The Plausibility of a String Quartet Performance in Virtual Reality

Ilias Bergström, *Member*, IEEE, Sérgio Azevedo, Panos Papiotis, Nuno Saldanha and Mel Slater

Fig. 1 Top view of the scenario (excluding the virtual body of the participant)

Abstract - We describe an experiment that explores the contribution of auditory and other features to the illusion of plausibility in a virtual environment that depicts the performance of a string quartet. 'Plausibility' refers to the component of presence that is the illusion that the perceived events in the virtual environment are really happening. The features studied were: Gaze (the musicians ignored the participant, the musicians sometimes looked towards and followed the participant's movements), Sound Spatialization (Mono, Stereo, Spatial), Auralization (no sound reflections, reflections corresponding to a room larger than the one perceived, reflections that exactly matched the virtual room), and Environment (no sound from outside of the room, birdsong and wind corresponding to the outside scene). We adopted the methodology based on color matching theory, where 20 participants were first able to assess their feeling of plausibility in the environment with each of the four features at their highest setting. Then five times participants started from a low setting on all features and were able to make transitions from one system configuration to another until they *matched* their original feeling of plausibility. From these transitions a Markov transition matrix was constructed, and also probabilities of a match conditional on feature configuration. The results show that Environment and Gaze were individually the most important factors influencing the level of plausibility. The highest probability transitions were to improve Environment and Gaze, and then Auralization and Spatialization. We present this work as both a contribution to the methodology of assessing presence without questionnaires, and showing how various aspects of a musical performance can influence plausibility.

Index Terms: presence, plausibility, place illusion, user studies, experimental methods, multimodal interaction, entertainment

1 Introduction

In the new popular wave of interest in virtual reality (VR) the concept of 'presence' has been rediscovered and is seen as central to the experience delivered to participants. This concept was elucidated in a classic set of papers in the early 1990s - e.g. [1-4] - as the sense of 'being there' in the place depicted by the virtual environment. In this paper we adopt the deconstruction of presence postulated in [5] into the concepts of 'being there', or Place Illusion (PI) as originally propounded, and Plausibility (Psi) which is the illusion that events in the virtual environment are really happening. In particular here we concentrate on Psi as the major issue of investigation. Research concentrating on eliciting the factors that contribute to presence - summarized in a recent meta study [6] - has concentrated largely on system properties and performance such as latency, rendering framerate and tracking, and most especially on visual characteristics of the VR displays. Here we focus mainly on the quality of auditory rendering, and how different auditory settings contribute to Psi. We adopt the approach to the quantification of presence based on an analogy to color matching theory [7], and we use this method in the assessment of Psi.

We carried out an experimental study with 20 participants to assess how varying a number of features of the VR would influence Psi. The scenario consisted of a string quartet, performing a piece of classical music, with the four musicians positioned in a circle surrounding the participant (Fig. 1), all in a realistically simulated virtual room. The scenario was intended to resemble a rehearsal or warm-up session, rather than a formal performance in front of an

- *Ilias Bergström is with KTH Royal Institute of Technology, Stockholm, Sweden. Email: onar3d@gmail.com.*
- *Sérgio Azevedo is with Microsoft Language Development Center, Lisbon, Portugal. Email: Antonio.Azevedo@microsoft.com.*
- *Panos Papiotis is with Universitat Pompeu Fabra, Barcelona, Spain. Email: panos.papiotis@upf.edu.*
- *Nuno Saldanha is with is with Microsoft Language Development Center, Lisbon, Portugal. Email: saldanha.nuno@gmail.com.*
- *Mel Slater is with ICREA and University of Barcelona, Spain and UCL, London. Email: melslater@ub.edu*

This work was carried out in the Event Lab, University of Barcelona.
Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org.
Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx/.

audience. To maximize the perceived realism of the quartet, we used recordings of motion and sound from a real string quartet performance, openly available as part of the QUARTET dataset[1] [8]. Motion capture data was used to animate the four virtual human characters and their instruments, while individual high-quality audio recordings of each instrument were captured using piezoelectric pickups and convolved with body impulse responses obtained from the same instruments in order to reproduce the 'dry' sound of each instrument without any environmental reflections.

Participants were in a virtual environment rendered with a set of features, each with a controllable number of levels of fidelity. Participants first experienced the environment at the greatest level of fidelity on all features, and paid attention to their quality of Psi (the *target*), i.e., how much they had the feeling that what they were experiencing was really happening. Then in 5 different trials they were able to choose transitions of individual features from lower to higher order fidelity, and keep doing this until they declared a *match* to the target feeling of Psi. In each transition they could change the levels of the following properties: (Gaze) whether virtual humans representing the musicians would react to the participant or not, the quality of sound Spatialization (sound direction), the amount of sound Auralization (sound reflection from the environment), and whether Environment sounds were heard from outside the virtual room.

The goal of the study was to understand how these features contribute to Psi. Based on previous work, where illumination realism was found to influence Psi (but not PI) [7], we expected the overall level of realism of the sound to be important. However, based on the principles of how Psi works, we also expected the non-auditory factor, that is whether the virtual characters paid attention or not to the participant to be important for Psi. A further contribution of this research is that we show how the method introduced in [7] was used to assess the relative influence of these four factors and their different levels, but without any reliance on questionnaires.

2 Background

2.1 Presence

While the illusion of 'being there' (PI) is the most stunning and remarked upon experience that VR delivers, it was argued in [5] that this misses another important aspect, the illusion that the events in the experience are really happening. For example, in our quartet scenario participants were placed in a virtual room with the four virtual characters around them. They were able to look in any direction via a head-tracked stereo, wide field-of-view, high resolution head-mounted display (HMD). When they looked down towards themselves they would see a virtual body substituting their own that also moved in synchrony with their own movements.

Now suppose that the quartet were playing, but its never responded to any actions of the participant - e.g., even if participants were blocking a player's view of her own instrument, or if attempting to interact with a player to draw his or her attention. Based on the argument in [5] PI would nevertheless occur because the head and body tracking affords perception through natural sensorimotor contingencies - that is, perception through body movement (head turns, leaning forward, back, looking around objects, bending down, stretching, and so on). This is based on the theory of active perception [9, 10]. PI can occur since it is a perceptual illusion that is a function of how well affordances for perception match those of perception in physical reality. To the extent that the sensorimotor contingencies supported by a VR system correspond to those of real perception, the simplest hypothesis for brain to adopt is that self-location is where it appears to be - in the place depicted by the virtual environment. However, in this scenario although participants would experience PI they would be unlikely to experience Psi, since the quartet members do not respond at all to the participant, so the reality of the situation is lost. On the other hand were the quartet to respond to actions of the participant then it is postulated that Psi could occur, depending also on other factors.

Hence in this argument PI is based on *how* participants are able to perceive the VR and Psi is based on *what* they perceive. PI is static (no events need to be taking place) whereas Psi is relative to events.

In the early 1990s factors that contributed to PI were conceptualized [1-3, 11, 12], and experimental studies have been carried out ever since to assess the impact of various possible contributors – a far from non-exhaustive list includes framerate [13], interaction and display methods [14, 15], pictorial realism [16], the role of self-representation with a virtual body [17, 18], illumination realism [19-21] and latency [22]. There have been many such studies, referred to in [23] and in the meta study [6]. However, Psi has rarely been studied, yet for many applications the illusion that events in the scenario are really occurring may be critical. For example, when using VR to assess how people might behave in emergency situations, application designers would want people to automatically behave realistically, which is more likely if they have the illusion that the unfolding events are really happening. We return to this in the Discussion and Conclusions.

There have been very few reported studies that concentrate specifically on the effect of sound on presence in the context of immersive virtual environments. Hendrix and Barfield [24] compared no sound, non-spatialized sound and spatialized sound. They found that spatialized sound positively influenced presence as being there (PI), but not the perceived realism of the environment. This is in line with the fact that spatialized sound is a sensorimotor contingency and therefore likely to contribute to the illusion of being there, but not necessarily be an important contributor to the reality of what is being perceived. Similarly in an experiment that examined the relationship between presence and task performance based on varying audio properties [25] it was found that spatial audio contributed to presence, with respect to a questionnaire that assessed presence as PI, but that with respect to a presence questionnaire that did not measure presence as 'being there' [26], this effect was not observed. It was also found in [27] that spatialized sound contributes to PI, comparing spatialized sound with no sound. Other studies have included sound as one of several factors thought to contribute to presence, and have concluded that sound is useful [28, 29].

2.2 Spatial hearing

Research on spatial audio reproduction technologies is richly varied and actively on-going – see [30] for a review, with solutions for virtual environments described, for example, in [31], and corresponding psychophysical studies reviewed in [32]. The methods employed in our experiment used state-of-the-art technology to simulate real-time three-dimensional audio spatialization and auralization [33]. These techniques are generally not known, or not employed in VR research[2] hence we include a brief review.

Much research has been devoted to understanding how we localize the sound sources that surround us, to the extent that we now have sufficient - though not complete - knowledge as to what aural percepts contribute to which type of information [34]. For determining azimuth sound source direction, we depend primarily on binaural cues, and subsequently also on spectral cues. Binaural cues are those that are determined from stimuli received at both ears, such as Interaural Time Difference (ITD) and Interaural Level Difference (ILD). Spectral cues are those derived from the spectral alterations sound has been subjected to by our body on its way to the eardrums. For detecting elevation information, we are restricted to relying only on spectral cues, leading to our perception of elevation being much less accurate than that of azimuth [34].

[1] *http://mtg.upf.edu/download/datasets/quartet-dataset*

[2] *A Google Scholar search of IEEE VR proceedings from 2000-2016 a tiny proportion that included 'audio' or 'sound' with at least one of the terms 'spatialization / spatialized', 'auralization / auralized', 'virtual acoustics', '3D audio / sound', 'binaural'.*

We also detect distance using our hearing from the sound's amplitude: the louder it is, the closer we perceive the source to be. To a lesser extent, distance can also be determined by high frequency energy content, since distance reduces it faster than sound at lower frequencies. Distance is additionally also determined from sound reflections, or reverberation as it is otherwise known: the louder the direct sound is in relation to its reflections coming from the surrounding environment, the closer it has to be to the listener. Also, the timing of the sound reflections provides numerous clues to the location and distance of the sound source.

The virtual recreation of realistically localizable direct sounds using ITD, ILD and Head Related Transfer Function (HRTF) / Head Related Impulse Response (HRIR) simulation [34], as discussed this far, is often referred to as spatialization, while the term virtual acoustics is employed when reflected sound is also taken into account in the simulation.

Spatial hearing cues are crucially perceived dynamically. Head movement is known to produce a significant increase in localization accuracy [35]. Front/back confusions that are common in static listening tests disappear when listeners are allowed to slightly turn their heads to help them in localizing sound [36]. Also, the temporal characteristics of the sound stimuli play an important role. A short sound is harder to localize than a sustained one, as is a continuously moving sound source compared to one that is static.

If sound directional cues, and sound reflections, are not correctly simulated when reproduction is over headphones, sound is likely to be perceived as coming from inside the head (in the literature referred to as in-head localization), as sounds from within the listeners own body are the only ones that naturally appear without any environmental reflections or directional cues [34].

3 MATERIALS AND METHODS

3.1 Recruitment

Twenty participants (7 of them male) were recruited through advertising using the lab database. Their average age was 25 ± 9 (SD) years. 18 had prior experience of VR. 7 considered themselves amateur musicians, of which 1 had formal music studies. None of the participants had any prior knowledge of the experiment. They were exposed to a virtual environment that consisted of a room which was sparsely furnished, and was populated by four string playing musicians, two seated and two standing, with the participant standing in the middle of the circle they formed (Figures 2, Video).

3.2 Materials

The HMD used was the NVIS nVision SX111. This displays a 3D scene in stereo with a horizontal field of view of 102 degrees and vertical field of view of 64 degrees by sending left-eye and right-eye images to left and right hand display screens. Its weight is 1.3Kg.

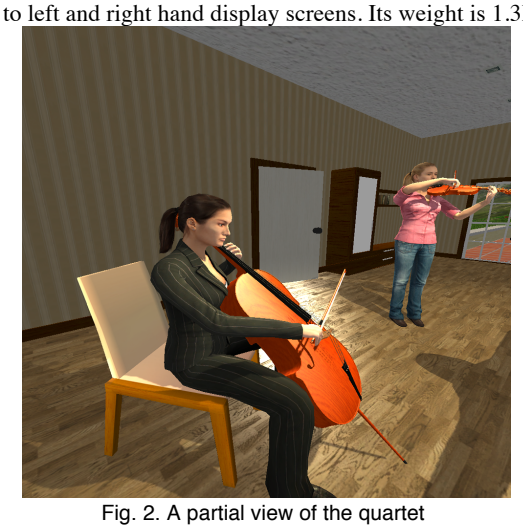

Fig. 2. A partial view of the quartet

Head tracking is with a six degree of freedom Intersense IS900 motion tracker. To track participants' whole body movements we used marker-based infrared tracking: a 12 camera Optitrack system from Naturalpoint, which in our configuration could track a volume of approximately 2.5m width x 2.5m length x 3m height. Participants wore a tight fitting Velcro suit that had 37 retroreflective markers attached. Movements were reconstructed by the Motive software at 100 Hz, with millimeter accuracy.

Auralization and Spatialization were achieved using the software library described in [33], which we integrated into our own custom software for the VR simulation. Spatialization is achieved through 'Virtual Ambisonics', which combines the benefits of Ambisonics 3D audio reproduction, with the ability to reproduce over headphones, by simulating virtual loudspeakers over headphones. This technique alleviates the issues otherwise present when reproducing 3D audio over headphones through HRIR convolution. The library also integrates an Auralization simulation, divided into two stages: it first calculates early reflections of first and second order, while late reverberation is implemented in a second stage, through reverberators embedded in feedback delayed network structures.

For audio reproduction we used the Sennheiser RS180 high-fidelity wireless headphones with uncompressed audio transmission. For participants to be able to switch between features while immersed in VR, we used a Nintendo Wii wireless Bluetooth remote control, held by participants in their dominant hand.

3.3 Features

The features that participants were able to manipulate are described by a vector S = <Gaze, Spatialization, Auralization, Environment> or <G,S,A,E>. Participants could choose to advance the fidelity of each of these by a button press on the Nintendo Wii (down, left, up, right buttons respectively). The following describes the levels of each feature, ranging from lowest to highest fidelity in each case:

Gaze:
(*Ignore, 0*) The virtual players do not respond at all to the participant.
(*Attend, 1*) The virtual players direct their head and gaze towards the participant occasionally, and their gaze follows the head location of the participant if she or he moves while gaze is in effect.

Spatialization:
(*Mono, 0*) No directional audio cues: only sound amplitude changes in relation to listener distance to sound source.
(*Stereo, 1*) Stereophonic directional sound cues. While the amplitude differences vary between left and right ear to reflect the sound source direction (thus one-dimensionally), there is no spectral filtering as would normally be caused by the head and the ear pinnae, filtering which in human hearing is relied on to determine sound direction in three dimensions.
(*Spatial, 2*) Full binaural three dimensional sound direction simulation.

Auralization:
(*Dry, 0*) No sound reflections. Only the direct sound signals are heard.
(*Large, 1*) Reflections that would correspond to a room larger than the one being visually perceived.
(*Real, 2*) Reflections that exactly match those that a room of the virtual room's dimensions would produce.

Environment:
(*None, 0*) No sound can be heard coming from outside of the room.
(*Birdsong & wind, 1*) Sounds corresponding to the environment visible through the windows (birdsong and slight wind noise) can be heard.

Altogether there were 36 possible configurations: 2 for Gaze × 3 types of Spatialization × 3 Auralization × 2 Environment. The accompanying Video illustrates all these various settings.

3.4 Procedures

3.4.1 Preparation and Familiarization with the Environment

The experiment was approved by the *Comisión de Bioética de la Universitat de Barcelona* and participants gave written informed consent. When participants arrived to the lab, they were given an information sheet to read, and the information was also explained to them verbally. They read and signed an informed consent form. They were helped to put on the full body tracking suit, the HMD - calibrated so that its two screens were symmetrically placed over the participants' eyes using the method described in [37] - and finally they put on the headphones. The motion capture area where the experiment took place was closed off from the rest of the laboratory by a black curtain, so that the participants were in darkness once the experiment started, to avoid possible light reflections leaking into the HMD. Upon starting, they were left to accustom themselves to the displayed environment for 1 minute. During this time they were asked to look around and describe what they saw.

They subsequently went through a training procedure, which consisted of the virtual musicians performing, while the participant changed each of the features, through pressing the buttons on the remote-control. They were encouraged to move freely by walking between the four musicians, while noting the differences that each button press made to the environment. They were allowed to go through as many training procedures as they wanted, until they were sure they were familiar with all features that could be changed, and the effect of pressing each button on the controller. Whenever they reached the maximum (highest fidelity) configuration they were left to experience it until they pressed the designated 'OK' remote-control button, which produced a distinct 'ping' sound. They were then asked if they wanted to reset all parameters, and do another training procedure. Normally 3-4 such procedures were required until participants reported that they understood the setup.

Table 1. The Starting conditions for the 5 trials, which were run in randomized order across the participants

Trial	Gaze	Spatialization	Auralization	Environment
1	0	1	0	0
2	1	0	0	0
3	0	0	0	0
4	0	0	0	1
5	0	0	1	0

Following the setup and training procedure, participants experienced the configuration <1, 2, 2, 1> (each feature at the highest level) for two minutes. They had been given the instruction: *"Pay attention to how real this feels. Later we will ask you to try to get that feeling of reality again"*. After experiencing the highest fidelity configuration, the instructions for the transitions were read to them once again, and they then proceeded to carry out the five experimental trials.

The trials started with configurations shown in Table 1, the order of presentation randomized across participants. To encourage participants to carefully consider each transition, and to avoid them immediately attempting a transition to the full configuration <1, 2, 2, 1>, we imposed the following rules for the 5 experimental trials:

Transitions could only be made in one direction - i.e., having chosen a higher level of one feature they could not go backwards. For example, if they had made the transition from monophonic to stereophonic sound (from 0 to 1 on Spatialization), they could not later go back to monophonic. This was to keep the task simple. Once participants reached the highest level of a property, subsequent button presses resulted in no change, but a 'beep' sound was produced to notify them of the fact.

Only one-step transitions could be made. For example, they could not choose to go directly from 0 to 2 under Spatialization or Auralization, but would need to first transition through the intermediate step 1.

In order to avoid participants transitioning randomly until no further transitions could be made reaching <1, 2, 2, 1>, we imposed a cost structure on transitions. We told them that they would start out with €15 (units of money). Every transition would cost 1. To stop, they had to press the designated 'OK' button on the remote. If they stopped too early i.e., before they were in the Psi state, they would lose 5. On the other hand if they reached the desired state they would get a bonus of 5. They were told that the final payment for the experiment was the maximum achieved amongst their 5 trials. We did not explain, and no participant asked, how we would know which state they were in - i.e., if they stopped 'too early'. Moreover, the final payments made were always €15 to all subjects regardless of their choices.

4 Results

4.1 Transitions

We denote the set of the 36 possible configurations that a participant could experience by C. The set of all possible transitions from configuration to configuration is therefore a subset of $C \times C$. Each transition is of the form $[G_t, S_t, A_t, E_t] \rightarrow [G_{t+1}, S_{t+1}, A_{t+1}, E_{t+1}]$ denoting the transition from the configuration that a participant was in at time t, to the configuration at time t+1. From the set of all such transitions we can construct the probabilities p_{ij} that a participant in configuration $i \in C$ would next choose configuration $j \in C$. This gives us the Markov transition matrix P. Then P^k is the k-step transition matrix, with elements that give the probability that a participant in configuration i would be in configuration j, k steps later. Let u be a 1×36 vector where u_j are the initial probabilities of being in configuration $j \in C$ (i.e., the probability of being in a particular starting configuration). Then uP^k are the probabilities of being in the configurations after k transitions. All of the above follows from Markov chain theory [38]. P was constructed from the 520 observed transitions. P is obviously a sparse matrix. However, the total number of possible transitions is not 36×36 but rather 84 given the restrictions described in 3.4.1.

Table 2. The 3 highest probability configurations after each transition shown in the corresponding row

After transition:	The three configurations with highest probability:					
	Config.	Prob.	Config.	Prob.	Config.	Prob.
1	0001	0.60	1000	0.25	0010	0.10
2	1001	0.44	0011	0.19	1010	0.17
3	1101	0.28	1011	0.26	0021	0.18
4	1021	0.32	1201	0.21	1111	0.20
5	1121	0.58	1211	0.25	0221	0.13

We are particularly interested in the u corresponding to the initial configuration being <0,0,0,0>, which is a vector of all 0 but 1 in the place corresponding to this configuration (for example, $u_1 = 1$ and $u_j = 0$ all $j > 1$). In this way we can consider the probabilities of configurations uP^k for successive k = 1,2,3,4,5. These would be the probabilities of being in the various configurations after k transitions having started in <0,0,0,0>. The configuration <1,2,2,1> is absorbing since it will always be reached after 6 transitions and then there are no more possible transitions.

From Table 2 it can be seen that after the first transition the highest probability configuration is <0,0,0,1> (sounds from outside the room could be heard), with probability 0.6. The next highest probability is 0.25 for configuration <1,0,0,0> (musicians looking at the participant). Similarly for the subsequent transitions. The unfolding pattern suggests that to maximize the chance of a match on

Psi participants first chose external environment sounds and then for the musicians to be noticing and responding to them. After having established these two, participants then moved to improve either Spatialization or Auralization although aiming at the higher fidelity Auralization by transition 4. By the time they had reached transition 5 the probability of having set Auralization to its highest level was more than double that for Spatialization.

4.2 Probability of a Match

From all the matching configurations chosen by participants we can compute the probabilities $P(match|<G,S,A,E>)$. This is the probability that a match would be declared having reached configuration <*G*,*S*,*A*,*E*>.

Table 3 shows these probabilities. This excludes <1,2,2,1> since by construction this would always be a match. The most interesting finding is comparing the first two rows. Adding the first level of Spatialized sound to the configuration <1,0,2,1> almost doubles the probability of a match. Rows 2-5 show a set of configurations that have nearly the same probability, and therefore show potential tradeoffs between the different features. For example, rows 2 and 3 suggest that including Gaze is the same as not having Gaze but the highest level of Spatialization, in the context where there are the highest levels of Auralization and Environment sounds. Or, comparing rows 3 and 4 if there are the highest levels of Spatialization and Auralization then having Environment sounds is not important.

Table 3. Probabilities (> 0.10) of a match in a configuration

	Config. <G,S,A,E>				Prob.
1	1	1	2	1	0.500
2	1	0	2	1	0.269
3	0	2	2	1	0.267
4	0	2	2	0	0.250
5	0	2	0	1	0.222
6	1	1	1	1	0.143
7	1	1	2	0	0.143
8	1	0	0	1	0.129
9	0	0	2	1	0.118
10	1	2	1	1	0.118
11	1	0	2	0	0.111

4.3 Marginal Probabilities

We can also compute the marginal probabilities for the individual features. In other words amongst all matching configurations, we compute the probability that they contain a particular setting for a feature. The results are shown in Table 4. The probability that a matching configuration would include the musicians taking notice of the participant is high (G=1), and the same for Environment noises (E=1). These are summed over all the settings of the other features.

Table 4. Probability that a matching configuration would contain the feature at the given level

Gaze		Spatial		Aural		Environ.	
0	0.09	0	0.14	0	0.08	0	0.03
1	0.91	1	0.24	1	0.04	1	0.97
		2	0.62	2	0.88		

We can compute various other interesting marginal probabilities. For example, the probability that a match has both Gaze and Environment set to the highest level is $P(G=1 \wedge E=1|match)=0.89$. The probability that at least one of Gaze or Environment has been set is $P(G>0 \vee E>0|match)=0.99$. On the other hand considering only the two sound features, the probability that both of these are at their highest level is $P(S=2 \wedge A=2|match)=0.57$. However, the probability that both features have at least one level of fidelity is $P(S\geq 1 \wedge A\geq 1|match)=0.82$. Finally, we can compute the probability that at least one of these features is set as $P(S>0 \vee A>0|match)=0.96$.

5 Discussion

In [5] three factors were postulated as contributing to Psi: (i) the environment responds to the actions participant (e.g., a virtual character moves out of the way as the participant moves into its personal space) (ii) there are events that relate personally to the participant (e.g., a character looks at or calls the name of the participant) (iii) the simulation has to match expectations where these are relevant (e.g., if it is a simulation of an event in reality then it had better conform to what would be expected to happen in reality). The last is the most difficult because it relies on detailed domain knowledge. The results of the quartet experiment are compatible with these ideas. The characters gazing at the participant and following his or her movements provide examples of (i) and (ii), and also support earlier findings on the effect of gaze behavior - e.g. [39-41]. In the video it can be seen that the room has windows and a door open to an outside country scene, and so environmental sounds from outside would help to foster the illusion that these events are actually taking place. In contrast the full level of Spatialization (S=2) helps to locate the participant with respect to the environment, being an important sensorimotor contingency. While important for PI it would be less important for Psi. On the other hand Auralization does not correspond to a sensorimotor contingency, and the sound reflections should be expected, and so add to the sense of reality of the situation. It is important to note that while PI and Psi are conceptually distinct there may be factors that contribute to both. So, for example, some level of Spatialization would be expected to occur (S=1), but the highest level (S=2) it would be likely to contribute more to PI.

On this point it is interesting to note that in spite of participants experiencing many different configurations of the environment their reported level of PI was universally high. For example, after all of the experiences participants were asked the extent to which they had the sensation of being in the virtual room, scored on a 1-7 Likert scale with 1 representing not at all, and 7 very much so. Out of the 20 participants 9 gave a score of 7, another 9 gave a score of 6, and the remaining 2 a score of 5. The invariant features were the head and body tracking, and the first person perspective virtual body that moved synchronously with participant movements. In other words there were strong visual sensorimotor contingencies, which were enough to contribute to a very high illusion of being there.

Features contributing differentially to PI and Psi were directly compared in [7] using the same method as in this paper. The features considered were visual field-of-view, illumination realism, display type (powerwall or HMD, both with head-tracking), and a virtual body seen from first person perspective that moved synchronously with real movements. Participants in two different groups were asked to find a configuration of features that matched their target feeling of either PI or Psi. It was found that those matching PI chose the wide visual field-of-view with a HMD with high probability, whereas illumination realism was chosen by those matching their target level of Psi. Illumination realism would be analogous to Auralization in the quartet experiment. The virtual body that moved synchronously with real body movements was important to both PI and Psi. Support for the importance of illumination realism for Psi was also found in [20, 21].

PI and Psi were also compared with respect to four features in [42] in the context of an outdoor scenario, though presented in a single screen with surround audio. The features were vision (none, colored lights, projection of the scene), sound (2D or 3D audio), haptics (none, simulated wind, simulated wind and heat) and olfaction (none, smells of the sea and a forest). The highest probability transitions configurations for PI were 3D audio, whereas

for Psi the wind. This makes sense because the scenario was a virtual journey through the sea and a forest, and realism would require wind. These results also fit with the current findings since the wind is analogous in the quartet study to the outside sounds. Just as in a room with windows and an open door to a country scene there would be an expectation of sounds from the outside, so moving through a forest and the sea would lead to an expectation of wind.

The study extends the methodology introduced in [7], but still has a number of limitations. The string quartet is a specific scenario and we cannot know the extent to which these findings would generalize to other scenarios. However, we have shown above that these findings do cohere with earlier theoretical and empirical work. The sample size is relatively small, although the total number of trials is large. Following [7] the method does not take into account potential intra-subject correlations, future work should do this, where a Bayesian approach to estimation would be most appropriate. The method is limited to the study of the range of configurations afforded by any particular immersive system - although this is always the case with any particular study. On the technical side it may be noticed that the hands of the virtual players did not correspond to the played notes. While the animations build on motion capture data, their creation involves a huge amount of manual craftwork in order to make the motion capture data presentable. Regarding the motion capture setup, attaching markers to the musicians' fingers would have significantly affected their performance. Accurate low-cost capture of fine-grained movements without massive manual intervention remains an important goal for motion capture technology.

6 Conclusion

While presence as 'being there' is a critical component of the VR experience, and there is no point to VR without it, it is not the only type of experience that VR can deliver and not the most difficult to attain. We would argue that the problem of attaining presence as Place Illusion has a solution with broad outlines known. As people are attesting with great surprise more and more these days - you put on a wide field-of-view head tracked (ideally 6 d.f.) HMD and you are 'there'. The more that 'real world' sensorimotor contingencies are afforded in VR the greater the likelihood that this will happen. For example, in [43] it was shown that adding greater sensorimotor contingencies (based on static haptics) made it more likely that participants would exhibit a stress response, as a behavioral correlate of presence. Of course research is still needed to understand the details and boundaries of this - with respect to latency, visual resolution (largely unexplored for technical reasons), the role of haptics, the role of sound, and so on. The parameters and their influence need to be fine-tuned, but the overall framework of knowledge is there. However, is there any utility in having the illusion of being in a place where nothing that happens is credible when it is supposed to be? For example, if you are supposed to be interacting with a virtual human but have no sense of plausibility of that character then the purpose of the VR scenario may be lost - irrespective of a strong sense of being in the virtual place. While a high degree of PI may be a necessary condition for the success of a VR scenario, it may not be sufficient in many applications. Hence it is important that resources are devoted to the study of Plausibility.

Including the study of Plausibility could help to unravel results that might otherwise be perplexing. For example, it was reported in [19] that contrary to expectations rendering quality (flat shading through to radiosity) did not influence a behavioral correlate of PI (a physiological response to stress). Two further studies then found that the difference in responses could be found with respect to Psi even though the levels of PI were unaffected by the differing rendering styles [20, 21].

Here we have specifically explored the space of a particular set of configurations of a system in order to see how different settings for sound and related features influence Psi. The scenario itself is unusual in VR (the quartet), and chosen because the sound is obviously a critical part of it. Concentrating on Psi we found that to deliver the illusion that the events were really happening participants tended to choose as most important two features that were not directly related to the quality of sound rendering - the gaze directions of the players following the participants, and sounds from outside the room. Moreover, amongst the two auditory parameters the one that corresponded more to realism was the more important than the one that corresponded more to a sensorimotor contingency (spatialized sound as a function of head movement).

Although the method we have used is not simple, we argue that its complexity reflects the complexity of what is being measured - a subjective illusion. It is possible to avoid the sole use of questionnaires by also employing physiological measures – e.g. [43] – but this requires introducing threat into the environment solely for the purpose of measurement. In using the method based on analogy with color matching theory we have concentrated only on plausibility, but the same method of course could be used for many other subjective correlates of a VR experience.

An additional point is that two decades ago Ellis [44] argued for a measure of presence that would include the notion of equivalence classes, so that designers could choose tradeoffs amongst features while maintaining a similar level of presence. Using questionnaires it is impossible to design such a measure, since '5' for one person might mean something completely different to '5' for another, and such ordinal measures cannot be combined with arithmetical operations. However, in the method introduced in [7] this problem does not arise, since even though each participant may have their own feeling for a matching configuration, it is precisely only *their* feeling that matters, and the method only relies on the *fact* that a match has been declared. If a person declares that their plausibility is the same in one configuration as in another, this is a fact that can be recorded. It does not require that someone else would make the same match, or even that the feelings that lead to a match need be similar across people. However, finding actual statistical regularities in the choices made across several participants provides evidence for a level of intersubjective agreement. Although we are all different, we tend to agree more than disagree on the types of configuration that might lead to a match. Table 3 provides an example of this, where equivalence classes naturally emerge from the matching probabilities. In that Table we can see 4 such classes: Row 1 is a singleton class, rows 2-5 with common probability of 0.25, rows 6-7 with common probability 0.14, and rows 8-11 with common probability close to 0.12. The method offers systems and application designers a tool to explore such tradeoffs.

One of the major applications of VR is in the field of entertainment and fantasy, where events and situations hardly conform to everyday expectations about how the world works. Participants will interact with prehistoric monsters and otherworldly aliens, thereby apparently violating the third requirement of Psi reviewed in the opening paragraphs of the Discussion. Yet Psi is not based on veracity or on the realism of the situation in itself. Each virtual world must establish its own set of rules and build new expectations. A new groundwork for Psi must be laid within the world itself. Problems can arise, however, when the virtual world is supposed to be a simulation of a real world situation. For example, in the scenario described in [45] participants witnessed a fight that broke out between two soccer supporters. In an early version of that scenario participants complained that soccer fans would never enter a bar decorated as shown in the virtual environment and therefore this reduced its plausibility. In the final version of the scenario the bar was made to look like a bar that would be attended by soccer fans. As another example, in [46] the participants were medical doctors who were confronted by virtual patients who inappropriately demanded antibiotics. The doctors complained that in reality their desk would always equipped with a computer screen displaying the medical record of the patient. In the scenario presented in this paper the situation was abnormal - a quartet playing in a room in a countryside setting, with a door open to the outside. Yet this did not trigger any queries or complaints about lack of realism amongst the

participants. Understanding Psi is not straightforward and there are many unanswered questions, in particular its boundaries.

There are tens of thousands of papers that include discussion of presence in the context of virtual reality or virtual environments and more than 200 with these terms in the title. In the overwhelming majority presence was restricted to the concept of 'Place Illusion'. Plausibility is a more challenging concept, representing highly complex interactions and relationships between the participants and events and situations in the virtual world. A similar effort of understanding and empirical work needs to be devoted to this concept.

ACKNOWLEDGMENTS

This research was supported by the European Union FP7 AAT project VR-HYPERSPACE (#285681) and by the European Union FP7-People project GOLEM (#251415).

REFERENCES

[1] R. M. Held and N. I. Durlach, "Telepresence," *Presence: Teleoperators and Virtual Environments,* vol. 1, pp. 109-112, 1992.

[2] T. B. Sheridan, "Musings on Telepresence and Virtual Presence," *Presence: Teleoperators and Virtual Environments,* vol. 1, pp. 120-126, 1992.

[3] J. M. Loomis, "Distal attribution and presence," *Presence: Teleoperators and virtual environments,* vol. 1, pp. 113-119, 1992.

[4] C. Heeter, "Being there: The subjective experience of presence," *Presence: Teleoperators and Virtual Environments,* vol. 1, pp. 262-271, 1992.

[5] M. Slater, "Place Illusion and Plausibility can lead to realistic behaviour in immersive virtual environments," *Philos Trans R Soc Lond,* vol. 364, pp. 3549-3557, 2009.

[6] J. J. Cummings and J. N. Bailenson, "How immersive is enough? A meta-analysis of the effect of immersive technology on user presence," *Media Psychology,* vol. 19, pp. 272-309, 2016.

[7] M. Slater, B. Spanlang, and D. Corominas, "Simulating virtual environments within virtual environments as the basis for a psychophysics of presence," *Acm Transactions on Graphics,* vol. 29, p. Paper: 92, 2010.

[8] E. Maestre, P. Papiotis, M. Marchini, Q. Llimona, O. Mayor, P. A., *et al.*, "Online Access and Visualization of Enriched Multimodal Representations of Music Performance Recordings: the Quartet Dataset and the Repovizz System," *IEEE Multimedia,* vol. in press, 2017.

[9] A. Noë, *Action In Perception*. Cambridge, MA: MIT Press, 2004.

[10] J. K. O'Regan and A. Noë, "A sensorimotor account of vision and visual consciousness," *Behav Brain Sci,* vol. 24, pp. 939-1031, 2001.

[11] T. B. Sheridan, "Further musings on the psychophysics of presence," *Presence: Teleoperators and Virtual Environments,* vol. 5, pp. 241-246, 1996.

[12] M. Slater and S. Wilbur, "A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments," *Presence-Teleoperators and Virtual Environments,* vol. 6, pp. 603-616, 1997.

[13] W. Barfield and C. Hendrix, "The Effect of Update Rate on the Sense of Presence within Virtual Environments," *Virtual Reality: The Journal of the Virtual Reality Society,* vol. 1, pp. 3-16, 1995.

[14] W. Barfield, K. M. Baird, and O. J. Bjorneseth, "Presence in virtual environments as a function of type of input device and display update rate," *Displays,* vol. 19, pp. 91-98, 1998.

[15] C. Hendrix and W. Barfield, "Presence within Virtual Environments as a Function of Visual Display Parameters," *Presence-Teleoperators and Virtual Environments,* vol. 5, pp. 274-289, 1996.

[16] R. B. Welch, T. T. Blackmon, A. Liu, B. A. Mellers, and L. W. Stark, "The effects of pictorial realism, delay of visual feedback, and observer interactivity an the subjective sense of presence," *Presence-Teleoperators and Virtual Environments,* vol. 5, pp. 263-273, 1996.

[17] M. Slater and M. Usoh, "Body Centred Interaction in Immersive Virtual Environments," ed: John Wiley and Sons, 1994, pp. 125-148.

[18] M. Slater, M. Usoh, and A. Steed, "Depth of Presence in Immersive Virtual Environments," *Presence-Teleoperators and Virtual Environments,* vol. 3, pp. 130-144, 1994.

[19] P. Zimmons and A. Panter, "The Influence of Rendering Quality on Presence And Task Performance in a Virtual Environment," in *Proceedings of IEEE Virtual Reality*, ed, 2003, pp. 293-294.

[20] M. Slater, P. Khanna, J. Mortensen, and I. Yu, "Visual realism enhances realistic response in an immersive virtual environment.," *IEEE computer graphics and applications,* vol. 29, pp. 76-84, 2009.

[21] I. Yu, J. Mortensen, P. Khanna, B. Spanlang, and M. Slater, "Visual realism enhances realistic response in an immersive virtual environment - Part 2," *IEEE Computer Graphics and Applications,* vol. 32, pp. 36-45, 2012.

[22] M. Meehan, S. Razzaque, M. C. Whitton, and F. P. Brooks Jr, "Effect of latency on presence in stressful virtual environments," *Virtual Reality, 2003. Proceedings. IEEE,* pp. 141-148, 2003.

[23] M. V. Sanchez-Vives and M. Slater, "From Presence to Consciousness Through Virtual Reality," *Nature Reviews Neuroscience,* vol. 6, pp. 332-339, 2005.

[24] C. Hendrix and W. Barfield, "The sense of presence within auditory virtual environments," *Presence-Teleoperators and Virtual Environments,* vol. 5, pp. 290-301, 1996.

[25] K. Bormann, "Presence and the utility of audio spatialization," *Presence: Teleoperators and Virtual Environments,* vol. 14, pp. 278-297, 2005.

[26] B. G. Witmer and M. J. Singer, "Measuring presence in virtual environments: a presence questionnaire," *Presence: Teleoperators and Virtual Environments,* vol. 7, pp. 225-240, 1998.

[27] S. Poeschl, K. Wall, and N. Doering, "Integration of spatial sound in immersive virtual environments an experimental study on effects of spatial sound on presence," in *IEEE Virtual Reality*, 2013, pp. 129-130.

[28] M. P. Snow and R. C. Williges, "Empirical models based on free-modulus magnitude estimation of perceived presence in virtual environments," *Human Factors: The Journal of the Human Factors and Ergonomics Society,* vol. 40, pp. 386-402, 1998.

[29] H. Q. Dinh, N. Walker, C. Song, A. Kobayashi, and L. F. Hodges, "Evaluating the Importance of Multi-sensory Input on Memory and the Sense of Presence in Virtual Environments," *Proceedings of the IEEE Virtual Reality,* pp. 222-228, 1999.

[30] C. Andre, J.-J. Embrechts, and G. V. Jacques, "Adding 3D sound to 3D cinema: Identification and evaluation of different reproduction techniques," in *Audio Language and Image Processing (ICALIP), 2010 International Conference on*, 2010, pp. 130-137.

[31] R. Mehra, A. Rungta, A. Golas, M. Lin, and D. Manocha, "WAVE: Interactive Wave-based Sound Propagation for Virtual Environments," *IEEE transactions on visualization and computer graphics,* vol. 21, pp. 434-442, 2015.

[32] A. Rungta, S. Rust, N. Morales, R. Klatzky, M. Lin, and D. Manocha, "Psychoacoustic characterization of propagation effects in virtual environments," *ACM Transactions on Applied Perception (TAP),* vol. 13, p. 21, 2016.

[33] T. Musil, M. Noisternig, and R. Höldrich, "A library for realtime 3d binaural sound reproduction in pure data (pd)," in *Proc. Int. Conf. on Digital Audio Effects (DAFX-05), Madrid, Spain*, 2005.

[34] F. Rumsey, "Spatial Audio. Music Technology Series," ed: Focal Press Oxford, 2001.

[35] T. Djelani, C. Pörschmann, J. Sahrhage, and J. Blauert, "An interactive virtual-environment generator for psychoacoustic research II: Collection of head-related impulse responses and evaluation of auditory localization," *Acta Acustica united with Acustica,* vol. 86, pp. 1046-1053, 2000.

[36] F. L. Wightman and D. J. Kistler, "Resolution of front–back ambiguity in spatial hearing by listener and source movement," *The Journal of the Acoustical Society of America,* vol. 105, pp. 2841-2853, 1999.

[37] J. A. Jones, J. E. Swan II, G. Singh, E. Kolstad, and S. R. Ellis, "The effects of virtual reality, augmented reality, and motion parallax on egocentric depth perception," in *Proceedings of the 5th symposium on Applied perception in graphics and visualization*, Los Angeles, CA, USA, 2008, pp. 9-14.

[38] S. Karlin, *A first course in stochastic processes*: Academic press, 2014.

[39] J. N. Bailenson, J. Blascovich, A. C. Beall, and J. Loomis, "Interpersonal Distance in Immersive Virtual Environments," *Personality and Social Psychology Bulletin,* vol. 29, pp. 1-15, 2003.

[40] M. Garau, M. Slater, V. Vinayagamoorthy, A. Brogni, A. Steed, and A. M. Sasse, "The Impact of Avatar Realism and Eye Gaze Control on the Perceived Quality of Communication in a Shared Immersive Virtual Environment," in *Proceedings of SIGCHI*, ed, 2003, pp. 529-536.

[41] W. Steptoe, R. Wolff, A. Murgia, E. Guimaraes, J. Rae, P. Sharkey, *et al*., "Eye-tracking for avatar eye-gaze and interactional analysis in immersive collaborative virtual environments," in *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, 2008, pp. 197-200.

[42] A. S. Azevedo, J. Jorge, and P. Campos, "Combining eeg data with place and plausibility responses as an approach to measuring presence in outdoor virtual environments," *PRESENCE: Teleoperators and Virtual Environments,* vol. 23, pp. 354-368, 2014.

[43] M. Meehan, B. Insko, M. C. Whitton, and F. P. Brooks, "Physiological measures of presence in stressful virtual environments," *Proceedings of SIGGRAPH,* vol. 21, pp. 645-653, 2002.

[44] S. R. Ellis, "Presence of mind: A reaction to Thomas Sheridan's "further musings on the psychophysics of presence"," *Presence-Teleoperators and Virtual Environments,* vol. 5, pp. 247-259, 1996.

[45] M. Slater, A. Rovira, R. Southern, D. Swapp, J. J. Zhang, C. Campbell, *et al*., "Bystander Responses to a Violent Incident in an Immersive Virtual Environment," *PLoS ONE,* vol. e52766. , p. doi:10.1371/journal.pone.0052766, 2013.

[46] X. Pan, M. Slater, A. Beacco, X. Navarro, D. Swapp, J. Hale, *et al*., "The Responses of Medical General Practitioners to Unreasonable Patient Demand for Antibiotics - A study of medical ethics using immersive virtual reality " *PLoS ONE,* vol. 11(2), 2016.