Disturbance and Plausibility in a Virtual Rock Concert: A Pilot Study

Alejandro Beacco* Ramon Oliva† Carlos Cabreira‡ Jaime Gallego# Mel Slater§
Event Lab, University of Barcelona
Barcelona, Spain

Figure 1: Two frames of the original DVD video footage of the concert (left), and two screenshots of our virtual recreation (right).

ABSTRACT

We present methods used to produce and study a first version of an attempt to reconstruct a 1983 live rock concert in virtual reality. An approximately 10 minute performance by the rock band Dire Straits was rendered in virtual reality, based on the use of computer vision techniques to extract the appearance and movements of the band, and crowd simulation for the audience. An online pilot study was conducted where participants experienced the scenario and freely wrote about their experience. The documents produced were analyzed using sentiment analysis, and groups of responses with similar sentiment scores were found and compared. The results showed that some participants were disturbed not by the band performance but by the accompanying virtual audience that surrounded them. The results point to a profound level of plausibility of the experience, though not in the way that the authors expected. The findings add to our understanding of plausibility of virtual environments.

Keywords: concert performance, computer vision, virtual reality, historical reconstruction, plausibility, presence.

Index Terms: I.3.7—Computer Graphics—Three-Dimensional Graphics and Realism—Virtual Reality

1 INTRODUCTION

Are events from the past forever lost, except through (often unreliable) memories? It is obvious that past events remain in the past and we can never directly experience them firsthand. Yet it may be possible to recreate past events, and live through them now. Virtual Reality (VR) offers that possibility through the extraction of 3D geometry, texture, motion and sound from extant video, and the exploitation of these to reproduce an immersive version of the past event. In this paper we report our first attempt to do this, concentrating on a rock concert that occurred in the 1980s. Our technical goal is to reproduce immersive versions of such concerts, although the methodology is general and may be applied to other events. The event we chose was Dire Straits performing 'Sultans of Swing' at the Alchemy concert in London in 1983 [1]. We report here on how we created the first version of this scenario and an exploratory study that aimed to provide early evaluation of participant experience.

Attempts to create concert experiences in VR go back to the 1990s. The creation of a virtual concert room is described in [2, 3], with accurate acoustics and sound auralization, that featured a flute player whose animation was controlled by an interpretation of MIDI output, where MIDI events were mapped into the motion of the player. There was also a version with 4 virtual musicians controlled by a real conductor. The user could experience the performance from different viewpoints, with appropriate changes in the auralization. Kragtwijk, et al. [4] also controlled a virtual drummer through the analysis of MIDI. Schertenleib, et al. [5] describe a system for the simulation of a virtual orchestra in a projection based VR system. A conductor could interactively modify the performance in real time. Zappi, et al. [6] created a multi-modal reality performance system featuring choreography that supported live performance. The system was hybrid in the sense that real and digital performances coexisted, and spectators could influence the performance through deployment of a basic audience tracking system. Bergström, et al. [7] displayed a string quartet where the characteristics of the auditory display and player behavior could be modified by participants in order to maximize aspects of their presence. Most recently Horie, et al. [8] described a virtual concert experience where participants influenced the course of the performance through their heart rate and brainwave signals, thus introducing a novel performance-audience physiologically based feedback loop.

This prior work has concentrated mostly on the display and interactive control of music performance, for example, controlling virtual characters playing musical instruments. The objective has

LEAVE 0.5 INCH SPACE AT BOTTOM OF LEFT COLUMN ON FIRST PAGE FOR COPYRIGHT BLOCK

* abeacco@ub.edu, equal contribution
† ramonoliva@ub.edu, equal contribution
‡ ccabreira@ub.edu, equal contribution
# jgallego@ub.edu, equal contribution
§ melslater@ub.edu, also with the Institute of Neurosciences, University of Barcelona, corresponding author

been mainly towards the means for creation or control of musical expression. The goal of our research is to recreate historical performances, with the prime aim of people having the experience of attending a concert. Hence in our case the music and performance are given, to the extent that there is sufficient recorded material of the event, and our goal is to create the illusion of presence in people of actually attending a live concert performance.

As was pointed out in [9] there is very little prior work that has attempted to reconstruct specific historical events in VR, although there are recent examples using augmented reality [10, 11]. However, the realizations of the scenario were manually constructed, even if based on photographs or paintings. In the work presented in this paper the reconstruction is of a scenario from recent history where video is available, and is partially automated.

Charron [12] in an analysis of why people continue to attend live concerts notwithstanding the vast amount of digital material available, argued that the prime reason is presence: being part of something special and unique with like-minded others who share the same interest in the band. Presence has been formalized as the sense of 'being there' in the situation and events depicted in the VR [13-15], and correspondingly the tendency to respond as if the situation were real [16]. It has been decomposed into two aspects: the perceptual illusion of 'being there' or 'Place Illusion' (PI) and the cognitive illusion that the events depicted are really happening, referred to as a 'Plausibility Illusion' (Psi) [7]. It was argued that PI depends on the VR affording natural sensorimotor contingencies [17], i.e., using the body for perception in much the same way as in physical reality (looking around, bending, reaching, moving the head), which has support in the meta study of [18], which found interaction through tracking to be a primary factor for presence. Psi is thought to depend on the extent to which the VR responds to participant actions, events that pertain personally to the participant, and how much the situation and events in the VR correspond to participant expectations based on real-world experiences, referred to as 'coherence' in [19].

There have now been several studies on Psi, for example [7, 19-24]. The distinction between PI and Psi is not clear cut, one study finding no substantial difference between them [21], while others have found that they are separable [20, 25, 26]. One problem is that different methodologies are used for assessment, so that results may not be comparable, and another is that participants may have completely different ways to express their experiences than the rigid boundaries imposed by a questionnaire. Since the concert experience we are attempting to evaluate is quite novel, with multiple aspects (the band, the setting, the audience, the sound, the fact that it represents a real event that had happened), rather than impose our own criterion on participants we asked them to write about their experiences suggesting some topics of focus – so that we did not impose categories such as PI or Psi on them with answers constrained to a Likert scale. Hence the study we undertook was exploratory, for us to find out how participants would respond, expressed in their own words.

In order to create the first VR version of the concert there were several aspects of technical development: (i) Starting from a DVD of the concert, appearance and movements of the performers had to be extracted in a form suitable for 3D display. (ii) A Crowd animation and rendering system had to be used to represent the audience. (iii) The concert hall had to be constructed. (iv) All had to be unified in one overall scenario combining the setting, the performers, the audience and the audio. (v) The participant had to be embodied in the scenario. Moreover, (vi) due to the coronavirus pandemic our participant-study had to be run remotely online, and therefore we specialized on the Oculus Quest 1, so that considerable effort had to be put into optimizing the scenario for this device. We discuss each of these in the following sections, followed by the participant-study.

2 Technical Methods

Initially we implemented the concert scenario with a high-end PC target in mind. But due to the coronavirus pandemic our pilot study had to be run remotely online on the Oculus Quest 1 device. Such hardware is not powerful enough to run our original reconstructed scene, so that optimization was necessary. In this section we describe how the concert scenario was constructed, focusing on the specific Quest 1 implementation. Figures 1-2 show images and video-q is available at https://youtu.be/ytuvfgCEWYw.

2.1 Visual Extraction

For the concert recreation we started from the video footage included in the live concert DVD "Alchemy" from Dire Straits (Figure 1 left). The members of the Dire Straits band were tracked using the method presented in [27]. This uses Convolutional Neural Networks (CNNs) to combine facial recognition, texture analysis and 2D poses and spatial information in order to track the main characters of the concert along the frames. Since the video footage is created from multi-camera video segments, we use a camera shot transitions detector to maintain the tracking even when the camera perspective changes.

Once the members of the band were tracked, we were able to reconstruct the displacement of the characters over the scenario, manually generating a root displacement curve for each one of them.

Along with the tracking process, we extracted the individual frames of each character. From these frames we chose the best ones that fulfilled the requirements of the 3D character reconstruction method that we used [28] that is, characters appearing completely with no cuts, from frontal and lateral views. Due to the poor resolution of the recorded footage, the resulting quality was still not high enough for the heads of the characters. We therefore used the AvatarSDK[1] software to generate the heads of the band members from high resolution pictures of the band members we found online. We then attached these high quality head meshes (that also included blendshapes) to the reconstructed bodies. Final appearance tweaks were carried out using the Substance Painter software[2], which resulted in a recognizable reconstruction of the Dire Straits band.

2.2 Crowd Animation and Rendering System for the Audience

In order to populate the concert venue with an audience, we designed a crowd animation and rendering system. While a close circle of characters around the participant were animated skinned meshes with reduced level-of-detail models, the remaining crowd was composed of impostors. We first used per-joint impostors [29] combined with instancing and palette skinning [30], but the Quest 1 hardware limitations made us switch to a more classic approach [31]. We added color variation to the different agents of the crowd.

Both impostors and close characters are animated by randomly playing animations from a curated selection of dance motions. In later versions we plan to adapt the speed of these animations to match the rhythm of the music, and include an extra up and down root motion to the characters based on the music beat.

[1] https://avatarsdk.com/

[2] https://www.substance3d.com/products/substance-painter/

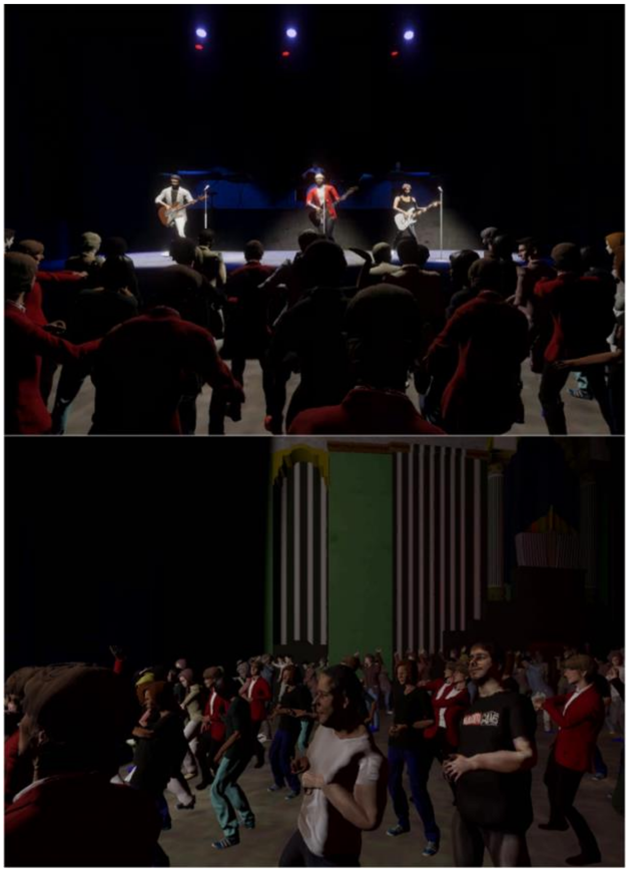

Figure 2: The concert (above) and the nearby audience (below).

2.3 The Concert Hall

To model the Hammersmith Odeon (now Apollo) concert hall, we used on-line photographs, and the DVD as a reference of how it was at the time of the concert. The model was effected by an artist using Blender[3] using approximate measures obtained online.

The final concert hall model consists of 15K triangles with very few materials using optimized texture atlases. Extended computer graphics techniques such as deferred lighting cannot be used on the Quest, so we had to stick to forward lighting and without any real time lights at all, leaving only 18 baked lights.

2.4 Integration

The whole concert scene was implemented with the Unity Engine (https://unity.com/). Only the Sultans of Swing track was processed. The voice of the lead singer was isolated and passed to the Oculus Lipsync module, to obtain visemes that were used to apply the corresponding blendshapes to the singer's facial animation.

2.5 Embodiment

The participant was embodied as a human avatar, so when looking down she or he would see a virtual body replacing their real one with synchronous correlated movements.

The pose of the upper body was estimated using Inverse Kinematics (IK) with the 6 DOF data provided by the Quest Head Mounted Display (HMD) and a tracked Quest controller on each hand. The position of the virtual elbow was estimated in a postprocessing stage, avoiding weird and impossible positions. The position of the knee was also estimated in a similar way.

[3] https://www.blender.org/

3 Exploratory Study

3.1 Goal

Since this type of scenario was new for us we wanted to discover the responses of people to being in the virtual rendition of the concert, so we did not deploy questionnaires. These would have reflected our own way of thinking rather than that of the people who naively experienced the scenario. Instead participants were asked to write a short description of their experience (ideally at least 150 words) concentrating on: Their feeling to be at a concert, their movements along with the crowd (e.g., dancing, clapping, cheering); aspects that drew them into the experience; aspects that drew them out of the experience, and any other comments.

3.2 Methods

3.2.1 Participants

A major problem was carrying out a study during a pandemic, where we could not meet participants face-to-face (indeed our laboratory was physically closed) and most especially we could not ask participants to use a head-mounted display shared by others. Through advertisement and personal knowledge we recruited people ('Helpers') who had their own Oculus Quest at home. Their instruction was not to take part in the experiment themselves, but rather recruit someone in their own home (or generally in their 'coronavirus bubble') to take part ('Participants'). The reason for this is that some of the volunteers were known to us and knew of our work, and we preferred people to take part who were naive to the application. Helpers who knew of our work and had discussed it with their potential Participants were excluded. In this way 22 people were recruited and results were obtained from 15. Reasons for not taking part by the deadline (one week after receiving the application) were eventual lack of time by the Participant, or technical problems with their Quest (unrelated to the application). A second method of recruitment was that one of the authors gave a class in an overseas institution, where all class members had their own Quest at home. The class was not a technical computer science class. The idea was that the students would have the experience, and then it would be discussed later in class. In this way a further 10 were recruited. We refer to the first sample as the Town and the second sample as the University.

The 25 participants were located in 5 different countries. However, 5 were excluded since they wrote under 100 words. The overall mean number of words for all 25 is 230 (SD 130). Hence the final sample size was 20, of whom 14 identified as female. Ages were 18-24 (7 participants), 24-34 (6), 35-44 (3) and 45-54 (4). There were 11 in the Town group (8 females) and 9 in the University group (6 females).

3.2.2 Procedures

An instruction sheet for download and installation of the application and further information about the study was sent by email. There was also a link to a Qualtrics (www.qualtrics.com) web page. This opened with an information sheet followed by a series of ethics questions giving participants the opportunity to withdraw or provide written informed consent. This was followed by questions about age and gender. The web page then asked the participants to don the HMD and start up the application and return to the questionnaire afterwards. Then the questionnaire asked participants whether they had completed the full experience (10:46 minutes), and asked them to write about it, ideally with at least 150 words as described above. Qualtrics automatically transmitted the responses to a secure server, with results visible only to the authors.

3.2.3 Sentiment Analysis

Since we had no fixed idea about how participants would respond in this novel virtual situation we chose a basic criteria to evaluate their responses: did they have more negative or more positive feelings during the experience, and why? Hence, the fundamental method of analysis was to derive a sentiment score for each document written by the participants using sentiment analysis [32, 33]. The sentiment scores were then automatically partitioned into clusters based on similarity between them, and then the documents in each subset examined to identify reasons for their similar sentiment scores.

The MATLAB 2020b 'Text Analytics Toolbox' was used to first obtain a pre-trained word-to-vector data base of 1M words representing each word as 300 dimensional vector [34]. Then a sentiment lexicon[4] is used to build a classifier of words to sentiment using a support vector machine (MATLAB function 'fitcsvm'). The classifier is then used to find the sentiment score for each word in a document (i.e., written by a participant) and the overall value of the document is taken as the mean of the sentiment scores of its constituent words. The sign and magnitude of a sentiment score indicates the strength and direction of its valence (i.e., positive scores indicate positive valence).

Having obtained the sentiment scores of all the documents, a k-means cluster analysis was used to partition the documents into subsets, where in each subset the score was similar. Finally documents with similar scores were inspected and compared for common themes.

Since this is an exploratory study we are not attempting to infer results to a wider population. Hence there are no significance tests, except as a way of indicating the strength of a relationship. Our interest focuses on how this sample of participants responded, and what lessons we can draw for future work.

4 Results

Bearing in mind that the scenario portrayed a rock concert, where the most that could go wrong might be participants picking out errors or inconsistencies, or finding it not believable, some of the results were quite surprising.

4.1 Sentiment Scores

Table 1. The sizes minima and maxima of the 4 clusters

Cluster	Size	Min	Max
1 Low	9	-0.034	0.189
2 Average	3	0.315	0.405
3 High	6	0.432	0.541
4 Very High	2	0.935	1.261

The sentiment scores (s) ranged from -0.034 to 1.261, with mean ± SD: 0.359 ± 0.317, median = 0.328, lower quartile = 0.141, upper quartile = 0.500. The sentiment scores are not significantly different between females and males, or between the Town and University groups, and are not correlated with age.

In order to subdivide the scores into subsets based on their similarity we used k-means clustering. The MATLAB function 'evalclusters' was used to choose the number of clusters, and two of the criteria ('silhouette', and 'DaviesBouldin') indicated 4 clusters as optimal, which is reasonable given the sample size, and by simple inspection of the scores.

Figure 3 shows summaries of the 4 clusters, illustrating clear separations between them, which is also shown in Table 1. Cluster 2 is referred to as 'Average' since the mid-point of the range (0.360) is almost exactly the mean of all the sentiment scores.

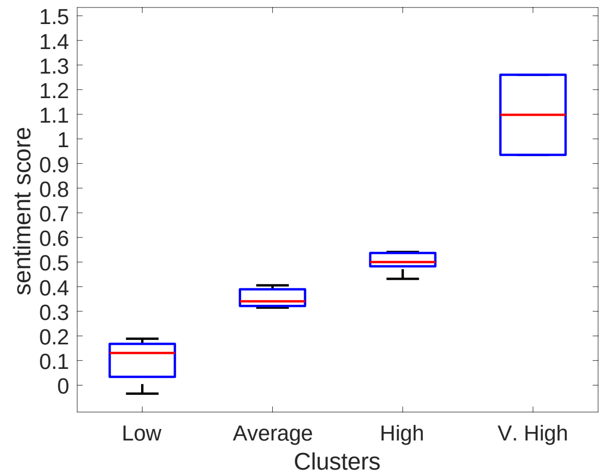

Figure 3: Box plots of Sentiment scores by cluster. The red lines are the medians, and the boxes the interquartile ranges (IQR). The whiskers show the ranges.

4.2 Analysis of Documents by Cluster

Each document was analyzed to identify general recurrent themes, and these are shown in Table 2. The proportions of words in each document corresponding to each theme were computed as a simple way to assess the predominance of the theme. These statistics were then classified by the clusters, in order to identify the major differences between them.

In order to find the most salient themes, we computed the proportion of words out of the total across all participants devoted to that theme, and also the Spearman rank correlation coefficients between the proportions in each theme and the sentiment scores. For example, 12% of all words written by all participants were classified under the theme *General+* shown in Table 2.

A theme is defined as *salient* if it contributed at least 10% of the total words, or if its correlation with the sentiment scores had significance at most 0.1 (note that these are not meant as significance tests but only as a way to assess the strengths of the correlations). The labels of the salient themes are highlighted in Table 2.

Table 2. Themes with % of total words in the theme, the Spearman correlation with sentiment score and significance.

Name	Example
General+: +ve statements **12%, 0.47, 0.04**	"The experience was certainly of being at a gig - it made me feel quite sad not to be able to go to any festivals this Summer."
Realism: In general, the lighting or movements (+ve) **4%, -0.05, 0.85**	"The people to my left and to my right were very real, looked very real. It was good, I thought it was the friend of a friend… in general the environment was very real".
Sounds: Crowd noise and movements (+ve) **6%, 0.40, 0.08**	"The audiences' applause. The music and when the vocalist is introducing the band."

[4] https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html#lexicon

Interaction: With characters (+ve) **3%, 0.26, 0.27**	"The ability to move was good, the interaction with other members of the crowd was interesting, those closest to me had individual ways of moving."
JoiningIn: Dancing, clapping, joining in (+ve) **6%, 0.00, 0.99**	"I was surprised at how immediately I started dancing and swaying like I was at a concert in real life. I listen to music all the time, I often sing along but don't tend to dance unless I'm seeing live music or out specifically for dancing."
General-: -ve statements **5%, -0.04, 0.86**	"Overall the whole concert experience was incredibly unnerving... It reminded me of how alone and strange I was to be placed in some nightmare like that."
Static: Static or repetitive movements of the characters (-ve) **9%, 0.15, 0.54**	"…the fact that everyone moved in a robotic fashion and the closest dancers looked like zombies drew me away from enjoying myself."
Hands: Participant's own virtual hands disturbing (-ve) **7%, -0.20, 0.40**	"A hand kept coming out of nowhere and it felt as something was trying to attack me. I think this was actually my own hands, but they were not mirroring what I was doing."
Disturbing: Crowd is disturbing (-ve) **13%, -0.56, 0.01**	"Two very creepy looking people on my right kept staring at me the whole time and that made me uncomfortable. It felt like they were watching me. They felt like 'predators'. I felt like if I kept dancing while looking at them they would come close to talk to me."
Expectation: Failure of expectation compared to real concert (-ve) **20%, 0.11, 0.66**	"I kept trying to interact with the people around me. I wish that I could have actually interacted with them somehow like high fiving or fist bumping or dancing with those around me…. there were times in the audio that it sounded like the audience was clapping but when I looked around, the audience was not visually clapping."
Errors: Implementation (negative) **1%, -0.38, 0.10**	"There was a lot of clipping issues with my hands and body, half of the time my wrist was distorting into weird positions or my hands got stuck mid-air. I also saw people being rendered on top of each other..."

Figure 4 shows the box plots of the salient themes of Table 2, and Table 3 the corresponding means and SEs. *Errors* is only non-zero in cluster 1 and therefore we do not show it in Figure 4. For cluster 1, which has the lowest sentiment scores, the predominant theme is *Disturbing*, where participants were disturbed by the crowd around them, some of them even fearing harassment. Cluster 1 has low scores on positive themes (*General+*, *Sounds*), and also high scores on failure of expectation (*Expectation*). It is the only cluster that includes the *Error* theme. Cluster 2 which has the next highest sentiment scores has scores on the positive themes, nearly 4 times lower on *Disturbing*, and about half of *Expectation* compared to cluster 1. Cluster 3 is about double cluster 2 on *General+*, but is also about double on *Expectation*. Finally cluster 4, the one with the highest sentiment scores, has the highest level of *General+*, equal highest level of *Sounds*, but is also the highest on *Expectation*.

What differentiates the lowest sentiment score (cluster 1) from the others is its high scores on *Disturbing* and *Expectation*. What differentiates the highest sentiment score (cluster 4) is the high *General+*, and *Sounds*, no *Disturbing*, but still high *Expectation*.

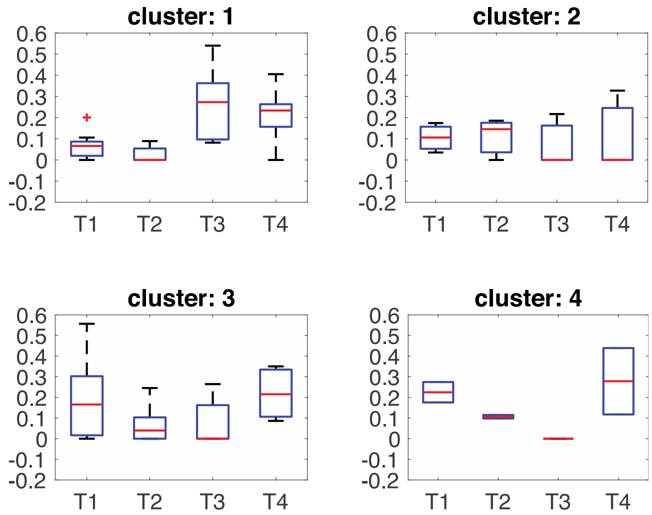

Figure 4: Box plots of the proportion scores of Table 2 by cluster. T1 = General+, T2 = Sounds, T3 = Disturbing, T4 = Expectation.

Table 3. Means ± Standard Errors of the proportions by cluster and theme

C	General+	Sounds	Disturbing	Expectation	Error
1	0.07 ± 0.014	0.02 ± 0.008	0.26 ± 0.039	0.21 ± 0.025	0.02 ± 0.007
2	0.20 ± 0.048	0.07 ± 0.022	0.07 ± 0.026	0.22 ± 0.025	0.00 ± 0.000
3	0.11 ± 0.016	0.11 ± 0.022	0.07 ± 0.028	0.11 ± 0.042	0.00 ± 0.000
4	0.23 ± 0.016	0.11 ± 0.003	0.00 ± 0.000	0.28 ± 0.051	0.00 ± 0.000

4.3 Further examples of the themes

Here we give more examples of the meaning of the themes through quotes from the documents.

General+: "The initial feeling of being in the room took me back. I felt quite immersed in the room and was excited to see the band come on stage. Once I saw the band come on, I enjoyed watching their movements and interactions."

"I think the project is really interesting in that famous concert audio can be enjoyed in a realistic environment. .... Classic historical concerts from top artists could be forever preserved and enjoyed for future generations with this technology."

"It was great! I liked the concert hall and the people in the crowd, everybody was partying…. I had a lot of fun. Thank you! And I liked it a lot that the musicians were dressed realistically, Knopfler with the red head band, and the suits, that was how I knew them. I appreciate that they looked like this!"

Sounds: "The most critical component that immersed me in the environment was the spatial audio. Audio sources were well-placed in the virtual space that I could accurately tell the position of each source. More importantly, when I rotated my head, they remained in the same place. It created a great sense of presence for me."

"The spatial sound actually makes it feel like at an actual concert, the audiences really helps this experience to be very interactive… The audiences' applause. The music and when the vocalist is introducing the band."

"I did enjoy the physical ambience, especially as I turned my head, the sound felt like it was always coming from the stage."

Disturbing: "Overall the whole concert experience was incredibly unnerving. Despite the lit up stage, my gaze was constantly looking around at the movement around me in the audience. I felt immersed, but not in that I was paying attention to the concert, but more so immersed in the weird nightmare I found myself in. … I just felt really creeped out by the out animated movements of those around me. The weird photo scanned people, all animating in seemingly random ways, further told me that I was alone. No natural thing would move or animate or look that way. In that sense the concert experience was quite lonely. It reminded me of how alone and strange I was to be placed in some nightmare like that."

"The feeling the experience gave me was representative of the actual feeling of being ALONE at a concert. … Even though I was standing pretty close the stage, the crowd around me was pretty sparse, and it felt intentional. It felt like the people around me (even though they didn't look it) were standing there to keep others away from me. … Four of the people surrounding me were wearing black suits and it was like they were my bodyguards."

"The solid eye contact of the man on my right was a little creepy, and as a woman, made me feel uncomfortable… The dancing of the two men seemed to resume when I looked at them but paused when I wasn't, which mimicked harassment episodes from the past, where men would pretend to be doing something else when I caught them gawking or ogling."

Expectation: "In addition, the performers didn't have any star power. They didn't interact with the crowd, and the crowd didn't interact with them, which was pretty weird. Also, there were some parts of the song when the drumming was more obvious, and it would have been nice to have the drummer synced up with that."

"The movements of the performers were a particular issue, since they are the focus at an event like this - they were too static, and the instrument playing animations were not right."

"The musicians' movements were not perfectly synced - my attention drew especially the drummer, who played much slower than the music. The biggest obstacle in enjoying the show as a real concert was probably the discrepancy between the high quality of music and the visual elements, especially agents around me."

5 LIMITATIONS AND FURTHER WORK

Since this was a pilot study there are several important issues to be considered that are beyond what we could tackle. First, we consider whether it the case that a virtual crowd is always as critical to participant responses as it is in this scenario. It is highly likely that in any scenario where there are crowds their appearance and behavior is will have important effects on individuals [35]. An interesting example concerns behavioral contagion. It was shown in [36] that participants are more likely to stare upwards when exploring a virtual environment if only 3 members of a crowd of 60 look upwards. However, it is important to consider which type of crowd behavior would be disturbing. In the rock concert the crowd were simply dancing, and not programmed to interact with the participants at all, they were doing 'normal' concert behavior, but still this generated negative affect for some participants. However, in another environment such as just walking through a street, in order to become particularly disturbing the crowd would have to be or act beyond normal expectancies. For example, imagine being in a street full of zombie-like creatures – even their just engaging in normal crowd behavior would already be disturbing. If the appearances of the characters were ordinarily human, then the actual behavior would have to be different to street norms – such as people repeatedly breaking personal distance. Hence considering the impact of crowds must take account of the actual context. We will specifically be studying crowd effects in future work.

Second, the situation we put participants in is a very rare one: attending a concert alone. Our current work is tackling this in two ways: using scanning technology we can make virtual characters that strongly resemble friends of the participants. These friends could either accompany the participant as automated virtual human agents, or be in a shared VR in real time while remotely located.

Third, such virtual humans that are close to the participants should respond personally to personally to them, such as smiling or even dancing in synchrony, a critical component of Psi [37] or generally of presence [38]. This was absent from the implementation used in this study, even though such interaction was inferred by some participants. This is considered further in the Discussion.

Fourth, consider whether the responses of participants might simply be due the Uncanny Valley effect [39], i.e., that the surrounding virtual characters were almost high fidelity human representations, since they were based on scans. We have no evidence of this either way, but of course it is a possibility that we cannot rule out. Some participants found that the characters were 'creepy' or like 'predators', or that they were 'staring', or had unnatural movements. It would be interesting to specifically investigate the Uncanny Valley effect, and this would be made possible nowadays with increasingly accurate scanning techniques.

Finally, given the nature of this pilot study, with a small sample size, and its exploratory nature, as mentioned we are not intending formal inferences and extension of our findings beyond this particular scenario and set of participants. For example, it is clear from Table 1, that although the clusters are very well separated in terms of their ranges of sentiment scores, some contain very few individuals. We saw this study as pointing the way for our future research, with lessons that may be of interest to VR researchers. Clearly we will not ourselves be able to follow up all the implications, but hope that others will be able to do so other scenarios.

6 DISCUSSION

The most interesting and useful studies are those where the results are surprising, when we can learn something new and unexpected. This is the case with the emergence of the *Disturbing* theme. From our point of view we had created our first ever version of a virtual rock concert based on a video record, where the virtual audience was natural part of the scenario. It never occurred to us that we had created what was for some a 'nightmare' (as one participant put it). Out of the 20 participants 12 devoted more than 8% of their answers to this theme, and 7 more than 20%. It accounted for 31% of the variance of the sentiment scores. If we had followed a standard design, with the VR exposure followed by a questionnaire, or even behavioral measures, we would have never discovered this. In VR studies we tend to think about measuring participant responses to our own pre-conceptions, such as presence. However, sometimes, especially when embarking on something quite new, we need to let participants express their experiences in their own way.

However, the fact that some participants were disturbed by the audience around them is itself a remarkable testimony to their presence in the scenario. Why should they be disturbed? Of course they all knew for sure that nothing real was happening, there was no band there, no audience, yet they responded in quite extreme ways – some women even fearing sexual harassment or being attacked, and the feeling of being alone in a crowd. Such presence requires both Place Illusion and Plausibility: the illusion of being in that place, and that the events happening around are real. It is difficult to conceive of how a sense of personal threat could occur without both of these illusions being active. This is very different to watching say a frightening movie. In such a case no one is likely to say, as one participant did: "In general terms, I felt at a concert but there was a moment when the avatars that were around me scared me a lot. I thought at some point they were going to attack me." Here Plausibility occurs because participants had the illusion, even if it was not programmed to be the case, that virtual audience members were interacting with (e.g., 'ogling') and paying attention to them personally.

The second most important observation is also concerned with Plausibility – the failure of expectation, what was referred to in [21] as 'coherence'. When participants have prior experience of how an event is supposed to be and this is broken, then Plausibility fails: "There wasn't a lot of action on the stage, so although I looked at the performers, they were not fully captivating." We had not completed extractions of all the animations of the performers from the video, so there were periods when they were relatively motionless. This does not match expectations of rock concerts. Conformity to expectations is the most difficult aspect of Plausibility, since it relies on detailed domain knowledge, and also is susceptible to individual differences. Designers of a scenario need to know in detail which particular aspects are crucial for a valid representation, and which are not so important. For example one participant wrote that the performers did not "… lean into the guitar licks", another complained that the light from the guitar strings was incorrect in relation to the emitted sound. As was found by [26] in a scenario with medical doctors, small discrepancies from practices in a consulting room (such as the unusual presence of a student in the room) can lead to a loss of Plausibility.

The third point concerns individual differences. Compare "I want to comment that this was great! I had a lot of fun", with "I felt immersed, but not in that I was paying attention to the concert, but more so immersed in the weird nightmare I found myself in". These are responses to identical stimuli. In order to build effective scenarios researchers will have to concentrate more on the issue of understanding individual differences and their impact – for example [40-42]. This probably involves building scenarios that adapt to individual differences that have been found to be important for the domain.

Related to this is that participants can perceive events and situations that did not happen, triggered by their own history and cued by the virtual surroundings. There were no black suited virtual characters standing around the participants. Or, "… the man on my right kept stopping dancing and it felt like he was just staring at me. Was this my friend I came with trying to dance with me? A creep at a concert, which is quite realistic as well." The virtual characters had not been programmed to look at let alone 'stare' at the participants. It could happen by chance, but just as in real life much of the interpretation of what we perceive, and even *what* we perceive, is internally generated, based on our own model of reality. As Stark [43] observed: virtual reality works because reality is virtual. Hence VR has been a very useful tool in the study and treatment of persecutory ideation [44], and generally in the domain of clinical psychology [45].

The Quest 1 version was developed out of necessity given the requirement to be able to run a study online due to the pandemic. However, our major work is for a version that is driven from the PC using any VR compatible GPU (starting from an NVidia GeForce 970) and any commercial headset (HTC Vive, Oculus Rift, or Oculus Quest through Oculus Link). This has several advanced features that are not possible on the Quest 1 - for example, a larger and more dense audience with higher quality animations, a higher quality dynamic lighting of the scene, and visual post-processing effects such as bloom and color grading. These are illustrated in Figure 5 and video-pc https://youtu.be/fOouvd75FIc.

Figure 5: The PC version of the concert program included higher quality graphics including dynamic lighting when the music starts with a large audience cheering. (The picture is dark because the scenario is).

7 Conclusion

We have reported on our first version of a system that has attempted to recreate a past rock concert in virtual reality. Our most important result is that we must expect the unexpected. From an ethical point of view we had thought that this was a completely safe environment, just about music. Some of the participants did not experience it that way. Although there were positive reactions to the experience, what stands out are the comments around the environment being disturbing, which has implications about how ethical cases for support should take such possibilities into account.

On the methodological side we have used a novel methodology for assessing the responses of participants to their experience. Instead of questionnaires, we encouraged participants to write about what happened, with only some suggestions about on what to concentrate. Through this way we were able to discover surprising results which would have been hidden had only Likert-based questionnaires or behavioral observations been used. We would argue that such an approach is particularly useful in the case of a new application or area of research.

Acknowledgments

This work is funded by the European Research Council (ERC) Advanced Grant Moments in Time in Immersive Virtual Environments (MoTIVE) #742989.

References

[1] Dire_Straits. Sultans of Swing. Accessed 22/10/2020. Available: https://www.youtube.com/watch?v=8Pa9x9fZBtY.

[2] R. Hänninen, L. Savioja, and T. Takala, "Virtual concert performance-synthetic animated musicians playing in an acoustically simulated room," *Proceedings of the International Computer Music Conference,* pp. 402-404, 1996.

[3] T. Takala *et al.*, "Concert performance in virtual reality," Acoustical Society of America, 1996.

[4] M. Kragtwijk, A. Nijholt, and J. Zwiers, "Implementation of a 3D virtual drummer," in *Computer Animation and Simulation 2001*: Springer, 2001, pp. 15-26.

[5] S. Schertenleib, M. Gutiérrez, F. Vexo, and D. Thalmann, "Conducting a virtual orchestra," *IEEE MultiMedia,* vol. 11, no. 3, pp. 40-49, 2004.

[6] V. Zappi, D. Mazzanti, A. Brogni, and D. G. Caldwell, "Design and Evaluation of a Hybrid Reality Performance," in *NIME*, 2011, vol. 11, pp. 355-360.

[7] I. Bergström, S. Azevedo, P. Papiotis, N. Saldanha, and M. Slater, "The plausibility of a string quartet performance in virtual reality," *IEEE transactions on visualization and computer graphics,* vol. 23, no. 4, pp. 1352-1359, 2017, doi: 10.1109/TVCG.2017.2657138.

[8] R. Horie, M. Wada, and E. Watanabe, "Participation in a virtual reality concert via brainwave and heartbeat," in *International Conference on Applied Human Factors and Ergonomics*, 2017: Springer, pp. 276-284.

[9] M. Slater *et al.*, "Virtually Being Lenin Enhances Presence and Engagement in a Scene from the Russian Revolution," *Frontiers in Robotics and AI,* vol. 5, p. 91, 2018.

[10] K. Jung, V. T. Nguyen, S.-C. Yoo, S. Kim, S. Park, and M. Currie, "PalmitoAR: The Last Battle of the US Civil War Reenacted Using Augmented Reality," *ISPRS International Journal of Geo-Information,* vol. 9, no. 2, p. 75, 2020.

[11] J. Challenor and M. Ma, "A Review of Augmented Reality Applications for History Education and Heritage Visualisation," *Multimodal Technologies and Interaction,* vol. 3, no. 2, p. 39, 2019, doi: doi.org/10.3390/mti3020039.

[12] J.-P. Charron, "Music audiences 3.0: concert-goers' psychological motivations at the dawn of virtual reality," *Frontiers in psychology,* vol. 8, p. 800, 2017.

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

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

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

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

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

[18] 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, no. 2, pp. 272-309, 2016.

[19] R. Skarbez, S. Neyret, F. P. Brooks, M. Slater, and M. C. Whitton, "A psychophysical experiment regarding components of the plausibility illusion," *IEEE transactions on visualization and computer graphics,* vol. 23, no. 4, pp. 1369-1378, 2017.

[20] 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, doi: 10.1145/1833349.1778829.

[21] R. Skarbez, F. P. Brooks Jr, and M. C. Whitton, "Immersion and coherence in a stressful virtual environment," in *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, 2018: ACM, p. 24.

[22] H. GalvanDebarba, S. Chague, and C. Charbonnier, "On the Plausibility of Virtual Body Animation Features in Virtual Reality," *IEEE Transactions on Visualization and Computer Graphics,* 2020.

[23] W. Bailey and B. Fazenda, "The effect of visual cues and binaural rendering method on plausibility in virtual environments," in *Audio Engineering Society Convention 144*, 2018: Audio Engineering Society.

[24] M. Hofer, T. Hartmann, A. Eden, R. Ratan, and L. Hahn, "The role of plausibility in the experience of spatial presence in virtual environments," *Frontiers in Virtual Reality,* vol. 1, p. 2, 2020.

[25] 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.

[26] X. Pan *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, no. 2, p. e0146837, 2016, doi: 10.1371/journal.pone.0146837.

[27] J. Gallego and M. Slater, "Multi-Person Re-Identification Based on Face, Pose and Texture Analysis in Unconstrained Videos," in *2020 IEEE 21st International Conference on Computational Problems of Electrical Engineering (CPEE)*, 2020: IEEE, pp. 1-4.

[28] A. Beacco, J. Gallego, and M. Slater, "Automatic 3D Character Reconstruction from Frontal and Lateral Monocular 2D RGB Views," in *2020 IEEE International Conference on Image Processing (ICIP)*, 2020: IEEE, pp. 2785-2789.

[29] A. Beacco, C. Andujar, N. Pelechano, and B. Spanlang, "Efficient rendering of animated characters through optimized per-joint impostors," *Comput. Animat. Virtual Worlds,* vol. 23, pp. 33-47, 2012.

[30] B. Dudash, "Skinned instancing," *NVIDIA White Paper,* 2007.

[31] F. Tecchia, C. Loscos, and Y. Chrysanthou, "Image-based crowd rendering," *IEEE Computer Graphics and Applications,* vol. 22, pp. 36-43, 2002.

[32] R. K. Bakshi, N. Kaur, R. Kaur, and G. Kaur, "Opinion mining and sentiment analysis," in *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, 2016: IEEE, pp. 452-455.

[33] B. Liu, "Sentiment analysis and opinion mining," *Synthesis lectures on human language technologies,* vol. 5, no. 1, pp. 1-167, 2012.

[34] T. Mikolov, E. Grave, P. Bojanowski, C. Puhrsch, and A. Joulin, "Advances in pre-training distributed word representations," *arXiv preprint arXiv:1712.09405,* 2017.

[35] J. Drury, "Recent developments in the psychology of crowds and collective behaviour," *Current opinion in psychology,* vol. 35, pp. 12-16, 2020, doi: doi.org/10.1016/j.copsyc.2020.02.005.

[36] E. M. Jorjafki, B. J. Sagarin, and S. Butail, "Drawing power of virtual crowds," *Journal of The Royal Society Interface,* vol. 15, no. 145, p. 20180335, 2018.

[37] A. Steed, Y. Pan, Z. Watson, and M. Slater, "“We Wait”—The Impact of Character Responsiveness and Self Embodiment on Presence and Interest in an Immersive News Experience," (in English), *Frontiers in Robotics and AI,* Original Research vol. 5, no. 112, 5, 2018-October-04 2018, doi: 10.3389/frobt.2018.00112.

[38] M. Kyriakou and Y. Chrysanthou, "How responsiveness, group membership and gender affect the feeling of presence in immersive virtual environments populated with virtual crowds," in *Proceedings of the 11th annual international conference on motion, interaction, and games*, 2018, pp. 1-9.

[39] M. Mori, "The Uncanny Valley," *Energy,* vol. 7, pp. 33-35, 1970.

[40] T. Iachini *et al.*, "The experience of virtual reality: are individual differences in mental imagery associated with sense of presence?," *Cognitive processing,* vol. 20, no. 3, pp. 291-298, 2019.

[41] M. Coxon, N. Kelly, and S. Page, "Individual differences in virtual reality: Are spatial presence and spatial ability linked?," *Virtual Reality,* vol. 20, no. 4, pp. 203-212, 2016.

[42] C. Chen, M. Czerwinski, and R. Macredie, "Individual differences in virtual environments—introduction and overview," *Journal of the American Society for Information Science,* vol. 51, no. 6, pp. 499-507, 2000.

[43] L. W. Stark, "How Virtual Reality Works! The Illusions of Vision in Real and Virtual Environments.," *Proc SPIE: Symposium on Electronic Imaging: Science and Technology,* pp. 5-10, 1995.

[44] S. Atherton *et al.*, "Self-Confidence and Paranoia: An Experimental Study Using an Immersive Virtual Reality Social Situation," *Behavioural and Cognitive Psychotherapy,* vol. 44, pp. 56-64, 2016, doi: 10.1017/S1352465814000496.

[45] D. Freeman *et al.*, "Virtual reality in the assessment, understanding, and treatment of mental health disorders," *Psychological Medicine,* pp. 1-8, 2017, doi: 10.1017/S003329171700040X.