Spatial Presence in Real and Remote Immersive Environments and the Effect of Multisensory Stimulation

Abstract

This article presents a user experiment that assesses the feeling of spatial presence, defined as the sense of “being there” in both a real and a remote environment (respectively the so-called “natural presence” and “telepresence”). Twenty-eight participants performed a 3D-pointing task while being either physically located in a real office or remotely transported by a teleoperation system. The evaluation also included the effect of combining audio and visual rendering. Spatial presence and its components were evaluated using the ITC-SOPI questionnaire (Lessiter, Freeman, Keogh, & Davidoff, 2001). In addition, objective metrics based on user performance and behavioral indicators were logged. Results indicate that participants experienced a higher sense of spatial presence in the remote environment (hyper-presence), and a higher ecological validity. In contrast, objective metrics prove higher in the real environment, which highlights the absence of correlation between spatial presence and the objective metrics used in the experiment. Moreover, results show the benefit of adding audio rendering in both environments to increase the sense of spatial presence, the performance of participants, and their engagement during the task.

1 Introduction

Spatial presence is commonly defined as the sense of “being there” (Biocca, 1997). The term “there” refers either to a virtual place or a real remote place (Wirth et al., 2007). Therefore, spatial presence encompasses the ability of users to experience a sense of presence in any environment in which they are transported. Namely, (1) if the environment is real and non-mediated, the user experiences a natural presence (Biocca, Burgoon, Harms, & Stoner, 2001); (2) if it is real and mediated, the user experiences a remote presence or “telepresence” (Steuer, 1992); and (3) if the environment is virtual, i.e., computer generated, the user experiences a virtual presence (Sheridan, 1992). This classification was proposed by Zhao (2002), who emphasized the importance of comparing the sense of presence between the different types of environments in order to better understand this phenomenon.

Thus, previous studies have focused on the evaluation of the sense of virtual presence. Usoh, Catena, Arman, and Slater (2000) carried out an experiment in which 10 participants had to find a box. The participants were located either in
Researchers in other areas of virtual reality (VR) have shown that the users’ actions during an experiment can be influenced by the perceptions associated with them (Slater et al., 2006). For example, a “perceived” agency can matter more during interactions with virtual humans than a “real” agency (Nowak & Biocca, 2003). Therefore, this awareness of a real place is essential. Users might be aware that all their actions have a tangible impact on the real world, which in turn may influence their experience of spatial presence. Consequently, an environment is considered remote as soon as it satisfies two criteria:

Criterion 1: It allows the remote perception of a real place, either through:

(a) The mediation of sensory stimulation of the space depending on its nature and the activity, and when the technologies afford it (e.g., manipulating objects with haptic force feedback gloves (Fernando, Furukawa, Minamiza, & Tachi, 2013)).

(b) The virtual representation of the space that may be highly realistic (3D reconstruction of the scene and/or physical modeling of real properties), or just symbolic depending on the focus (learning, task efficiency, entertainment (Mestre, Maíano, Dagonneau, & Mercier, 2011), etc.).

(c) Combination of (1a) and (1b). This case is more common in augmented reality environments when the goal is to provide remote assistance (e.g., surgical operations in Shenai et al., 2011), or when there is no real-time data flowing (e.g., space operations in NASA Robonaut: Ambrose et al., 2000), or when the nature of the space makes it difficult to perceive (e.g., submarine and subterranean operations in Remote Operated Vehicles: Lichiardopol, 2007).

Criterion 2: Users’ actions must have real consequences on the real place, with varying
immediacy according to the constraints of teleoperation systems. In addition, users have to be aware of this real impact.

If the first criterion is usually met, the second is often overlooked, creating confusion between remote and virtual environments. For instance, in McCall, O’Neill, and Carroll’s (2004) study on presence, participants were either physically located in a real garden or in a computer-generated environment representing the same garden (viewed inside an HMD). This environment was generated through the rendering of photographic images of the real garden (criterion 1b met). However, the participants had no impact on the actual garden because the captured images were processed before the experiment with no real-time rendering (criterion 2 not met). Consequently, the photorealistic environment was virtual, not remote.

Another reason that may explain the lack of studies on spatial presence in remote environments is the low level of sensory fidelity that systems could provide at that time. The sensory fidelity, also referred to as immersion (Slater, Linakis, Usoh, & Kooper, 1996), is determined by the quality, the richness, and the consistency of each sensory channel used by the system (Schubert, Friedmann, & Regenbrecht, 1999a). Its relationship with spatial presence has been a topic of considerable theoretical discussion and empirical investigations for over 25 years. The reader is referred to Cummings and Bailenson’s paper (2016) for a detailed review of the effect of immersive system technology on presence.

Thus, it is now generally accepted that immersion plays a significant role in the emergence of spatial presence through the construction of a mental representation of the environment in which users are located (Bowman & McMahan, 2007; Wirth et al., 2007). In other words, whether in virtual or remote environments, systems of higher immersive quality elicit a greater sense of presence. Yet, the impact of immersion is more valuable when assessing spatial presence in remote environments. This is because of the importance of providing high quality natural sensory channels to allow users to experience an immersive situation similar to that in a real (non-remote) environment (Sheridan, 1992). However, tools were missing that enable researchers to develop remote environments facilitating a high sensory fidelity experience. With recent technological advances, including visual quality of HMDs, sound spatialization, and overall system latency reduction, researchers have been able to develop modern remote environments with high sensory fidelity. Such environments could provide a high perception of the real world. Therefore, assessing spatial presence in real and remote highly immersive environments could be very interesting. By comparing them, one can isolate the effect of modern VR technology on the sense of presence and truly assess the impact of immersion.

Hence, the main goal of the present study is to assess the user’s feeling of spatial presence in a remote immersive environment (presented with an HMD, following the criteria (1a) and (2)) and compare it with the sense of spatial presence in a real environment. Also, the study aimed to evaluate the effect of immersion by comparing the environments in almost similar immersive (visual and auditory) conditions.

The remainder of this article is divided into three main sections. The first section provides an overview of related work highlighting the research focus of the article. The second section describes the design of the user experiment, followed by the evaluation protocol and the results. The third and last section concludes with an interpretation of these results and proposes new perspectives.

2 Related Work

2.1 Spatial Presence in Remote Environments: Telepresence

Spatial presence in remote environments or remote presence is about the user’s perception of a real environment and real consequences when it is mediated and/or virtually represented by means of technologies. The term takes its origin from “telepresence” (Steuer, 1992) which was first employed in 1980 in the context of teleoperation systems, designating remote task manipulation as the “illusion of being transported via
telecommunication systems to a real, physical location, which could then be experienced synchronously from afar” (Minsky, 1980). Later, Akin, Minsky, Thiel, and Kurtzman (1983, p. 3) used telepresence to describe the specific situation when “the manipulators have the dexterity to allow the operator to perform normal human functions” and argued that the feeling of presence is reached when “the operator receives sufficient quantity and quality of sensory feedback.” Since then, telepresence has been of particular importance to robotics researchers.

Many studies highlighted its value in the transmission of information, especially concerning the spatial perception of remote sites (Sheridan, 1992; Schloerb, 1995). Also, they emphasized the importance of providing teleoperators with spatial cues to enhance their sense of telepresence. It is within this context that Muhlbach, Bocker, and Prussog (1995) first coined the term “spatial presence” in remote environments. They assumed that telepresence was affected by two variables: (1) communicative presence, defined as the degree to which reciprocal signals (verbal and nonverbal) are transmitted, and (2) spatial presence, defined as the degree to which spatial cues about the remote site are transmitted. They conducted an experiment to investigate how spatial (auditory and visual) cues affect the impression of spatial presence and telepresence in a video conference system. They collected questionnaire data and concluded that spatial cues enhance spatial presence as well as telepresence.

Similarly, in the context of virtual environments, Biocca (1997) introduced the term “physical presence” and defined it as the sense of being physically transported in virtual or mediated environments. This definition was quickly adopted in the VR community as referring to spatial presence in virtual environments (Schubert et al., 1999a).

Consequently, spatial presence is now referring to the sense of “being there” in both virtual and remote environments (Regenbrecht & Schubert, 2002; Schuemie, Van Der Straaten, Krijn, & Van Der Mast, 2001; Wirth et al., 2007). Furthermore, telepresence, which was initially the origin of the concept, is now used to define spatial presence in the specific context of remote environments as remote spatial presence.

### 2.2 Assessing the Degree of Spatial Presence

Research in remote environments has mainly focused on optimizing the technological aspects of teleoperation systems to improve user performance (Yamaashi, Cooperstock, Narine, & Buxton, 1996). These systems were designed in the robotic field (e.g., space operations, Ambrose et al., 2000), the medical field (e.g., surgical applications, Taylor, Menciassi, Fichtinger, & Dario, 2016), and the industrial domain (e.g., assembly operations, Radi et al., 2010) to name a few. The evaluation of spatial presence as a subjective feeling was instead the focus of researchers in virtual environments. Indeed, a large part of these studies was concerned about identifying the factors influencing presence, its usefulness in VR applications, and especially, the tools for assessing this feeling (Mestre, 2015). Therefore, the measures proposed in the literature are virtual context-related.

Moreover, the way presence is assessed depends on the framework used to define it (Schuemie et al., 2001). As the notion of spatial presence has been redefined multiple times, the methods of evaluation have also changed. Consequently, several methods to measure presence in the context of virtual environments exist. These methods can be divided into subjective (mostly using questionnaires) and objective measures, with post-experiment questionnaires being the most common approach for assessment.

The following subsections summarize the most widely used methods and measurements; for more details, see Lombard, Biocca, Freeman, IJsselsteijn, and Schaevitz (2015, pp. 139–185). The idea of this article is to use the same measurements to assess spatial presence in remote environments.

#### 2.2.1 Subjective Questionnaires

Many presence questionnaires have been proposed since the early days of presence research. Slater, Usoh, and Chrysanthou (1995) proposed a one-dimensional questionnaire in which the presence score is taken as the number of answers that have a high score. Kim and Biocca (1997) designed a questionnaire based on their metaphor of transportation comprising two dimensions: (i) arrival, being
present in the mediated environment, and (ii) departure, not being present in the unmediated environment.

Shortly after, Witmer and Singer (1998) designed the Presence Questionnaire (PQ), based on three factors: (i) involvement, (ii) behavioral fidelity of interactions and control of locomotion and (iii) the user’s ability to concentrate on the tasks. Schubert, Friedmann, and Regenbrecht (1999b) constructed the Igroup Presence Questionnaire (IPQ) by combining both the Slater et al.’s (1995) and Kim and Biocca’s questionnaire. It is based on eight factors. Three of them are directly concerned with presence: (1) the sense of spatial presence, (2) the involvement into the environment, and (3) the sense of reality attributed to the virtual space. The others are considered as immersion and interaction variables that may influence presence. Usoh et al. (2000) developed the SUS (Slater-Usoh-Steed) questionnaire based on several questions all variations on one of three themes: (1) the sense of being in the virtual environment, (2) the extent to which the environment becomes the dominant reality, and (3) the extent to which the environment is remembered as a “place.”

However, one of the most validated questionnaires is the Independent Television Commission Sense of Presence Inventory (ITC-SOPI) by Lessister et al. (2001). It has been widely used within the literature and was shown to produce reliable results (Busch et al., 2014; Gorini, Capideville, De Leo, Mantovani, & Riva, 2011; Tang, Biocca, & Lim, 2004; Usoh et al., 2000). It is based on four factors: (1) the sense of spatial presence, (2) the user’s engagement, (3) the ecological validity of the environment, and (4) the negative effects (such as cybersickness). This questionnaire has the advantage to be quite easy to administer and score (Lombard et al., 2015).

Despite having proved their usefulness for measuring the sense of presence, questionnaires are subjective measures highly dependent on users’ experiences. In addition, they do not provide a continuous measurement of presence during the experiment because users complete questionnaires at the end of their experience, so as not to cause breaks that reduce the sense of presence. Therefore, objective non-invasive metrics have been designed to assess the sense of presence continuously during the experiment. These metrics are generally combined with questionnaires.

2.2.2 Physiological Measures. In parallel to subjective questionnaires, an approach based on physiological indicators sought to establish reliable and validated measures for spatial presence. Thus, studies on the reliability of physiological indicators to measure presence such as changes in Heart Rate (HR) (Mechan, Insko, Whitton, & Brooks, 2002) and skin temperature and conductance-based on Electrodermal Activity (EDA) (Wiederhold et al., 2001; Wiederhold, Gevirtz, & Wiederhold, 1998) provided promising results. For example, Meehan and colleagues (Meehan, Razzaque, Insko, Whitton, & Brooks, 2005) showed that in a stressful virtual environment depicting a pit room, changes in HR correlated positively with self-reported presence.

However, these measurements require a baseline comparison for each participant, which means a considerable effort in study design. Besides, it has been shown that additional equipment to measure physiological responses, such as Electroencephalography (EEG) to measure brain responses (Baumgartner, Valko, Esslen, & Jäncke, 2006), can be a cause of breaks in presence (Brogni, Slater, & Steed, 2003). Moreover, this equipment is more efficient when participants do not move (Nalivaiko, Davis, Blackmore, Vakulin, & Nesbitt, 2015), which reduces the scope of possible experiments.

Thus, physiological indicators exist that could become potentially reliable measures for presence (for more details, see Lombard et al., 2015, pp. 150–185). Nevertheless, the correlation between them and the sense of presence is not yet firmly established (Freeman, Lessiter, Pugh, & Keogh, 2005). A common approach to achieve this goal is to compare results from these kinds of measures with outcomes from presence questionnaires (Bracken, Pettey, & Wu, 2014; Freeman, Avons, Meddis, Pearson, & IJsselsteijn, 2000).

2.2.3 Behavioral Measures. Researchers also investigated the relationship of behavioral indicators with presence (Blascovich et al., 2002; Slater, 2003). In a survey,
Schuemie et al. (2001) showed that virtual experiences could evoke the same reactions as real experiences.

These reactions can be represented by postural adjustments and body movements. For example, Freeman et al. (2000) measured postural responses (lateral movement) to a video sequence filmed from the hood of a car traversing a rally track. Results showed no significant correlation between the postural response and presence. However, the degree of presence was determined by a single question. In addition, presence may have been biased by the novelty of the technology used (stereoscopic display consisting of two-color monitors equipped with polarized filters). They concluded that such postural responses could be useful in corroborating subjective ratings of presence, but not in replacing them.

Similarly, Lepecq, Bringoux, Pergandi, Coyle, and Mestre (2009) measured body movement during an experiment in which participants had to walk through either a virtual or a real aperture. Results showed that participants swiveled their bodies similarly in both real and virtual situations. Moreover, the body rotation was a function of aperture and shoulder width.

Bailenson, Blascovich, Beall, and Loomis (2003) conducted two experiments. In the first experiment, participants were standing while a walking virtual human approached them and invaded their personal and body space. In the second experiment, the virtual human was stationary, while the participants had to approach him. The metric used was participants’ movements and posture changes as the virtual human invaded their space (how far participants moved away from the virtual human in the first experiment), or as they were approaching it (in the second experiment). More precisely, researchers examined the degree of interpersonal distance that participants maintained between themselves and virtual humans in both situations. Results from both experiments showed that participants exhibited patterns of interpersonal distance behavior relative to virtual humans similar to those from research with actual humans. However, these measurements are more related to the user’s sense of social presence (the sense of “being with,” Biocca, 1997), which is beyond the scope of this article. For more information about interpersonal space in virtual environments, see Lachini et al. (2016).

The reactions evoked by the environment can also be reflex responses to virtual danger (Slater, 2009). Maïano, Therme, and Mestre (2011) found a correlation between aversive stimuli (fire, smokescreen, and warning alarm) appearing in a virtual environment, the degree of self-reported anxiety of participants navigating in this environment, and the way they moved away from “danger.”

Other measures that have been studied as potential behavioral indicators are the attention-based measures. In their studies, Bracken, Petey, and Wu (2011, 2014) investigated the relationship between participants’ sense of presence measured by a questionnaire and their reaction time to some audio distraction cues while watching a movie. A significant correlation was found between this reaction time and two subcomponents of presence: immersion and attention.

Therefore, observational methods to measure presence are promising. Yet, more investigation is still needed to establish reliable measures based on people’s reaction and behavior.

2.2.4 Performance Measures. Close to behavioral measures (and, sometimes even considered as a category of the latter), performance measures may provide an objective indicator of spatial presence in immersive environments (Cummings & Bailenson, 2016; Laha, Sensharma, Schiffbauer, & Bowman, 2012). However, care must be taken to compare environments with the same degree of similarity.

Thus, the degree of similarity, also referred to as “the fidelity of the environment” (Dalley, Robinson, Weller, & Caldwell, 2004) can strongly affect performance. For example, McMahan, Gorton, Gresock, McConnell, & Bowman (2006) found that object manipulation can be successfully performed with lower fidelity VR technologies, such as less costly displays, with no loss of efficiency. In particular, it is essential to use similar interactive techniques across environments as the way people interact in the environment can influence their performance (Carvalho, Bessa, & Magalhães, 2014; McMahan et al., 2006). Indeed, some interaction paradigms make better use of humans’ innate skills rather than imposing...
new learning processes (such natural gesture-based interaction), and others rely on acquired skills and previous users’ experiences to improve performances (such as joystick interaction in games), without affecting the sense of presence.

In addition, different types of tasks can be used to evaluate performance. The most common types are the memory/learning tasks and the manipulation tasks (for a taxonomy, see Poupyrev, Weghorst, Billinghurst, & Ichikawa, 1997). Depending on the task, some interaction paradigms may be more efficient, which may increase task performance.

Furthermore, the relationship between presence and performance may also be highly dependent on the measures of performance used (Picciano, 2002). Many studies investigated this relationship with different measures of performance. For example, Stanney, Kingdon, Graeber, and Kennedy (2002) conducted a study in which participants located in a virtual environment had to complete the maximum amount of basic tasks. Results showed that best performances—measured by the number of tasks completed and the time needed to complete each task—resulted in a higher sense of presence. Youngblut and Huie (2003) found a significant relationship between presence measured by the SUS questionnaire and user’s performance in a learning task.

In contrast, Mania and Chalmers’ study (2004) showed that the sense of presence was not correlated with task performance measured by the ability to acquire knowledge during a lecture. More recently, Stevens and Kincaid (2015) studied the relationship between presence and performance in virtual simulation-based training. Participants had to destroy as many virtual enemy forces as possible. The time taken to complete the task was used as an objective performance measure. Results indicated a moderate relationship between the performance measure and presence.

Thus, studies on presence and performance provided promising results. With technological advances in high fidelity user-interfaces in terms of interaction, the difference between real and remote environments is narrowing, leading researchers to further explore the relationship between presence in remote sites and task performance.

2.3 Multisensory Stimulation Effect on Spatial Presence

Another particularly relevant aspect of presence research for immersive environments is the evaluation of potential positive effects of multisensory stimulation. For instance, Mania and Chalmers (2001) compared the sense of presence—using Slater’s questionnaire—between a real environment, an immersive (visual and audio) virtual environment, and an audio-only virtual environment. Participants \( n = 18 \) reported higher levels of presence for the real environment as compared to the immersive virtual environment, and a higher level of presence for the immersive (visual and audio) virtual environment as compared with the audio-only virtual environment.

Larsson, Västfjäll, Olsson, and Kleiner (2007) hypothesized that consistency across sensory renderings in terms of matching the visual space to the auditory space is crucial for the sense of presence. They carried out an experiment in which 30 participants were exposed to four different audiovisual conditions with varying degrees of auditory-visual consistency (one purely visual and three auditory-visual). In a presence questionnaire, participants rated the auditory-visual conditions as inducing significantly higher presence than the condition with only visual information. However, no significant differences in presence ratings between the three auditory-visual conditions were found. Cooper et al. (2018) also examined the effect of the different sensory cues (audio, visual, and haptic cues) on subjective ratings of presence. Results indicated significant main effects of audio and tactile cues on presence.
Thus, the benefits of multisensory cues have been demonstrated in virtual environments. The next step should be to investigate to what extent the results hold for real and remote environments, and whether the multisensory stimulation effect is different between such environments to improve presence.

2.4 Objective of the Study

As previously mentioned, this article aims at evaluating the sense of spatial presence in remote environments. A user experiment is presented that compares the sense of spatial presence in a real place with the presence experienced in a remote mediation of the same place. Real-time sensory restitution in the remote location is achieved using an HMD coupled with a 360° camera and a 3D audio system. Besides, the article investigates the effect of sensory channels on spatial presence by comparing a "visual-only" and a "visual and audio" immersive condition in such real and remote environments. The task performed during the experiment is a 3D pointing task.

Spatial presence and three subcomponents (ecological validity, engagement, and negative effects) are evaluated using the ITC-SOPI questionnaire (Lessiter et al., 2001). A performance measure and behavioral observations on users’ activity are also logged to study their relationship with the feeling of spatial presence and whether they can be conceivable tools for objective assessment of this sense.

The main hypotheses are that the environments and sensory renderings will affect the sense of spatial presence and its components, as well as the task performance and user’s behavior, with a positive effect of the real environment and the "visual and audio" condition. The experiment is described in more detail below.

3 Experiment

3.1 Experimental Design

3.1.1 The Real and the Remote Environment. Two rooms with a very similar layout (dimension = 4.25 m × 3.85 m) were used to represent the real and the remote environment in the experiment:

- An “operating room” (see Figure 1), representing a rectangular office where 12 tablets were attached to the four walls at fixed positions throughout the experiment. These positions were determined using a uniform sampling procedure.
- A “teleoperating room” where a teleoperation system was located that allowed participants to be remotely transported in the operating room.

3.1.2 The Pointing Task. Pointing tasks have been the focus of many studies that showed their effectiveness to evaluate users’ behavior and capture performance (see Raynal, Dubois, & Schmitt, 2013, for a detailed review). Consequently, a 3D pointing task was chosen for the experiment. More specifically, the task was to point as fast as possible a sequence of images that were displayed sequentially (i.e., one image at a time) on the tablets. The order of the images displayed was determined randomly; nevertheless, it remained the same for all participants. In addition, the images randomly depicted different types of worldwide well-known animals: cat, dog, duck, and pig, and sounds corresponding to the animal displayed were played from the tablets.
Thus, the participants were seated in the middle of one of the two rooms on a swivel chair. This chair allowed them to see the different walls of the room (by swiveling), and therefore see the tablets. Then, the participants had to perform the pointing task (see Figure 2):

- Either directly if they were located in the operating room, in which case they were experiencing a sense of natural presence.
- Or remotely through the teleoperation system if they were in the tele-operating room, in which case they were experiencing a sense of telepresence.

### 3.2 Method

#### 3.2.1 Participants

In total, 28 participants took part in the evaluation (19 men, 9 women, $M_{\text{age}} = 27.4$ years, SD = 4.4 years, age ranging 22–39 years). All participants had a normal or corrected view and non-impaired hearing. In addition, they all came from academic or scientific fields: 18 participants were students, while the remaining 12 came from different scientific fields of the university. Regarding the level of experience in VR, only 4 participants had never used an HMD before and reported no experience in VR systems, 17 reported that they were beginners, 4 reported intermediate expertise, and 5 considered themselves experts.

#### 3.2.2 Materials

The general setting is described in Figure 3.

The visual set-up. In the operating room, the participants experienced a natural visual and auditory stimulation. For the teleoperating system, remote visual capturing was obtained with a Ricoh Theta-V\(^1\) 360º panoramic camera placed at the center of the operating room. Images were streamed at a resolution of 3840 × 1920 at 30 Frames Per Second (FPS) to Unity’s real-time, the real-time rendering engine\(^2\) v2018.1.0 (consisting basically in a large spherical texture), then visualized in an HTC Vive Virtual Reality headset,\(^3\) provided a 90-Hz refresh rate. Remote action was made possible by tracking participants’ hand with a Leap Motion sensor attached to the headset. Then, a corresponding hand avatar was displayed to visually simulate the real hand as it was acting in the scene of the operating room (see Figure 3(d)). Informal interviews were collected from pilot tests on five participants. These interviews showed that all participants considered the avatar hand as their personal hand.

Because the field of view of the VR headset was limited to about 110º, similar viewing conditions were created in the operating room by making the user wear a headset mockup. The mask of the headset mockup was configured using a calibration procedure summarized in Figure 4. With the head still and one eye closed, a

---

1. [https://theta360.com/tr/about/theta/v.html](https://theta360.com/tr/about/theta/v.html)
2. [https://unity.com/](https://unity.com/)
Figure 3. The general setting of participants (a, b); their corresponding first-person view (c, d); in the operating room (a, c); in the tele-operating room (b, d).

trained operator looked at a calibration target consisting of vertical lines drawn on a vertical board, positioned at a fixed distance, thus providing tangible support for measuring the field of view. Looking straight to the zero-line, different sizes of eye apertures were trialed, until the perceived field of view matched the $55^\circ$-line on each side. The measure is subjective, but actually closely matches the one achieved in an identical virtual setup with the Vive HMD. The procedure was repeated for each eye, to create the final mockup of Figure 5. This calibration procedure ensured that the potential FOV effects were controlled. In particular, the number of tablets that were visible at any given point of the evaluation was the same regardless of the room in which the participants were located. Again, the informal interviews showed that the mockup suited the participants well.

The tracking system. In order to record the motion of each participant in the operating room, an infrared tracking system was set up, consisting of a six-camera ARTTrack5\textsuperscript{4}-tracking network. Participants wore infrared markers on both their pointing hand and the headset (see Figure 3(a,c)). The pointing gestures were recorded at 60 Hz with millimetric accuracy. In the tele-operating setting, head tracking was obtained with the Vive lighthouse system associated with the HTC-Vive headset, while hand tracking was provided by the Leap

\textsuperscript{4}https://ar-tracking.com/products/tracking-systems/arttrack5/
Motion sensor attached to the headset that recorded the hand motion in the headset coordinate frame, and then translated into the world coordinates.

To determine where the participants were pointing, a vector was calculated based on two virtual points: a first point located just in front of the participants’ eyes and a second point located on the top of their pointing hand. In order to avoid the misinterpretation of the pointing gesture and to obtain the correct pointing vector (Herbort & Kunde, 2016), participants were instructed to close their hands while targeting the images so they do not use one of their fingers to point. Then, pointing to an image was considered by the system as successful as soon as an intersection between the pointing vector and an invisible virtual sphere representing the location of the corresponding tablet was detected.

The audio setup. In the operating room, the participants experienced a natural 3D sound, as each tablet surrounding them was effectively emitting sound. To reproduce a similar auditory stimulation in the teleoperating room, a 1st-order Ambisonic microphone (Tetramic) was used to capture the sound in the operating room. This sound was then rendered binaurally over headphones (Closed Sennheiser HD 280) in the teleoperating room. A patch designed in the Max environment managed the entire audio processing pipeline. In particular, the SPAT library (Carpentier, Noisternig, & Warusfel, 2015) was used to perform the conversion from the Tetramic’s recording format (A-format) to the widely accepted Ambix format, and subsequently the decoding from ambisonic to binaural rendering using the virtual speaker array approach (Noisternig, Musil, Sontacchi, & Holdrich, 2003). The SPAT also added a small amount of reverberation to simulate the acoustics of the operating room (1.2 second of reverberation). Besides, the source aperture was fixed to 90°. Finally, the RMS level was set equal in both rooms to 75 dB (A).

3.2.3 Procedure. Independent variables. A mixed-design study was run with two fixed factors:

ENV: The type of environment with two modalities labeled REAL and REMOTE, representing respectively the situation where a participant is performing the task while being physically located in the operating room (i.e., experience natural presence), and the situation where a participant is operating remotely from the teleoperating room (i.e., experience

\[ \text{http://www.core-sound.com/TetraMic/1.php} \]

\[ \text{https://cycling74.com/products/max/} \]
telepresence). This factor was run following a between-subject design, first to minimize the learning effect on the task between the two conditions, but also to avoid tiring the participants.

SEN: The sensory mode with two modalities labeled WITH SOUND and NO SOUND, representing respectively the multisensory condition (visual with audio rendering) and the (visual-only) silent condition. This factor was run following a within-subject design.

The order of both factors was counterbalanced across participants using a Latin Square Design\(^7\) to control the effects of the assignment of participants and condition order.

**User evaluation.** The evaluation was approved by the local ethics and research committee of the university (CER\(^8\)). Each participant performed the task in only one of the two environments (REAL or REMOTE condition) depending on its order. However, they had to go successively through the two sensory modes WITH SOUND and NO SOUND, no matter the environment in which they operated. Therefore, all participants performed the task twice, which consisted of three parts:

- **The training:** The participants took a training session in order to get familiar with the setup, the instructions of the evaluation, the pointing task, and the constraints (keeping a fixed position and the use of the same closed hand to point all along with the evaluation).
- **The experiment:** The participants had to find and point on the images displayed on the tablets one after the other (as soon as an image was pointed at, it disappeared, and another one appeared in another location). They had to point to the maximum number of images, in a time limit of 3 minutes.
- **The post-assessment (questionnaire):** Once the evaluation was over, the users’ feeling of presence was assessed by administering the ITC-SOPI questionnaire.

The duration of the evaluation (training, experiment, and post-questionnaire) was approximately 20 min. As the participants repeated the task twice (one for each sensory modes), the total duration to complete the experiment was 20 × 2 ~ 40 min.

**Data collection.** In total, 56 trials were registered: 1 [ENV] × 2 [SEN] × 28 participants. For each trial, the total number of images pointed, and the trajectory path of both head and hand of participants were logged in the room coordinate frame. Finally, aside from the ITC-SOPI questionnaire, personal demographic variables (age, gender, profession/education, and experience with VR systems) were also collected. Therefore, the study comprised seven different dependent variables which can be grouped as follows:

- **Four subjective variables:**
  Responses to the 5-point Likert scale ITC-SOPI (1 = strongly disagree, 5 = strongly agree) yielded scores for each of the four components of the questionnaire:
  
  - SP: perceived spatial presence.
  - ENG: personal engagement.
  - ECO: ecological validity of the environment.
  - NEG: negative effects.

- **Three objective variables:**

  - **SCORE:** the participant’s performance that reports the total number of images pointed during the evaluation.
  - **HRHI:** the head-related hesitation indicator that reports how many times a participant swiveled in the wrong direction while trying to find the location of an image, and then decided to turn in the opposite direction (i.e., participant’s turnaround). A smaller value means that the participant had less hesitation (or very little) to locate the displayed images.

\(^7\)https://www.statisticshowto.datasciencecentral.com/latin-square-design/

\(^8\)https://www.universite-paris-saclay.fr/recherche/poletthis-ethique-et-integrite/comite-dethique-pour-la-recherche
STCI: the smoothness of the trajectory curve indicator that reports how much the participant pointing gesture was “fluid.”

The SCORE was directly computed from the participants’ results during the pointing task.

Concerning HRHI, it was extracted from participants’ head trajectory. Precisely, the indicator was processed—for each tablet to be localized—by calculating the number of times the head trajectory exceeded an angle higher than 90° in one direction, and then higher than 90° in the other direction.

Finally, STCI was extracted by processing the data cloud of participants hand trajectory as follows:

• The 3D data set was projected on each wall of the room to get 2D data.
• A principal component analysis (PCA) was computed on the raw 2D data set to get the variance of each principal components (one for the coordinates on the x-axis and the other for the coordinates on the y-axis).
• Finally, the ratio of these variances was calculated. The latter represents the smoothness of the trajectory curve indicator.

**Hypotheses.** The hypotheses were as follows:

**H1:** The environment will significantly affect the reported responses to the spatial presence questionnaire, the performance measure, and the behavioral indicators. More precisely, it is expected that SP, ENG, ECO scales will be higher in the *REAL* condition, while the NEG scale will be higher in the *REMOTE* condition (H1a). SCORE will be higher in the *REAL* condition (H1b). HRHI will be lower in the *REAL* condition (H1c). And STCI will be higher in the *REAL* condition (H1d).

**H2:** The sensory channels will significantly affect the reported responses to the spatial presence questionnaire, the performance measure, and the behavioral indicators. More precisely, SP, ENG, and ECO scales will be higher in the *WITH SOUND* condition, while the NEG scale will be higher in the *NO SOUND* condition (H2a). SCORE will be higher in the *WITH SOUND* condition (H2b). HRHI will be lower in the *WITH SOUND* condition (H2c). STCI will be higher in the *WITH SOUND* condition (H2d).

### 3.3 Results

This section is divided into four parts. The first part describes the interaction effects of the environments [ENV] and sensory modes [SEN] factors. The second and the third parts describe, respectively, the environments and the effect of the sensory modes on each measure. The last part describes the correlation between subjective responses to the questionnaire and objective performance and behavioral measures.

The result of the statistical parametric and nonparametric tests of each measure is reported. For statistically significant effects (*p* < .05), the effect size estimate *r* is given, which is interpreted as follows: a value under ±0.30 represents a small effect, a value over ±0.30 and under ±0.50 represents a medium effect, and a value over ±0.50 represents a large effect. The same goes for the correlation coefficients. All the analyses were performed using RStudio version 3.4.1.

#### 3.3.1 Interaction Effect

In order to analyze the interaction effect of ENV and SEN, a two-way analysis of variance (ANOVA) was computed on each measure. An aligned rank transform was used on nonparametric measures (followed by (*) before running the analysis. The results showed no significant interaction effect of ENV and SEN on any dependent variables: SP*, F(1,52) = 1; *p* = .32, ENG*, F(1,52) = .21; *p* = .64, ECO*, F(1,52) = .6; *p* = .44, NEG*, F(1,52) = .12; *p* = .73, SCORE, F(1,52) = 3.7; *p* = 0.6, HRHI*, F(1,52) = .67; *p* = .42, STCI*, F(1,52) = 1.44; *p* = .24. The focus was then given to the effect of each factor regardless of the other factor.

#### 3.3.2 Environments Effect

**Subjective questionnaire:** *SP, ENG, ECO, and NEG scales.* The means of the four scales were higher in the *REMOTE* condition than in the *REAL* condition, especially the SP- and ECO-scale (see Figure 6). In absolute terms, mean levels in the *REMOTE* condition of the SP-, the ENG-, and the
ECO-scale were high (above 3). A Wilcoxon signed-rank independent test was performed for every scale between the REAL and the REMOTE condition. The one-tailed significance values are reported. There were no significant differences in the ENG-scale, \( W = 77; \ p = .35 \) and the NEG-scale, \( W = 84; \ p = .53 \). However, statistically significant differences were found in the SP-scale, \( W = 39; \ p < .01; \ r = 1.18 \) and in the ECO-scale, \( W = 29; \ p < .01; \ r = 1.61 \), with unexpectedly, the REMOTE condition outperforming the REAL condition. These results go against (H1a), which expected that REAL condition would be better than REMOTE one.

**Objective measures**: SCORE, HRHI, STCI indicators. A Shapiro–Wilk normality test was performed on the data set of each measure. Results indicated that only SCORE and HRHI followed a normal distribution (SCORE: \( W = .96; \ p = .44 \), HRHI: \( W = .98; \ p = .9 \), STCI: \( W = .92; \ p = .04 \)). Therefore, a Welch two-sample independent \( t \)-test was performed on the average SCORE and HRHI, and a Wilcoxon signed-rank independent test was performed on the average STCI. The tests indicated statistically significant results with REAL condition overcoming REMOTE one for each measure: SCORE, \( t = 3.89; \ df = 25.87; \ p < .01; \ r = 1.47; \ CI = [-21.61; -6.67] \), HRHI, \( t = -2.72; \ df = 25.99; \ p < .05; \ r = 1.03; \ CI = [-3.45; -0.48] \), STCI, \( W = 170; \ p < .000; \ r = 1.4 \), which validates H1{b,c,d} hypotheses (see Figure 7).

**3.3.3 Sensory Channels Effect. ITC-SOPI questionnaire**: SP, ENG, ECO, and NEG scales. Except for the NEG-scale, the means of SP-, ENG-, and ECO-scales were all higher in the WITH SOUND condition (see Figure 8). A Wilcoxon signed-rank dependent test between the NO SOUND and the WITH SOUND condition was performed on each scale. There were significant
differences in the SP-scale, $V = 14.5; p < .000; r = .9$ and the ENG-scale, $V = 43; p < .000; r = 1.1$, which validate the (H2a) hypothesis. However, no significant statistical differences were found in the ECO-scale, $V = 88; p = .13$, and the NEG-scale, $V = 139; p = .99$.

Objective measures: SCORE, HRHI, STCI indicators. A Shapiro–Wilk normality test was performed on the data set of each measure. Results indicated that only SCORE followed a normal distribution (SCORE: $W = .97; p = .12$, HRHI: $W = .95, p = .02$, STCI: $W = .82; p = .000$). Therefore, a Welch two-sample dependent $t$-test was performed on the average SCORE, and a Wilcoxon signed-rank dependent test was performed on the average HRHI and STCI. The tests indicated statistically significant results with WITH SOUND condition overcoming NO SOUND one for each measure: $t(54) = 4.15; df = 27; p < .001; r = .52$; $CI = [3.29; 9.71]$, HRHI, $V = 301.5; p < .000; r = 1.23$, STCI, $W = 209; p < .01; r = .84$, which validates H2[b,c,d] hypotheses. (See Figure 9.)

Correlation between the subjective presence questionnaire and the objective measures. A Pearson correlation test was run between SP-, ENG-, ECO-, NEG-scales of the questionnaire, and the objective metrics. Results indicated no statistically significant correlation of four scales with the performance measure SCORE: SP, $t(54) = -1.32; p = .19$, ENG, $t(54) = .42; p = .67$, ECO, $t(54) = -1.72; p = .09$, NEG, $t(54) = 1.0; p = .32$. Regarding behavioral indicators, a significant correlation was found only between HRHI and (ENG)-scale, $t(54) = -2.3; p < .05; CI = [-0.52; -0.04]$, and STCI and the (NEG)-scale, $t(54) = 3.38; p < .01; CI = [0.17; 0.61]$.

3.4 Discussion

The first part of this article highlighted the importance of the immersive fidelity between environments in order to evaluate the potential of objective measures to assess spatial presence. In particular, the reproduction of accurate remote auditory component was of major concern to reach renderings comparable to the real condition (Postma & Katz, 2016). Nevertheless, the spatial resolution of the First-order Ambisonic microphone used for the teleoperation system was lower than the quality of the natural hearing in the real environment. This disparity between the environments could bias the results as hypothesized in Khenak, Vézien, Thery, and Bourdot (2019). In order to explore this possibility, the interaction effect of both environments and sensory renderings factors was analyzed. It showed no significant effects of sensory renderings regarding the real and the remote environment, meaning that the audio setup (empiric evaluation of reverberation, first-order 3D microphone), despite its shortcomings, provided sufficient accuracy localization.

Consequently, the analysis has centered on evaluating the effect of the two factors independently of each other. This analysis was performed on the four scales of the ITC-SOPI questionnaire (spatial presence, personal engagement, ecological validity, and negative effects), the average score of participants, their degree of head hesitation, and the smoothness of their hand trajectory during the pointing task.

Concerning the effect of the environments, the responses to the questionnaire indicated a significantly higher level of spatial presence and ecological validity in the remote environment compared with the real environment. This sense of a higher presence in the remote situation than in reality can be viewed as “hyper-presence,” which was mentioned by Biocca (1997). A
possible cause could be a “technological effect” resulting from the lack of familiarity of participants with the VR HMD (HTC-Vive), especially in such a telepresence configuration. This effect could have led participants to appreciate more the remote experiment and consequently give better ratings. Indeed, most of them were beginners (66.7%) in the use of VR headsets and some informal interviews showed the great enthusiasm of participants in the experiment. This means that the responses to the questionnaire could be highly dependent on prior users’ experiences (whether during the experiment at hand or beforehand) and interests. It encourages then the use of additional measures that are more objective.

This “hyper-presence” may also be related to the fact that the environment was a between-subject factor. A between-subject design was chosen to mitigate the learning effect on the task. Since the sense of spatial presence is a subjective user’s feeling, every participant could perceive it and rate it differently based on their personal interpretation of the feeling and their “inner” scale. Nevertheless, the use of ITC-SOPI questionnaire, which has already proved to be a valid questionnaire (Lessiter et al., 2001), allows reducing this potential bias.

In contrast, the engagement scale showed no significant difference between the environments, which can be explained by the similarity of the task and content in both environments, as the goal of the remote environment was precisely to remotely allow the participants to operate in the real environment. The same applies to the negative effects scale that showed no statistically significant difference between the environments, which demonstrates the potential growth of such a teleoperation system in terms of visual quality, tracking precision and latency reduction.

Thus, the hypothesis (H1.a) that the real environment provides a higher sense of spatial presence, ecological validity, and engagement, and a lower negative effect was not confirmed and even took an opposite direction through the emergence of a sense of “hyper-presence” in the remote environment.

On the other hand, objective results related to participants’ performance showed a higher task efficiency in the real environment, which meets the expectation of the hypothesis (H1.b). The most likely explanation is that in the remote environment, the participants felt they had fewer facilities to move because of the equipment of the teleoperation system (headset, headphones, wires, etc.). Therefore, they moved more slowly and had more restricted gestures while performing the pointing task. This assumption tends to be confirmed by the two ad-hoc indicators extracted from the head and hand participants’ trajectory:

- a higher level of hesitation of participants when turning their head to find the location of images;
- and less smoothness when pointing on images in the remote environment.

Concerning the effect of sensory renderings, the sense of spatial presence and the degree of ecological validity were higher when providing audio rendering, while the negative effects related to the disorientation of participants were lower. In particular, the results of the experiment provided strong evidence on the benefits of audio rendering in increasing the engagement of participants. These results which support the hypothesis (H2.a) are in agreement with previous studies that showed the usefulness of sensory cues in the improvement of presence (Lipscomb, 1999; Mania & Chalmers, 2004), and user satisfaction (Lee, Billinghurst, Baek, Green, & Woo, 2013) in virtual environments. Similarly, the participants’ performance and the smoothness of their pointing gesture trajectory was significantly higher when the audio sensory channel was provided, reflecting less hesitation. It should be noted that these results were found regardless of the type of environment (as no interaction effect was found).

Furthermore, previous papers raised the question of an existing correlation between the sense of presence and performance (see Section 2). Indeed, a common assumption is that immersive environments increase both presence and performance (Cummings & Bailenson, 2016; Tan, Gergle, Scupelli, & Pausch, 2006). This assumption suggests a positive link between presence and performance (Lombard et al., 2015, pp. 139-185). In this study, no correlation was found between the perceived feeling of spatial presence and the actual pointing
task performance. This result echoes some evaluations (Picciano, 2002; Welch, 1999) that tended to argue that presence and performance are not related. Yet, it contradicts more recent studies (Cooper et al., 2018) and encourages further evaluations. Besides, no correlation was found between spatial presence and the participants’ behavior represented by the two ad-hoc indicators. However, the different body representations in each environment (the participant’s hand in the real environment, and the 3D representation of participant’s hand in the remote environment) may explain the results as it was proved that body representation could affect task performance (Medeiros et al., 2018).

Therefore, more investigation is needed to understand the reliability of performance and behavioral analysis as potential tools to measure spatial presence.

4 Conclusion

This article presented a user evaluation assessing the feeling of spatial presence, by comparing a real and a remote immersive environment representing respectively an actual office setting and its remote representation captured by telesensing technologies. In an experiment, participants performed a 3D pointing task while being directly located in the office or being remotely transported in the same office through a teleoperation system. The study evaluated the sense of spatial presence and three subcomponents (engagement, ecological validity, and negative effects) using the ITC-SOPI questionnaire (Lessiter et al., 2001). In addition, objective metrics were recorded based on the users’ performance to achieve the task and their behavior obtained by extracting ad-hoc indicators that represented the hesitation and smoothness of users’ pointing gestures.

Results from the questionnaire highlighted an unexpected higher sense of spatial presence, or “hyper-presence,” in the remote environment compared to the real environment. A potential explanation could be the “technological effect” resulting from the lack of familiarity of participants with the telepresence technologies used. Therefore, future work will address this assumption by comparing the sense of spatial presence between different environments with the same telepresence technologies. More precisely, the next studies will evaluate spatial presence between remote and virtual environments. Particular attention will be paid to rely on similar VR user-interface technologies for the rendering of both types of environments in order to mitigate the influence of technological effects. Such studies will also allow investigating the effect of users’ awareness of real consequences on the physical place that characterizes remote environments over virtual ones (Slater et al., 2006). Therefore, it will be essential to design tasks that have a significant impact on the physical place in order to increase this awareness.

Moreover, the future comparisons between environments will have to be carried out following a within-subject design in order to ensure that there is no influence of the user’s interpretation of the sense of spatial presence. Results collected from these comparisons will help to clarify the boundaries between remote and virtual environments.

On the other hand, task performance and behavioral indicators provided results that were clearly in favor of the real environment. So that no correlation between the objective measures and the feeling of spatial presence was found, this result calls for a more in-depth investigation on the use of reliable, objective behavioral tools to measure spatial presence. More specifically, the next studies will aim to develop behavioral indicators related to kinematic metrics extracted from users’ trajectories (e.g., user’s backtrack during a navigation task, Martin, Férey, Clavel, Darses, & Bourdot, 2012). Thus, a possible experiment would be to analyze the trajectory of participants when traveling on a circuit and their reaction when facing obstacles in a remote or virtual environment (i.e., avoidance of obstacles, Lynch et al., 2017). In this kind of experiment, it would also be very interesting to study the influence of participants’ awareness of the real consequences of their actions on the physical location (in the case of remote situations). Thus, one could wonder whether the participants will break or bypass the obstacles depending on the environment in which they are located.

This article also evaluated the effect of providing spatialized audio rendering consistent with visual rendering
in both real and remote environments. Results provided evidence of the importance of audio rendering to enhance the sense of spatial presence and its components. More specifically, the engagement component proved to be highly affected by audio rendering. Performance and behavioral indicators also confirmed this assumption regardless of the type of environment.

To summarize, this article is a first step in understanding the sense of spatial presence in remote environments. It compared a real and a remote environment and raised the question over the difference between virtual and remote environments. In addition, it sought to introduce new behavioral and performance tools to measure spatial presence, and it identified new issues about hyperpresence and potential effects of VR technologies (the technological aspect is a significant difference between real and remote environments). Such studies will certainly benefit the presence research community as well as the designers of teleoperation technology.

Acknowledgments

This research was partially supported by RTA Digiteo and Labex DigiCosme (Idex Paris-Saclay ANR-11-IDEX-0003-02), and by EquipEx DIGISCOPE (ANR-10-EQPX-26-01) operated by the French National Research Agency (ANR), as part of the program Investissements d’Avenir.

References


Freeman, J., Lessiter, J., Pugh, K., & Keogh, E. (2005). When presence and emotion are related, and when they are not. 8th Annual International Workshop on Presence, 21–23.


