Virtual Reality
https://doi.org/10.1007/s10055-020-00472-4

ORIGINAL ARTICLE

Social viewing in cinematic virtual reality: a design space for social movie applications

Sylvia Rothe[1] · Alexander Schmidt[1] · Mario Montagud[2] · Daniel Buschek[3] · Heinrich Hußmann[1]

Received: 6 July 2019 / Accepted: 24 September 2020

Abstract
Since watching movies is a social experience for most people, it is important to know how an application should be designed for enabling shared cinematic virtual reality (CVR) experiences via head-mounted displays (HMDs). Viewers can feel isolated when watching omnidirectional movies with HMDs. Even if they are watching the movie simultaneously, they do not automatically see the same field of view, since they can freely choose their viewing direction. Our goal is to explore interaction techniques to efficiently support social viewing and to improve social movie experiences in CVR. Based on the literature review and insights from earlier work, we identify seven challenges that need to be addressed: communication, field-of-view (FoV) awareness, togetherness, accessibility, interaction techniques, synchronization, and multiuser environments. We investigate four aspects (voice chat, sending emotion states, FoV indication, and video chat) to address some of the challenges and report the results of four user studies. Finally, we present and discuss a design space for CVR social movie applications and highlight directions for future work.

Keywords Cinematic virtual reality · Omnidirectional video · 360° video · Social viewing · Interactive TV

1 Introduction

Omnidirectional movies are attracting widespread interest and have many possible applications, e.g. telling stories about exciting locations and experiences in the world, or documenting places of historic interest. In cinematic virtual reality (CVR) the viewer watches omnidirectional movies using head-mounted displays (HMD) or other VR devices. Thus, the viewer can feel immersed in the scenes and freely choose the viewing direction. The drawback of these systems is the associated visual and mental separation from other people. Natural elements of discussion, like pointing at interesting objects in the video or keeping the awareness about what others are focusing on, is impeded by the HMD.

In contrast to spatial presence (the "sense of being there"), social presence describes the "sense of being together" (De Greef and IJsselsteijn 2001). Several definitions for both terms are used in the literature and they can be measured in different ways (Skarbez et al. 2017). Social Presence depends on communicative signals such as voices and non-verbal cues (IJsselsteijn et al. 2000), on unfocused and focused interaction (Schultze and Brooks 2019), as well as on the task type and several other aspects (Oh et al. 2018).

In this work, we identify key challenges and related design dimensions that are crucial for efficiently supporting social presence when watching CVR movies together. We provide an overview of the current state-of-the-art in this area (Sect. 2) and identify seven open challenges (Sect. 3). For these challenges, we propose potential approaches (Sect. 4), evaluate them in user studies (Sect. 5) and discuss the obtained results (Sect. 6). In our user studies, paired

✉ Sylvia Rothe
sylvia.rothe@ifi.lmu.de

Alexander Schmidt
alexander.x.schmidt@gmail.com; schmidt.al@campus.lmu.de

Mario Montagud
mario.montagud@i2cat.net

Daniel Buschek
daniel.buschek@ifi.lmu.de

Heinrich Hußmann
hussmann@ifi.lmu.de

[1] Institute of Informatics, Ludwig-Maximilians-University Munich, Munich, Germany

[2] Universitat de València & i2CAT Foundation, Valencia, Spain

[3] Research Group HCI + AI, Department of Computer Science, University of Bayreuth, Bayreuth, Germany

Published online: 17 October 2020

participants simultaneously watched omnidirectional movies via HMD. Four different communication *components* were added and compared: voice chat, video chat, sending emotion states and informing about each other's Field of View (FoV). For each of these components, two different approaches (*methods)* were implemented and compared to each other. The results provide insights into which components are important and how they can be combined. In our experiments, the most important components for the social experience were "sending emotion states" and "voice chat". "FoV indication" and "video chat" were perceived as less important.

Based on our user studies, we derive a design space for social viewing applications for CVR, which describes several dimensions and sub-dimensions (Sect. 7). With this design space, we discuss our approaches and suggest future research directions. The design space supports the development of applications for social viewing in CVR and assists in finding important issues for the development of social movie players and experiences for CVR.

2 Related work

Social viewing in CVR is very close to the research topic of collaboration in VR. It is important to inspect to what extent the results of VR research can be transferred to CVR and which of the methods of collaboration can also be used for social viewing. Furthermore, questions and results for social viewing of traditional films have to be checked for their applicability.

2.1 Social viewing of movies/watching movies together

Much research in recent years has been focused on the relevance of social viewing scenarios (Harboe et al. 2008b; Nathan et al. 2008; Shin and Kim 2015; Voorveld and Viswanathan 2015). These studies indicate a need for further research to efficiently enable such scenarios. Previous research works have investigated social aspects in shared video watching scenarios (Kim et al. 2018; Shin and Kim 2015). A number of different approaches explored the communication between people who are watching television together in different locations. For example, 2BeOn (Abreu et al. 2002) provides TV viewers with online communication services such as instant messaging and videoconferencing. Amigo TV enables shared TV watching in different locations complemented by voice chat, text chat, and individual emoticons (pictures, audio, video). SocialTV (Harboe et al. 2008a) and SocialTV 2 (Harboe et al. 2008b) indicate which TV show the other group members are watching and allows the exchange of messages between the members.

Weisz et al. (2007) integrated text chat in a social viewing scenario. This had a positive influence on the social relationships between the viewers, but they got distracted while chatting and watching a video simultaneously. Adding natural break periods in the video could reduce the feeling of distraction. In our work, we want to restrict the chat possibilities to voice and video chat, because our input device should be simpler than a keyboard, due to the use of HMDs. Furthermore, our assumption is that the displayed text could reduce the enjoyment of the CVR social viewing experience. Geerts et al. (Geerts et al. 2011) investigated the influence of voice and text chat modalities. They found out that participants feel closer together when using voice chat. Inspired by this, we want to examine how different communication channels such as video and voice chats can be adapted to CVR applications.

Watching omnidirectional videos together was investigated by Tang et al. (Tang and Fakourfar 2017) for close-by viewers. In their experiments, participants used tablets for watching the movie in the same room. It was discovered that participants observed others' physical movements to infer the viewing direction. This strategy is not applicable when wearing HMDs.

2.2 Synchronization

Synchronization of the media playout across the involved devices is a key requirement in social viewing to time-align the playout processes of all involved devices (Boronat et al. 2018; Montagud et al. 2012). This includes designing and adopting the appropriate communication and control protocols, monitoring algorithms, reference selection strategies, and adjustment techniques. Likewise, media synchronization must be preserved after issuing navigation control commands, e.g. play, pause, seek, in a shared session. Wersync (Belda et al. 2015; Montagud et al. 2015) is a web-based platform for distributed media consumption, integrating synchronization and social interaction features between remote users. An integrated text chat tool was implemented for communication in Wersync. Since typing text is more difficult in VR devices, other communication techniques need to be investigated.

2.3 Video players for omnidirectional movies

In CVR viewers see only a part of the omnidirectional movie on the display. Therefore, in some cases guiding methods are necessary so the viewer does not miss important details (Mateer 2017; Nielsen et al. 2016; Rothe et al. 2019). For this established filmmaking techniques can be used, for example sounds, lights and movements (Rothe and Hußmann 2018).

Video players for watching omnidirectional movies on monitors of desktop PCs were studied by Chambel et al. (2011) and Neng and Chambel (2010). They introduced techniques to display the position of the own FoV in the full omnidirectional image for a better orientation. In CVR, the orientation is easier, since the viewers have information about their own direction by sensor information. However, the used techniques could also be suitable to indicate the FoV of the co-watchers in social viewing scenarios in CVR.

Matos et al. (2018) implemented a 360° video player with several dynamic annotation methods: marker, subtitle, miniature, arrow, vignette, lateral light, and minimap. Depending on the context of the video, the methods had resulted in different effects. All of these methods are worth checking for their transferability to social viewing. Montagud et al. (2018) developed an accessible-enabled 360° player. It allows the consumption of 360° video and spatial audio augmented with a hyper-personalized presentation of access services (subtitles, audio description, and sign language interpreting). Different guiding methods are included to assist the users in finding the target speaker or action in the 360° area when access services are enabled, like arrows, a radar, and an auto-positioning mode.

2.4 Virtual togetherness, social presence

In contrast to the sense of being part of the virtual environment (spatial presence), the sense of being together in a virtual world (social presence, virtual togetherness) assumes the presence of other persons. Virtual togetherness is influenced by the sense of being in the virtual world and the communication between the users in the virtual world (Durlach and Slater 2000a). Oh et al. (2018) present an overview of definitions and concepts of social presence. Based on 152 studies, several factors are categorized and discussed in that work. The results show that depth cues, audio quality, haptic feedback, and interactivity can increase social presence. It is emphasized that social presence not only depends on the environment but also on the persons involved in the process and the task. Haptic communication in shared virtual environments can improve the sense of being together (Ho et al. 1998). In contrast, the studies in this paper pay particular attention to the impact of aural and visual communication on the social experience of watching a CVR video together.

For visual communication body postures play an important role (Durlach and Slater 2000b). De Simone et al. (2019) compared watching videos together in three conditions: face-to-face, facebook space (Facebook 2019), where the user is presented as cartoon-like customizable avatar, and a video-based social VR system, where the user is presented by a 2D real video-based image. In their study, the video-based condition could provide an experience comparable to the face-to-face one, for the subjective quality of interaction and for social togetherness. The perceived quality of interaction was lower for the avatars. However, presence and the quality of communication depends on the realism of the avatar (Heidicker et al. 2017; Smith and Neff 2018; Waltemate et al. 2018).

2.5 Collaboration in VR

The topic of collaboration in VR was explored already in the 1990s (Carlsson and Hagsand 1993; Margery et al. 1999; Normand et al. 1999). Projects like DIVE (Distributed Interactive Virtual Environment) (Carlsson and Hagsand 1993) and COVEN (COllaborative Virtual ENvironments) (Normand et al. 1999) laid the foundation for today's research on collaboration in VR.

Cordeil et al. (2017) found no major differences between CAVE and HMD regarding verbal communication and shared focus when checking 3D network data together. Leap Motion sensors were used for showing points of interest (*PoI*), so the collaborators could see their partners' fingers. Additionally, the *FoV* of each user was displayed. In their experiments, users solved tasks faster using HMD. Nguyen et al. (Nguyen et al. 2017) introduced CollaVR, a tool for filmmakers, that allows a shared inspection of 360° scenes via HMDs. Voices and visualization of each other's FoV are used for interaction. A rectangularly framed FoV is visible if the gaze directions of the viewers are close enough, which means the two FoV overlap. Otherwise, an arrow is displayed for indicating the direction of the framed field of the collaborator's view. CollaVR is implemented for the communication of people working on the movie, not for the end user's consumption experience. Mouse and keyboard are used as input devices, and a graphical interface shows the timeline and buttons for the included features. In our work, we apply various notification methods for indicating *PoIs*, with a less intrusive visualization, in order to minimally interfere with the viewing experience. Another example of collaboration is VR video conferencing, which was investigated by Gunkel et al. (2017, 2018).

Dorta et al. (2016) compared the social experience of watching a movie together using a walk-in system and VR headsets. They concluded that headsets induce a higher sense of presence, but make the communication between the viewers more complicated. One reason for this was the difficulty of knowing where the other person is looking at. Even if walk-in systems seem more suitable for social CVR, they are rarely available and only applicable to public or exhibition-based spaces, not to domestic scenarios.

The above findings from other research fields are important to understand the behaviour of the audience and to identify and address the challenges for social viewing in CVR.

3 Challenges for social viewing in CVR

In contrast to traditional cinema or TV, each CVR viewer has its own display and gets isolated from the surrounding environment when watching a movie via HMD. Based on the literature review and on our own expertise in this field, we have identified seven key challenges to support social viewing in CVR, which we plan to investigate. While further challenges may exist, these seven challenges are important for a first design approach.

Challenge 1 - Communication: A key issue in a social viewing scenario for traditional movies is the communication channel used for users' interaction. This also applies within the CVR landscape. Voice interaction is essential in social viewing, as explicated in Sect. 2. However, even if co-located viewers can communicate via voice without any additional implementations, no one knows where the other viewers are looking at within the 360º environment. How can a viewer indicate details in the movie?

Challenge 2 - FoV Awareness: One of the main problems for social viewing via HMDs is the difference between the users' FoV and the missing awareness of the other's FoV. What FoV is a user's comment referred to? Why is the co-watcher laughing? How can a viewer indicate details in the movie that are not necessarily in the others FoV? Being unaware of where co-watchers are looking at within the omnidirectional scenes can be a communication barrier and lead to difficulties of understanding.

Challenge 3 - Togetherness: Another challenge for social viewing via HMDs is to provoke the feeling of "being together"—i.e. of not watching a movie in isolation. When watching a movie together in the cinema or TV, the co-watcher is perceived in the periphery of the view. Even though "silent" feelings (e.g. sadness) cannot be heard by the other user, they can be recognized or inferred by postures, gestures or facial expressions. This is currently not possible in CVR.

Challenge 4 - Accessibility: Watching CVR movies together should also be possible for people requiring add-on access services, e.g. subtitles/sign language for the deaf and hearing-impaired or audio description for blind or visually-impaired persons. A social CVR movie application should provide these possibilities.

Challenge 5 – Interaction Techniques: For being aware of each other, some interaction methods are necessary. It is important to keep in mind that these methods should not destroy the viewing experience. For example, graphical menus or keyboards in the virtual world can reduce the feeling of presence and immersion, as well as interfere with the media consumption experience. The main form of interaction in CVR, looking around with the HMD to select the image section, results in a very natural interaction mechanism. However, for social viewing, further techniques are needed, which should also be natural and not disturbing. Non-intrusive input techniques are required for communication and navigation.

Challenge 6 - Synchronization: When viewing movies together, each user has to be on the same timecode at every moment. Even if one of them plays the movie back or forward, the other person has to instantaneously see and hear the same to provide a consistent experience. By providing this, all involved users will perceive the same events at the same time, thus preventing inconsistent interaction and frustrating situations, e.g. cheering of a remote friend when a goal is scored, before the goal sequence is actually shown on the local display. Similarly, shared navigation, i.e. execution of playout control commands, will enable more interactive experiences, e.g. watching together the repetition of a scene, pausing the video to discuss specific issues.

Challenge 7 - Multiuser Scenarios: There are various scenarios for social viewing in CVR. The most obvious—two persons are watching an omnidirectional movie together via HMDs—is examined more closely in our user studies. However, there are use cases with more than two persons, e.g. educational experiences or presentations in museums. A situation similar to a traditional cinema is conceivable as well. In such a constellation, visual and aural information of all users can cause overloading. Other issues arise from asymmetric settings: In social viewing, the involved participants may use different types of VR devices or even participate by using a desktop. In addition, the heterogeneity of devices has to be considered when designing interactive and social CVR platforms and experiences. There are possible scenarios where not all viewers watch the movie via an HMD.

4 Approaches to address the challenges

Each of the identified challenges requires appropriate design guidelines and insights to enable satisfactory social CVR experiences. In this section, we present four approaches to address some of the challenges. For each approach, we discuss several methods. For some of them, we present user studies in Sect. 5 and discuss how they meet the challenges (Sect. 6). These are important steps for determining the dimensions of the design space for social viewing in CVR (Sect. 7).

4.1 Voice chat

Speech can transfer texts, the direction of the speaker, moods, and feelings. Although voice chat increases togetherness (Geerts et al. 2011), it can negatively impact the viewing experience because of distraction.

Voice chat is one way to communicate in *remote* environments. The voices can be non-spatial or spatial, where the direction of the voices is referenced to the virtual world. We considered two obvious spatial voice options: (1) The direction of the sound is a fixed position in the virtual world next to the viewer—or (2) it comes from the region where the other person is looking at. The first approach replicates the real situation of sitting next to each other in the virtual world. The advantage of the second variant is that the sound comes from the PoI and people are used to looking in the direction where a sound is coming from. After an informal pilot test, we decided to place the virtual sound source at the position where the speaking person is looking at, even though this direction does not match the speaking direction (Fig. 1a). In the pilot test, the participants found it helpful if the other person's voice comes from the PoI.

Since the voices do not belong to the movie experience, it is possible that spatial sound confuses the participants, since it works as an element of the virtual world. In our user study (Sect. 5), we investigated in more detail the comparison of the *spatial* with the *non-spatial* technique.

In *co-located* environments, the involved participants can speak to each other directly. This could be an advantage for togetherness, whereas it could be difficult to supplement the approach with additional information, such as the viewing direction.

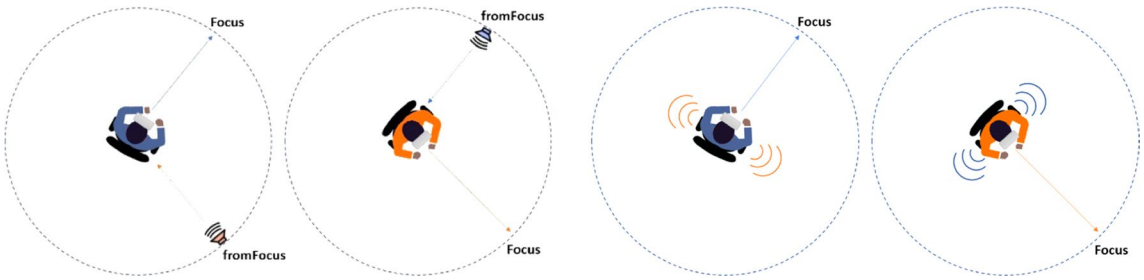
4.2 Sending emotion states

For indicating emotional states to each other, a simple sign language can be used, realized by showing icons to the co-watcher. We implemented methods by sending smileys (Fig. 2a) or photos of faces with various expressions (Fig. 2b). A smiley might be captured easier and faster, a photo might increase the togetherness.

In the user studies, we investigate if the mentioned communication methods make the viewing experience more social and which of both methods is preferred by the participants.

4.3 FoV indication

Knowing about other's current FoV is essential for a coherent interaction in social viewing. One method is to frame the FoV of the co-watcher (Nguyen et al. 2017). When using such a method the FoV is visible if the FoV overlap. In the case the FoV of the co-watcher is off-screen, an arrow can



(a) *Spatial*: The source of the sound is in the center of the speaker's viewport (fromFocus)

(b) *Non-spatial*: The sound does not depend on the viewing direction.

Fig. 1 Two voice chat methods: Two persons are watching a movie together via HMD. The blue symbols belong to the blue person, the orange symbols to the orange person

(a) *Smileys* show the feeling of the co-watcher

(b) *Photos* show the feeling of the co-watcher

Fig. 2 Use of smileys or photos to inform the co-watcher emotional states

be used for finding it. This method is suitable for working tasks such as movie editing (Nguyen et al. 2017).

Another possibility is the picture-in-picture (PiP) method, where a video window shows the co-watchers FoV (Fig. 3a). This has the advantage of visually showing the other's FoV, independent on the own viewing direction, but the disadvantage of covering a larger area of the display. Therefore, it should be possible to switch it off. The PiP-video can be placed on that side of the display that is closer to the co-watcher's FoV.

Both methods mentioned so far indicate the exact FoV of the co-watcher. A further option would be to show the viewing direction of the co-watcher, not the exact FoV. Methods used by gliders for collision avoidance systems could be applied. Such systems show from which direction another glider is coming. We used it to display the viewing direction of the co-watcher. An example is shown in Fig. 3b. The slide bar at the bottom shows if the PoI is on the right or on the left side. The slider on the right shows if the PoI is higher or lower than the own viewing direction.

For our user study, we decided to investigate one method that shows the FoV and one method showing the viewing direction. Since the frame method needs additional guiding in the cases the FoVs of the viewers do not overlap, we opted for the PiP-method (Fig. 3a) and compared it with the bar-method (Fig. 3b)

4.4 Video chat

In the previous approaches, the viewers cannot see each other. For enabling this, we include a video chat window via PiP. Figure 4 shows two examples of video chat methods. In the first one, the front-view of the co-watcher is displayed at the bottom centre of the screen (Fig. 4a), even if the viewer turns the head. The second one is very similar to the situation of viewing a movie together in cinema or TV. The PiP is placed on the side of the viewer (Fig. 4b). The viewer can only see the co-watcher if she/he looks to this side.

Since both users are watching the movie via HMD, a large part of the face is not visible. Accordingly, it needs to be

(a) Picture-in-Picture (*PiP*): The window shows the FoV of the co-watcher.

(b) The red *bar* at the bottom shows if the viewing direction of the other is on the right or on the left side. The bar on the right shows if the PoI is higher or lower.

Fig. 3 Different methods for showing the FoV of the co-watcher

(a) *Front*-method: Chat window in front of the viewer, connected to the display. The co-watcher is always centered in the FoV.

(b) *Side*-method: Chat window on the side, connected to the virtual world. The co-watcher can be seen when the head is turned to this region.

Fig. 4 Two methods for voice chat

explored whether this issue hinders natural communication. Methods such as "Headset Removal" (Burgos-Artizzu et al. 2015; Google Research and Daydream Labs 2017; Thies et al. 2016) can solve this problem in the future, but we did not explore it in this paper.

In a first step, we investigated the position of the chat window and compared the two described methods (Fig. 4a, b). In the *front*-method, the PiP window is fixed on the screen, in the *side*-method, it is fixed in the virtual world. We did not remove the background of the persons, as they should be separated from the movie picture.

5 User studies

In our user studies, we investigated in more detail the four approaches that were introduced in the preceding section: voice chat, sending emotion states, FoV indication, and video chat. In each of the four experiments, two methods were compared to each other, by using a within-subject methodology. The independent variable was the method; the dependent variables were presence, sickness, usability, and togetherness. In the second part, the components were compared to each other (between-subject) to find out which of them make the biggest contribution to social viewing experiences, and which ones play a minor role and may possibly be omitted.

6 Participants and material

In our studies, we focused on shared viewing by two persons in symmetrical environments (both using HMDs) to gain initial experiences. A total of 86 participants took part in the entire study (see Table 1). Among them were both VR beginners and experienced VR users. Some of the participants knew each other beforehand, others did not. We did not investigate the dependence of the results on these characteristics. For two experiments, the second person was not available and a person of the team took on the role of the co-watcher the data of which was not included in the data set. All methods were implemented in a way that they can be used for remote and co-located environments. For all tests, an Oculus Rift with headphones was used. The used films were nature documentaries, similar in style and pace, and had an approximate length of 8 min.

Except for the voice chat case, all tests were conducted in a large room, where a remote environment was simulated: both participants sat far from each other and were only connected by the network. They did not speak during the study. We chose this one-room setting to observe both participants in parallel. This was only possible for the visual methods since the HMD blocks the visual real environment. However, voices were not blocked, even if the participants used headphones. Therefore, for the voice chat test, the participants were in neighbour rooms.

For each of the components, two methods were compared, as indicated in Table 1. The aim of our study was to find out which method is more suitable for social viewing in CVR, and on the other hand to learn more about the advantages and limitations of each method.

Voice Chat: Two voice chat methods were tested. The first one used spatial sound, while the second one used normal stereo sound, which did not depend on the viewer's line of sight. The spatial sound came from the direction in which the speaking participants looked ("fromFocus" in Fig. 1).

Video Chat: The two video chat methods (Fig. 4) differed in the position of the chat window. In the front-method, the other person was in front of the view and the chat window was fixed at the bottom of the display and turned with the line of sight. This case is similar to the situation where two people are sitting opposite each other. However, the window is always in front even when turning around. For the second method (side-method), the video chat window was fixed beside the speaker, in our study on the right side. In such a case, it depends on the viewing direction whether the person can be seen in the display. This case corresponds to the situation in a cinema or sitting on the couch and watching a movie together. Even if in a real application a video chat is combined with voice chat, the method was tested without spoken language, since we were interested in the influence of each component separately.

FoV Indication: To investigate FoV awareness, we chose two methods: the bar-method and the PiP-method. For the PiP-method, the FoV of the other person could be seen on a little video window that was fixed to the display and did

Table 1 Components for social viewing in CVR investigated in this paper

Component	Participants	Method 1	Method 2
Voice chat	23 participants (4 female, 19 male, age: mean = 27.26, SD = 6.4)	Non-spatial	Spatial
Video chat	22 participants (7 female, 15 male, age: mean = 30.7, SD = 11.4)	Front (Fig. 4a)	Side (Fig. 4b)
FoV indication	21 participants (8 female, 13 male, age: mean = 23.9, SD = 2.5)	PiP (Fig. 3a)	Bar (Fig. 3b)
Emotions	20 participants (6 female, 14 male, age: mean = 25.6, SD = 6)	Smiley (Fig. 2a)	Photo (Fig. 2b)

For each component, we implemented two methods and compared them to each other

not change its position during turning around (Fig. 3a). The bar-method is inspired by a glider warning system. At the bottom of the screen and on the right side, there are bars along a line (Fig. 3b). One of the bars is coloured red and indicates where the other person is looking at.

Sending Emotion States: For sending information about feelings, two visual methods were compared. Smileys were sent in the first and photos of facial expressions were sent in the second method. For both methods, four pictures were available: two affectively positive and two negative ones. A hand-held controller was used for sending the pictures. The position of the picture was always the same, fixed on the edge of the display.

7 Procedure

A within-subject test design was used to compare the two methods for each component. Each participant watched two films, each with a different method. The films and methods were counterbalanced in order and assignment. After each film, a part of the questionnaire about simulator sickness, presence, and togetherness was filled out.

To measure simulator sickness we applied the Simulator Sickness Questionnaire (SSQ) of Kennedy et al. (1993). For each item one of the sickness levels (none, slight, moderate, severe) could be chosen and the answers were transformed to a scale from 0 (none) to 3 (severe). To investigate the presence, we used the IPQ presence questionnaire (Schubert et al. 2002). The questions were answered on a seven-level Likert scale. Since not all questions of the SSQ and IPQ are appropriate for CVR, we did not include all items. In this way it was not possible to calculate the total score exactly as originally defined for the scale. However, we received enough information to compare the different test options. For the togetherness part, the following questions from the ABC questionnaire (IJsselsteijn et al. 2009) were chosen:

(S1) I feel part of a group because of the contacts.
(S2) I know what the other feels during contact.
(S3) Because of the contact, I can relate to the other person.

Again, we used a seven-level Likert scale for the answers. At the end of the study, the participants answered questions for comparing the methods:

(C1) Which method do you find more comfortable? (usability)
(C2) Would you use the method for a longer time? (usability)
(C3) With which method do you feel more connected with the other? (togetherness)

For the questions (C1)–(C3), the participants were asked to justify their answers.

8 Implementation

The methods are implemented in Unity3D 2017. For the implementation of the synchronization features, the *Multiplayer High Level API* (HLAPI) was used ("Unity —Manual: The Multiplayer High Level API," 2018). With this API no additional server is necessary, one of the participant computers can simultaneously act as client and server. For each co-watcher, a player prefab was defined, which includes the network identity and contains the necessary properties and functions. This prefab is invisible and presents the viewer in the VR environment. It was placed in the centre of the sphere on which the omnidirectional movie is projected.

The Unity *NetworkManager* component manages the communication and the synchronization of the scenes. For avoiding network problems, the movies were locally stored on the client and just positions, directions, and meta information were transferred via the network. In case of poor network quality, in real-life applications, the video should also be uploaded in advance, to avoid interfering with the experience.

Voice Chat: The voice communication was realized by the Unity plug-in *Dissonance Voice Chat*; the spatial condition, additionally by the *Oculus Spatializer* Plugin.

Video Chat: The webcam frame was implemented as an object that could be switched on/off by the user. Additionally, transparency could be customized. For the front-method, a webcam was placed in front of the viewer, for the side-method at the side. In order to convey the feeling of sitting side by side and to have the opportunity to look at each other, we positioned the camera for one user on the left side, for the other one on the right side.

FoV Indication: For both methods, the PiP as well as the bar, the *Unity Raw Image* was used. It shows non-interactive images to the user and can display any texture.

Sending Emotion States: The smileys and picture elements were realized by the *Unity Raw Image* component since a RawImage element is well suited for the representation of 2D graphics.

9 Results -Part 1: comparison of the methods for each of the components

For comparing the two methods of each component, we performed a two-sample t test (alpha = 5%) for each Likert-item, which showed nearly no significant differences regarding presence, sickness, and togetherness. The only difference was in the video chat component for question (S2), where the score for the side-method (mean = 4.41, SD = 1.87) was significantly higher than for the front-method (mean = 3.27, SD = 1.78, $p = 0.04$).

We used the exact Fisher test to calculate the p-values and to find significant differences in the results of the

comparative questions (C1-C3). Additionally, we analysed the qualitative answers. In this way, we found preferences, advantages, and disadvantages, which we present in detail below.

Voice Chat: Both methods were accepted by the participants. Most of them would like to use it, even for longer videos (78.2%, 87%). When asking about the preferred method, the spatial-method received a higher score for all questions (Table 2).

Most participants preferred voice chat with spatial sound. P14 compared both methods in this way: "In the first method (stereo), the bulk of the conversation was about the gaze direction. With the second method (spatial), I could tell that just by the direction from which his voice came, and you could talk about the pictures right away instead of having to find them first." P22 preferred the spatial-method: "It has felt more integrated into the video through the different locations of the sound. This made the experience more interesting." Some participants mentioned the advantages and disadvantages of the spatial-method: "The spatial-method was helpful in finding the view. However, it was also a bit more distracting." (P17). "The stereo-method distracts less; with the spatial-method, you also want to look where the other is looking and thus you always seek the right view".

For a better understanding of social viewing, the following reasons are relevant.

- "helped to find the view from the other one" (P8, P11, P14, P17, P18, P19, P21)
- "more spatial and more real" (P10, P21, P23)
- "feeling of being in the same room" (P11, P22)
- "closer to the real experience" (P14)
- "more like exploring an environment together" (P11)
- "more interaction" (P13)

Some of the participants preferred the stereo-method:

- "more familiar" (P15, P16)
- "less confusing" (P16)
- "less distractive" (P16, P17)
- "can hear well" (P5, P6)

Video Chat: In this study, the scores for usability and togetherness were very similar for both methods (Table 3).

Analysing the qualitative data, we could recognize important findings for both methods. P18 remarked: „It depends on the content. For watching a cinematic movie, I find the side-method better".

The front view was superior in the following aspects:

- "more comfortable" (P2, P6)
- "better for communication" (P11, P17, P20, P21)

Reasons for preferring the side view were:

- "similar to cinema/TV" (P1)
- "can see more of the partner when he looks ahead" (P1)
- "feel addressed when he turns to one" (P1)
- "disturbs less, is more realistic" (P4, P7, P10, P11, P14, P19)

FoV Indication: In the FoV awareness part, we could not find any significant differences between the two methods. (Table 4).

In the qualitative answers, we found advantages of both methods:

Benefits of PiP:

- "easier to understand and faster" (P3, P4, P18)
- "both views can be seen simultaneously" (P8, P12, P16, P18)
- "I see what the other one sees" (P19)
- "more personal "(P10), "more connected" (P15)

Table 2 Comparison of the two methods regarding usability and togetherness for the voice chat methods

	Find more comfortable?	Use for longer time	Feel more connected with the other?
Stereo	30.4%	78.2%	26.1%
Spatial	69.6%	87%	73.9%
p value	0.02	0.7	0.003

For the second question multiple answers were admitted

Table 3 Comparison of the two methods regarding usability and togetherness for the video chat methods

	Find more comfortable?	Use for longer time	Feel more connected to the other?
Front	45.5%	68.2%	50.0%
Side	45.5%	68.2%	45.5%
p value	1	1	1

Some participants did not decide for one of the methods

Table 4 Comparison of the two methods regarding usability and togetherness for FoV indication

	Find more comfortable?	Use for longer time	Feel more connected to the other?
PiP	47.6%	71.4%	52.4%
Bar	42.4%	71.4%	28.6%
p value	1	1	0.21

Some participants did not decide for one of the methods

- "does not have to turn the head" (P10)

Benefits of bar:

- "concealed less" (P1)
- "less intrusive/discreet" (P5, P6, P7, P11, P20, P21)
- "easier to understand" (P5, P16, P20), "better orientation" (P14)

Sending Emotion States: For most participants, the smiley-method was more comfortable (75%). More participants would like to use the smiley-method for a longer time (85%). However, the feeling of togetherness differed just slightly (Table 5).

For some participants, the smiley-method seemed familiar, but for others it was the face. P20 pointed out that images of faces could create expectations for communication. Several participants answered the second question in another way than the first one with the substantiation that the face is more realistic (P3, P13, P16).

Benefits of smileys:

Table 5 Comparison of the two methods regarding usability and togetherness for the emotion methods

	Find more comfortable?	Use for longer time	Feel more connected to the other?
Smiley	75%	85%	55%
Face	20%	55%	40%
p value	0.03	0.06	0.53

Some participants did not decide for one of the methods

- "faster and easier to recognize" (P2, P3, P5, P6, P8, P11, P16)
- "familiar "(P2, P18)
- "anonymous in the case of unknown people" (P7, P12)
- "distract less" (P13, P18)

Benefits of face:

- „more authentic" (P4)
- „easier to understand" (P9)
- „familiar" (P19)

10 Results -Part 2: comparison of the components

In the second part of the study, we investigated which components are important for feeling togetherness. For this, we compared the favourite method of each component from part 1 to each other:

- *Voice chat*: spatial-method
- *Video chat*: front-method
- *FoV indication*: PiP-method
- *Sending emotion states*: smiley-method

Using the Shapiro–Wilk test, homogeneity of variances was checked which showed that equal variances could not be assumed for all components. Therefore, the Kruskal–Wallis test was used for finding significant differences between the components. There were no significant differences for presence and sickness. Using the post hoc pairwise Mann–Whitney tests, we found significant differences for the togetherness aspect (questions S1–S3). In Fig. 5 the means and *p*

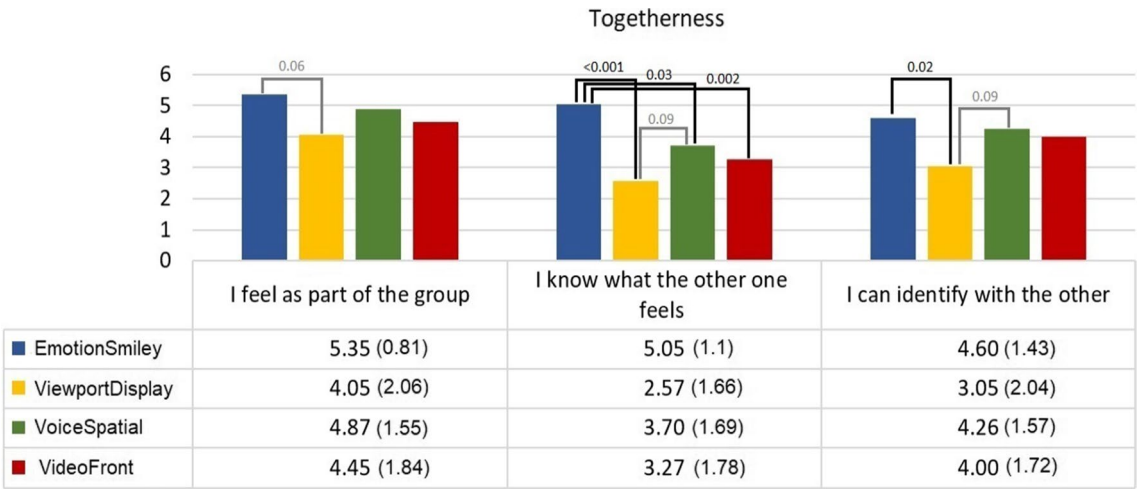


Fig. 5 Means and standard deviations for the social questions. For significant differences ($p < 0.05$) the p-values are given in black, for weak significant differences ($0.05 <= p < 0.1$) in grey (colour figure online)

values are summarized and illustrated. For significant differences ($p < 0.05$) the p values are added in black, for weak significant differences ($0.05 <= p < 0.1$) in grey.

As Fig. 5 shows the smiley-method was most important for togetherness, followed by the spatial voice chat and the video chat. Knowing the FoV of the co-watcher was less important for the participants of our study. We discuss the more thoroughly in the next section.

11 Discussion of the results

In general, all four methods were well accepted by the participants. In this section, we discuss the results for each user study by comparing the methods and evaluating how the approaches meet the challenges.

11.1 Components

Sending emotion states reached the highest score for togetherness, with the smiley-method rated as the preferred one. Additionally, most participants (87%) would use the smiley-method for a longer time (C2). The smiley was less distractive compared to the photo and gave fast enough information about the feelings of the co-watcher. The photos of the faces draw more attention and generate an expectation for communication. However, the participants wanted a larger selection of smiley types.

Voice Chat was also an important component for the participants to create a sense of togetherness. Additionally, most participants (87%) would use the spatial-method for a longer time (C2). Not only spoken remarks are relevant. In addition, a laugh or a sigh shows the presence of another person. The spatial-method was preferred regarding comfort and togetherness. Even if we could not find any significant difference in the answers of the presence questionnaire, several participants mentioned that the spatial voice chat method improved the spatial experience.

There was no preferred *video chat* method in our tests. In the front view, the co-watcher is always visible. However, the HMD covered a big part of the face. The HMD is less present in the side-method and the co-watcher is only visible if the viewer is looking in this direction. Video chat was tested in our study without spoken language, as we were interested in the impact of the visual content to compare the two video chat methods. In a real application, this component should be supplemented with the auditory component.

The *FoV indication* was the component with the lowest score regarding togetherness. For togetherness, it seems to be less important to be informed about the exact FoV of the partner. Since the simultaneous application of all components would lead to an overcrowding of the display and overstraining of the watchers, this component seems most suitable to cut in the presence of the others. However, when inspecting an omnidirectional movie, working on it or sharing information about it, this component can be very important if not even necessary (Nguyen et al. 2017).

It is known from multimedia psychology that the combination of visual and aural information is easier to comprehend than the combination of two forms of visual information (Moreno and Mayer 1999). Therefore, only one method per sense should be used. Voice chat could be extended by video chat or smileys. Which of the above components are advisable depend also on the type of movie and the aim of watching (enjoyment, information or learning). What is considered useful and helpful is influenced by how much dialogue is desired with the partner.

11.2 How the approaches meet the challenges

Each of the investigated approaches meets more than one challenge. Table 6 gives an overview of which approach can support which challenge, some of them as a result of the user study (marked by ✓). Others are a result of the discussion and should be checked in future work (marked by +). In the following section, we describe how the investigated methods affect each challenge. For some of the challenges, additional approaches are needed, which we discuss in Sect. 4.5.

Communication: Communication while watching a movie together can be realized in various ways: verbal

Table 6 Table shows how the approaches meet the challenges. ✓: the approach meets the challenge as a result of the user study; +: the approach needs revision to check if it meets the challenge

	Voice chat	Video chat	FoV indication	Sending emotion states
Communication	✓	✓		✓
FoV awareness	+		✓	+
Togetherness	✓	✓		✓
Accessibility	+	+		+
Interaction techniques	✓	✓	✓	+
Synchronization				
Multiuser environments	+			+

communication by voice chat, visual communication by video chat or sending pictures (smileys, photos). Each of the visual communication methods can be complemented by voice chat. Based on the results of our experiments, a combination of spatial voice chat and sending smileys should be examined. In a combination of these two methods, the smiley could be positioned near the region of interest. Even if it is outside of the co-watcher's FoV, the position can be found, since it is the source of the spatial sound. Such a combination should be investigated in the future.

FoV Awareness: Knowing the other's FoV seems to be less important for togetherness. However, this could also be caused by the implemented methods. Both methods are not very natural and need time for familiarization. Compared with that, the spatial voice chat, which gives also information about the other's viewing direction, was preferred by most participants. A similar way would be possible with the smiley- method while positioning the smileys in the viewing direction. However, this would need an additional guiding method in case this area is not in the other's FoV. The same combination as for communication could be a solution and should be verified: spatial voice chats with sending smileys in the current viewing direction.

Togetherness: All of the methods of our user study were chosen for improving togetherness. The results in Fig. 5 show that sending emotion information has the biggest impact on togetherness. The FoV awareness resulted in the lowest score. To provide togetherness, it seems to be more important to know what the other person feels than to know what they see. In addition, voice chat is important for togetherness—more than video chat. This can be influenced by the fact that the HMD covers a big part of the face during the chat. Additionally, video chat is more distracting during a movie watching experience. For addressing this challenge, spatial voice chat combined with smileys should also be investigated.

Accessibility: People using a social movie player for CVR can have different requirements for the application. A person with visual impairments will prefer the voice chat. For people with hearing problems, the smiley-method or the video chat may be more important. Besides, some of the implemented approaches are suitable to meet requirements of accessibility outside of social viewing. A social movie player with the proposed features could support the viewing experience: voice chat can be used for audio description or explaining movies in easy language. A language interpreter in the front video chat window is able to support deaf people with sign language. Sending signs can support people with mental impairments to follow the story.

Interaction Techniques: Our approaches are head- and controller-based methods. All the head-based inputs (direction for spatial voice chat and FoV indication) are natural ways of interaction. In our first tests, to send smileys/photos, controllers were used to be sure the actions were triggered on purpose. When wearing an HMD, only simple controllers can be used, which limits the number of different smileys. Participants of our user study wished to have a larger choice, which would require tools or devices that allow triggering more different smiley types. Gestures could be an alternative approach to be tested in the future.

Synchronization/Navigation: For our user studies, we used movies with 4 K resolution. The synchronization via the Unity NetworkManager was working well. However, the movies were provided at the local computer and only the viewing direction and other metadata were transferred. This needs preparation by providing the movie in advance. For spontaneous social CVR experiences in remote environments, advanced and adaptive FoV-based delivery and synchronization techniques should be provided. In the present paper, we focused on Human–Computer Interaction (HCI) aspects, and do not go deeper into these technical research topics.

Multiuser Environments: All the approaches were tested with two persons. For adapting them to more participants, some extensions are required. On the one side, it is difficult to assign notifications to different members of the group. On the other side, the application has to be protected against overloading. During voice chat, the members could be distinguished by their voices. However, sometimes voices are similar and it can be helpful to see who is speaking through a visual sign, for example using coloured loudspeakers or by combining the chat with the emotion or video chat component. Nevertheless, it is difficult to realize multiuser environments for a group without an additional role concept, which describes the channels and rights for every user. More than one chat window would overload the screen. The same goes for FoV indication with the bar-method. The frame method could also be applied for more than two, but not too many, participants. Smileys can be assigned to different viewers easily by colours. Even if there are possibilities for adapting the approaches to groups, a concept to avoid overloading the screen and overstrain the viewer is needed. Not always, all items of all group members should be shown simultaneously.

11.3 Limitations

Some of the participants had never watched a movie via HMD before. Even for the others, viewing behaviour can change over time, as they consume more CVR videos.

In our study, we used only one film genre (nature documentaries) and the content may have affected the results. Other types of movies or content genres may involve more or less interaction. Further research is necessary to determine the impact of the social viewing techniques on the content genre, underlying context and relationship between the participants, as well as their number. We only investigated

the case where two persons share the experience. For more participants, it takes more effort to avoid overcrowding the screen. Mapping mechanisms are needed to identify which information belongs to which person. Another approach would be a role concept.

Since in our study all components were investigated separately, the next step should identify how the components can be combined in an optimal way. One of the results of our study was that voice chat is very helpful for togetherness and can be added by enabling visual communication. If the video chat includes voices, it has to be investigated how the direction of the voice can be used as it was done in the spatial voice chat approach.

We did not consider the relationship between the participants. Some of them knew each other, others did not. Further experiments to analyse this aspect can lead to more detailed results.

12 Design space

12.1 Terminology

The exploration of the tested approaches resulted in a design space for social viewing in CVR (Table 7). Three main dimensions have been identified: the *viewers*, the *notifications* and the *devices* that are to be considered. Each of these major dimensions has sub-dimensions that describe attributes of the dimension and can take on multiple values.

Viewers: For designing a CVR social movie application, several characteristics regarding the viewers are relevant. Two different scenarios with respect to the **location** are possible: viewers can be *co-located* in the same room or participate *remotely* in different locations. Additionally, it is important to know the **number** of participants and their **roles**. If there are only two viewers (*paired*), both can have the same rights for interacting with the application (*coequal*). For *groups* of participants, this could cause conflicts and overload the display. Defining **roles** and allocating rights can avoid such problems. Examples for roles are *guide* (who can send information to all members), *follower* (who can receive information from the guide) and *slave* (who is synchronized in time and viewing direction to the guide). Furthermore, the **relationship** between the participants has to be taken into account.

Notifications: The second dimension highlights the need for a careful design of the used notifications. It is relevant to know which communication **channels** will be used (*auditory, visual, haptic* or *sensor info*), and how they will be *triggered* (**trigger**). The space also considers which **type** of notifications should be exchanged: *information* about the movie content, own *emotion* states or *directions/positions* of gaze or point of interest. Additionally, it is important if the notification is located on the screen/HMD or in the virtual world (**reference**). *Screen-referenced* items are connected to the display and move along with it in case the viewer is turning the head. *World-referenced* items are connected to the virtual world, in our case to the movie. They stay fixed at their place in the movie world, even if the viewer turns the head (Yeh et al. 1999). Screen-referenced methods have the advantage that they do not depend on the viewing direction, while a world-referenced visual cue can be visible or not depending on the viewing direction. However, the world-referenced cues are often better for the VR experience. The disadvantage of visual world-referenced cues (not always visible) does not exist for aural cues. They can be heard for all viewing directions.

Devices: Social VR experiences are influenced by several device characteristics. As **input** for the communication, *controllers, speech, gestures, haptic signals,* and others are possible. Various **display** types are possible. The used

Table 7 Design space for social viewing in CVR

	Sub-Dimension	Option 1	Option 2	Option 3	Option 4
Viewers	Location	Co-located	Remote		
	Number	Paired	*Group* (> 2)		
	Roles	Coequal	*Guide*	*Follower*	*Slave*
	Relationship	Know	Unknown		
Notifications	Channel	Audio	Visual	*Haptic*	*Sensor info*
	Trigger	Continuous	By sender	*By receiver*	*By sensors*
	Type	Information	Emotion	Direction, positions	
	Reference	Screen-referenced	World-Referenced	*Switching*	
Devices	Input	Controller	*Speech*	*Gesture*	Gaze/head
	Display	HMD	*Monitor*	*Mobile device*	*Cave*
	Symmetry	Symmetric	*Asymmetric*		

For the three core dimensions—viewers, notifications, devices—sub-dimensions and possible values are shown. The italic styled values were not involved in the studies presented in this work. All the other values were implemented, but not all of them compared to each other

devices define the **symmetry** of the application, which can be *symmetric* (the participants are using the same devices) or *asymmetric* (the participants are using different devices).

From our perspective, based on the literature, our study and prototype experiences, the mentioned dimensions are the most important ones for social CVR experiences. However, special use cases may require slightly different or further considerations.

12.2 Application of the design space to the methods of the user studies

To explain the dimensions, in this section we classify the methods which were investigated in the user studies and discuss the above results in relation to the dimensions of the design space. Afterwards, Sect. 7.3 presents further possibilities and combinations revealed by the design space.

Viewers: In our user studies, the participants were co-located except for the voice chat component *(location*). This was done since we wanted to start in co-located environments but do not exclude the audio component. The remote condition was chosen since no complete noise-cancelling for the headphones was possible. Nevertheless, the score for togetherness was relatively high for the voice component.

All other approaches were tested in co-located environments, which simulated a remote environment. The participants were sitting in a large room far from each other. From the technical side, there is no difference between a remote and a co-located environment in these approaches since a network was used for data exchange and synchronization. However, it could be possible that there is an impact on the feeling of togetherness when the viewers are sitting close to each other. This location aspect should be investigated in the future and we added it as a dimension to the design space.

In all experiments of this paper, the viewers were coequal since we did not use any role model (*roles*). In all tests, pairs of participants viewed the movie together (*number*). Some of them were familiar with each other, but not all of them (*relationship*). We did not focus on this aspect and we avoided influencing the results by mixing the participants regarding this dimension for each component. Regarding the emotion component, participants mentioned that in the case of unknown co-watchers, the face-method could violate privacy.

Notifications: We investigated two *channels*: the aural (voice chat) and the visual (video chat, emotion, FoV indication). We found that the aural component has a big influence on togetherness. It can be used stand-alone or complemented by a visual channel (video chat or FoV indication). However, when using the visual component alone, the participants missed the aural component.

The impact of *triggering* was not investigated in depth in our user studies. Voice chat and video chat were continuous and switched on during the full video experience. However, for the video chat, some participants wished to have the possibility of switching off the chat window temporarily, especially in the front-method. The same request was made by the participants of the FoV indication component. The participants did not comment on this for the voice component. An explanation for this is that in the voice case there is nothing to hear as long as nobody needs the contact, but in the visual case, the window is there nevertheless. So, for the visual methods, a triggering concept for notifications would be needed, which is not necessary for the aural component. Both emotion methods (smileys, faces) were triggered by the sender. There was no complaint or wish for temporarily switching it off. It seems that visual continuous methods are too distractive and co-watchers have the feeling that they should react.

Comparing the *types* of notifications, we investigated information, emotions, and viewing directions. The participants preferred to infer the viewing direction of the co-watcher from the direction of the voice. The FoV indication achieved the lowest level concerning togetherness. This is a difference with respect works on collaborative scenarios, as described in CollaVR (Nguyen et al. 2017). For collaboration in movie cutting, the user has to know the exact FoV of the other person. Social viewing is focused on enjoyment and additional information on the screen can destroy it. Knowing the direction of the other viewer is not important all the time. However, in the case the other person tells something via voice chat, it is beneficial for togetherness to know the others' viewing direction.

For techniques in virtual environments, the *reference* is an important attribute. In our study, the voice chat with spatial sound from the direction of the PoI is world-referenced. The front video chat is screen-referenced, as well as the two investigated FoV indication methods (bar, PiP). In contrast, the framing method for the FoV, which is used for CollaVR (Nguyen et al. 2017), is world-referenced. The two methods for sending the emotion status investigated in this study are screen-referenced. However, both methods could be adapted to world-referenced methods for supporting in more easily finding the cause of the emotion. In that case, the picture should be placed at the PoI. We did not use this approach in our study since the cue would not be always visible and it would need additional methods to find it.

Devices: In our studies, we used two sorts of devices. On the one hand, we used HMDs as a *display* and as an *input* device for head-based methods. In our approaches, the direction of the spatial voice chat is given by the HMD direction. On the other hand, we used a controller for sending the emotion pictures. This was less distractive than a menu or a virtual keyboard. However, the number of smileys was limited by the device type. In our first approach, all participants

used the same display: an HMD. So, all test cases were in a *symmetric* environment.

12.3 Impact of the design space: required concepts and future work

Beyond our investigations in the study part of this paper, we share our ideas about other options for future work. This discussion is based on the findings from the conducted experiments and the presented design space. With the considered approaches, several challenges of social viewing could be met. Additionally, with the definition of the design space, we were able to identify gaps that could initiate further research.

Viewers: For designing a social viewing application for CVR it is important to consider if the viewers are co-located in the same room. In this case, the co-watchers can directly communicate via voice and touches are possible. For using the spatial voice chat method in co-located environments, headphones are needed which are able to completely cancel the ambient noise. However, further research is needed to determine which condition is preferred for togetherness in co-located environments: the voice from the real world or the voice in the virtual world. Additionally, the environment (temperature, smell, airflow, sounds in the room) is the same for co-located viewers, which could increase the feeling of togetherness. Further research is needed to determine if the location influences togetherness and if additional components for remote user scenarios are needed.

In our approaches, we did not investigate if viewers, who are *familiar* with each other, need other options or parameters than unknown users. Examples of parameters to be examined are the position and distance of the video chat window. The communication for *unknown* viewers will be influenced by the fact whether anonymity should be preserved or the communication is open for becoming familiar. Such facts influence the usage of photos or videos.

For expanding the *numbers* of viewers, further research is also necessary. Communicating in a group of users in VR environments can be problematic, as it will result in overloading and negatively affecting the VR experience. One approach could be a role concept that differentiates between two roles: the guide (one user) and the followers (the other users). The guide (e.g. the instructor in learning scenarios) will be taken as the reference for communication and synchronization and will be the only participant with the navigation functionalities enabled. Other interaction, collaboration, and guiding features can be also provided, such as shared volume control and pointers (as in Montagud et al. 2015) and automatic sharing and adaptation of the FoV. To allow more interactive and flexible sessions, the roles of guide and followers can be changed dynamically. A slave mode, where the follower is synchronized in time and viewing direction to the guide, causes simulator sickness (Nguyen et al. 2017). However, it can be helpful in asymmetric environments with non-VR collaborators (e.g. on a desktop).

Likewise, a novel social viewing concept is considered in (Núñez et al. 2018), consisting of a multi-screen scenario in which different users play a different role: observer (TV), assistant (tablet) and inspector (HMD). The inspector's FoV is streamed to the TV to allow the remaining users being aware of the 360° scenes, thus overcoming isolation and stimulating interaction. Defining the roles is an important step for social CVR experiences. However, also other components are important and should be considered.

Notifications: Not only audio and video channels can be used. Informing the partner about heart rate or other sensor information could be a new approach for sharing emotions (Hassib et al. 2017). Such information could be visible all the time (continuous), triggered by the system, or just displayed only if values above/below a certain threshold are reached.

For our emotion method, we used the screen-referenced variant. In this way, the smileys/pictures were always visible. Further research is needed to investigate whether placing the smileys near the PoI will make it easier to understand why a person is sad or happy. However, the smiley could cover parts of the PoI. Additionally, the picture would not be visible if the PoI is not in the FoV. A combination could be an alternative: screen-referenced for off-screen PoI and world-referenced for on-screen PoI. If the viewer is turning the head and changing the viewing direction, the method switches accordingly.

Devices: The aim of this work was to investigate the possibilities of social viewing in CVR for viewers using an HMD. A next step should be the exploration of options for different devices and for asymmetric environments, e.g. if one viewer is watching the movie via an HMD while the other is using a desktop or mobile device. For communication as well as for navigation, non-disturbing input methods are needed. We think graphical elements on the display or in the virtual world or operating via speech can disturb the viewing experience. Several input methods can be used for interaction in VR. Holderied (2017) compared controller-based and head-based gaze pointers in VR. The users of the gaze pointer completed the tasks faster and rated them with higher usability scores. Head and gaze movements are natural techniques for changing the viewing direction or selecting a region in the virtual world. However, they are not sufficient for all interactions. For example, choosing between different smileys and sending one of them needs a simple selection mechanism, which can be realized by controllers or gestures. Gestures are a natural method of interaction in VR (O'Hagan et al. 2002). Pakkanen et al. (2017) compared remote control, pointing with head orientation and hand gestures. In their experiments, head and controller-based methods achieved the best results, since there were some

Fig. 6 Sending smileys by gestures. Two examples of simple smileys and the corresponding gesture

technical problems in gesture recognition. Gesture recognition is still not a mature technique and accuracy is expected to improve in the near future. Since the input device was not in the focus of our studies, we used a controller to avoid technical problems. Using a controller as an input device for the smileys/photos restricts the number of input possibilities. Replacing the controller by easy-to-remember gestures (e.g. the ones shown in Fig. 6) allows increasing this number.

Social viewing in CVR should also be possible in *Asymmetric Environments*. For all our user studies, the participants used the same devices. However, co-watchers in social viewing scenarios do not necessarily have the same hardware. For example, the FoV may differ, because different HMDs have different FoV sizes. Participants can also use desktops or walk-in systems. Gugenheimer et al. (2017) implemented ShareVR, which enables users of the real world to interact with users in a virtual world. They studied asymmetry in visualization and interaction. The derived guidelines have to be verified for CVR.

13 Conclusion

With our research, we contribute to finding ways for using CVR as a social activity and support the social aspect of CVR movie watching. A close cooperation between other fields of Virtual Reality and Social TV is necessary for enabling CVR as a social experience. We reviewed the literature for these fields and identified seven key challenges to enable social viewing in CVR: communication, FoV awareness, togetherness, accessibility, interaction, synchronization, and multi-user environments. Several approaches were proposed and discussed, and the most promising methods were evaluated through user tests. In our studies, the preferred techniques for supporting social viewing were sending smileys and voice chat. However, all the results depend on the movie content and the desired intensity of the contact.

Based on the collected knowledge, we presented a design space for social CVR applications with three main dimensions: viewers, notifications, and devices. We discussed this design space with reference to the investigated approaches and found additional aspects for future work. The presented design space supports further research and designing applications for social viewing in CVR: It provides a structured approach for developing new ideas and concepts for social movie applications, e.g. a social movie player for CVR, and assists to identify under-represented and unexplored aspects.

The field of social viewing in CVR is relatively new and more knowledge about the viewers' requirements and behaviour is needed. We believe that the design space and the described approaches are initial, but valuable, steps for the exploration of this relevant and timely field.

Acknowledgements The work by Mario Montagud has been funded by the Spanish Ministry of Science, Innovation, and Universities with a Juan de la Cierva - Incorporación grant, with reference IJCI-2017-34611, and by European Union's Horizon 2020 program, under agreement no 762111 (VRTogether).

Funding Open Access funding enabled and organized by Projekt DEAL.

References

Abreu J, Almeida P, Branco V (2002) 2BeOn—interactive television supporting interpersonal communication. In: Multimedia 2001. Eurographics. Springer, Vienna, pp 199–208. https://doi.org/10.1007/978-3-7091-6103-6_20

Belda J, Montagud M, Boronat F, Martinez M, Pastor J (2015) Wersync: a WEB-based platform for distributed media synchronization and social interaction. In: Proceedings of ACM international conference on interactive experiences for television and online video 2015. ACM, New York, NY, USA, pp 9–10

Boronat F, Montagud M, Marfil D, Luzon C (2018) Hybrid broadcast/broadband tv services and media synchronization: demands, preferences and expectations of Spanish consumers. IEEE Trans Broadcast 64:52–69. https://doi.org/10.1109/TBC.2017.2737819

Burgos-Artizzu XP, Fleureau J, Dumas O, Tapie T, LeClerc F, Mollet N (2015) Real-time expression-sensitive HMD face reconstruction. In: SIGGRAPH ASIA 2015 technical briefs on—SA'15. ACM, New York, NY, USA, pp 1–4. https://doi.org/10.1145/2820903.2820910

Carlsson C, Hagsand O (1993) DIVE—a multi-user virtual reality system. In: Proceedings of IEEE virtual reality annual international symposium, 1993. IEEE, pp 394–400. https://doi.org/10.1109/VRAIS.1993.380753

Chambel T, Chhaganlal MN, Neng LAR (2011) Towards immersive interactive video. Through 360° hypervideo. In: Proceedings of the 8th international conference on advances in computer entertainment technology—ACE'11. ACM, New York, NY, USA, pp 1–2. https://doi.org/10.1145/2071423.2071518

Cordeil M, Dwyer T, Klein K, Laha B, Marriott K, Thomas BH (2017) Immersive collaborative analysis of network connectivity:

CAVE-style or head-mounted display? IEEE Trans Vis Comput Graph 23:441–450. https://doi.org/10.1109/TVCG.2016.2599107
De Greef P, IJsselsteijn WA (2001) Social presence in a home tele-application. CyberPsychol Behav 4:307–315. https://doi.org/10.1089/109493101300117974
De Simone F, Li J, Galvan Debarba H, El Ali A, Gunkel SN, Cesar P (2019) Watching videos together in social virtual reality: an experimental study on user's QoE. In: IEEE conference on virtual reality and 3D user interfaces (VR). IEEE, pp 890–891. https://doi.org/10.5281/zenodo.2572867
Dorta T, Pierini D, Boudhraâ S (2016) Why 360° and VR headsets for movies?: Exploratory study of social VR via Hyve-3D. In: Actes de La 28ième Conférence Francophone Sur l'Interaction Homme-Machine. ACM, New York, NY, USA, pp 211–220. https://doi.org/10.1145/3004107.3004117
Durlach N, Slater M (2000) Presence in shared virtual environments and virtual togetherness. Presence Teleoper Virtual Environ 9:214–217. https://doi.org/10.1162/105474600566736
Facebook (2019) Facebook spaces [WWW Document]. https://www.facebook.com/spaces. Accessed 6 Sept 2020
Geerts D, Vaishnavi I, Mekuria R, Van Deventer O, Cesar P (2011) Are we in sync? Synchronization requirements for watching online video together. In: Proceedings of the SIGCHI conference on human factors in computing systems—CHI'11. ACM, New York, NY, USA, pp 311–314. https://doi.org/10.1145/1978942.1978986
Google Research and Daydream Labs (2017) Headset removal. YouTube, San Mateo
Gugenheimer J, Stemasov E, Frommel J, Rukzio E (2017) ShareVR: enabling co-located experiences for virtual reality between HMD and non-HMD users. In: Proceedings of the 2017 CHI conference on human factors in computing systems—CHI'17. ACM, New York, NY, USA, pp 4021–4033. https://doi.org/10.1145/3025453.3025683
Gunkel SNB, Prins M, Stokking H, Niamut O (2017) Social VR platform: building 360-degree shared VR spaces. In: Adjunct publication of the 2017 ACM international conference on interactive experiences for TV and online video—TVX'17 adjunct. ACM, New York, NY, USA, pp 83–84. https://doi.org/10.1145/3084289.3089914
Gunkel SNB, Stokking HM, Prins MJ, van der Stap N. ter Haar FB, Niamut OA (2018) Virtual reality conferencing: multi-user immersive VR experiences on the web. In: Proceedings of the 9th ACM multimedia systems conference on—MMSys'18. ACM, New York, NY, USA, pp 498–501. https://doi.org/10.1145/3204949.3208115
Harboe G, Massey N, Metcalf C, Wheatley D, Romano G (2008a) The uses of social television. Comput Entertain 6:1–15. https://doi.org/10.1145/1350843.1350851
Harboe G, Metcalf CJ, Bentley F, Tullio J, Massey N, Romano G, (2008b) Ambient social TV: drawing people into a shared experience. In: Proceedings of the SIGCHI conference on human factors in computing systems—CHI'08. ACM, New York, NY, USA, pp 1–10. https://doi.org/10.1145/1357054.1357056
Hassib M, Pfeiffer M, Schneegass S, Rohs M, Alt F (2017) Emotion actuator: embodied emotional feedback through electroencephalography and electrical muscle stimulation. In: Proceedings of the 2017 CHI conference on human factors in computing systems—CHI'17. ACM, New York, NY, USA, pp 6133–6146. https://doi.org/10.1145/3025453.3025953
Heidicker P, Langbehn E, Steinicke F (2017) Influence of avatar appearance on presence in social VR. In: 2017 IEEE symposium on 3D user interfaces, 3DUI 2017—proceedings. Institute of Electrical and Electronics Engineers Inc., pp 233–234. https://doi.org/10.1109/3DUI.2017.7893357
Ho C, Basdogan C, Slater M, Durlach NI, Srinivasan MA (1998) An experiment on the influence of haptic communication on the sense of being together. In: Proceedings of the British telecom workshop on presence in shared virtual environments. pp 10–11
Holderied H (2017) Evaluation of interaction concepts in virtual reality applications. In: Informatik 2017. Gesellschaft für Informatik, Bonn, pp 2511–2523. https://doi.org/10.18420/in2017_254
IJsselsteijn WA, de Ridder H, Freeman J, Avons SE (2000) Presence: concept, determinants and measurement. In: Proceedings SPIE 3959, human vision and electronic imaging V. SPIE, pp 520–529. https://doi.org/10.1117/12.387188
IJsselsteijn W, van Baren J, Markopoulos P, Romero N, de Ruyter B, (2009) Measuring affective benefits and costs of mediated awareness: development and validation of the ABC-questionnaire. In: Awareness systems. Springer, London, pp 473–488. https://doi.org/10.1007/978-1-84882-477-5_20
Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG (1993) Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. Int J Aviat Psychol 3:203–220. https://doi.org/10.1207/s15327108ijap0303_3
Kim J, Song H, Lee S (2018) Extrovert and lonely individuals' social TV viewing experiences: a mediating and moderating role of social presence. Mass Commun Soc 21:50–70. https://doi.org/10.1080/15205436.2017.1350715
Margery D, Arnaldi B, Plouzeau N (1999) A general framework for cooperative manipulation in virtual environments. In: Virtual environments' 99. Springer, pp 169–178. https://doi.org/10.1007/978-3-7091-6805-9_17
Mateer J (2017) Directing for cinematic virtual reality: how the traditional film director's craft applies to immersive environments and notions of presence. J Media Pract 18:14–25. https://doi.org/10.1080/14682753.2017.1305838
Matos T, Nóbrega R, Rodrigues R, Pinheiro M (2018) Dynamic annotations on an interactive web-based 360° video player. In: Proceedings of the 23rd international ACM conference on 3D web technology—Web3D'18. ACM Press, New York, New York, USA, pp 1–4. https://doi.org/10.1145/3208806.3208818
Montagud M, Boronat F, Stokking H, van Brandenburg R (2012) Inter-destination multimedia synchronization: schemes, use cases and standardization. Multimed Syst 18:459–482. https://doi.org/10.1007/s00530-012-0278-9
Montagud M, Cesar Garcia PS, Boronat F, Marfil D (2015) Social media usage combined with TV/video watching: opportunities and associated challenges. IEEE Comput Soc STCSN E-Lett 3:1–6
Montagud M, Fraile I, Meyerson E, Genís M, Fernández S (2018) ImAc: enabling immersive, accessible and personalized media experiences. In: Proceedings of the 2018 ACM international conference on interactive experiences for TV and online video—TVX'18. ACM
Moreno R, Mayer RE (1999) Cognitive principles of multimedia learning: the role of modality and contiguity. J Educ Psychol 91:358–368. https://doi.org/10.1037/0022-0663.91.2.358
Nathan M, Harrison C, Yarosh S, Terveen L, Stead L, Amento B (2008) CollaboraTV: making television viewing social again. In: Proceeding of the 1st international conference on designing interactive user experiences for TV and video—UXTV'08. ACM, New York, NY, USA, pp 85–94. https://doi.org/10.1145/1453805.1453824
Neng LAR, Chambel T (2010) Get around 360° hypervideo. In: Proceedings of the 14th international academic mindtrek conference on envisioning future media environments—MindTrek'10. ACM, New York, NY, USA, pp 119–122. https://doi.org/10.1145/1930488.1930512
Nguyen C, DiVerdi S, Hertzmann A, Liu F (2017) CollaVR: collaborative in-headset review for VR video. In: Proceedings of the 30th

annual ACM symposium on user interface software and technology—UIST'17. ACM Press, New York, New York, USA, pp 267–277. https://doi.org/10.1145/3126594.3126659

Nielsen LT, Møller MB, Hartmeyer SD, Ljung TCM, Nilsson NC, Nordahl R, Serafin S (2016) Missing the point: an exploration of how to guide users' attention during cinematic virtual reality. In: Proceedings of the 22nd ACM conference on virtual reality software and technology—VRST'16. ACM, New York, NY, USA, pp 229–232. https://doi.org/10.1145/2993369.2993405

Normand V, Babski C, Benford S, Bullock A, Carion S, Chrysanthou Y, Farcet N, Frécon E, Harvey J, Kuijpers N et al (1999) The COVEN project: exploring applicative, technical, and usage dimensions of collaborative virtual environments. Presence Teleoper Virtual Environ 8:218–236. https://doi.org/10.1162/105474699566189

Núñez A, Montagud M, Fraile I, Gómez D, Fernández S (2018) ImmersiaTV: an end-to-end toolset to enable customizable and immersive multi-screen TV experiences. In: Workshop on virtual reality, co-located with ACM TVX 2018, Seoul (South Korea)

O'Hagan RG, Zelinsky A, Rougeaux S (2002) Visual gesture interfaces for virtual environments. Interact Comput 14:231–250. https://doi.org/10.1016/S0953-5438(01)00050-9

Oh CS, Bailenson JN, Welch GF (2018) A systematic review of social presence: definition, antecedents, and implications. Front Robot AI 5:1–35. https://doi.org/10.3389/frobt.2018.00114

Pakkanen T, Hakulinen J, Jokela T, Rakkolainen I, Kangas J, Piippo P, Raisamo R, Salmimaa M (2017) Interaction with WebVR 360° video player: comparing three interaction paradigms. In: 2017 IEEE virtual reality conference—IEEEVR'17. IEEE, pp 279–280. https://doi.org/10.1109/VR.2017.7892285

Rothe S, Hußmann H (2018) Guiding the viewer in cinematic virtual reality by diegetic cues. In: International conference on augmented reality, virtual reality and computer graphics. Springer, Cham, pp 101–117. https://doi.org/10.1007/978-3-319-95270-3_7

Rothe S, Buschek D, Hußmann H (2019) Guidance in cinematic virtual reality-taxonomy, research status and challenges. Multimodal Technol Interact 3:19–42. https://doi.org/10.3390/mti3010019

Schubert T, Friedmann F, Regenbrecht H (2002) igroup presence questionnaire (IPQ) [WWW Document]. http://www.igroup.org/pq/ipq/index.php. Accessed 30 June 18

Schultze U, Brooks JAM (2019) An interactional view of social presence: making the virtual other "real". Inf Syst J 29:707–737. https://doi.org/10.1111/isj.12230

Shin D-H, Kim J (2015) Social viewing behavior in social TV: proposing a new concept of socio-usability. Online Inf Rev 39:416–434. https://doi.org/10.1108/OIR-12-2014-0299

Skarbez R, Brooks FP Jr, Whitton MC (2017) A survey of presence and related concepts. ACM Comput Surv 50:1–39. https://doi.org/10.1145/3134301

Smith HJ, Neff M (2018) Communication behavior in embodied virtual reality. In: Conference on human factors in computing systems—proceedings. Association for Computing Machinery, New York, New York, USA, pp 1–12. https://doi.org/10.1145/3173574.3173863

Tang A, Fakourfar O (2017) Watching 360° videos together. In: Proceedings of the 2017 CHI conference on human factors in computing systems—CHI'17. ACM, New York, NY, USA, pp 4501–4506. https://doi.org/10.1145/3025453.3025519

Thies J, Zollhöfer M, Stamminger M, Theobalt C, Nießner M (2016) FaceVR: real-time facial reenactment and eye gaze control in virtual reality. ACM Trans Graph 37:1–15. https://doi.org/10.1145/3182644

Unity - Manual: The Multiplayer High Level API [WWW Document], (2018). https://docs.unity3d.com/Manual/UNetUsingHLAPI.html. Accessed 2 May 19

Voorveld HAM, Viswanathan V (2015) An observational study on how situational factors influence media multitasking with TV: the role of genres, dayparts, and social viewing. Media Psychol 18:499–526. https://doi.org/10.1080/15213269.2013.872038

Waltemate T, Gall D, Roth D, Botsch M, Latoschik ME (2018) The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response. IEEE Trans Vis Comput Graph 24:1643–1652. https://doi.org/10.1109/TVCG.2018.2794629

Weisz JD, Kiesler S, Zhang H, Ren Y, Kraut RE, Konstan JA (2007) Watching together: integrating text chat with video. In: Proceedings of the SIGCHI conference on human factors in computing systems—CHI'07. ACM, New York, NY, USA, pp 877–886. https://doi.org/10.1145/1240624.1240756

Yeh M, Wickens CD, Seagull FJ (1999) Target cuing in visual search: the effects of conformality and display location on the allocation of visual attention. Hum Factors J Hum Factors Ergon Soc 41:524–542. https://doi.org/10.1518/001872099779656752