue in the building process of the visually perceived space [8][9][10].

From a more basic geometrical point of view, visual cues are 2D entities obtained by a central projection of 3D features. With one eye (one projection), one has monocular cues, including perspective, motion parallax, optic flow, occlusions, lighting, shading, accommodation, etc. When two projections are combined, humans perform an oculomotor convergence-accommodation process and use disparities to deduce the 3D representation of the observed scene.

Both monocular and binocular cues can be characterized by the following two subsets:

- Geometrically and graphically based cues which can be produced by any computer graphics framework such as OpenGL,
- Oculomotor based cues: accommodation and convergence.

As mentioned before, the fusion process is not well understood. Nevertheless, some findings give partial explanations regarding the integration process. Oculomotor and stereopsis cues are known to be inter-dependent (disparity and stereo-acuity), however this dependency is variable. One cue can dominate the other two functions of the distance between observer and the observed object [11]. Depth acuity has been shown to be high within the peripersonal space, defined as the space that can be reached by our hands, typically 1m or less, and low for the extrapersonal space [11][12]. Others [13] have defined these spaces differently using 5m as the defining limit, based on the fact that oculomotor factors become negligible beyond this point. In our experiments, we defined the peripersonal space as less than 1m, which is within both definitions.

B. SWD and HMD visual display technologies

There are two main types of implementations of an immersive virtual display. The first consists of projection widescreens, loud speakers, shutter glasses and/or polarized 3D glasses to provide a stereoscopic stimulation, allowing the user to observe the VE not as a projection on 2D surface, but as 3D solid structures within the experimental site. The second, and the most widely used involves the use of the HMD. With both HMDs or SWDs, binocular visual imagery provides convergence and retinal disparity cues that contribute to the perception of egocentric and allocentric distances in depth.

C. Methods used to assess depth perception cues

Depth and distances are euclidean quantities expressed in meters. Measuring effectiveness of depth cues is equivalent to establishing a relationship between a visual cue (or a set of visual cues) and an euclidean distance. Unfortunately, this input-output scheme is not so valid because humans are weak "instruments" for measuring distances [14][15][16]. In fact, depth perception is considered to be a process leading to an invisible cognitive state [7] and thus inaccessible directly. To overpass this limitation and to take advantage of the human ability to compare, researchers use allocentric or egocentric distances comparison to assess depth cues [8][9][10][17]. In the first case, the observer compares the respective distances between objects and a reference point. In the second case, the observer compares distances between objects relatively to himself. Using this strategy, one can indirectly access the visual cues involved in depth perception process.

In general, there are three methods used to judge egocentric distances: *verbal answers*, *perceptual matching*, and *action-based tasks* [7][10]. In the *verbal answer* case, subject verbally report the comparison between two perceived locations. The task is a forced-choice test and subjects must give answers such as "near" or "far"; "same depths" or "different depths", etc.

For *perceptual matching* tasks, the subjects directly act on the objects position through interfaces such as a mouse, keyboard or a joystick in order to move the target to a position that matches a reference object [15][18][19].

Finally, for *action-based tasks*, observers are asked to indicate the perceived distance while performing a physical action [7][20][19][21][22]. *Blind walking tasks* is the most common action-based task [7][10][19][20] in which the observer perceives an object at certain distance, then walks with covered eyes until they reach the perceived egocentric distance. This method's limitation is that large errors are made when subjects attempt to estimate distances beyond 10m [23][24]. Thomson [23] attributed these larger errors to a decay of the spatial memory for target position while walking to the target.

D. Depth misestimation in VE

The visual realism components is highly related to absolute distance judgment or depth perception. It is one of the major issues in VR because rapid understanding and accurate motor-based interactions (like grasping, reaching, intercepting or pointing) in virtual 3D depends on depth perception. Visual richness is positively influenced by monocular cues like dynamic shadows, textured objects, motion parallax, etc. and as visual richness improves so does motor-based interactions [25]. That said, a recurrent issue in both poor and rich VEs is distance misestimation [26]. Systematic underestimation of distances was seen when HMDs were used compared to the same estimation in the real world [20][27]. Similarly, studies on distance perception using the HMD [16][28][29] have found significant underestimation of egocentric distances.

Depth misestimation seems to be a mystery that many authors have proposed explanations for the literature includes several hypotheses to explain the phenomena of depth misestimation in general, and the egocentric underestimation using HMD in particular. In [20], underestimation phenomena was not attributed to the limited field of view of a user while using the HMD. Willemsen and colleagues [30] argue that mechanical properties play a role in the underestimation phenomena. Other explanations include a lack of graphical based-realism does not causes that phenomena [29] or mismatches between the viewed world and the experimental site (e.g., subjects are aware that the viewed scene does not correspond to the place where the experiment is performed) cause the phenomena of distance misestimation [31]. Similarly, some researchers have shown that other factors like visual cues (such as accommodation and convergence) and situations (visually directed actions) may affect distance or depth estimations [22]. Other studies revealed that the misestimation of depth also exists when SWDs are used [32].

In summary, the following hypothesis have been proposed in the literature to explain the misestimation of egocentric distances:

- the reduction of the field of view;
- the weight of the HMD;
- the difference between the viewed world and the experimental place;
- but that monocular versus stereo viewing does not cause it;
- this effect exist in VR when is displayed in SWD.

The egocentric distances perception in VR immersive displays is not fully understood. Moreover, the identification of sources leading to then distance misestimation effects in VE remains an open question, and no specific study has been conducted to directly compare depth perception performances between the two main display technologies.

The first purpose of this our to compare and to evaluate human depth perception in VEs using HMD and SWD. The second point is to confirm whether the depth misestimation in VEs is a reality. We consider this work as a necessary step before attempting to understand and determine the factors underlying this phenomena.

III. Experimental Protocol and Setup

The aim of our experiment is twofolds:

- Determine the effect of display technologies on depth perception
- Study the effects of some visual cues on depth perception

To this end we ran the same experiment using both SWD and HMD. In addition, we varied the background and the shape of the used objects.

In the following we detail our protocol including the used stimuli, hardware and participants.

A. Protocol, stimuli and expectations

The designed experiment was inspired by the works described in [33] and [34] who respectively assessed depth perception in the real world and correlation between size and distance estimation. For both tasks, subjects verbally answered forced choice questions concerning depths of two objects.

In our experiment, we constructed a set of four comparisons types. Each comparison is composed of two stimuli that differ in size and/or in position as reported on Figure 1:

- *Comparison I:* both objects are the same size, but at different depths (apparent sizes are different),
- *Comparison II:* both objects are the same size, and at the same depth (apparent sizes are equal),
- *Comparison III:* both objects are different sizes and at the same depth (apparent sizes are different),
- *Comparison IV:* both objects are different sizes, but at different depths (such that apparent sizes are equal),

This set of pair-comparisons takes into account the findings of [33] and [34]. In the first paper it was found that an object which was consistently overestimated in size was consistently overestimated in distance (size-distance paradox) which strongly support the hypothesis about the size-distance covariance in depth estimation. In addition, Berryhill and colleagues [34] recently reported a high degree of accuracy by healthy subjects in judging either the size or the distance of real objects even though both perceptive variables were not covarying and were the only cues available.

Naively, the comparisons I and II might be sufficient for our experiment, however, subjects could potentially rely on apparent size which could make a difficulties in understanding the results: the ability to discriminate between those

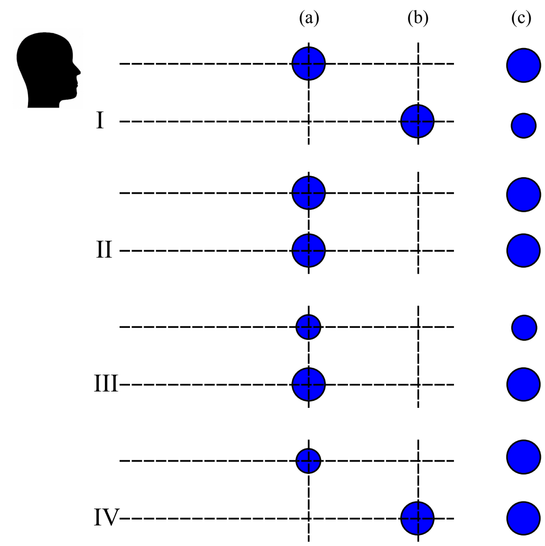

Figure 1: Four comparisons of two objects, (a) and (b) represent the virtual objects presented to subjects and (c) corresponding retinal size

perceiving depth correctly and those relying on apparent size to give a correct answer. Therefore, two other comparisons were added, III and IV. These comparisons amplify the size/distance conflict leading a higher ambiguity for subjects relying on apparent size. To limit the visual cues subjects could apply during the task, we used an impoverished environment (uniform background). Likewise, lighting and shading conditions remained constant. We also used a spherical object to force subjects to rely only on stereopsis and convergence. Finally, the stimuli were displayed directly in front of the subjects to avoid any parasitical motor activity (head movement).

The second part of the experiment dealt with the effect of some visual cues on depth estimation. To that purpose we varied both background and object shapes. For the backgrounds, we used two uniform colors, black and white (BB and WB). These conditions were strictly followed and applied for both SWD and HMD.

B. Task and conditions

1) Objects: The presented objects were two spheres of diameters 7.5cm and 10cm and two cubes of edge lengths 7.5cm and 10cm as shown on Figure 2.

2) Scene organization: The virtual objects were displayed exactly 60cm and 80cm in front of subjects' eyes. These distances were chosen as a trade-off imposed by two opposite constraints. First that the virtual objects be within the subject's peripersonal space because the present work is not only focused on depth perception but also is the first step in more complex study dealing with motor action during reaching tasks. Nevertheless, virtual objects cannot

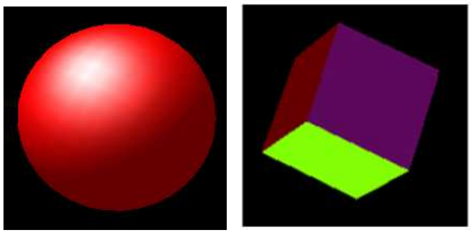

Figure 2: Displayed virtual sphere and cube

be displayed too close to the subject without creating visual discomfort and stress.

3) Timing: Each object was individually presented to avoid the apparent motion cues that might alter subjects' perceptual judgments. The first for 3s, following by a 2s pause without virtual objects before the second object was presented for a total of 8s. These 8s sequences were repeated 10 times (i.e., 80 trials per conditions described in Figure 1). Both the order of virtual objects presented (sphere/cube) and use of HMD/SWD was randomized. Likewise, the order of the conditions was randomly assigned to avoid bias due to learning process. The each experimental condition took approximately 20 min with a brief pause at the half way point.

4) Question: For all conditions, two alternative forced choice (2 AFC) was used. Subjects were asked to verbally answer the following question after each comparison of the two displayed virtual objects: "*Are the two objects you just saw located at the same position?*" They had to answer "*yes*" or "*no*" before the next comparison started.

C. Hardware

The experiment was carried out in TEle Robotics and Applications department (TERA) VR room.

The SWD system includes two videoprojectors, two polarization filters and one widescreen. The videoprojector model is the evo22sx+ from Projectiondesign with luminance of 3000 ANSI lumens. They are placed side-by-side on the ceiling of the VR room and are equipped with two orthogonal circular polarization filters. Both beams are oriented to the widescreen and the projection distortion is corrected until getting two perfectly overlayed rectangular images of dimensions 1.805×1.535 m^2. The screen is polarized such that wearing light passive polarized glasses allow subjects' eyes to see two different images (Figure 3).

The HMD is a binocular Cybermind Visette45. The optical system designed for subjects' eyes to accommodate on a plane located 2m in front for a diagonal field of view of 45°. It weighs approximately 750 g (Figure 4).

In both cases, left and right images of resolution 1280×1024 were generated using the OpenGL library on a PC Dell bi-Xeon 3GHz with 8Gb of RAM running GNU/Linux and displayed simultaneously at 50Hz. The subjects were

Figure 3: A subject wearing polarization filter glasses and head tracking device performing the task in SWD presentation

Figure 4: A subject performing the task in HMD presentation

seated approximately 1.8m in front of the projection screen (see Figure 3) in order to allow them to accommodate approximately at the same distance for both display device. The subjects were asked to maintain a fixed position and to move as little as possible during the trials of the experiment.

D. Participants

Eight observers took part in the study, five males and three females of age 42 ± 10 years old. All were naive to

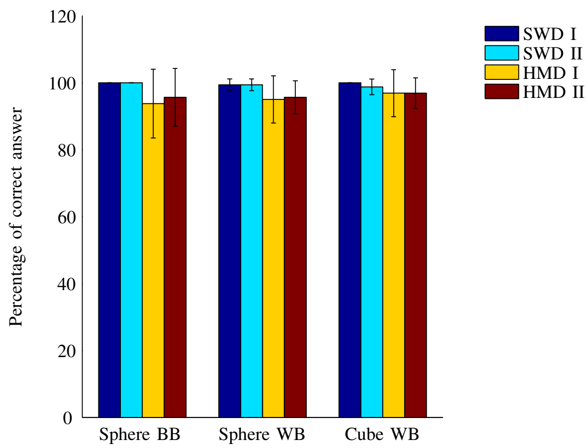

Figure 5: Average rate of correct answers in SWD and HMD conditions considering the comparisons I & II

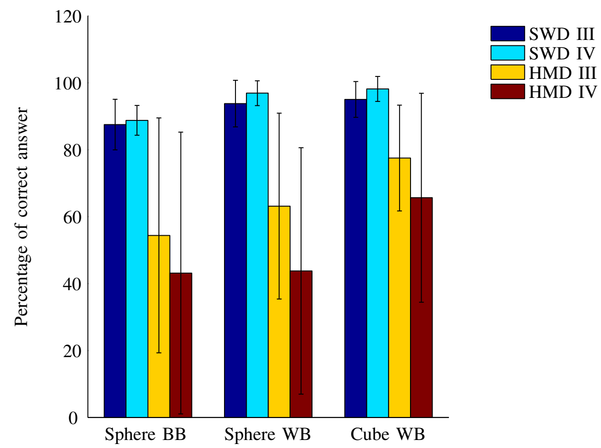

Figure 6: Average rate of correct answers in SWD and HMD conditions considering the comparisons III & IV

the purpose of the experiment and had normal or corrected to normal visual acuity.

IV. Preliminary analysis of the results

In this section we present the obtained results with respect to the two main questions:

- SWD vs HMD comparison
- Effects of specific visual cues on depth perception

In the first part we present subjects' success rates in different conditions using SWD and HMD. To that end, we start with a global analysis based on mean and standard deviation. To refine, an ANOVA was completed to determine whether there were any significant effects regarding presentations, conditions and/or subjects' personal performances.

A. Global analysis

1) SWD vs HMD: To compare depth perception performances in both SWD and HMD conditions, we calculated means and standard deviations for the four comparisons (see Figures 5 and 6). For the SWD presentation, subjects showed good performances in all comparisons. However, we noticed larger variability in the comparisons III and IV. In the HMD presentation, the subjects' rates of success were greater than 85% in the comparisons I and II. For comparisons III and IV, performances were lower than 75%. In addition standard deviations for these two comparisons were larger than any other comparison.

2) Effects of object's shape and background: For the SWD, we noticed that there was no specific effect due to the background nor the object shape for comparisons I and II. For comparisons III and IV, the success rate gradually increased (black background $<$ white background $<$ cube). For the HMD, the same tendency although with a greater slope. In other words, the enhancement effect was more pronounced with HMD.

B. ANOVA analysis

ANOVA tests revealed that the order of presentation within the four comparisons had absolutely no effect ($p > 0.05$). Thus, in all analysis we used four comparisons instead of the eight sub-comparisons of the designed protocol and effectively performed during the experiment.

1) SWD vs HMD: When we compared the results obtained for the HMD compared with those of SWD, ANOVA tests revealed significant differences, particularly for the comparisons III and IV (see Table I).

2) Effects of object's shape and background: A one-way ANOVA test for different conditions reveals no significant difference between the three comparisons I, II, and III for the three conditions in the SWD presentation (I: $F[2, 21] = 1$ and $p > 0.39$, II: $F[2, 21] = 1.11$ and $p > 0.35$, III: $F[2, 21] = 2.89$ and $p > 0.08$). As for comparison IV, the ANOVA test revealed a significant difference between the "sphere-black background" condition compared to the "sphere-white background" and "cube-white background" conditions ($F[1, 14] = 15.77$ and $p < 0.0014$, $F[1, 14] = 21$ and $p < 0.0001$).

3) Variability between subjects: The ANOVA test revealed a significant difference between subject performances only for the HMD condition in comparisons III and IV (III: $F[7, 23] = 7.14$ and $p < 0.00001$, IV: $F[7, 23] = 16.29$ and $p < 0.000001$). In fact, there was one subject who had good performances in both presentations SWD and HMD. Furthermore, two subjects had an average rate of success more than 50% in the HMD condition for the comparisons III and IV but less compared to their performances in the SWD. For the remaining subjects, had a noteworthy decrease in the accuracy rates for the HMD condition in all stimulus comparisons.

ANOVA test revealed no significant difference between comparisons III and IV for the HMD presentation. One can

Table I: ANOVA tests result for the four comparisons across three conditions including F-values and p-values, comparison between SWD and HMD.

ANOVA test	Comp.	Conditions		
		Sphere BB	Sphere WB	Cube WB
p-values	I	$p > 0.11$	$p > 0.11$	$p > 0.23$
	II	$p > 0.11$	$p > 0.06$	$p > 0.32$
	III	$p < 0.02$	$p < 0.01$	$p < 0.01$
	IV	$p < 0.01$	$p < 0.001$	$p < 0.01$
F-value: $F(1,14)$	I	2.97	2.88	1.58
	II	2.97	4.06	1.07
	III	6.81	9.16	8.79
	IV	9.30	16.49	8.55

see from the Figures 5 and 6 that the cube condition was better than the others two conditions for the HMD in the comparisons III and IV. This results were confirmed by using the d' method in the next section.

V. Results and Data Analysis within the SDT framework

In our first analysis revealed a large variability between subjects. ANOVA tests are not reliable enough in such situations. To overcome this limitation and to more thoroughly analyze the data, we choose the SDT framework to refine the analysis and to better understand the found results.

A. SDT description

Signal detection theory (SDT) is used to analyze experimental data with categorization tasks using ambiguous stimuli. Tanner and Swets [35] proposed a statistical decision theory and specific ideas about electronic signal-detecting devices to build a model that closely approximates of how people actually behave that in a such situations. The model was described in detail and named "*signal detection theory*" by Green and Swets [36].

We choose this framework to evaluate and analyze our data because the task is based on ambiguous stimuli. Moreover, the data in our experiments are binary answers,"*yes*" and "*no*". The SDT-based analysis can provide estimations of subjects capabilities in terms of discriminative behavior or sensitivity regarding the presented stimuli d' [36].

B. SDT in our experiment

The task in our experiment is to judge whether two stimuli in comparison are at the same depth ("*yes*" or "*no*" answers). Let us consider the comparisons as new stimuli or meta-stimuli (the presentation of two successive objects within the same comparison) to be analyzed with the SDT. Disambiguation between two meta-stimuli may be related to relative apparent sizes and depth displayed during each comparison. For instance, for comparison I and III, we have the same configuration regarding apparent sizes and different depths configuration (Figure 1). Likewise, we have the inverse situation in comparisons II and IV: the same configuration for apparent sizes and different depths. The disambiguation and the consequent indirect questions that we used in the SDT are the following:

- Given different apparent sizes, are the depths judged to be equal or not? (disambiguation between I and III)
- Given the equal apparent sizes, are the depths judged to be equal or not? (disambiguation between II and IV)

Correct estimations judgments in the previous two conditions means that subjects overcome the size-distance paradox and rely only on the actual depth perception they have. In the remainder of the section, the meta-stimuli word is replaced by stimuli. If a "*yes*" answer to a presented stimulus is a correct answer, it is called a hit (H); but if a "*yes*" answer to a stimulus is a mistake, it is called a false alarm (FA). If a "*no*" answer is the correct response, it is called a correct rejection; but if a "*no*" answer is incorrect, it is called a miss. The proportions of hits and false alarms reflect the effect of two underlying parameters. The first parameter reflects the separation between the comparison (e.g., I) and the ambiguous comparison (e.g., III) of the stimulus. The second parameter is the strategy of the participants. The expected SDT models are expected to quantify subjects' perceptive sensibility in detecting environmental changes. Specifically, the performances when using SWD versus HMD devices and the effects of changing the background or the object shape.

Individual d' values were extracted from the differences between the normalized percentage of correct hit answers and the normalized percentage of false alarm answers. The hit rate is simply the proportion of "same apparent size and different position" responses that occurs for comparison IV. The false alarm rate is the proportion of "same apparent size, same position" for comparison II.

C. SDT analysis

In SDT framework, the value of d' is given as follows

$$d' = z_{\mathrm{H}} - z_{\mathrm{FA}}, \tag{1}$$

where z_{H} and z_{FA} are respectively the normalized probabilities of hit and false alarm rates.

In our analysis, we computed the normal distribution for all subjects by using the bootstrap procedure in order to estimate an accurate mean and variation. The aim of this analysis is to clearly observe how the subjects distinguish the comparisons I/II from III/IV respectively. Indeed, we statistically reinforced the obtained data since our set is small and contain a variability in some conditions. The best approach is to apply a bootstap procedure to extract a idealized models and assumptions, as introduced by Efron [37], in order to estimate and approximate a realistic model with normalized Gaussian distributions. The obtained mean values and variances characterize subjects' decision making behavior.

D. SWD vs HMD

The variance of the normal distributions describes the standard deviation of the population obtained by bootstrapping. The accuracy of the subjects' performances can be explained by the different variances characterized the normal distributions (Figures 9, 10). For the obtained representations, we found subjects' performances with SWD are better than those obtained with HMD across all four comparisons. Indeed, the obtained normal distributions related to SWD presentation (Figures 9) had smaller variances were clearly separated compared to those obtained for the HMD (Figure 10).

This latter observation confirms the results given by the one-way ANOVA test. Moreover, SDT framework helped us to quantify the differences between these two presentations. Indeed, the d' mean value for the HMD presentation was approximately half compared to the SWD presentation as shown by Figure 7.

$$d'_{\text{HMD}} \approx 0.5 \times d'_{\text{SWD}} \tag{2}$$

This indicates that there was more confusion and subjects had difficulty distinguishing between stimulus comparisons using HMD.

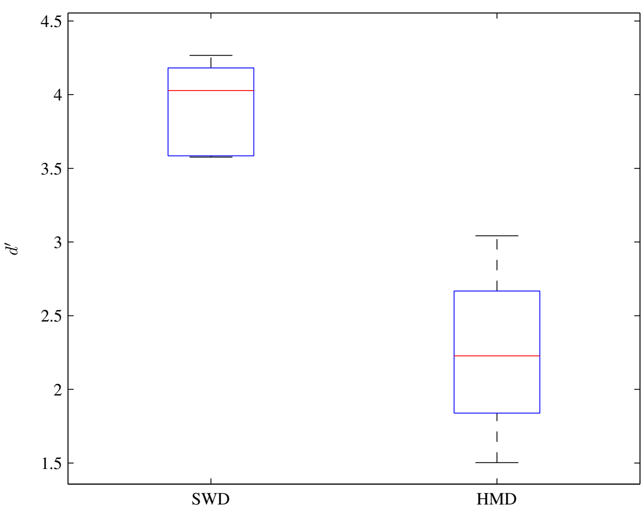

Figure 7: Subjects' d' average for SWD and HMD

E. Object shape effects on depth perception

It is obvious from Figures 9 that varying object shape had no effect on performances. Indeed, the mean values of the normal distributions are identical as shown in Figures 9b and 9c. On the contrary, for the HMD presentation there was clearly a difference between shapes since the two obtained normal distributions were more evidently separated with a cube compared to a sphere (see Figures 10b and 10c).

F. Comparison between pair-comparisons

From Figure 9, we observe that the variances of the normal distributions of comparisons I, II and IV using SWD are very low. This reflects the small variability of the population in these cases. On the contrary, we recorded a higher variability with the stimulus comparison III. Therefore, in the SWD presentation this comparison characterized by different apparent sizes but same position seems to be the most ambiguous to the subjects .

For the HMD presentation, we observed that the variances of the normal distributions are low for comparisons I and II (see Figure 10). As for the two other comparisons, the variability for comparison III is higher than those for comparisons I an II, but lesser than for comparison IV. Thus, the comparison IV, characterized by constant apparent size with different depths, is the most difficult to judge using a HMD.

VI. Discussion

A. Influence of the presentation: SWD vs HMD

The first goal of this work was to determine if observer performance in a depth perception task varies with respect to display technology. Two different statistical methods, ANOVA and d'-based method, revealed a difference. Indeed, while estimating relative depths is almost perfectly achieved for all the four presented comparisons with SWD, this is not true with HMD, especially when ambiguous situations are presented. Indeed, the d' mean value for the HMD presentation is approximately the half compared to the one of the SWD presentation (see Figure 7).

More specifically, for the SWD presentation the observers showed good performances even in the ambiguous comparisons. This contradicts with [32] which reported that subjects misestimate depth even by using widescreen. On the contrary, for the HMD presentation our finding is coherent with several studies showing that observers misestimate egocentric distances in VEs when they wear a HMD [16][28][29]; this remain true regardless of experimental methodology in these studies. Others have shown that distance misestimation in VEs by using HMD is not due to the limited field of view [20]. On the contrary the field of view restrictions of HMDs, in addition to other parameters that constraint head movements such as the weight, may have an influence on the accuracy of distance estimations [27].

After the experiment, all subjects were asked about the strategy they used to achieve the task. In SWD presentation, subjects answered that the task was more realistic and the virtual objects seemed to be reachable by their hands, in other words, the objects were within their peripersonal space. They did not state strategies using apparent size. On the contrary in the HMD condition, they were particularly less accurate in comparisons where apparent size does not reflect the correct depth. During post-experiment interviews, all

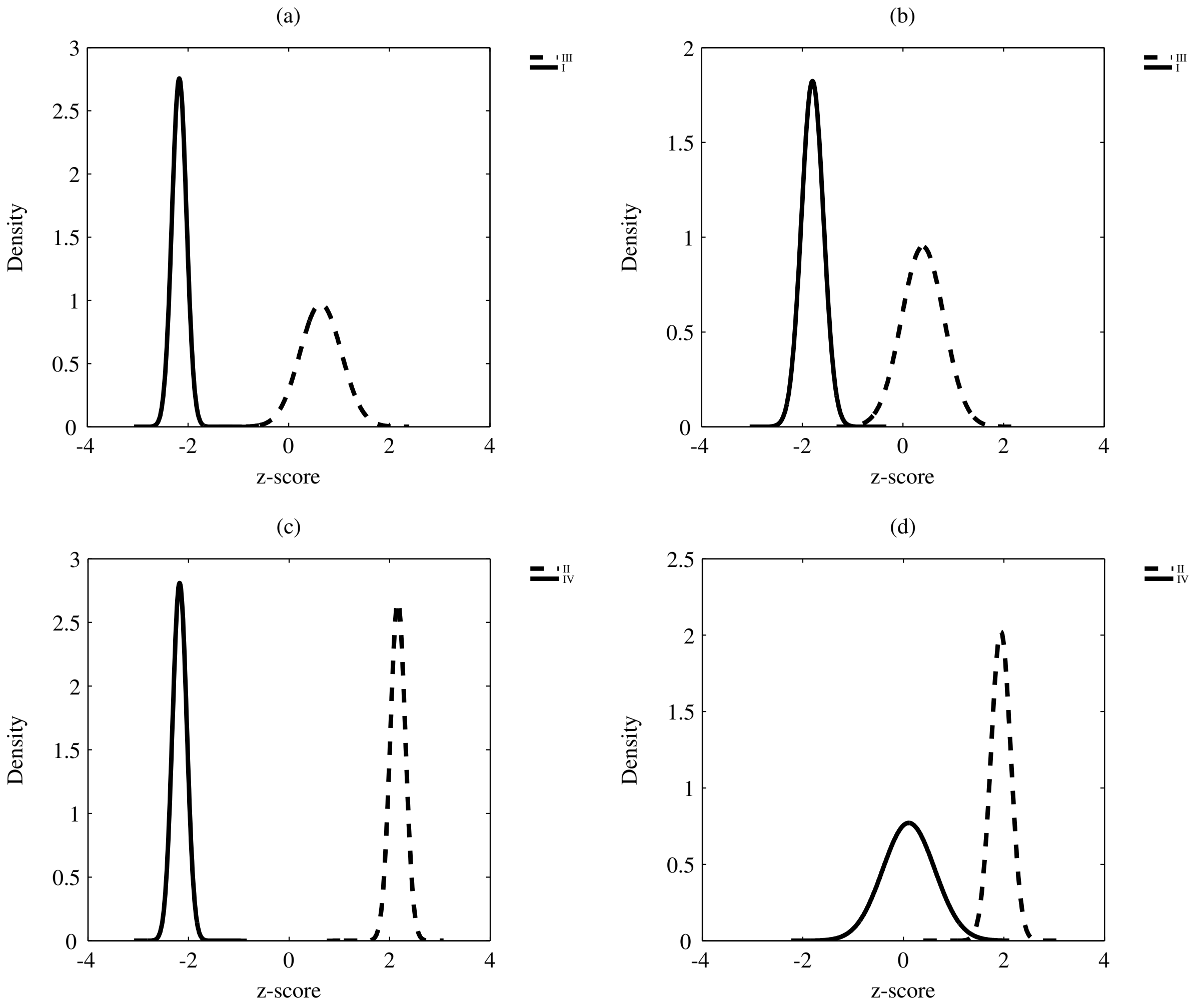

Figure 8: Normal distribution of subjects' z-score for the condition sphere with black background, (a,c) for SWD and (b,d) for HMD

subjects reported that the task was more difficult with the HMD than the SWD.

B. Influence of the background and shapes changes

In the data analysis part, object shape had an influence on subjects performances for the HMD but not for the SWD presentation. Specifically, for the HMD, performances were better with the cube than with the sphere. These differences were particularly visible with the d'-based analysis. This suggests that subjects relied in this case more on disparity cues because the cube contains edges, vertices and perspective that give more information than the sphere for depth evaluation. These latter cues helped subjects estimate depth and overcome ambiguous comparisons in some trials.

Therefore, one can presume that subjects missed some effective cues in the HMD presentation that were present and used with SWD. One hypothesis might be that wearing HMD isolate subjects' visual channel from all other stimuli, leading to misestimate of egocentric distances with regard to the body as reference. By definition egocentric frames of reference based on the body or specific parts of it, to define spatial positions [38][39].

As for background effects, we found that changing color influenced the observers' performance: (comparison IV, see ANOVA test). This finding contradicts reports that showed different VEs conditions did not impact observers' depth estimation [14].

C. Stimulus comparisons

The first analysis shows clearly that the performances of subjects were better for the SWD: the accuracy rates were higher and the variability lower. More explicitly, subjects more effectively resolved conflicts present in stimuli comparisons III and IV. Indeed, they did not rely on angular

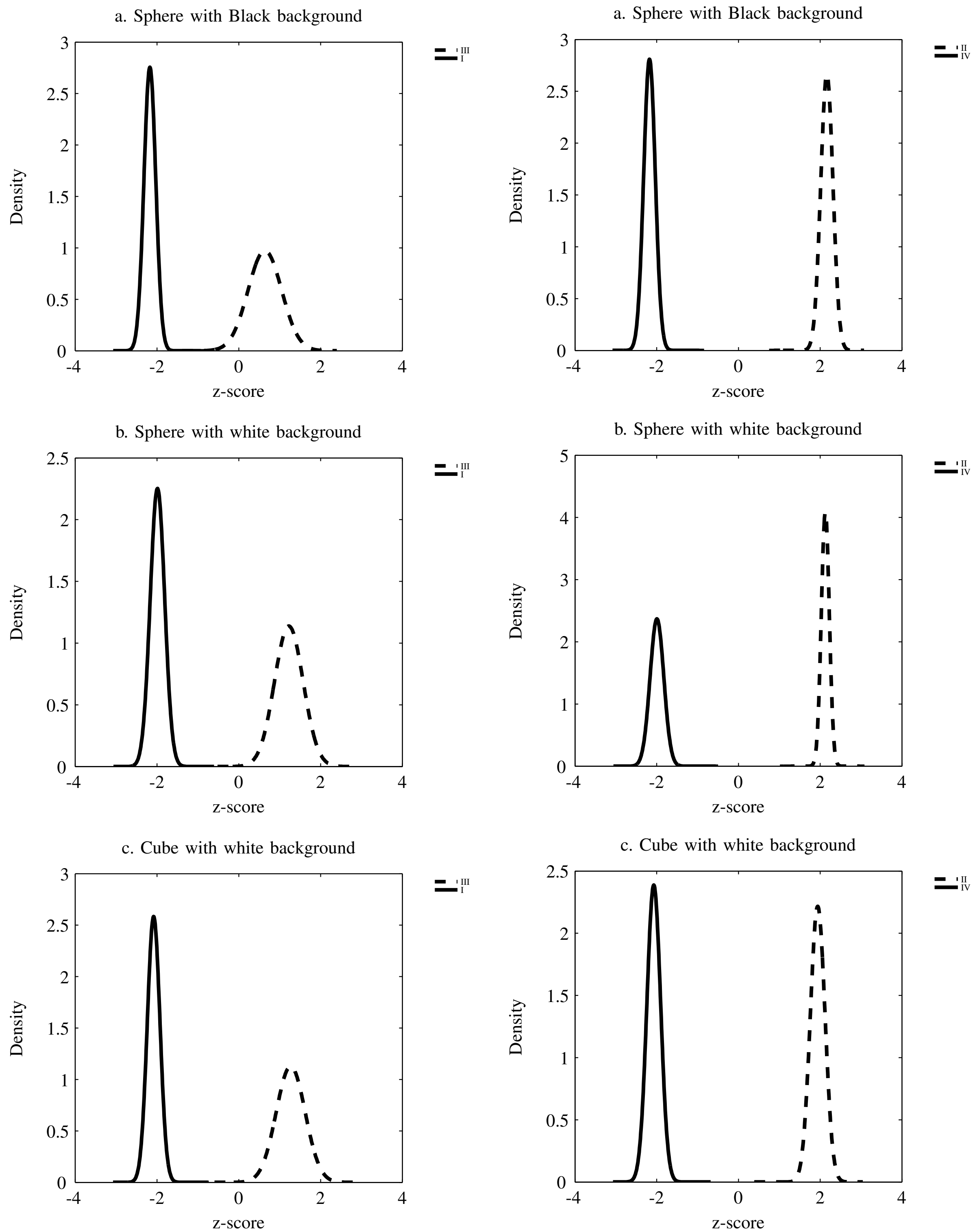

Figure 9: Normal distribution of subjects' z-score for the three conditions using the SWD

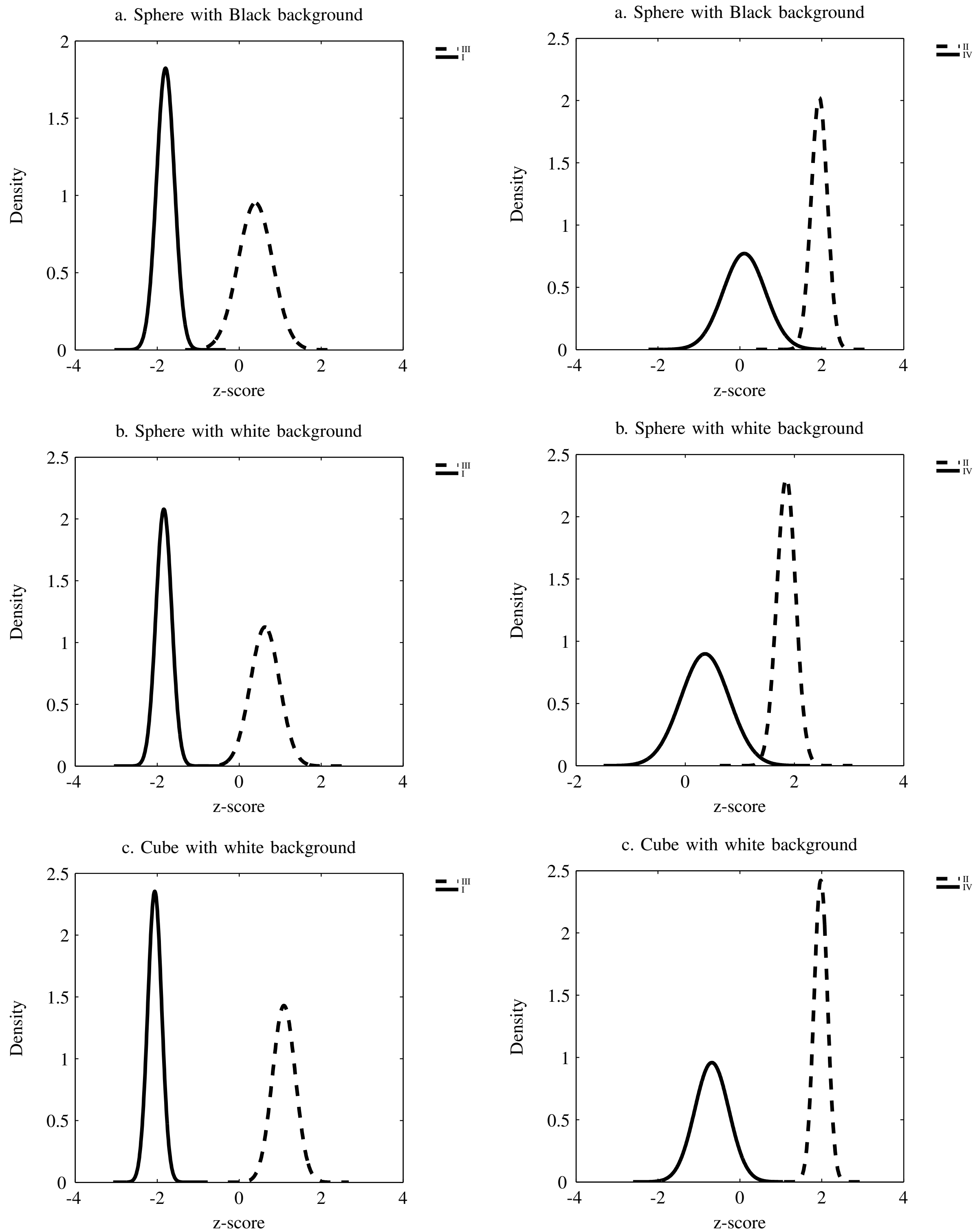

Figure 10: Normal distribution of subjects' z-score for the three conditions using the HMD

size to perceive depths in the SWD comparisons. With the HMD, there was more confusion and ambiguity, especially for the comparisons III and IV. For the comparisons I and II we obtained similar results for both conditions. Furthermore, the effect of the shape of the stimuli had a noticeable impact.

When analyzing the different distributions, it appears that the comparison III was the worst in terms of variance that characterize subject performance for SWD presentation. In this comparison, two objects with different apparent size were presented at the same position. Moreover, the distribution was large and close to zero. This means that subjects react with great variability and most of them answered randomly. This, suggests that subjects do not rely on apparent size nor on the disparity within this type of display. The same phenomena occurred for the comparison IV (same apparent size, different depths) but only when the HMD was used suggesting that in this comparison subjects rely more on the apparent size than on disparity. This fact is confirmed in the cube condition (the disparity is more effective with cube vertex): the performances were less variable and the answers were more accurate.

VII. Conclusion and future work

In this study, we qualitatively evaluated human depth perception in VR systems. This issue is fundamental to many areas particularly for brain and cognitive science research. Indeed, the use of VR has been increasingly used for simulation allowing researchers to create a variety of realistic stimuli under experimental conditions. To fulfill these needs, VR tools must be perfect, or at least well understood, to avoid co-lateral effects and biases, otherwise experimental results and interpretations will suffer from VR-induced distortions and illusions.

Observers were instructed to estimate the depth of virtual objects in four assigned comparison tasks which varied object shape, background color and display technology. Given the relatively small sample size (n of 8) of this study, we realize that care must be taken when drawing statistical conclusions. To address this concern, we confirmed the results obtained through our first statistical analysis by doing a second round of analysis. Specifically, we employed the d' method combined with bootstrap statistics. The d' method allows us to derive additional statistics and provide additional information to confirm or question the original conclusions. The bootstrap method augmented the data in our data set allowing us to overcome the variability observed with our subjects.

More specifically we investigated factors leading users to misestimate the egocentric distances within the peripersonal space when wearing HMDs. To do this, we conducted an experiment comparing human performance on a variety of depth perception tasks when using HMD versus SWD. We had two noteworthy findings. The first is that subjects are able to correctly compare depths using both systems when objects are of the same physical size, however, when objects are of different physical sizes this capability only persisted when subjects used a SWD (performance decreased with a HMD). This allows us to conclude our second finding, which is that subjects rely on apparent size when making depth comparison using a HMD, but not with a SWD. One key difference between the HMD and SWD experience is that a subject is unable to see his or her own body, which suggests that humans may use relevant visual cues from their body's position to judge depth and distance.

The possibility that seeing one's own body provides important visual cues for depth perception will be investigated in future research of this lab. An experimental protocol is currently being developed to that allows researchers to vary the presence of visual cues from subjects' own bodies in the context of a SWD. In other words, we will next explore whether seeing one's own body influences a subject's depth perception abilities with virtual objects while excluding any possible variables that inherently exist between HMDs and SWDs. Additional questions remain regarding the possible importance of physiological properties, specifically accommodation, convergence, or eye movement, that might be investigated with eye tracking technology.

References

[1] A. Naceri, R. Chellali, F. Dionnet, and S. Toma, "Depth perception within virtual environments: A comparative study between wide screen stereoscopic displays and head mounted devices," in *ComputationWorld'09*. IEEE Computer Society, 2009, pp. 460–466.

[2] I. Sutherland, "The ultimate display," in *Proceedings of the International Federation of Information Processing Congress*, 1965.

[3] M. Slater, A. Steed, J. McCarthy, and F. Maringelli, "The influence of body movement on subjective presence in virtual environments." *Hum Factors*, vol. 40, no. 3, pp. 469–477, Sep 1998.

[4] A. J. Parker, "Binocular depth perception and the cerebral cortex." *Nat Rev Neurosci*, vol. 8, no. 5, pp. 379–391, May 2007.

[5] M. Zago, G. Bosco, V. Maffei, M. Iosa, Y. P. Ivanenko, and F. Lacquaniti, "Internal models of target motion: expected dynamics overrides measured kinematics in timing manual interceptions." *J Neurophysiol*, vol. 91, no. 4, pp. 1620–1634, Apr 2004.

[6] P. Senot, M. Zago, F. Lacquaniti, and J. McIntyre, "Anticipating the effects of gravity when intercepting moving objects: differentiating up and down based on nonvisual cues." *J Neurophysiol*, vol. 94, no. 6, pp. 4471–4480, Dec 2005.

[7] J. E. Swan, A. Jones, E. Kolstad, M. A. Livingston, and H. S. Smallman, "Egocentric depth judgments in optical, see-through augmented reality." *IEEE Trans Vis Comput Graph*, vol. 13, no. 3, pp. 429–442, 2007.

[8] J. Loomis, "Distal attribution and presence," *Presence*, vol. 1, pp. 113–119, 1992.

[9] J. M. Loomis, J. A. Da Silva, J. W. Philbeck, and S. S. Fukusima, "Visual perception of location and distance," *Current Directions in Psychological Science*, vol. 5, pp. 72–77, 1996.

[10] J. M. Loomis and J. M. Knapp, "Visual perception of egocentric distance in real and virtual environments," *In L. J. Hettinger and M. W. Haas (Eds.), Virtual and Adaptive Environments*, pp. 21–46, 2003.

[11] J. Cutting and P. Vishton, "Perceiving layout: The integration, relative dominance, and contextual use of different information about depth," *In Epstein, W., & S. Rogers (Eds.), Handbook of Perception and Cognition: Perception of Space and Motion*, vol. 5, pp. 69–117, 1995.

[12] F. H. Previc, "The neuropsychology of 3-d space." *Psychol Bull*, vol. 124, no. 2, pp. 123–164, Sep 1998.

[13] S. Nagata, "How to reinforce perception of depth in single two-dimensional pictures," *Pictorial communication in virtual and real environments*, pp. 527–545, 1991.

[14] C. Armbrster, M. Wolter, T. Kuhlen, W. Spijkers, and B. Fimm, "Depth perception in virtual reality: distance estimations in peri- and extrapersonal space." *Cyberpsychol Behav*, vol. 11, no. 1, pp. 9–15, Feb 2008.

[15] T. Grossman and R. Balakrishnan, "An evaluation of depth perception on volumetric displays," in *AVI '06: Proceedings of the working conference on Advanced visual interfaces*. New York, NY, USA: ACM, 2006, pp. 193–200.

[16] R. Messing and F. H. Durgin, "Distance perception and the visual horizon in head-mounted displays," *ACM Trans. Appl. Percept.*, vol. 2, no. 3, pp. 234–250, 2005.

[17] J. A. Aznar-Casanova, E. H. Matsushima, J. A. D. Silva, and N. P. Ribeiro-Filho, "Can exocentric direction be dissociated from its exocentric distance in virtual environments?" *Percept Psychophys*, vol. 70, no. 3, pp. 541–550, Apr 2008.

[18] S. R. Ellis and B. M. Menges, "Localization of virtual objects in the near visual field." *Hum Factors*, vol. 40, no. 3, pp. 415–431, Sep 1998.

[19] B. Wu, T. L. Ooi, and Z. J. He, "Perceiving distance accurately by a directional process of integrating ground information." *Nature*, vol. 428, no. 6978, pp. 73–77, Mar 2004.

[20] J. M. Knapp and J. M. Loomis, "Limited field of view of head-mounted displays is not the cause of distance underestimation in virtual environments," *Presence*, vol. 13, pp. 572–577, 2004.

[21] S. H. Creem-Regehr, P. Willemsen, A. A. Gooch, and W. B. Thompson, "The influence of restricted viewing conditions on egocentric distance perception: implications for real and virtual indoor environments." *Perception*, vol. 34, no. 2, pp. 191–204, 2005.

[22] M. Mon-Williams and J. R. Tresilian, "Ordinal depth information from accommodation?" *Ergonomics*, vol. 43, no. 3, pp. 391–404, Mar 2000.

[23] J. A. Thomson, "Is continuous visual monitoring necessary in visually guided locomotion?" *J Exp Psychol Hum Percept Perform*, vol. 9, no. 3, pp. 427–443, Jun 1983.

[24] J. Decety, M. Jeannerod, and C. Prablanc, "The timing of mentally represented actions." *Behav Brain Res*, vol. 34, no. 1-2, pp. 35–42, Aug 1989.

[25] J. Gibson, *The Ecological Approach to Visual Perception*. London: Lawrence Erlbaum Associates, 1986.

[26] A. Murgia and P. M. Sharkey, "Estimation of distances in virtual environments using size constancy," *The International Journal of Virtual Reality*, vol. 1, pp. 67–74, 2009.

[27] P. Willemsen, M. B. Colton, S. H. Creem-Regehr, and W. B. Thompson, "The effects of head-mounted display mechanical properties and field of view on distance judgments in virtual environments," *ACM Trans. Appl. Percept.*, vol. 6, no. 2, pp. 1–14, 2009.

[28] B. G. Witmer and P. B. Kline, "Judging perceived and traversed distance in virtual environments," *Presence: Teleoper. Virtual Environ.*, vol. 7, no. 2, pp. 144–167, 1998.

[29] W. B. Thompson, P. Willemsen, A. A. Gooch, S. H. Creem-regehr, J. M. Loomis, and A. C. Beall, "Does the quality of the computer graphics matter when judging distances in visually immersive environments?" 2002.

[30] P. Willemsen, A. A. Gooch, W. B. Thompson, and S. H. Creem-Regehr, "Effects of stereo viewing conditions on distance perception in virtual environments," *Presence: Teleoper. Virtual Environ.*, vol. 17, no. 1, pp. 91–101, 2008.

[31] V. Interrante, B. Ries, and L. Anderson, "Distance perception in immersive virtual environments, revisited," in *Proc. Virtual Reality Conference*, 2006, pp. 3–10.

[32] J. M. Plumert, J. K. Kearney, J. F. Cremer, and K. Recker, "Distance perception in real and virtual environments," *ACM Trans. Appl. Percept.*, vol. 2, no. 3, pp. 216–233, 2005.

[33] H. E. Gruber, "The relation of perceived size to perceived distance," *The American Journal of Psychology*, vol. 67, pp. 411–426, 1954.

[34] M. E. Berryhill, R. Fendrich, and I. R. Olson, "Impaired distance perception and size constancy following bilateral occipitoparietal damage." *Exp Brain Res*, vol. 194, no. 3, pp. 381–393, Apr 2009.

[35] W. P. Tanner and J. A. Swets, "A decision-making theory of visual detection." *Psychol Rev*, vol. 61, no. 6, pp. 401–409, Nov 1954.

[36] D. Green and J. Swets, *Signal detection theory and psychophysics*. Wiley, New York, 1966.

[37] B. Efron, "Bootstrap methods: Another look at the jackknife," *Ann. Statist.*, vol. 7, pp. 1–26, 1979.

[38] K. Ball, D. Smith, A. Ellison, and T. Schenk, “Both egocentric and allocentric cues support spatial priming in visual search.” *Neuropsychologia*, vol. 47, no. 6, pp. 1585–1591, May 2009.

[39] G. Committeri, G. Galati, A.-L. Paradis, L. Pizzamiglio, A. Berthoz, and D. Lebihan, “Reference frames for spatial cognition: different brain areas are involved in viewer-, object-, and landmark-centered judgments about object location.” *J Cogn Neurosci*, vol. 16, no. 9, pp. 1517–1535, Nov 2004.