Analyzing the Adequacy of Interaction Paradigms in Artificial Reality Experiences

Narcís Parés¹ and David Altimira¹

¹Universitat Pompeu Fabra

In Artificial Reality experiences, that is, interactive, unencumbered, full-body, 2D vision-based virtual reality (VR) experiences (heirs of the seminal Videoplace by Myron Krueger), there are two possible interaction paradigms, namely, first-person and third-person—which differ significantly from the classic VR first- and third-person notions. Up until now, these two paradigms had not been compared or objectively analyzed in such systems. Moreover, most systems are based on the third-person paradigm without a specific justification, most probably due to the influence of the original Videoplace system and because it is the only paradigm available in commercial development tools and leisure systems. For example, many rehabilitation projects have chosen to use these VR systems because of their many advantages. However, most of these projects and research have blindly adopted the third-person paradigm. Hence, the field of virtual rehabilitation has analyzed the beneficial properties of these systems without considering the first-person paradigm that could potentially present better adequacy. To find and understand potential differences between the two paradigms, we have defined an application categorization from which we developed two full-body interactive games and set up an experiment to analyze each game in both paradigms. We studied how 39 participants played these games and we quantitatively and qualitatively analyzed how each paradigm influenced the experience, the activity and the behavior of the users and the efficiency in accomplishing the required goals. We present the results of these experiments and their general implications, and especially for virtual rehabilitation due to the potential impact these systems may have in the well-being of many people.

Narcís Parés is a computer scientist in audiovisual communication with an interest in AMVR, embodied interaction and interactive playgrounds; he is an Associate professor in the Cognitive Media Technologies group of the ICT Department of Universitat Pompeu Fabra (Barcelona, Spain) and is coordinator of the Interdisciplinary Master in Cognitive Systems and Interactive Media. David Altimira is a computer scientist with an interest in human-computer interaction, entertainment computing and responsive environments; he is a PhD candidate at the ICT Department of Universitat Pompeu Fabra.
1. INTRODUCTION

Virtual Reality (VR) is increasingly becoming an important tool in many areas such as rehabilitation (Kizony, Rand, Weiss, & Katz, 2004; Rizzo & Joungyoon, 2005), providing some clear advantages over rehabilitation in real environments as
explained by Holden (2005). Among the VR technologies that are being used, the systems based on artificial vision (or camera capture) are especially popular. These have existed since the mid-1970s when Myron Krueger defined his first Videoplace system (Krueger, Gionfriddo, & Hinrichsen, 1985). In this article we refer to them as Artificial Reality systems following Krueger’s term and use the acronym ArtR to differentiate it from AR, which usually refers to “augmented reality.”

The last decade has seen ArtR systems explode in leisure and rehabilitation as technology has become more accessible. In this article we provide references from both fields; however, we pay special attention to rehabilitation because of three reasons: First, there has been quite a lot of work with ArtR systems in this field with very interesting results; second, from all fields of application of ArtR systems, rehabilitation is socially a very important field in which our findings can have a strong impact; third, we pay attention because one of the paradigms that we define emerged within this field.

Some ArtR systems, such as the Vivid GX, are costly and need programming knowledge, which makes them more difficult for psychologists and therapists to use, although they are powerful and adaptable to each specific project. Other systems in the form of videogame console extensions, such as the Sony PlayStation 2 EyeToy, have been broadly used as rehabilitation tools thanks to their low cost and off-the-shelf nature, although they are closed systems that cannot be customized. Researchers have been able to test several proposals in rehabilitation (Brooks & Petersson, 2005; Burke, Morrow, McNeill, McDonough, & Charles, 2008; Flynn, Palma, & Bender, 2007; Haik et al., 2006; Kizony, Katz, & Weiss, 2004; Kizony, Rand, & Weiss, 2004; Parés et al., 2006; Parés, Masri, van Wolferen, & Creed, 2005), taking advantage of the important properties of these systems, namely, they are unencumbered systems that do not invade the user’s body, they are robust because users do not need to manipulate any physical interface, and they propose quite natural interaction to the users and are hence relatively easy to use. However, almost all systems used in rehabilitation have used only one of the two possible interaction paradigms available in these systems (i.e., third-person) without even being aware that another exists (first-person). This could be missing opportunities for better systems or hiding issues that hinder rehabilitation potential.

In this article we first describe the technological differences between the two paradigms to understand how they are configured. Next, we introduce the factors that differentiate their interaction (i.e., those that define each paradigm). We clarify how and why these paradigms differ from the classic definitions of first- and third-person paradigms in VR. We then describe the state of the art of applications of ArtR, especially concentrating on rehabilitation and disabilities, so as to provide a clear view on how these have been biased toward the third-person paradigm and how first-person has been historically ignored. We describe the emergence of the first-person configuration and justify the need to analyze and compare both paradigms in order to have a complete scientific view and sufficient decision elements to choose the best in future applications. This comparison is addressed first from an analytic and reasoned approach, which then leads to the description of our hypotheses and
FIGURE 1. Typical system configuration of a third-person vision-based virtual reality system composed of (a) camera, (b) computer system, (c) visual feedback display, (d) user, and (e) representation of user within virtual environment. Here the user is situated away from the display. Her image is integrated in the virtual environment (VE) to provide a means with which to interact (f) with the virtual objects within the VE.

to the experimental design that we have defined to verify or refute them. We then describe and discuss the results. These results provide a validation of previous work with ArtR systems and design guidelines for the future.

2. ARTIFICIAL REALITY AND ITS INTERACTION PARADIGMS

2.1. Basic Artificial Reality System Configuration

Typically, these systems are based on a video camera that captures images (shape and gestures) of the body of the user (Figures 1 and 2). The captured images are related to the virtual elements generated and controlled by the system. Visual feedback is presented to the user in a display in front of her, where she can perceive the consequences of her actions on the virtual environment (VE).

It is important to stress that these systems use a single camera to capture user activity, and hence this activity is captured on a single plane (i.e., it is a two-dimensional [2D] capture system).\(^1\) If the user moves toward or away from the camera, changing the distance at which the system expects the user to be, the system does not detect this

---

\(^1\)In contrast to motion capture systems that use multiple cameras to capture the movements of the body of the user in three-dimensional [3D] space but that need markers on the body of the user and an uncomfortable calibration process.
as a forward/backward displacement. Rather, if the user moves toward the camera, the system will see a “larger user” that can eventually cover all the visual field of the camera and impede any gesture detection. On the other hand, if the user moves away from the camera, the system will see a “smaller user” and hence the user will have difficulties in reaching the extremes of the VE.

The 2D nature of these capture systems must not be confused with the dimensions of the VE in the experience. For example, despite the user is sensed only in 2D, the physical right and left movements of the user could be mapped onto right and left turns during the navigation of a 3D virtual environment.

These systems may be found in two distinct hardware configurations that have given rise to two types of interactive media for full-body interaction: those that use a third-person interaction paradigm and those that use a first-person paradigm. However, we must not confuse hardware configuration with the specificities of the interaction paradigms.

2.2. First-Person versus Third-Person Configurations

The two interaction paradigms just mentioned emerge from the way in which the hardware is configured with respect to the user. However, as said before, hardware

---

2These two cases can actually be used to cheat in game experiences. In the first, the user can better reach the entire environment and/or better collide with targets, whereas in the second the user can better avoid colliding with forbidden virtual elements. However, as the instructions video for the Sony PlayStation 2 EyeToy says, it is not fun if you cheat.
configuration must not be confused with the specificities of the paradigm. Rather, the hardware configuration is only part of the equation, and there are other design decisions that come into play. We first start by describing each of the two configurations and later we explain their interaction specificities.

In the present research, the third-person paradigm is always related to a system that is configured as shown in Figure 1 and that is the configuration used by the original Videoplace system (Krueger et al., 1985). Typically, the camera (a) is placed on the same side as the visual display (c). This forces the user (d) to stand a distance away from both camera and display such that the camera correctly captures the entire body of the user. Because of this, to be able to provide a clear feedback to the user regarding her interaction with the virtual objects, the system integrates the captured image of the user (e) into the VE. As we detail later, this has an important impact on how the experience is mediated to the user and, hence, on the type of interaction paradigm. For now we say only that there is an indirect interaction of the user with the VE. Hence, we call this indirect way of interaction a third-person interaction paradigm.

On the other hand, the first-person paradigm is related to a system configured as shown in Figure 2. Here the camera (a) is placed opposite the display (c). This places the user (d) between the camera and the display. This is interesting because, as opposed to the third-person configuration, the user can stand close to the display. We later go more in depth about the impact of this on the type of interaction and mediation of the experience, but for now we say that this physical proximity of the user to the VE provides a direct interaction. Hence, we call this direct way of interacting with the VE a first-person interaction paradigm. We would like to clarify that there is no need to touch the screen in this first-person configuration. We are not referring to a multitouch system.

2.3. Difference Factors Between First- and Third-Person Paradigms

On analyzing the experience of the user, the difference between these two paradigms goes beyond the mere hardware configuration or the fact of seeing, or not, one’s image in the display. The difference actually defines distinct media altogether because the way the experience is mediated to the user is completely different. If we analyze it from a media studies point of view we can see that the different perception and cognitive processes in the user modify her understanding of the experience.

Specifically, in third-person the user must mentally map, at a distance, her body movements onto her representation within the environment to calibrate her actions and achieve the desired goals while interacting with the virtual objects. It is important to stress this point. If the image of the user were not integrated into the VE, the user, being placed a certain distance away from the display, would have no reference whatsoever on which part of the VE she would be acting upon. By integrating her
image in the VE the system behaves like a digital mirror. The image of the user can be a simple silhouette, extracted from the real-time video capture (as in Krueger’s, 1985, Videoplace), or a live video feed of the detailed image of the user. In any case, the user can recognize her image and can interact with the virtual objects through it in the VE (Figure 1).

We have found no literature on this self-calibration or adaptation process of the user with respect to her image in the VE. However, we could understand it as analogous to when a child learns how to comb her hair in front of a mirror and has trouble with coordination gestures to place the comb in the right position with respect to her head. It would be interesting to formally study this in future research. In the meantime, we take it as an issue that may potentially make interaction more difficult to the users when in third-person paradigm.

It is important to observe that in third-person, although the user is moving in physical space, all interaction activity is occurring in virtual space. In other words, because the user is placed a significant distance away from the display and it is her captured image in the VE that allows her to act upon the virtual objects, we can effectively say that all collisions, displacements, manipulation, and so on, are occurring in virtual space only. This mediates the experience of the user in a very specific manner because the user must understand she is “inside” the VE, despite seeing it from “outside.” We can say this is a fully virtual experience.

On the other hand, in first-person the user does not need to do this mental mapping because she is very close to the display and, although she does not need to touch the display, she can see her body affecting and acting upon the virtual objects directly by physical adjacency (Figure 2). For example, the user can swipe her arm in front of the screen, in physical space, and observe how her arm “collides” with a virtual object and displaces it in virtual space. This means that there is no need for a representation of her body within the VE, that is, there is no need for an extra visual feedback.

However, now the user must be aware that the physical space she moves in—this thin layer that is very close to the display—is in tight coupling with the virtual space of the VE “in” the display. In other words, she must be aware that her body movements, occurring in physical space, have an impact on the virtual objects, which are moving and behaving in virtual space. She must merge both worlds, physical and virtual, in a co-placed relationship. We can say this is an archetypal mixed reality experience (Milgram, Takemura, Utsumi, & Kishino, 1994).

Because the user can directly see her body physically adjacent to the virtual objects and influencing the VE, first-person could appear as a more natural way of interaction than third-person, for example, in the previous example of the user swiping her arm and impacting a virtual object, the situation is closer to our habitual way of experiencing manipulation of objects in physical space. However, the mental effort that the user must make in merging physical and virtual space might not be so obvious. Although this has been an important research topic in VR ever since Milgram et al. (1994) defined the notion of mixed reality, it has not been studied before in ArtR.
We take this as an issue that may potentially hinder interaction of the users when in first-person paradigm.

One further configuration difference is that in third-person, the display can range from a large-scale projection to a desktop computer monitor. It is the case, for example, of the Sony PlayStation 2 EyeToy, which, being a home system, is usually used on a home television (Figure 3). Instead, in first-person configuration it does not seem to make sense to have small displays. This is because in first-person the display is meant to present visual stimuli at a human scale, such that the user can relate her physical body to the virtual world in a 1:1 relationship. This 1:1 scale relationship is not necessary in third-person where the representation of the user within the VE already provides the user with a relation of scale with respect to the virtual objects regardless of the size of the display (Figure 3a). This is further explained in the next section.

At this point, we can see that the differences between the two paradigms do not only depend on having the camera in front or behind, or a larger or smaller display. These configurations have an impact beyond technological reasons only. They impact on interaction design decisions such as the need, or otherwise, of including a silhouette that represents the user. They impact also on the described perceptual and cognitive issues that generate different meanings and understanding of the experience by the user. Hence, the differences between first- and third-person paradigms are a conglomerate of properties that are tightly interrelated defining different media that yield very specific experiences in each case.

We can therefore summarize the factors that holistically identify and differentiate first- and third-person paradigms as shown in Figure 4. These factors have never before been pinpointed, described, or addressed in this manner, and our research tries to gain some insight on them. This article does so by comparing the two paradigms through the experiments described next.

**FIGURE 3.** In third-person configuration, the size of the display (a) is irrelevant because the virtual image of the user is the element that sets the scale relation with respect to the virtual objects.
FIGURE 4. Summary of factors that together identify and differentiate first- and third-person interaction paradigms.

<table>
<thead>
<tr>
<th>Third-person paradigm</th>
<th>First-person paradigm</th>
</tr>
</thead>
<tbody>
<tr>
<td>User away from display.</td>
<td>User close to display.</td>
</tr>
<tr>
<td>System presents visual representation of user’s body within VE as interaction feedback.</td>
<td>System does not present any visual feedback of the user’s body.</td>
</tr>
<tr>
<td>User has an indirect interaction with virtual objects through her representation within the VE.</td>
<td>User has a direct interaction of her body with the virtual objects (although there is no need to touch display; see text for details).</td>
</tr>
<tr>
<td>User has a fully virtual experience in virtual space (her virtual image with the virtual objects), although it is a view from “outside” (physical space) to the “inside” (virtual space).</td>
<td>User has a mixed reality experience that relates her physical body in physical space to the virtual objects in virtual space. Therefore, the user does not have a view from “outside” to the “inside”. She rather has to merge both “outside” and “inside” into a single co-placed activity.</td>
</tr>
<tr>
<td>The scale relation between the user and the virtual objects is provided by the proportion of the presented image of the user in the VE with respect to the virtual objects. This is therefore independent of the size of the display.</td>
<td>The scale relation between the user and the virtual objects is a 1:1, or human-sized, relation provided by the size of the display.</td>
</tr>
</tbody>
</table>

*Note. VE = virtual environment.*

2.4. Differentiating These Paradigms From Classic VR First- and Third-Person

Work done in VR comparing and testing first- and third-person interaction has been mostly done on head-mounted displays (HMDs) or desktop systems (e.g., for a recent example, see Salamin, Tadi, Blanke, Vexo, & Thalmann, 2010). However, we assert that these versions of first- and third-person interaction are not the same concepts as the ones we are using in this article. An initial hint of this difference is that they are more often referred to as egocentric or subjective view, and exocentric or follow-up view, respectively. To our understanding, the classic approach to these concepts comes from the fact that they refer only to virtual space.

Let us look first at egocentric or subjective view. In classic VR applications (i.e., not ArtR) this type of view aims to put the user “inside” the VE looking at the surrounding virtual objects. The user is therefore preoccupied with virtual space and nothing relates her to physical space. Hence, many applications allow the user to look “down” and see part of her (virtual) body (legs, feet, etc.). In other cases, such as in
first-person shooters, the point of view shows the tip of the weapon that the user is supposedly carrying. In driving games the front of the car (or vehicle) that the user is supposedly driving may be seen in the foreground as a reference of this subjective viewpoint. It is clear that under this paradigm, the experience of the user is meant to be fully in virtual space.

In contrast, as described earlier, in ArtR the first-person interaction paradigm blends virtual and physical space into a single mixed reality experience. Both, the physical body—and surrounding physical space—of the user and the virtual objects, and virtual space that they inhabit, are important in the final experience that is defined. To have an impact on the virtual objects, the user must be able to put her physical body in relation to them in the border that is created between virtual and physical space. Therefore, ArtR first-person and VR first-person are not comparable under the same terms and actually refer to two different media types and interaction paradigms.

Let us now turn to exocentric or follow-up view in classic VR applications. Although this is an uncommon view in scientific or simulation experiences—where egocentric or subjective view reigns—and has been almost only used in videogames, it is getting more interest in the last years, for example, in training applications (Salamin et al., 2010). This type of view is composed of two basic parts: an interaction element and a (separate) point of view. The interaction element, often called an avatar, which can be either a human figure or a vehicle or any other element, is that with which the user is identified in the VE. It sets the user’s position and orientation within the VE, and the user can control it to have an impact on the virtual objects. The other part is the point of view that, as opposed to egocentric systems in which the user controls it to navigate the VE, it now only follows the avatar at a distance behind and (often) slightly above it. Therefore, the user must control the avatar to change the point of view (hence the name “follow-up” view). Because of this, exocentric systems are based on 3D navigation and are solely worried about the virtual world. The user gets no reference of the physical world, and it is irrelevant for the experience.

In contrast, in the ArtR third-person interaction paradigm, we clearly have the user “outside” in physical space looking “into” the virtual space at her mirrored image that allows her to impact the virtual objects. This notion of the “mirrored image” is important because it is not an artificially created control element, avatar, but rather it is a manifestation of the user’s physical actions mapped within virtual space. Hence, the physical space in which the user moves is important not only at a conceptual level but also at a practical level because the movements of the user are being captured on a plane (i.e., in 2D), and this implies that in most applications the action occurs also in 2D. In some cases, physical space may even help contextualize the activity as in the SIDEshow system (Davis & Bobick, 1998a).

3. RELATED WORK

We have not found any previous attempts to formally understand the aforementioned factors, similarities, and differences between interaction paradigms in ArtR.
Even more surprising, very few works that use either one of the paradigms have reflected at all on why they use one or the other, one of these few being the MEDIATE project (Parés et al., 2006; Parés et al., 2005), as described later.

We now present the state of the art of ArtR systems and applications to be able to set the grounds of what has been done, to see how it compares to what we are describing and proposing, to show how in the history of these systems the third-person paradigm has dominated the field without a justified or explicit reason, to describe how the first-person paradigm emerged within this context, and to provide a good overview of how much work has been done in the field of rehabilitation and special needs.

3.1. Third-Person Systems and Applications

With Krueger’s Videoplace system the third-person configuration was established as the default configuration of ArtR systems.

After Krueger’s system, many other have been important throughout the years. The earliest one was the Mandala Gesture Extreme System by Vivid Group that started in 1984 in Canada (and which later became the company GestureTek, in 2000). They have provided many location-based entertainment experiences as well as an important set of libraries for developers. The configuration was also clearly third-person, although they used a live video feed of the image of the user as opposed to only the silhouette that Krueger used. Because of this, they used chroma key background behind the user to facilitate segmentation (as they were using color cameras), whereas Krueger used a light screen behind the user to capture the dark silhouette.

Davis and Bobick defined the SIDEshow system (Davis & Bobick, 1998b) with a dual-screen configuration in which the user stands between the two screens (one in front of the user and another behind the user). They were among the first to use near Infrared lighting and cameras to capture the silhouette of the user. Their system was again third-person.

Reality Fusion’s GameCam was the first home system based on a webcam that was commercially available in 1999 (Roberts, 2000). It was immediately followed that same year by the Intel’s Me2Cam Virtual Game System (D’Hooge & Goldsmith, 2001) also based on a webcam. Although not very successful, they set the grounds for the launching of the Sony PlayStation 2 EyeToy in 2001, which has recently been updated with the PlayStation 3 Eye.

Booth-Simpson defined a system in which the user casted a shadow over the projection of the VE by standing between the projector and the screen. The user interacted with the VE through her shadow (Mine Control; Simpson, 2001). We describe this system in more detail in a later section.

Some examples of experiences that use the third-person paradigm are Critter and Digital drawing (Krueger et al., 1985), the ALIVE system (Maes, Darrell, Blumberg, & Pentland, 1995), the Virtual PAT (Davis & Bobick, 1998b), SURVIVE (Russell,

We believe that, because the systems that are easily available—such as the Vivid GX or the Sony PlayStation 2 EyeToy—use only the third-person paradigm and no commercial systems use first-person, the projects that have used these systems have not had any other option but to use third-person. Although they have achieved good results, doubts remain about what would have happened had they used a first-person paradigm. Of interest, none of these systems have questioned this.

### 3.2. First-Person Systems and Applications

The first project of which we are aware that used the first-person paradigm was the MEDIATE project back in 2002 (Parés et al., 2006; Parés et al., 2005). This European-funded project (IST FP5) designed and developed an interactive multimodal space for children with autism. The project wished to provide a sense of control and a sense of agency to these children (something they do not usually experience in their daily lives). Interaction within the space had to be unencumbered, autonomous, and free, and therefore the authors decided to use ArtR systems. However, in the design process they feared that the third-person paradigm could hinder the experience of the children with autism. They posed the reasonable doubt that perhaps these children could not recognize their image in the VE, or that they could even be put off when seeing themselves within the VE. They also disliked having to impose on the children the physical limitation of not approaching the screens and visuals. The interaction had to be free, and children should be allowed to move around the space without any special constraints. They therefore decided to devise a system that allowed the children to approach or even touch the screens, as well as to see their own bodies physically in relation to the virtual objects. This way they configured the system as in Figure 2 and defined the first-person paradigm.

In 2006 the Remote Impact System also used a first-person system configuration in order to allow the users to hit the screen in a sort of “tele-boxing” experience (Mueller, Agamanolis, Vetere, & Gibbs, 2007). We analyze this system in more detail in the next section.

More recently, many systems have appeared with different configurations, for example, the Interactive Floor systems (e.g., GestureTek; Luminvision; Vertigo Systems GMBH; ArcstreamAV; FeedTank; HB-Laser; etc.). These systems project the VE onto the floor and detect the users from above, as they interact over it. Another system is that of the Interactive Slide (Parés, Soler, & Ferrer, 2009), a large inflatable slide augmented with a projection and computer vision system. Other examples are Motionscapes (Lieberman, n.d.), Robot Factory (Parés et al., 2009), and Connections (Parés & Carreras, 2007).
3.3. Comparing and Analyzing Particular Systems

Concerning similar configurations to the ones we propose as being general, we must mention a few and situate them in relation to the two that are being used in this article due to their apparently unique configuration or properties.

On one hand, we have the new Kinect system for the Microsoft Xbox videogame console. Although the technology of the camera that captures user activity is very different from the previously referenced systems—as it captures a depth map of the user and her surrounding physical space—the camera itself could, in principle, be used for any configuration. However, a close analysis of the arrangement of the components in a typical home environment shows that the configuration is that of a third-person arrangement (Figures 1 or 5). The experiences include some sort of representation of the user within the VE and hence, the interaction of the user is indirect.

On the other hand, we have the Front Shadow system by Zach Booth Simpson (2001). His system is based on the fact that the user is asked to stand between the projector and the projection screen, hence casting a shadow on the projected images. The camera is not placed on the same side of the display (or screen), because instead of capturing the body of the user directly, it captures the silhouette of the user.

FIGURE 5. Categorization of Artificial Reality applications, based on object- versus space-based interaction and on gesture- versus reach out-based interaction (see text for details). (Color figure available online.)
shadow of the user. To do so the camera must be placed on a side with respect to the installation’s longitudinal axis and pointing toward the screen to capture the shadow without having the user in the way. What the system does is to use this shadow as the interaction element provided to the user to act upon the virtual objects. Therefore, it can be clearly seen that this is again a third-person configuration where the interaction of the user with the virtual objects is again indirect (in this case, through her shadow).

We would also like to analyze the configuration used in the Remote Impact System (Mueller et al., 2007). This system uses a hardware configuration that is clearly analogous to the first-person configuration just presented (Figure 2). However, the system also includes the image of the user (captured by the camera behind her) in the VE. This would seem to be a hybrid between a first- and a third-person system. It could also seem to be redundant because the user is already seeing her body physically adjacent to the projected virtual objects and hence not need that extra visual feedback. After a closer analysis of the system and the specific game mechanics that the authors designed, we can see that the image of the user is actually not the element of interaction with the other virtual objects, that is, the represented image of the user cannot alter the other virtual objects. Therefore, it is rather only another virtual object in the same hierarchal level as the others. The reason is that in the game, which is based around a telepresence or teleconference boxing game, the user must hit the projected image of a remote user to win points but must avoid hitting her own image to avoid losing points. In other words, the user still has a direct interaction of her physical body with respect to the virtual objects (which in this case include the image of the user herself) and the image of the user cannot act upon the other virtual objects. Hence, we can assert it is a “pure” first-person configuration.

Although this is not meant to be a comprehensive review of all existent ArtR systems, we believe the ones mentioned throughout this article are representative of the range of existing systems. Therefore, we believe that the two general configurations that we previously describe (first- and third-person) correctly and fully represent the totality of the 2D camera capture, or ArtR systems.

### 3.4. Rehabilitation and Disabilities Applications Examples

Several ArtR applications for rehabilitation have appeared in the past 10 years. We must start from the work by Kizony et al. because they have set the ground with important research on these technologies, especially in the field of motor rehabilitation of stroke patients. They started by analyzing the feasibility of these technologies in the field of virtual rehabilitation via the Vivid GX system (Kizony, Katz, et al., 2004; Kizony, Rand, Weiss, & Katz, 2004). Kizony et al. (2004) also proposed a model that consists of three nested circles “Interaction Space”, the “Transfer Phase” and the “Real World.” This model tries to explain the process of rehabilitation using VR, which starts from the user interaction within the virtual environment (“Interaction Space”) to the transfer of the acquired skills to the real world.
However, the cost of such systems made it difficult to achieve a widespread use in rehabilitation centers and much less at the patients’ homes. They therefore explored the possibilities of using low-cost systems—namely, the Sony PlayStation 2 EyeToy—and compared the results of rehabilitation with respect to the Vivid GX system (Kizony et al., 2004). They found that when using the EyeToy, the inability to grade the difficulty level of the environments frustrated some acute stroke patients. Frustration is an undesired effect that may appear if patients cannot manage to interact correctly. This fact could influence the user involvement in the task and, in general, the effectiveness of the use of the VR systems in rehabilitation.

They also compared rehabilitation using these systems with VR experiences that use HMDs (Kizony et al., 2005). According to their research there is a large potential and feasibility of video-capture (ArtR) systems with respect to the VR first-person HMD-based systems. Their experimental findings report that HMD-based systems seem to cause side effects more frequently whereas ArtR systems prove a similar performance, are cheaper, and are less encumbering.

After Kizony’s work, other researchers have also tested the feasibility of using the Sony PlayStation 2 EyeToy game system with quite successful results. Brooks and Petersson (2005) tested this system as a supplementary activity to regular therapy in hospitals contributing to the rehabilitation process as a motivational tool in children. Flynn et al. (2007) tested the EyeToy system on an individual with chronic stroke. One final example is that of Haik et al. (2006), who tested the system in burn rehabilitation.

Some projects have opted to develop their own systems. For example, Burke et al. (2008) developed a system that was applied to upper-limb stroke rehabilitation. Another example is that by Herbelin, Ciger, and Brooks (2008), in which they use a number of technologies to better adapt to each type of disability or special need, and among these technologies they use their proprietary third-person system.

They have all achieved very encouraging results. However, they have only used the third-person paradigm and have not questioned this decision. In the case of those using commercial systems, it is not surprising because these systems are based on the third-person paradigm and do not allow changing to first-person. However, it is interesting that those who developed their own system also chose the third-person paradigm.

The only project that we are aware of having used the first-person paradigm in rehabilitation and special needs is the MEDIATE project previously described (Parés et al., 2006; Parés et al., 2005).

MEDIATE achieved very good results and was clearly accepted by the 90 children with autism who used it. However, the project did not have sufficient time to test whether a third-person paradigm would have achieved equal or better results.

We believe that one of the factors that influence user interaction within the virtual environment is the interaction paradigm, which in turn might influence the level of enjoyment; the ability to interact correctly; the performance; and, in a later stage, the transfer of the skills to the real world. Better understanding the role of the interaction paradigm in such systems should help system developers and clinicians to find more effective rehabilitation experiences.
3.5. Other Work on the Notions of First- and Third-Person

In trying to better understand the differences between the notions of first- and third-person interaction we searched for other works in the fields of VR and computer games. We found the following assertions:

- In VR applications where the movements of the characters are important such as fighting games (e.g., martial arts; Hämäläinen et al., 2005) or in applications used for teaching a physical skill (e.g., Virtual PAT; Davis & Bobick, 1998b), Maes argued that the application is best developed with a third-person approach because it allows users to see (and correct) their position and movements (Maes, Darrell, Blumberg, & Pentland, 1995).
- The ALIVE system (Maes et al., 1995) was developed using third-person because they argued that with this paradigm the user tends to be less disoriented about her own position and orientation in the virtual space than using first-person.
- Oshita (2006) stated that for applications where users can control directly the full body motion of a character in 3D virtual environments (using a 3D motion capture system) or when users need to sense the movements and interactions with virtual objects, the third-person approach is more suitable.
- Warren, in his thesis (Warren, 2003) on ArtR systems, stated that an advantage of using computer vision in full-body interactive experiences is that it can promote expressiveness. However, he does not differentiate between first- and third-person paradigms.
- Designers of Intel Me2Cam reported that showing the faces of the children in the 2D games was a large portion of the appeal of the experience (D’Hooge & Goldsmith, 2001).³
- Traditionally VR applications use the first-person approach to make the users feel a sense of Presence, that is, as if they were in the virtual environments (Baumgartner et al., 2008; Slater, Linkis, Usoh, & Kooper, 1996).

However, none of these design decisions have been fully substantiated, and many are mere intuitive design guidelines. For example, the notion of Presence is controversial, having been contested by several authors as to its validity as an evaluation tool for VR. We detail this further in a later section.

4. PROBLEM STATEMENT

The factors that identify and differentiate the two ArtR interaction paradigms have made us reflect upon the adequacy of using either one of the two paradigms, in any context, but especially in rehabilitation because of the potential impact that

³Our belief is that this was probably due to the novelty factor, which should disappear over time as technology becomes more available.
such differences could have on people that present some type of physical or cognitive
disability. As pointed out earlier, the use of either one may create opportunities or
pose difficulties.

Through some informal play sessions with the Sony PlayStation 2 EyeToy, we
observed that some users in some games unconsciously began to move toward the
display (a TV set in this case). Although this hindered their play, because at a certain
distance the camera system cannot detect gestures anymore, the users tended to
continue their approach to the TV.

This suggested to us the possibility that users were experiencing a low sense of
control. One possible explanation for this would be that distance from the user to
the display in the third-person paradigm was not helping to achieve a good mapping
between the representation of the user and the virtual objects. It occurred to us that
perhaps users needed to see their hands “touching” the virtual objects to have a better
sense of control. In other words, they seemed to quest for a first-person paradigm
although the game was conceived in third-person.

We therefore thought it was time to analyze all these differences between
paradigms and the observed situations in a formal and systematic manner.

5. METHODS

To understand the differences between the two paradigms we defined the
following steps:

- We believe that the effect of one particular interaction paradigm on the expe-
  rience of a user depends on the type of application. We therefore started by
classifying the possible range of applications into categories to be able to decide
which types we would use in our experiments.
- We then chose two categories from two extreme types, focused our experiments
  on these two types of applications, and analyzed the results through quantitative
  and qualitative analysis comparing the same application in first- and third-
  person mode. For the quantitative analysis we used specific system parameters
  that would describe the different attitudes that users adopt during interactive
  experiences implemented in both paradigms. We based qualitative analysis on
  some simple user questionnaires.

5.1. Categorization of the Applications

We analyzed a large set of applications from different ArtR systems (especially
games) and we classified them into different categories (Figure 5).

We differentiated between those applications that are based on interaction with
virtual objects and those that are more based on the user moving her body in space.
We also differentiated between those that asked for specific gestures of the user and
those that asked the user to use the body to reach out to specific places in the display or objects in the VE. Therefore, the criterion we followed for the classification was based on the interactive task required from the user in the application.

From this analysis we obtained the following six categories (clockwise starting at top right in Figure 5):

- **Activation Zones**: Users “touch” certain defined parts of the environment with their body at specified instants to activate effects or behaviors in the VE. (Note that we are not referring to physically touching the display, but rather virtually reaching to access some points in the VE that can trigger events.) Some examples are Drumming from Vivid Group (n.d.) to make music and Groove from EyeToy (“EyeToy: Groove,” n.d.) to dance.

- **Navigation/Displacement through VE**: Users move within the VE through body gestures. Some examples are Survive and Sharkbait from Vivid Group (n.d.) and Athletics from EyeToy (Sony, 2011).

- **Body Shapes**: Users are asked to assume a specific pose of their body or gesticulate in specific ways dictated by the game. Some examples are Boot Camp from EyeToy (“EyeToy: Play 3,” n.d.) and Virtual PAT (Davis & Bobick, 1998a).

- **Real-Time Generation from Gestures**: Users generate virtual elements or objects such as particles as they move their body. Some examples are Digital drawing (Krueger & Gionfriddo, 1985) and Hand Painting from Vivid Group (n.d.).

- **Control of Objects**: Users interact with and control one or more objects of the virtual environment with their body (e.g., a ball). Some examples are Volleyball from Vivid Group (n.d.) and Pinball from Intel’s Me2Cam (D’Hooge & Goldsmith, 2001).

- **Intercept Objects**: Users intercept certain moving virtual elements (moving targets) with their body. Some examples are Orbosity from Vivid Group (n.d.) and Ghost grab from EyeToy (“EyeToy: Play 3,” n.d.).

As may be seen in Figure 5, opposing categories of applications around the center of the diagram are opposing game philosophies. Because of this we chose the two that were clearly the extremes of the object-based to space-based criterion, namely, Control of Objects and Navigation through VE. With this we wished to capture a wide range of possibilities during the experiments.

5.2. Definition of Our Applications

Two applications were designed and developed for this study, one for each of the two chosen categories. For each application (game) we developed two versions, one in first-person and the other in third-person paradigm. In addition, the first type of application was based on a 2D VE, whereas the second on a 3D VE. The games were dubbed “Ball game” and “Athletics game,” classified under the categories of
Control of Objects and Navigation through VE, respectively. We now describe each game.

**Ball Game**

Ball game (Figure 6) is a game where the player uses her body to control a virtual ball in a 2D space. The space is populated by colored squares arranged in a regular matrix. The virtual ball starts to move as soon as the game begins. The ball maintains its direction of movement until it finds an edge of the screen and then bounces against it. Along its movement, the ball can hit the squares and displace them from their current position. These collisions of the ball with the squares do not affect the trajectory of the ball. The length of the displacement of the squares depends on the speed of the virtual ball and on the angle of collision. The virtual ball also changes its trajectory on colliding with the body of the user. However, the body of the user does not directly affect the colored squares. Hence, the goal of the game is to direct the virtual ball so as to sweep as many squares as possible out of the game area in 3 min of play time.

As mentioned earlier, we developed two versions of the same game: in first-person interaction and in third-person. In first-person (Figure 6a), the user must be aware that her physical body (all of it, “from head to toes”) becomes a barrier for the virtual ball in the thin border of mixed reality where the body meets the screen (although there is absolutely no need to touch the screen). Therefore, by standing close to the screen the user may see how the virtual ball bounces against her body. On the other hand, in third-person paradigm (Figure 6b), the user stands away from the screen and sees her image inserted in the VE. Because the image of the user does not affect the squares, the game places the image of the user graphically behind these.

**FIGURE 6.** Ball game: (a) user playing in first-person paradigm, standing very close to the display and having no visual representation in the virtual environment (VE); (b) user playing in third-person paradigm, standing away from the display and having his image in the VE (behind the squares, in shades of gray) on the right-hand side of the display. (Color figure available online.)
The user must move to place her image in the VE such that the ball bounces against it and moves in the desired direction.

**Athletics Game**

Athletics game (Figure 7) is a game where the users must run through a virtual path represented in 3D. Along the path, the user finds obstacles (trees) and rewards (diamonds). The goal of the game is to move along the path and obtain as many diamonds as possible in 3 min of play.

Again we have two versions of the game. In first-person (Figure 7a), the user stands up close to the display and there is no visual representation of her body. In third-person (Figure 7b), the user stands away from the display and can see her visual representation in the VE.

Users can control the virtual player using body movements and gestures:

- **Run:** The user must run in place standing in front of the screen. The number of oscillations per minute of the center of gravity of the user determines the speed (see upcoming Physical Activity section). The higher the number of oscillations, the faster the speed.
- **Turn:** The user may turn right or left while running by extending her arm outward to the right or left, respectively. The difference between the right and left extension of the arms determines the speed of rotation of the virtual player. The larger the difference of the extension of each arm, the faster the rotation.
- **Lateral movement:** Physical lateral movements of the user are mapped onto virtual lateral movements to allow for rapid collision avoidance.

**FIGURE 7.** Athletics game: (a) user playing in first-person paradigm, standing very close to the display and having no visual representation in the virtual environment (VE); (b) user playing in third-person paradigm, standing away from the display and having his image in the VE (here only a dark rectangle may be appreciated) at the center of the display. (Color figure available online.)
5.3. Definition of the Hypotheses

Our hypotheses were based on the design decisions of the MEDIATE project (Parés et al., 2006; Parés et al., 2005) and on our informal observations as described before. In other words, we wanted to find whether there were any differences in the way that users interacted in the two paradigms.

Therefore, we finally formulated the following hypotheses for the two games:

- **H1**: There are significant differences in the presented and reported sense of control while interacting in first-person paradigm with respect to interacting in third-person paradigm.

- **H2**: There are significant differences in the amount of physical activity while interacting in first-person paradigm with respect to interacting in third-person paradigm.

As explained earlier on, we also decided to formulate a third and fourth hypotheses to assess whether the size of the display influenced the experience and performance of the user in third-person. On the other hand, as pointed out earlier, it makes no sense to define first-person systems with small displays because we need to display visual stimuli at a human scale.

- **H3**: There are significant differences in terms of presented and reported sense of control between the game configurations in third-person on a large display and third-person TV, on a small TV-like display.

- **H4**: There are significant differences in terms of presented physical activity between the game configurations in third-person on a large display and third-person TV, on a small TV-like display.

5.4. Description of the Evaluated Parameters

The following are the parameters that we decided to quantitatively and qualitatively analyze to find differences between first- and third-person paradigms in both applications. These parameters were chosen as a result of our observations and of issues posed in previous work—such as the MEDIATE project—as described in the previous sections.

**Sense of Control**

We operationalized *sense of control* as both the objective and subjective degree of competence or mastering in playing the game.

- The subjective degree of sense of control was measured using a user questionnaire (please refer to section 5.6, Data-Gathering Methods).
- The objective degree of sense of control was based on the scores obtained in the game.
**Score.** The score should be indicative of the performance or competence of the user and we used it as a measure of the sense of control as previously stated. In the case of the Ball game, the score is obtained by calculating the percentage of squares that the user sweeps out of the virtual space in 3 min of play. In the Athletics game, the system calculates the number of diamonds the user obtains in 3 min of play.

**Physical Activity**

We suggested the possibility that each interaction paradigm could induce a different amount and type of physical activity on the user. We divided physical activity into two categories, namely, *Gesticulation* and *Physical displacement*.

**Gesticulation.** To measure gesticulation of the user we took into account (Figure SM1):

- Horizontal gesticulation: The differences in width of the bounding box of the captured image of the user through time.
- Vertical gesticulation: The vertical variations of the user’s center of gravity through time.

We checked the state of the bounding box and the center of gravity at 6 Hz because we found this frequency to be high enough to capture the gesticulations with sufficient fidelity without being flooded by data. For each play we calculated the sum of the differences of the width of the bounding box and the total variation of the vertical center of gravity for the 3 min of play.

**Physical Displacement**

We evaluated only the lateral displacement because these were the meaningful displacements in the mechanics of the games (Figure SM2). To measure the physical displacement we used the bounding box. We were interested in the lower left and lower right points: p and q, respectively (Figure SM2). The criterion of movement we took is the following: A movement is produced when p and q increase or decrease at the same time. The movement at time t is calculated as

\[
\text{movement}(t) = \min (|p(t) - p(t - 1)|, |q(t) - q(t - 1)|).
\]

In this case we also checked the state of the bounding box at 6 Hz. For each play we calculated the amount of displacement of the user during the 3 min of play.

---

*The SM figures are in the Supplementary Materials file; see the Notes section at the end of this article.*
5.5. Design of the Experiment

The study was based on finding how the different game configurations influenced the physical activity and the experience of the participants. This was done for each of the two games. In addition, we were also interested in the effect of the size of the visualization interface in the case of third-person interaction paradigm. Therefore, we used a $3 \times 2$ m ($\sim 10 \times 7$ ft) back-projected screen for first- and third-person, and a 21-in. computer monitor for the second configuration of third-person.

We analyzed 39 healthy subjects. There were 14 female and 25 male subjects within an age range of 12 to 65 years ($M = 30.13$). We imposed no further restrictions on the subjects. Figure 8 summarizes the experimental design which was based on a $1 \times 3$ factorial design.

We decided that the best strategy for the study was to use repeated measures, that is, test each subject in all three conditions in both games. Hence, for each trial each user had to do six tests. We randomized the order of the different tests within each user trial to counterbalance the order effect when using the repeated measures method. A user would always alternate games in her tests, and her first game would always be alternated with respect to the previous user (Figure SM3).

The users were situated in a predefined position at the beginning of each test which depended on the interaction paradigm. In first-person (FP) paradigm, they were placed at 40 cm in front of the display. In third-person (TP) and TP-TV paradigm, they were placed 320 cm away from the display. From that moment on, they were allowed to move freely around the physical space during play. Each full trial lasted for approximately 25 min.

### FIGURE 8. Summary of the experimental design undertaken with two Artificial Reality games from two extreme categories, played by users in two interaction paradigms, with one of them presenting two modes.

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>FP: First-person (large display) (Figure 4)</th>
<th>TP: Third-person (large display) (Figure 3)</th>
<th>TP-TV: Third-person (small display) (Figure 5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ball Game Category: Control of Objects</td>
<td>Dependent variables:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Sense of Control</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>o Score</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>o Declared sense of control</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Physical activity:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>o Gesticulation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>o Displacements of the body</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Athletics Game Category: Navigation through VE</td>
<td>Dependent variables:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Sense of Control</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>o Score</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>o Declared sense of control.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Physical activity:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>o Gesticulation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>o Displacements of the body</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. VE = virtual environment.*
5.6. Data-Gathering Methods

**Still Camera.** Cameras situated in front of the user, for TP and TP-TV game configurations, and behind the user, for FP game configuration, were used to calculate the physical activity parameters, these being the part of the objective quantitative measures of the experiment.

**System Log files.** These files contained the data from the physical activity parameters as well as the scores of each user experience used for the quantitative objective measure for sense of control.

**Questionnaires.** The questionnaires were used essentially to gather, on one hand, general information of the users and, on the other hand, the subjective perception of the users with respect to the parameters we have already described. Although there exist some questionnaires that are often used in VR to measure the experience of the user, such as the *presence questionnaire* (Witmer & Singer, 1998), the notion of presence has been contested by several researchers and in a number of ways, providing sufficient doubts to the validity of the concept and of the questionnaires (Ellis, 1996; Slater, 1999, 2004; Slater & Wilbur, 1997). We therefore tend not to enter this controversy and work independent of the notion of presence. On the other hand, standard questionnaires include information that we would not use and do not include some parameters we needed such as the subjective level of effort exerted during the experience. Finally, because we were to apply repeated measures, the users would have to fill in many questionnaires and therefore we wanted to keep them as short as possible.

The first set of questions was designed to categorize the user according to age, gender, frequency of playing videogames, and frequency of doing physical exercise. These questions were answered at the end of each trial.

The second set of questions had to be answered at the end of each test. These questions were designed to measure, from a subjective point of view of the user, the level of effort exerted during the task, the sense of control and the degree of enjoyment of the experience (essentially through Likert scales).

**Observations and Verbal Reports of User Activity.** We took notes of our observations of the users during play and their verbal reports during and after play, as a relevant source of information to complete the analysis. This information not only completed the rest of data gathered but also helped us better understand why some events occurred and what were the results on our hypotheses.

6. RESULTS

With the gathered data for each variable and game configuration, our objective was to compare the three distributions for each variable to see their relation in each game configuration (FP, TP, and TP-TV).
The steps we followed to obtain the results were as follows:

1. We checked for normal distribution for each set of data applying the Kolmogorov-Smirnov test.
2. For each variable we checked whether at least two of the means of the distributions obtained in each game configuration were significantly different. To do so we applied:
   a. Repeated measures analysis of variance (ANOVA) for normally distributed data.
   b. Friedman test for non-normally distributed data.
   In both cases we used a two-tailed approach because we are only testing for inequality.
3. In the cases in which we found significant differences between the means ($p < .01$ in ANOVA or Friedman test), we then compared the distributions pair to pair applying:
   a. Bonferroni post hoc for normally distributed data.
   b. Wilcoxon test for non-normally distributed data.

6.1. Results of the Ball Game

Sense of Control

**Scores.** Means: FP = 63.54 ($\sigma = 7.23$); TP = 70.44 ($\sigma = 11.62$); TP-TV = 68.9 ($\sigma = 7.69$) (Figure SM4$^5$)

The ANOVA showed that at least two means of the FP, TP, and TP-TV distributions significantly differ, $F(2) = 7.919, p < .001$.

Observing the box plot (Figure SM4), the largest overlap is found between the TP and TP-TV distributions. Bonferroni post hoc showed that the means of TP and TP-TV are not significantly different ($p < 1.000$).

On the other hand, there is not much overlap between the FP distribution and those from TP and TP-TV. Both median and mean of the FP distribution are smaller than their respective in TP and TP-TV. Moreover, Bonferroni post hoc showed that the mean of the scores in FP is significantly different from TP ($p < .003$) and TP-TV ($p < .004$).

**Declaration of Sense of Control.** Means, range (1 [*no sense of control at all*] … 5 [*maximum sense of control*]): FP = 3.00 ($\sigma = 0.86$); TP = 3.92 ($\sigma = 0.87$); TP-TV = 3.77 ($\sigma = 0.84$)

The relations between the distributions (Figure SM5) are analogous to those found in the distributions of the scores of the Ball game. The Friedman test showed

$^5$The prefix SM refers to the Supplementary Material document that may be found on the journal’s web site.
that at least two means of the FP, TP, and TP-TV distributions are significantly different \((p < .001)\). Moreover, the Wilcoxon test showed there are no significant differences between the means of the TP and TP-TV distributions \((p < .260)\) and that the mean from the FP is significantly different from the other two \((p < .001\) in both cases).

**Physical Activity**

**Physical Displacement.** Means: \(FP = 1,810 (\sigma = 619.21); TP = 2,669 (\sigma = 1059.24); TP-TV = 2,114 (\sigma = 769.27)\) (Figure SM6 left)

The ANOVA showed that at least two means of the FP, TP, and TP-TV distributions significantly differ, \(F(2) = 13.208, p < .001\).

Observing the three distributions in the box plot we can see there is an overlap in all of them. However, Bonferroni post hoc showed that the mean of the TP significantly differs from the FP \((p < .001)\) and the TP-TV \((p < .001)\).

**Gesticulation.** Means horizontal gesticulation: \(FP = 4,187 (\sigma = 1,214.41); TP = 5,562 (\sigma = 1,226.36); TP-TV = 5,518 (\sigma = 1,371.45)\) (Figure SM6 center)

Means vertical gesticulation: \(FP = 1,330 (\sigma = 461.34); TP = 1,949 (\sigma = 596.29); TP-TV = 1,825 (\sigma = 563.66)\) (Figure SM6 right)

In both types of gesticulation, the relations between the distributions are very similar. If we look at the box plot (Figure SM6), there is a high overlap between TP and TP-TV and almost no overlap between FP and the others. Moreover, in both vertical and horizontal gesticulation, the ANOVA showed that at least two means of the FP, TP, and TP-TV distributions significantly differ: horizontal gesticulation, \(F(2) = 20.248, p < .001\); vertical gesticulation, \(F(2) = 31.023, p < .001\).

In both types the mean of FP is the smallest one. Moreover, Bonferroni post hoc showed that for:

- Horizontal gesticulation:
  - The mean of FP is significantly different from TP \((p < .001)\) and TP-TV \((p < .001)\).
  - The mean of TP does not significantly differ from TP-TV \((p < 1.000)\).

- Vertical gesticulation:
  - The mean of FP is significantly different from TP \((p < .001)\) and TP-TV \((p < .001)\).
  - The mean of TP does not significantly differ from TP-TV \((p < .292)\).

**6.2. Results of the Athletics Game**

**Sense of Control**

**Scores.** Means: \(FP = 14.78 (\sigma = 8.71); TP = 16.24 (\sigma = 7.58); TP-TV = 16.76 (\sigma = 8.20)\) (Figure SM7)
The ANOVA showed that there are no pairs of means significantly different between the FP, TP, and TP-TV distributions, $F(2) = 1.484, p < .241$.

We can see in Figure SM7 that the means do not differ very much and all distributions in the box plot show a high overlap.

*Declaration of Sense of Control.* Means, range (1 [no sense of control at all] ... 5 [maximum sense of control]): FP $= 2.73$ ($\sigma = 1.10$); TP $= 3.14$ ($\sigma = 1.21$); TP-TV $= 3.19$ ($\sigma = 1.13$)

Looking at the means of the distributions (Figure SM8), it might seem that the participants experienced slightly less sense of control playing in the FP game configuration than in the other two configurations. However, the ANOVA showed that there are no pairs of means significantly different between the FP, TP and TP-TV distributions, $F(2) = 2.333, p < .104$.

**Physical Activity**

*Physical Displacement.* Means: FP $= 1,318$ ($\sigma = 355.90$); TP $= 1,293$ ($\sigma = 452.45$); TP-TV $= 1,296$ ($\sigma = 479.84$) (Figure SM9 left)

The means of the three distributions do not differ very much, and we can see in the box plot (Figure SM9 left) that there is a high overlap between the three distributions.

The ANOVA showed that there are no pairs of means significantly different between the FP, TP, and TP-TV distributions, $F(2) = 0.060, p < .942$.

*Gesticulation.* Means horizontal gesticulation: FP $= 7,133$ ($\sigma = 2,261.28$); TP $= 6,683$ ($\sigma = 1,744.28$); TP-TV $= 7,067$ ($\sigma = 1,618.00$) (Figure SM9 center)

Means vertical gesticulation: FP $= 2,040$ ($\sigma = 775.05$); TP $= 2,110$ ($\sigma = 846.73$); TP-TV $= 2,115$ ($\sigma = 766.52$) (Figure SM9 right)

In each type of gesticulation, the means of the distributions are very similar, and if we look at the box plot (Figure SM9) there is a high overlap between the three distributions. Moreover, in both vertical and horizontal gesticulation, the ANOVA showed there are no pairs of means significantly different between the FP, TP, and TP-TV distributions: horizontal gesticulation, $F(2) = 2.377, p < .108$; vertical gesticulation, $F(2) = 0.280, p < .758$.

**6.3. Summary of Results**

Figures 20 and 21 summarize the results presented in the previous sections, for both games and all the configurations.
6.4. Latency

One may argue that the latency or the time delay between the generation of an event and the resulting update of the game state, including the feedback to the users, could have influenced the users in their experience. Because of this, we assessed the delay of our system.

It is well known that high delays can influence the players although, in some cases, users may adapt to that delay. The adaptation to it might be easier when interacting in third-person than in first-person because in the former the users have a visual feedback of the delay. There are many studies about how delay affects the performance and the behavior of users, especially in collaborative tasks. Vaghi, Greenhalgh, and Benford (1999) performed a study of a collaborative virtual ball game where they provide evidence that strategic adaptations occur in delays ranging from 150 ms to 1,000 ms. Gergle, Kraut, and Fussell (2006) studied the impact of delayed visual information on the coordination of pairs in a collaborative physical task and they tried to typify the function that could describe the relationship between visual delay and task performance. They found that the range of

![FIGURE 9. Summary of the results of the experiments for the Ball Game.](image)

![FIGURE 10. Summary of the results of the experiments for the Athletics Game.](image)
delay that can be tolerated before a collaborative task performance begins to suffer depends on many factors such as the complexity of the task and the granularity of the interaction. Another example is the studies of first-person shooter games where delay above 100 ms can significantly reduce the overall user performance (Beigbeder et al., 2004). However, shooting is affected with even modest 75 to 100 ms of latency. During fast-pace interaction, smaller amounts of delay are more likely to cause problems. Simulator-like games, such as racing games, flight simulators, and sports games, are those where the delay affects most. Pantel and Wolf (2002) studied the impact of delay in a car racing simulator. They show that a delay up to 50 ms is uncritical. They state that with this delay the driving behavior is unmodified and it is hardly noticed. A delay of 100 ms can be noticed but hardly optically seen (acceptable if no high demands with respect to realism is needed).

To calculate the latency of our system we took into account the frame rate of the camera (87 frames per s), the speed of the refresh of the visual output (60 fps), and the update rate of the system (30 fps for the Ball game and 27 fps for the Athletics game). The maximum delay (worst case scenario) for the Ball game and Athletics game was 62 ms and 66 ms, respectively. These delays could be acceptable even for the simulators and racing games. Therefore, both applications run with a range of delay unnoticeable by the users, and we can claim that they needed no strategic adaptation to the latency. Moreover, no users verbally reported the presence of latency or any related discomfort.

7. DISCUSSION

The two games showed different behaviors of the users during their interaction with respect to the two paradigms. On the Ball game (from the Control of Objects category), we have found significant differences between paradigms, whereas in the Athletics game (from the Navigation through VE category) we have found none. However, this is interesting because the categories of applications that we have defined seem to be meaningful with respect to interaction, and this opens new doors for future research.

Let us now analyze the results obtained in each of the games in relation to the hypotheses we defined and the referenced literature.

7.1. Sense of Control

Considering the two variables, subjective feeling of sense of control and scores, we can claim for the Ball game application that

- There was a statistically significant difference in the means of the sense of control variables between first-person and the other two configurations and that
in first-person, in absolute values, sense of control was lower than in the other two cases (Figure 9).

Therefore, we have evidence to think that our hypothesis can be accepted showing that the two paradigms affect the notion of sense of control in the users. However, contrary to what we had thought at the beginning of this research, first-person does not seem to provide a higher sense of control but actually a lower sense. This must be confirmed in the future with other applications in the same category of Control of Objects, but it can already be an informative guideline for the design of future applications.

For the Athletics game, quite surprisingly, we have obtained no significant differences in any of the analyzed parameters. Therefore, we cannot reject our null hypothesis (Figure 10). This could indicate that in this category (Navigation through VE) the paradigms do not affect the interaction of the users. However, we do not have sufficient evidence to confirm this. This is surprising because the means of scores and declared sense of control were apparently clearly lower in first-person than in third-person. This would in principle indicate a similar trend as in the Ball game.

Oshita (2006) stated that for applications in which users can directly control the full body motion of a character in virtual environments or when users need to sense the movements and interactions with other objects, the third-person paradigm is more suitable. Although Oshita worked in 3D motion capture, there is an interesting relationship between what he claimed and our results.

If we analyze both our games, in the Ball game, users need to perceive a high correlation between their body actions and the virtual ball in order to control it. On the other hand, in the Athletics game, users do not need such a tight interaction correlation with the different virtual elements. Verbal reports from the users and our observations suggest that users playing in first-person in the Ball game partially “lost” this tight correlation. Some players in first-person tried to interact only with their hands and forgot that the ball could collide with the rest of their physical body. This is one of the reasons we believe that users experienced less sense of control in first-person than in third-person. This “lost” interaction correlation may have occurred because people are not used to interacting with the virtual elements directly with their physical body. Apparently, the “thin layer” that the user must deal with in first-person and that provides the Mixed Reality experience presents some cognitive difficulties to the users in understanding the coplaced relationship between physical and virtual worlds.

On the other hand, having a virtual representation of themselves or an avatar “in” the VE, as the third-person paradigm offers, seems to indeed help users understand how they may affect the virtual objects. Moreover, in the Ball game, the position of all the extremities of the user is very important, whereas in the Athletics game, the movements of the body are more important than the exact position of the extremities. Therefore, we believe that having a visual feedback of the user in the third-person Ball game was useful for the participants to correctly interact and would explain the relations we found between the distributions coming from each of the configurations in terms of sense of control.
7.2. Physical Activity

Considering the variables for physical activity, we can claim that in the Ball game:

- There was a statistically significant difference between the means of physical displacement in first-person and third-person and that the participants tended to be less active in the first-person configuration than in third-person (Figure 9).
- There was also a statistically significant difference between the means of gesticulation in first-person and third-person and the participants in first-person tended to gesticulate less than in the other configurations (Figure 9).

Hence, we have evidence to support that our alternative hypothesis can be accepted, showing that in this game the two paradigms affect the physical activity of the users. Moreover, the differences between means in the two paradigms, apart from being statistically significant in this game, they are numerically substantial. Comparing the means in absolute values, we see that when using third-person paradigm we obtain an important increase with respect to first-person, specifically: ~47% increase for physical displacement, ~33% increase for horizontal gesticulation, and ~46% increase for vertical gesticulation.

Similarly to the previous case of sense of control, and contrary to what we had initially thought, first-person seems to generate less physical activity than third-person. This must be confirmed in the future with other applications from this same category.

In the case of the Athletics game, again it is surprising that we have obtained no significant differences in any of the physical activity parameters (Figure 10). Therefore, we cannot reject our null hypothesis. In this case, though, the comparison of the means is not as clear as for sense of control. Hence, we will have to further our research in this area.

Physical activity is also important in leisure activities, and therefore the results just presented can guide in the design of more active and engaging experiences. As introduced previously, Warren’s (2003) work on expressiveness can also be a referent in this differentiation between first-person and third-person paradigms. “Expressiveness” was not an object of study of our work. However, the greater gesticulation and displacement found in third-person could suggest that this paradigm offers greater expressiveness than first-person. This opens the door to experimenting with the expressive nature of these paradigms and application categories to better motivate, engage, and help users in their experiences. Perhaps some applications can be enhanced in expressive nature by using the first-person paradigm and others by third-person. This is left for future work.

7.3. Large versus Small Display in Third-Person Paradigm

Analyzing now our third and fourth hypotheses, that is, those that refer to the difference between third-person deployed on a large projection screen (TP) and that deployed on a small TV-like screen (TP-TV), we must be cautious because our results are not clear.
The results tell us there are no significant differences between the TP and TP-TV configurations in terms of presented or reported sense of control, neither in the Ball game nor in the Athletics game (Figures 20 and 21). This could mean that the change in size of the display does not make a difference in the interaction of the users. However, we do not have sufficient evidence to say so, and therefore we will have to proceed with new experiments in the future to better analyze this.

In the case of physical activity, the hypothesis cannot be rejected or accepted either. For the Ball game, the results tell us that there is a significant difference for physical displacement (Figure 9; and hence we could in principle reject its null hypothesis) but, on the other hand, there is no significant difference for gesticulation. In the case of the Athletics game, there differences are not significant in any variable.

Therefore, although this could indicate that there is no need to have large-scale installations for ArtR experiences—that use large projection screens—in applications such as for rehabilitation, the analysis of the change in size of display for the TP paradigm demands further attention in the future in these and the rest of categories of applications.

7.4. Impact on Rehabilitation

The results on sense of control are interesting to contrast with the working hypotheses of the MEDIATE project that took its researchers to choose a first-person paradigm. They wanted to avoid two potential negative results: that children with autism would not understand or recognize their silhouette in third-person paradigm and that, even if they recognized it, they would not understand that they could interact with the virtual objects through that silhouette. Because the main goal of this project was to provide children with autism with as high a sense of control as possible, they thought that these potential risks could hinder this sense of control. Hence, they decided to use a first-person paradigm thinking that by not having the disturbing third element (the silhouette) and seeing their own physical body generating the changes in the virtual world, children with autism would more clearly understand they were the cause of the changes and hence acquire a higher sense of control.

The fact is that according to the psychologists in the team, children did achieve a good sense of control. Now, contrasting this with the lower sense of control found in the present experiments in first-person in the Ball game, it would seem reasonable to think that the researchers in MEDIATE would have achieved even better results with a third-person paradigm. However, we see no significant difference in the Athletics game, and therefore it becomes necessary to discriminate between the types of applications. MEDIATE was based on an application that falls in the Real-Time Generation from Gestures category from our categorization. This category cannot be easily compared to the two we have investigated in this article. Therefore, it becomes an important future work to analyze this type of applications under the same controlled conditions as we have done here to be able to better understand what the differences are between first- and third-person. Another important point would
be to try to analyze the potential differences with children with autism to understand how their specific intellectual disorder can possibly have an influence in how the experience is being understood by them.

Regarding physical activity, this was not an issue in the MEDIATE project, although the researchers did want to foster the children to move about in the interactive multimodal space to avoid passive attitudes.

The results in the present study are also interesting to contrast with other work done in rehabilitation. Most of the applications that we have seen are being used in this field belong to the Intercept Objects category. For example, Kizony and colleagues used the game of Birds and Balls and the game of Soccer from the Gesture Xtreme VR system, and the game Kung Foo and the game Wishy-Washy from the Eye Toy system, where the first three belong to the Intercept Objects category and the fourth belongs to the Activation Zones category (Kizony, Katz, & Weiss, 2004; Kizony et al., 2004). Brooks and Petersson (2005) used the game Keep Up from Eye Toy that belongs to the Control Objects category. Finally, Flynn et al. (2007) used mostly the following games from the Eye Toy: Knockout (Intercept Objects), Pool (Control Objects), DIY (Intercept Objects), and Colors (Activation Zones).

Most of these games are found in the far extreme of the Reach Out axis of our categorization, and all have chosen the third-person paradigm. Although most are indeed different from the categories we have investigated, the Intercept Objects category seems very close to the Control Objects that we have examined through the Ball game. This could give an idea that these rehabilitation approaches have a good chance of having better results in the third-person paradigm—which they have used—than if they had chosen a first-person paradigm. In a sense these works and experiments have been validated by our results. Nonetheless, it remains an important task to specifically analyze the two categories of Intercept Objects and Activation Zones in the same manner as we have done in the present study.

On the other hand, for these rehabilitation applications physical activity can be crucial, especially for those in motor rehabilitation. For example, in applications that help in the rehabilitation process of a parietic arm of a person who has suffered from stroke. In these cases it is clear that the knowledge gained in the present study can better help them in designing experiences for their rehabilitation strategies as well as evaluate them. Once we have analyzed all the types of applications in our categorization, researchers and developers will have a comprehensive view on how to fine-tune their applications for their specific goals.

8. CONCLUSIONS

The present work has, for the first time, analyzed, evaluated, and discussed the potential differences of ArtR experiences. This has been done by formally defining the two interaction paradigms that are possible and a categorization of types of applications. Formal controlled experiments have been undertaken to quantitatively and qualitatively evaluate the differences.
This work has opened a research field that will provide very important information for all application areas of ArtR enhancing different properties that can be crucial for their users. One of these areas is rehabilitation, which can have a very significant impact on society.

As we suspected before we started this research, in the evaluated parameters and interactive tasks the differences between interaction paradigms may vary according to the category of application chosen. Quite different behaviors have been experimentally observed in the interaction of the users in the categories of Control of Objects (with the Ball game) and Navigation through VE (with the Athletics game). This suggests that the categories we have defined not only reflect clear conceptual criteria but also are meaningful in practice. We must now extend the experiments to the other categories to explore the full space of possibilities.

Having experimentally found that in general the third-person paradigm generates more physical activity and sense of control in the users than first-person, it partially reinforces previous experiences (in all fields, but especially) in rehabilitation, as most of these experiences have been developed using the third-person paradigm. This is especially true also because most of the rehabilitation applications that have used ArtR are classified under the Intercept Objects category that is quite close to Control of Objects, this latter being the category that has provided stronger evidence of the difference between paradigms.

The findings in this article can also be seen as an initial set of design guidelines for future design and development of experiences. In other words, from the differences we have found between the two application categories that we have studied, with respect to the two paradigms, we could now better predict how a new application would influence the sense of control and physical activity of the users. However, we must now develop applications for the rest of the defined categories and repeat the experiments to have a broader and deeper understanding of the interaction paradigms in those categories. Once we have explored the rest of categories, we will be able to better guide designers.

We hope that our contribution in the present study will open many new doors to research. We would like to encourage the community to help us move forward in this research.

---

**NOTES**

*Authors’ Present Addresses.* Narcís Parés, Cognitive Media Technologies Group, ICT Department, Universitat Pompeu Fabra, c. Roc Boronat, 138, 08018 Barcelona, Spain. E-mail: narcis.pares@upf.edu. David Altimira, Cognitive Media Technologies Group, ICT Department, Universitat Pompeu Fabra, c. Roc Boronat, 138, 08018 Barcelona, Spain. E-mail: davidaltimira@gmail.com.

*Supplementary Materials.* A supplementary file is available for this article. This supplementary material provides diagrams to better describe our operationalization of the measured parameters of the user’s body, a diagram to describe the randomization of the user trials and
all the box plots of our statistical results for a comprehensive description of the statistics. Each figure has a short text describing the section of the article to which it is related. Free access to this file is available at the online edition of this article on the HCI publisher’s website (http://www.tandfonline.com/loi/hhci20).

**HCI Editorial Record.** First manuscript received August 11, 2010. Revisions received February 14, 2011, and June 4, 2011. Accepted by Rob Jacob. Final version received December 22, 2011. — Editor

**REFERENCES**


Luminision Ltd. http://www.luminvision.co.uk/interactive.htm


Sony PlayStation Eye: http://es.playstation.com/ps3/

Sony PlayStation Eye-Toy: http://www.eyetoy.com


Vertigo Systems GMBH. http://www.vertigo-systems.com/startseite.html?&L=1#ul


