

# Watching videos together in social Virtual Reality: an experimental study on user's QoE

Francesca De Simone\*<sup>1</sup>

Jie Li<sup>1</sup>

Henrique Galvan Debarba<sup>2</sup>

Abdallah El Ali<sup>1</sup>

Simon N.B Gunke<sup>3</sup>

Pablo Cesar<sup>1</sup>

<sup>1</sup> Centrum Wiskunde & Informatica, Amsterdam, the Netherlands

<sup>2</sup> Artanim Foundation, Meyrin, Switzerland

<sup>3</sup> TNO, The Hague, the Netherlands

## ABSTRACT

In this paper, we describe a user study in which pairs of users watch a video trailer and interact with each other, using two social Virtual Reality (sVR) systems, as well as in a face-to-face condition. The sVR systems are: *Facebook Spaces*, based on puppet-like customized avatars, and a *video-based* sVR system using photo-realistic virtual user representations. We collect subjective and objective data to analyze users' Quality of Experience (QoE) and compare their interaction in VR to that observed during the real-life scenario. Our results show that the experience delivered by the video-based sVR system is comparable with real-life settings, while the puppet-based avatars limit the perceived quality of the interaction. Our protocol for QoE assessment is fully documented to allow replication in similar experiments.

**Index Terms:** I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Virtual reality H.5.2 [Information Interfaces and Presentation]: User Interfaces—Evaluation/methodology

## 1 INTRODUCTION

A social Virtual Reality (sVR) system is an application that allows multiple users to join a collaborative Virtual Environment (VE) and communicate with each other, usually by means of audio and visual cues. The VE can be a computer-generated 3D scene or a 360-degree natural scene. Each user is represented in the VE as a computer-generated avatar or, in recently proposed systems, with a virtual representation based on live captures [6].

A variety of studies exists in the literature, where the impact of specific factors of the sVR system design on user's Quality of Experience (QoE) has been analyzed [2, 7, 8]. Nevertheless, at the best of our knowledge, no previous study analyzed the communication behaviour between users in a VE that is more than a simple computer-generated space, in combination to customized avatars.

## 2 EXPERIMENT

In our experiment, pairs of users watch a video trailer using two different sVR systems, as well as sitting together in an actual room. The scope of the experiment is twofold: to collect users' QoE feedback and to explore the interactions taking place between two users watching videos together via a sVR platform.

**sVR systems** In the VR case, the two users sit in the same VE, where they can have audio and visual interactions while watching a video together on a virtual screen. Two systems are considered:

- The *Facebook Spaces* system (FS) [1], where each user is depicted as a half-body cartoon-like customizable avatar. The

\*e-mail: f.de.simone@cw.nl



Figure 1: View by one user in (a) the Facebook Spaces and (b) the video-based sVR system [6].

avatar is personalized by the facilitator before the experiment starts, to look like the user (Fig. 1(a)).

- A *video-based* sVR system (VB), where each user's 2D real image is captured and segmented, by means of a Kinect sensor, while the user is wearing an Head Mounted Display (HMD) (Fig. 1(b)). The system is described at [6].

The same 360-degree background image is used as VE in both systems, while some system design differences cannot be modified. In the FS system: users wear hand controllers that allow seeing their own virtual hands; the avatars appear as sitting on virtual chairs around a virtual table; the virtual screen appears at a fixed position on the other side of the table; the video playback is started by one user, using a virtual touch player interface. In the VB system: users do not hold/wear any controller and do not see any part of their own body; the users' representation appears as sitting on a couch; the virtual screen is covering the wall in front of the users; in the virtual user representation the HMD is visible and occluding a large part of the user's face; the video playback is started by an operator external to the VE. In both cases, each user sits on a chair fixed to the floor in two separate rooms, wearing an Oculus Rift HMD and noise-cancelling headphones. The two rooms are isolated controlled environments, with no background noise, each operated by one facilitator.

**Face-to-face** In the face-to-face case (F2F), the users sit on two separate chairs in the same conference room with a screen in front of them. Both users are recorded using a webcam. The room layout is similar to that of the 360-degree background image used in the sVR systems.

**Experimental design** We apply a within-subjects design: one pair of users experiences all three conditions, i.e., FS, VB and F2F, watching a different action/science fiction movie trailer in each condition. The three trailers have approximately the same number of views on YouTube. A fully counter-balanced test design is applied so that each condition is experienced first the same amount of times across the user set and the associated trailer-condition is also balanced (i.e. each trailer is associated to each system the same number of times). The users' audio and visual interactions,

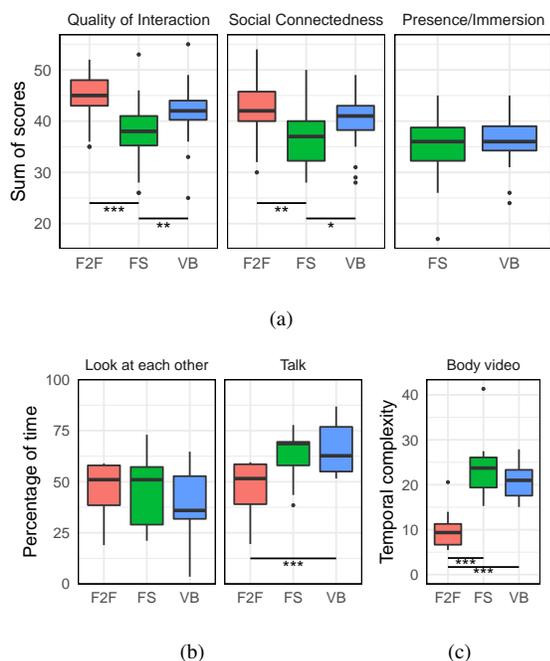


Figure 2: Box plot of: (a) the questionnaire subjective scores for face2face (F2F), Facebook Spaces (FS) and video-based sVR (VB); (b) the percentage of time spent talking to and looking at each other; (c) the temporal complexity [4] of the videos recording users’ bodies.

including the HMD viewports for the sVR sessions, and recordings of their body movements, are captured. After testing each system, users are asked to fill in a questionnaire designed for the purpose of our test based on an adaptation of the questionnaire described and evaluated at [5]. The questionnaire focuses on three main factors: the *quality of interaction*, the *social connectedness*, and the sense of *presence/immersion* experienced by the users. At the end of the entire experiment, users undergo a semi-structured interview. Our questionnaire and all forms used to collect feedback from users can be found at doi:10.5281/zenodo.2572867.

### 3 RESULTS

We recruited 16 pairs of paid users (average age 31.06, std 7.39, 17 women and 15 men). Users in each pair knew each other.

**Questionnaire** Fig. 2(a) shows the boxplot of the subjective scores collected via the questionnaire for the factors quality of interaction, social connectedness and presence/immersion. Presence/immersion was only assessed in the two sVR conditions. Via Friedman test and Wilcoxon signed-rank test we found a significant effect of the system condition on Quality of Interaction ( $p < .001$ ), with FS < F2F and FS < VB ( $p < .001$  and  $p < .003$  respectively) and on Social Connectedness ( $p < .01$ ), with FS < F2F and FS < VB ( $p < .009$  and  $p < .05$  respectively).

**Semi-structured interviews** Audio recordings of the semi-structured interviews were transcribed and coded by two researchers, following an open coding approach. As main outcomes, almost half of the participants (47%) expressed concerns that the avatars in the Facebook system were not realistic. Half of the participants (50%) preferred the video based system for activities such as watching a movie. Most of the participants notified that the sVR systems should to be improved by: providing better body representations (22%); a more comfortable HMD (28%); including multi-sensory experience (9%) and expanding the field of view of the HMD (13%).

**Objective data** Fig. 2(b)-(c) shows the box plots for the percent-

age of time spent looking at each other and talking to each other, over the entire duration of the experience, as well as the boxplot of the temporal complexity [4] of each video recording of the users’ bodies, for each condition. We conducted statistical analysis per dependent variable. The Shapiro-wilk test failed to reject the normality of the distribution of the residuals of the models. Thus, the data was analyzed using the repeated-measures ANOVA and multiple comparison test, to check if statistical significance is found. We found a significant effect of condition on the percentage of time spent talking at each other ( $p < .003$ ), with F2F < VB ( $p < .003$ ). In terms of body movements, we found a significant effect of condition, with F2F < VB and F2F < FS ( $p < .001$  in both cases).

### 4 CONCLUSION

Our results (Fig. 2(a)) indicate that, in terms of subjective quality of interaction as well as social connectedness, the video-based sVR system could provide an experience comparable to the actual face-to-face one. The puppet-based avatars limited the perceived quality of interaction. This finding appears to be in-line with the results of existing studies in literature that correlate avatar’s realism to the quality of the communication in mediated communications [3, 7]. From the collected objective data (Fig. 2(b)-(c)), it appears that users talked more in the sVR conditions with respect to the face to face condition. This could be explained by the fact that the novelty of the sVR gave them another reason to talk. Indeed by listening to their conversations, in the sVR conditions the users did talk about the VE and their own avatar appearance. Also, users moved significantly more in the sVR conditions rather than in the F2F experience. This could be explained by the fact that having a limited field of view when wearing the HMD, they did move more their heads. The novelty effect of being in a VE, bringing the user to visually explore the surrounding scene, and the discomfort linked to wearing the HMD could also be causes for more body movements. Future work will focus on extending the analysis of the collected data, as well as studying the correlation between objective and subjective results.

### ACKNOWLEDGEMENTS

This work was funded by the European Commission as part of the H2020 program, under the grant agreement 762111, *VRTogether*.

### REFERENCES

- [1] Facebook spaces. <https://www.facebook.com/spaces>. Accessed: 2019-02-19.
- [2] R. Fribourg, F. Argelaguet, L. Hoyet, and A. Lécuyer. Studying the sense of embodiment in VR shared experiences. In *IEEE VR 2018*, pp. 1–8, 2018. doi: 10.1109/VR.2018.8448293
- [3] P. Heidicker, E. Langbehn, and F. Steinicke. Influence of avatar appearance on presence in social VR. In *IEEE 3DUI 2017*, pp. 233–234, 2017. doi: 10.1109/3DUI.2017.7893357
- [4] ITU-T Rec. P.910. Subjective video quality assessment methods for multimedia applications. 1999.
- [5] J. Li, Y. Kong, T. Roggla, F. De Simone, S. Ananthanarayan, H. de Ridder, A. El Ali, and P. Cesar. Measuring and understanding photo sharing experiences in social Virtual Reality. In *(to appear) Proc. of ACM CHI 2019*.
- [6] M. J. Prins, S. N. B. Gunkel, H. M. Stokking, and O. A. Niamut. TogetherVR: A framework for photorealistic shared media experiences in 360-degree VR. *SMPTE Motion Imaging Journal*, 2018.
- [7] H. J. Smith and M. Neff. Communication behavior in embodied Virtual Reality. In *Proc. of ACM CHI 2018*, pp. 289:1–289:12, 2018. doi: 10.1145/3173574.3173863
- [8] T. Waltemate, D. Gall, D. Roth, M. Botsch, and M. E. Latoschik. The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response. *IEEE Trans. on Visualization and Computer Graphics*, 24(4): pp. 1643–1652, 2018. doi: 10.1109/TVCG.2018.2794629