CollaVR: Collaborative In-Headset Review for VR Video

Cuong Nguyen[1] **Stephen DiVerdi**[2] **Aaron Hertzmann**[2] **Feng Liu**[1]

[1]Portland State University
Portland, OR, USA
{cuong3, fliu}@pdx.edu

[2]Adobe Research
San Francisco, CA, USA
{diverdi, hertzmann}@adobe.com

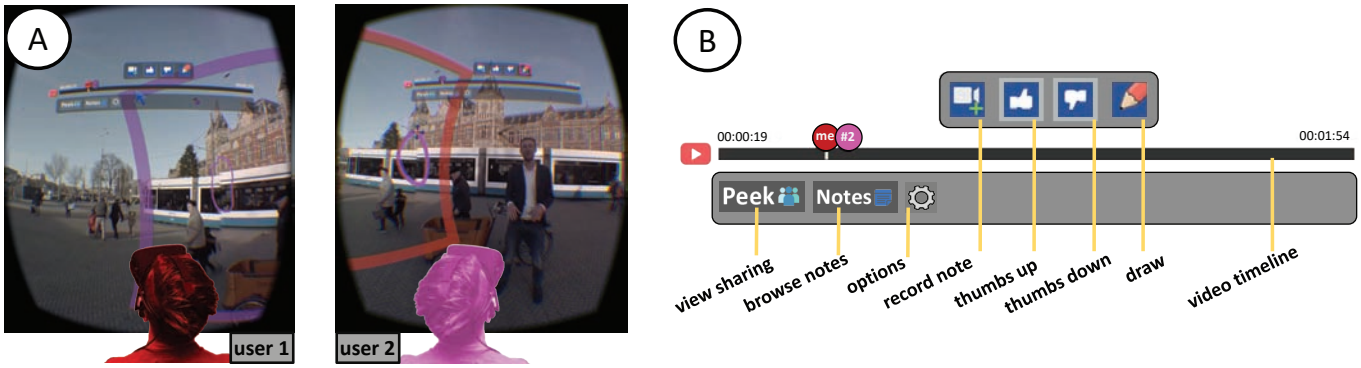


Figure 1: (A): In-headset views of two CollaVR users watching a video together. The clients are connected through a local network. In the headset, users see visualizations of each other's viewport. User 2 (purple) circles a stitching artifact on the video and the drawing immediately appears on user 1's (red) screen. (B): The timeline interface includes features to support communication, view sharing, and notetaking for VR video. © EU2016NL

ABSTRACT

Collaborative review and feedback is an important part of conventional filmmaking and now Virtual Reality (VR) video production as well. However, conventional collaborative review practices do not easily translate to VR video because VR video is normally viewed in a headset, which makes it difficult to align gaze, share context, and take notes. This paper presents CollaVR, an application that enables multiple users to review a VR video together while wearing headsets. We interviewed VR video professionals to distill key considerations in reviewing VR video. Based on these insights, we developed a set of networked tools that enable filmmakers to collaborate and review video in real-time. We conducted a preliminary expert study to solicit feedback from VR video professionals about our system and assess their usage of the system with and without collaboration features.

ACM Classification Keywords

H.5.1 Information Interfaces and Presentation: Multimedia Information Systems

Author Keywords

Virtual reality; collaboration; video reviewing; video editing

UIST 2017, October 22-25, 2017, Quebec City, QC, Canada
 ISBN 978-1-4503-4981-9/17/10 ...$15.00.
https://doi.org/10.1145/3126594.3126659

INTRODUCTION

Collaboration and review are integral parts of the conventional filmmaking process. For instance, editors and directors frequently work in front of the same computer monitor, discussing edits while referring to the display [1]. Likewise, in the "dailies" process, filmmakers gather to view individual shots on a large screen, discussing and giving feedback, while a coordinator controls playback. A producer or client may also review video on their own and then later give written notes to the editor or director.

Virtual Reality (VR) video[1] is an emerging art form with creative practices adapted from conventional filmmaking [23, 27]. However, collaborative review and feedback are much more difficult for VR video because fully experiencing VR video requires wearing a headset (e.g., immersive head-mounted display) that blocks all view of the outside world. This interferes with almost every type of collaboration, because there are no affordances for awareness of others' actions [33]. Two participants in the same room cannot easily point to a specific element in a video. Because each viewer effectively has their own video player, even synchronizing timing and view direction are difficult. The headset makes it difficult to use notetaking devices, such as paper or keyboard. In formative interviews with professionals, we found that editors suffer many of these issues and that current tools fail to create an effective shared environment for video reviewing.

[1]VR video specifically means monoscopic 360° video in this work. We focus on this format since it is the most commonly available for cinematic VR experience; there are also other formats [23].

We present CollaVR, an application that enables multiple users to review a VR video together while wearing headsets. We focus on supporting synchronous collaboration scenarios that are central to film production, in which collaborators can exchange feedback in-person or in dailies review sessions. Through formative interviews with professionals, we distilled several high-level goals for our system and developed a set of networked tools to achieve these goals (Figure 1). In particular, CollaVR's tools enable headset users to 1) quickly understand collaborators' activities and view directions in the headset to streamline discussion in the video, 2) share and synchronize video playback through interactions that resemble physical reviewing activities, and 3) record and review multimodal feedback directly in the headset.

We conducted a preliminary expert study to solicit feedback from VR video professionals about CollaVR. Our results show that experts are positive about the collaboration potentials of our system over a baseline interface that does not support collaboration. Our system allows experts to actively and collaboratively review video in the VR headset using natural interactions that are often seen in face-to-face collaboration. Groups of experts discussed video issues using more implicit than explicit verbal cues, spent more time viewing video together, and engaged in collaborative notetaking.

Our contribution is thus to understand and address the specific use case requirements for collaborative VR video review, and to carefully combine existing technology into a working system to solve specific problems in VR.

RELATED WORK

Collaboration in VR

Research on VR collaboration primarily focuses on 3D environments [4] while VR video is less thoroughly explored. The CU-SeeMe VR system [15] enables teleconferencing in "desktop VR" (a 3D virtual environment viewed on a desktop computer monitor, similar to a modern first-person video game), including an early form of spatialized audio voice chat. Fraser et al. [12] explore visualizations that support awareness in VR collaborations. For collaborative analytic tasks, Cordeil et al. [9] find that VR headsets can be good alternatives to CAVE systems and can support both co-located and remote collaboration. Perhaps most closely relevant to our work, Henrikson et al. [17] propose a storyboarding system to allow an artist and a filmmaker to plan VR stories together, though they do not focus on simultaneous headset viewing scenarios. In contrast to these works, we focus on the review stage of video production and provide tools designed specifically for fine-grained collaboration on VR video.

Watching VR video together

At the core of collaborative video review is the social act of watching a video together. Watching video in a VR headset is normally an isolated experience. Recent explorations have experimented with capturing and rendering humanoid avatars in VR, varying fidelity from simplified 3D models, as seen in the Facebook 360 demo [31], to textured meshes from full body scans [25]. Avatars can convey body language and a sense of co-presence but require specialized capture equipment, such as depth cameras [20]. Moreover, they do not necessarily support fine-grained collaboration required for film production. For example, they may not accurately convey where other participants are looking or pointing. McGill et al. find users are unsure whether to look at the avatar or the video and the lack of shared cues about where the other participant was looking reduced their enjoyment [25]. In contrast, we designed our awareness visualizations for professional users to review video: they convey instantly what other people are doing and where they are looking in the video. These non-verbal cues are important for video reviewing as they help collaborators refer to objects and ground conversation quickly [16]. Our visualization could also aid social viewing.

Collaborative video reviewing

Some previous work focuses on asynchronous review of conventional 2D video. Phalip et al. [30] describe a remote reviewing system for film scores. Pavel et al. [29] describe a system that allows collaborators to record and exchange feedback, and include video recording and browsing features to make asynchronous collaboration similar to in-person reviewing. The interfaces of these systems are feature-rich and are mainly designed to be used on a desktop computer. While these techniques could be used in our system, we emphasize issues central to the problem of in-headset collaboration for VR video. To this end, we focus on synchronous review and address issues of awareness, synchronization, and notetaking that are specific to the in-headset experience. Nguyen et al. [27] present an in-headset tool for VR video editing, but they not provide tools for collaboration. Some commercial review systems such as Lookat.io [21] support adding annotations to VR video on a web browser but do not allow users to review or discuss in real-time in VR headsets.

FORMATIVE INTERVIEWS

We interviewed professionals from four different VR video studios to inform our design. These professionals include a VR editor/filmmaker, a VR technical art lead, a VR director, and two VR editors. We asked participants to describe their current VR video reviewing workflow.

Video production is a collaboration among multiple stakeholders [29]. VR video's uniquely immersive nature requires that all participants review footage in-headset. For example, clients can ensure their brand message is recognizable, editors can find jarring scenes that may cause discomfort, and colorists can spot bad lighting. All the professionals we spoke with felt it was important to review footage in VR.

Unfortunately, our interviews also confirmed a pervasive issue with current reviewing practices: the benefits of face-to-face interaction are curtailed when collaborators wear VR headsets, and simple co-located reviewing tasks such as watching a video together or discussing an issue become very difficult. As a result, each studio has devised their own compromise review solution and there is no one standard "current workflow" in practice. We will now elaborate on the specific needs of the interviewees and how their workflows are adapted accordingly.

Social awareness

When reviewing a VR video together, editors need to understand where everyone is looking, both to confirm everyone saw important details in the video and to be able to have informed discussions. However, once someone is watching a video in-headset, there is no way to directly see what is in their current field of view. One editor mentioned that she frequently has to ask clients "did you see it?" Similarly, when the client gives feedback, the editor does not know what the client's viewing experience was, which inhibits understanding. High-end headsets (e.g., Oculus Rift, HTC Vive) can mirror the in-headset view on a desktop monitor, but video review frequently happens on mobile headsets (e.g., Google Cardboard, Samsung GearVR) which lack this capability. View mirroring is not currently supported remotely.

Common context

Video production involves people with different expertise and domain languages, making communication difficult. It can be even harder in VR video when everyone has to describe their experiences in an immersive environment. Thus, our interviewees stressed the importance of discussing and annotating the video together. One participant explained that when everybody is in a room together and looking at the same video, they can use pronouns like "this" and "that" (i.e., deictic references) and can visually indicate what's on the video to describe changes. With these natural social interactions, an editor can discuss feedback directly with collaborators and there is ample opportunity to articulate or clarify comments. However, when users put on VR headsets, their activities are not shared and they must constantly ground the conversation by cumbersome techniques such as referring to timestamps or specific objects and events in the video (e.g., "at time 31 seconds, the person on the left in the blue shirt").

View sharing

We identified several activities that concern sharing or controlling the video view.

Looking over shoulder. One editor described an ad-hoc reviewing setup: each team member would load the VR video on a phone and watch separately by holding the phone in their hands and rotating it. Users could "peek" at other peoples' screens to gather information or exchange feedback while retaining control of their own video. However, the editor pointed out this method is only good for high-level feedback and coordination, because the video is not viewed in headsets.

Watching together. Several participants mentioned they frequently try to watch video simultaneously in their headsets. Watching together is a common technique in conventional workflows, where editors use a synchronization service such as Cinesync to control video playback [8]. Watching together in the headset allows each viewer to look around independently while synchronizing playback among peers. One editor explained that, when more people look at a video together, there are more opportunities to spot issues. However, controlling the video playback is difficult when each viewer has a separate player, particularly if viewers pause or rewind the playback. As a result, collaborators may not know if they are viewing the same events at the same time, making detailed discussion difficult.

Coordinated viewing. One participant described a "dailies" reviewing process in which the shot animator would select a predefined view direction and render a regular 2D perspective video. The conventional dailies process was then used for the rendered 2D video: a coordinator would control the video playback and everyone else in the room would view the shot and discuss it. This shared viewing and control process does not currently exist in a headset and rendering to a regular 2D perspective video loses the VR viewing experience.

Notetaking

Capturing feedback from a discussion is an important part of the review process. Editors often use review notes as a task list for the next day. Collaborators can share notes for asynchronous reviewing. Written notes, however, are hard to produce when wearing a VR headset. One participant shared that she writes notes blindly on paper. Moreover, VR video can require the reviewer to describe complex in-headset experiences with information about gaze direction, semantics, and visuals that may be difficult to capture in writing. Two editors mentioned that most notes they receive from clients are high-level and do not capture the nuances of the clients' feedback. Another editor said that notetaking during viewing is cumbersome because it requires pausing the video and interrupts the story's flow, so there is a disincentive to provide comments.

DESIGN GOALS

Our formative interviews make it clear that while shared experiences are a common part of traditional video collaboration, they are largely nonexistent for VR video review. As a first step to remedy this, we derived high-level design goals from the interview feedback:

1. **Awareness.** Enable natural social interactions such as observing gaze, listening to voice, and being able to indicate what's on the video.
2. **View sharing.** Establish a common context for discussion via view sharing techniques such as peeking, watching together, and coordinated viewing.
3. **Notetaking.** Enable recording and browsing feedback that captures the full VR video viewing experience.

For the case of synchronous, co-located review, we envision satisfying these design goals in a VR application user interface that allows several users wearing headsets to view a VR video, communicate with one another, and take and share notes. The VR interactions should resemble the natural, intuitive social behaviors that are often used by co-located professionals in film studios for regular videos. These interactions could also generalize to remote or asynchronous collaboration, which we leave as future work.

SYSTEM

Based on our design goals, we developed CollaVR, a system that connects headsets on a local network, allowing multiple users to review VR video together. Figure 1 shows an

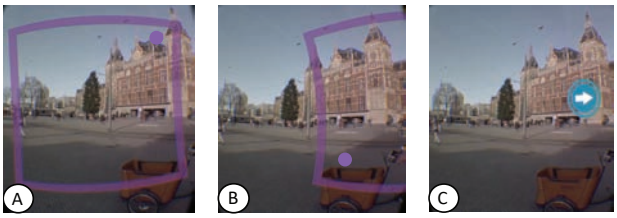


Figure 2: Left: Our system renders the viewport of a collaborator as a color-coded rectangle. The extent of the rectangle depicts the field of view. The 3D cursor pointer is also rendered inside the viewport, so users can point to a video scene element. Middle and Right: When the viewport travels outside the current field of view, an arrow appears on the periphery to provide a directional cue toward the off-screen position. © EU2016NL

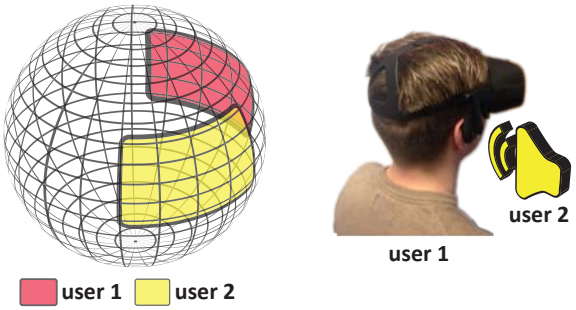


Figure 3: Illustration of spatialized voice chat between user 1 and user 2. As shown on the sphere, user 2 looks to the right of user 1. We spatialize user 2's audio stream so that user 1 will hear user 2's voice as coming from the right.

overview of CollaVR. We employ a client/server architecture. Each client is an instance of an in-headset VR video reviewing application that supports watching VR video and sharing feedback. The server is a separate process that connects clients; we normally run it in the background on one of the client computers. It receives and broadcasts state (e.g., current time and gaze direction) among all client systems and performs audio mixing for voice chat.

CollaVR currently supports the Oculus Rift CV1 and DK2 headsets with orientation tracking. Users interact with CollaVR using a standard desktop mouse and keyboard; handheld VR controllers can also be supported by emulating a mouse input. A microphone and headphones are required for voice chat. Since we focus on studio workflows where multiple people review video synchronously in-person, we assume our system is connected to a fast local network and multimedia files are stored locally on each client computer to reduce network bandwidth.

The client interface (Figure 1b) includes several tools and visualizations to facilitate collaborative video review. Specifically, our "awareness visualization" allows collaborators to quickly understand each other's status, our "view sharing" tools help people share views to establish common context for discussion, and our "feedback recording and browsing" tools allow users to capture and review multimodal notes. We will now detail each of the features of our client system.

Awareness visualization

CollaVR employs a combination of visual and auditory cues to help users quickly understand what their collaborators are doing. In particular, we want to reproduce some of the benefits of face-to-face interactions that people often use in studio, such as observing gaze direction, listening to voice, and gesturing at the video. We convey this information through viewport visualization, spatialized voice chat, and activity visualization.

Viewport visualization

We visualize a collaborator's view as a rectangular viewport (Figure 2a), rendered by projecting a rectangle centered on the collaborator's gaze onto the view sphere. The rectangle size matches the field of view of the collaborator's headset, so the user can understand what video elements are visible to the collaborator. The rectangle is rendered with a thick border, and is uniquely colored for each user.

When two users are discussing a video, if their gaze directions are not close enough, their fields of view may not overlap and the viewport visualization may be entirely offscreen, e.g., if the collaborator has turned too far to the right (Figure 2b). In this case, we show an arrow icon in the periphery of the user's view (Figure 2c) pointing in the direction of the collaborator's viewport. If the user turns to follow the arrow, when the collaborator's viewport is visible again the arrow goes away.

This visualization is only useful for collaborators viewing the same part of the video; in other cases, the visualization may just be distracting [14]. For this reason, we only enable this visualization when the collaborator is viewing video within seven seconds before or after the viewer, and the visualization can also be turned off in the settings as well.

Spatialized voice chat

When immersed in VR with headset and headphones, even though users are co-located, they may not be able to hear one another clearly as they speak. Therefore, CollaVR supports voice chat by streaming audio from the headset microphone to all collaborators, mixed with the VR video's audio.

To further support spatial awareness of collaborators, we spatialize the microphone audio in 3D by making it seem as though a collaborator's audio is emanating from a position on the view sphere coincident with the collaborator's gaze direction (Figure 3). Spatial audio is perceptible even when the source is off-screen [14], and research has shown its effectiveness in aiding visual search [5] and video conferencing [3]. 3D spatial rendering is implemented in the server by transforming the user's single channel microphone audio stream using an off-the-shelf higher-order ambisonics audio library [32]. For each client, the server spatializes the audio streams of the other clients based on their view direction, mixes them to a single stereo audio stream, and sends the mixed audio to the client over the network. Sometimes the VR video sound can interfere with voice chat, so we let users mute the video's audio with a toggle in the "Option" panel.

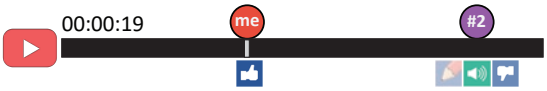


Figure 4: Activity icons such as thumbs up, thumbs down, speaking, and drawing are attached to the user's timeline position. These icons indicate user actions to all collaborators. They fade out over 3 seconds.

Even when the viewport visualization is turned off to minimize visual distractions, spatial audio voice chat can maintain spatial awareness of collaborators. These two cues provide redundant information through complementary modalities.

Activity visualization

During in-person collaboration, people benefit from observing each other's reaction to the video. These reactions are perceived through implicit expressions (e.g., gasping, nodding, facial expressions) and explicit interactions (e.g., dragging the timeline, gesturing towards the video). The headset blocks awareness of others' reactions, so we provide a minimal interface for users to share their activities during review. More complex tracking systems could also be used to capture users' facial expressions and body language [20, 24].

Figure 4 shows our activity visualization. On the timeline, each user is represented as a color-coded icon, labeled with the user ID. Below each user icon are activity icons for speaking, thumbs up, thumbs down, and drawing. These icons appear when the activity occurs and fade out over three seconds. Each client sends its activity events to the server, so they can be shared with all other clients. The speaking event is detected by comparing the root mean square (RMS) energy [28] of the audio buffer with a threshold RMS which is calibrated by measuring a 5-seconds quiet period in the environment.

The buttons for drawing, thumbs up, and thumbs down are located above the timeline for easy access (Figure 1b). A user can draw directly on the video to pinpoint specific details or explain spatial feedback to other collaborators. The thumbs up and down buttons complement drawing by providing a low-effort way to convey affect during playback. As described in our formative interviews, giving feedback on VR video can be difficult for non-expert reviewers, especially when feedback occurs simultaneously with viewing. Moreover, observing thumbs icons on the timeline could also help collaborators monitor and coordinate effort in group work. Previous research found that similar features were well-received by users in a real-time collaborative search setting [26].

View sharing

As mentioned in our formative interviews, editors rely on sharing views in different ways during review. CollaVR provides specific tools to support these behaviors: the "peek" tool supports "looking over the shoulder," the "follow in time" tool enables "watching together," and the "slave" tool supports "coordinated viewing." These features are available on the "Peek" panel below the timeline (Figure 5).

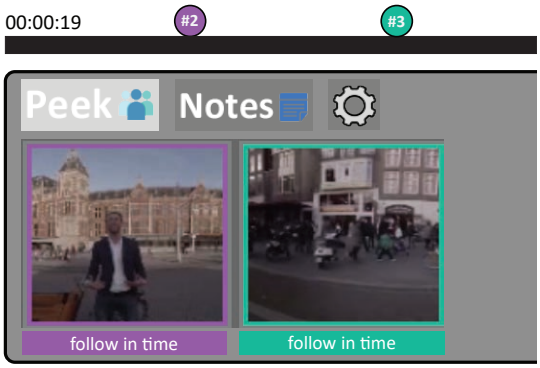


Figure 5: View sharing tools. When "Peek" is enabled, we display color-coded thumbnails showing each user's (e.g., #2 and #3) current video view. The user can further click "follow in time" or click on the thumbnail to trigger different view sharing modes. © EU2016NL

Peek tool

Peeking enables users to quickly see each other's in-headset view. As shown in Figure 5, the Peek panel contains one thumbnail per collaborator, showing each persons' current view of the video. The server broadcasts every users' current timeline position and view direction, and each client renders the thumbnail images accordingly. When a user hovers the mouse over a thumbnail, the corresponding user icon on the timeline is enlarged to emphasize that user's current temporal location.

Follow-in-time tool

Below each thumbnail is a "follow in time" button which allows a user to relinquish timeline control to a collaborator. Importantly, the user retains independent control of their gaze direction, so collaborators can coordinate and divide tasks [10]. For example, several editors can watch different directions of a video together to check for artifacts before publishing; searching for artifacts in a VR video can be tedious when done alone. When follow in time is active, a "Cancel" button appears above the timeline to restore normal viewing.

Slave tool

We also allow users to slave to a collaborator's view to emulate "coordinated viewing." Slaving synchronizes both the time and view direction of the user to a collaborator who is called the "master user" during this interaction. Thus, the slave user sees the exact same video image as the master user, allowing them to discuss fine-grained details or semantics of VR video such as peripheral vision, audience attention, or story [17]. To activate slaving, the user clicks on the thumbnail image and exits this mode with the "Cancel" button.

In our early experiments, we found that slaving could quickly cause simulator sickness [19] in the slave user. The slave user does not have control of their view, so when the head motion of the slave and master users differs, it creates conflicting motion cues for the slave user, leading to discomfort. This effect is related to using a gamepad to control the headset's orientation when playing VR games [36].

Figure 6: Slave visualization. The master user looks at the bicycle rider (right). The slave user (left) uses "Slaving' to watch the master user's view. In this mode, the slave's own view is dimmed and a reduced-size copy of the master full field of view is rendered opaquely on top (middle). This allows the slave user to observe the master's actions while retaining peripheral self-motion perception. © EU2016NL

View vignetting can reduce discomfort [11, 27] but vignettes can occlude the peripheral details of the master user's view. We experimented with a variation of previous techniques [6, 36] in which we render semi-transparent moving particles over the scene, locked to the slave user's head motion. However, in the VR video review context, our pilot users found these particles too distracting and were often seen as video artifacts.

CollaVR employs a new visualization technique to reduce discomfort during slaving, while allowing the slave user to maintain a complete spatial awareness of the master user's view. The slave user's current view is dimmed, and a slightly scaled-down copy of the master user's entire field of view is rendered opaquely on top (Figure 6). This ensures the peripheral motion cues reinforce the slave user's self-motion perception [6], while the main visualization shows the slave user all of the master user's view.

Feedback recording and browsing

Because normal notetaking and feedback techniques are difficult in the headset, CollaVR supports capture and playback of feedback in the headset while also recording the context of the feedback. When the user presses the Record Note button on the timeline (Figure 1b), the user's gaze direction, speech, stroke drawing, and timeline control are all recorded until the user stops the recording. Automatic speech recognition is used to transcribe the user's speech when recording is finished, as this is much more convenient than typing in VR. Always-on recording could also be used, though this may entail privacy issues [37]. The recorded feedback is stored in a SQL database and the server signals clients to retrieve the newly added data, allowing collaborators to easily share feedback.

The "Note" panel (Figure 7) displays all recorded notes, color-coded by user. Hovering the cursor over a note shows the speech-to-text transcription of the user's comments, as well as the time interval that the note spans on the timeline. Clicking the note causes it to play back, and the viewer will see a recording of the original user's interactions. To reduce discomfort, the slaving visualization is also used here.

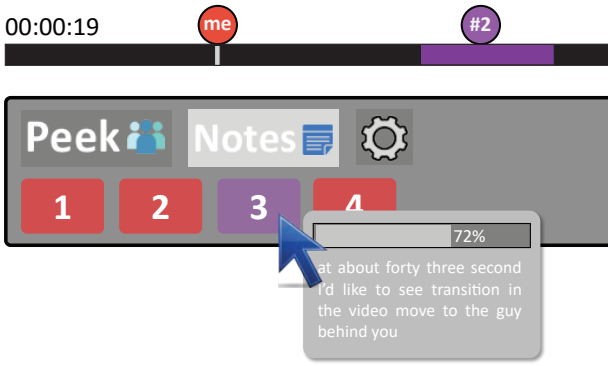


Figure 7: Recorded feedback is listed on the "Notes" panel, color-coded by author. A user can hover over a note to see its transcription and timeline interval. It also shows a progress bar when the user clicks on the feedback to review it.

USER STUDY

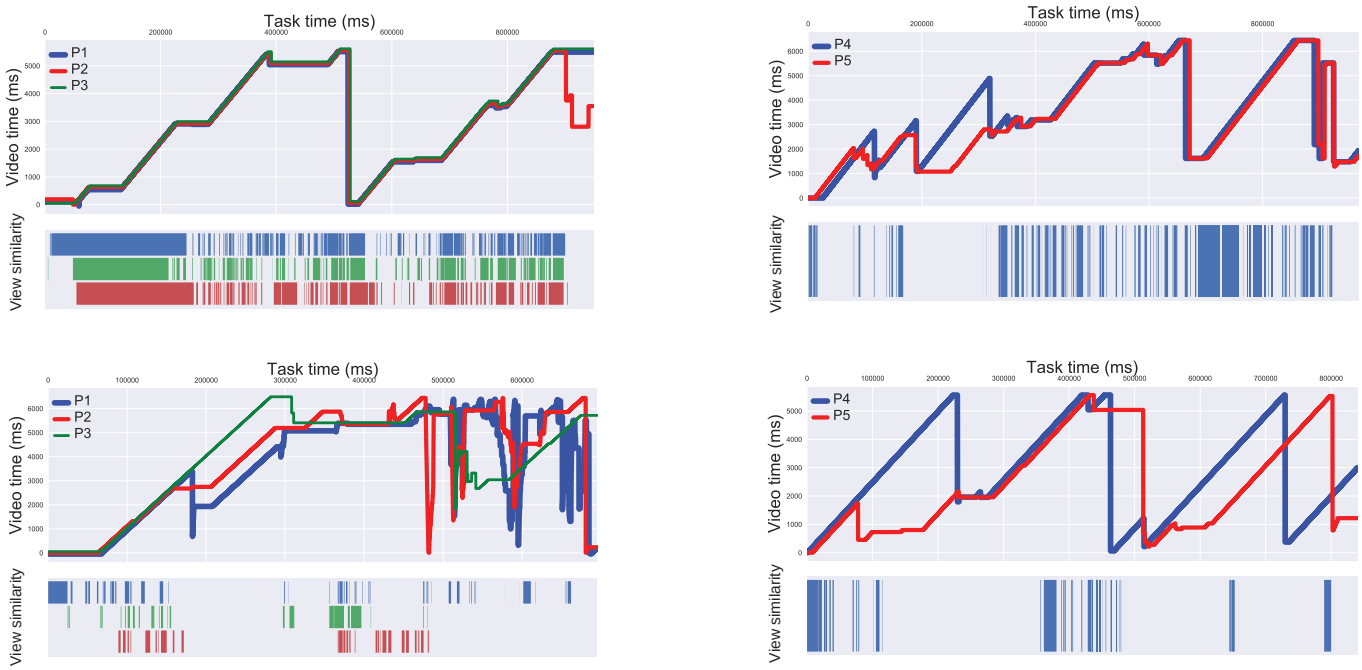
We conducted a preliminary expert review study to solicit feedback on how CollaVR supports collaborative VR video review. Our main research question is to explore how our system can support multiple users to collaborate, discuss, and review VR video together. We chose a 2×2 within-subject design, comparing CollaVR with a baseline in a reviewing task.

Our study asked participants to critique VR video together and provide feedback for an editor. To reduce learning effects, we designed two tasks with two different videos. The videos were two documentaries, one depicting the story of a kite-surfer, and the other, a biking trip through Norway. Both videos were produced by the same aspiring videographer and were roughly equivalent in time (3:30 minutes) and editing style. Both videos also contained several technical issues such as shaky motion and stitching artifacts that motivate discussion. The video resolution is 1920 x 1024 pixels.

The baseline condition is a stripped down version of CollaVR with all the collaboration tools removed (awareness visualization, view sharing, and notetaking). The remaining timeline interface allows viewing a video in a headset and also mirrored on a desktop monitor. Its design is similar to the current VR video players such as the GoPro player[2]. Although CollaVR could be used with a desktop monitor, participants were encouraged to focus the discussion on the VR aspects of the video which require in-headset viewing. Since the baseline does not support notetaking, we gave participants pen and paper in that condition.

We invited five VR video professionals from three studios. These participants were not part of the formative interviews. They were divided into two groups of two to three people. Group 1 consists of a senior producer/panoramic imaging expert (P1), an editor/engineer (P2), and another editor/sound designer (P3) from the same studio. Group 2 consists of an award-winning editor/cinematographer (P4) and a director (P5) from two different studios, but who have collaborated

[2]http://www.kolor.com/gopro-vr-player/download/



(a) Group 1: ours (top), baseline (bottom)

(b) Group 2: ours (top), baseline (bottom)

Figure 8: Video reviewing patterns (Video browsing time and Headset view similarity vs. Task time) of Group 1 and 2. CollaVR enables useres to spend more time watching the video together (at the same time and in similar view directions). Note, in (a), view similarity is color-coded between pairs of users: blue (P1-P2), green (P2-P3), and red (P1-P3), but all three users worked together.

extensively before. Each group was assigned the task (kite-surfing or biking) and the system (CollaVR or baseline) following a counter balance order. Participants were given adequate training with the assigned system using a surfing video that is not used in the study tasks. Participants took 5 minute breaks after each task to alleviate discomfort.

The study was conducted in a university research lab. Participants sat at computer desks in different corners of the room. The computers were connected via Gigabit Ethernet. One Oculus Rift CV1 and two DK2 headsets were used. The CV1 includes headphones and a microphone. The DK2s were paired with Logitech H390 headsets. The latency of the spatialized audio and note recording features was measured end-to-end at 0.36 second and 4 seconds, respectively. Participants were each compensated with a $25 gift card for their time.

Measures

We logged users' activities including voice chat, head tracking data, and tool usage. During the task, two researchers coded participants' conversation into a voice chat log. There is one entry per discussion, recording the timestamp, topic, and whether they used deictic or detailed references [16]. We only note spatial and temporal references. Deictic references use pronouns (e.g., "this," "that") or gaze-centered cues (e.g., "right where I'm looking"), while detailed references explicitly describe scene elements or timestamps.

After each task, we logged the total time and the number of recorded notes, and asked participants to fill out questionnaires individually. The questionnaire asked participants to rate self-perception of collaboration and how well the system supported them. Finally, participants listed their favorite features of the system and described their collaboration strategies.

We analyzed the log data to find times where two users shared the same video context, which is a basis for collaboration [33]. We adapted the *measured shared focus* metric in collaborative VR analytics research [9] to determine users' *headset view similarity*. Two views are considered similar if the angle between the head orientations is less than 40° (half the minimum horizontal field-of-view of the three headsets used in the study) and the difference between the timeline positions is less than seven seconds (determined empirically for the selected video materials). Then the *headset view similarity* is computed by dividing the total time two users have similar views by the total task time.

Results

All groups were able to complete both tasks. Table 1 reports several performance statistics. Overall, groups using CollaVR spent more time on the task, engaged in more discussions, and aligned views more often. Participants were impressed by our interface, commenting that it would help alleviate collaboration problems they currently suffer in many stages of their work, including: having editors resolve issues with clients on the spot (P1, P2, P3, P4), discussing edits with other filmmakers (P3, P5), pitching a new story idea to a client (P3, P4), and dailies review (P5). We now examine participants' interactions in more detail.

	Group 1 (P1, P2, P3)		Group 2 (P4, P5)	
Measures	*Ours*	*Baseline*	*Ours*	*Baseline*
Task time (minutes)	15.8	11.6	16.2	14.00
Number of discussions	26	16	15	12
Headset view similarity (%)	35.5	11.4	19.8	7.1
Notes produced	6	0	3	0

Table 1: Results of the study. The headset view similarity of group 1 was computed by averaging the view similarity of each pair in the group.

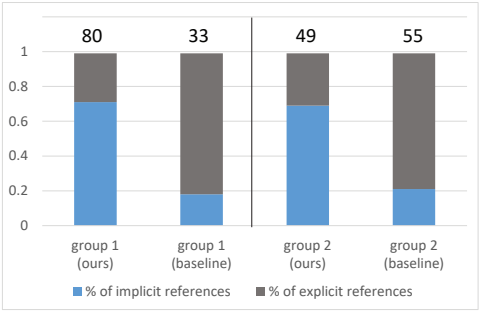


Figure 9: Communication patterns, shown as fraction of implicit and explicit references uttered by study participants. The numbers above the chart show the total number of references made in that condition.

In-headset video reviewing

Figure 8 visualizes the participants' reviewing behaviors. The *Video time* × *Task time* charts show that Group 1 (Figure 8a) mostly watched the video together. They explained that their collaboration strategy involves discussing while watching together. CollaVR supports this via the "follow in time" feature, and the group assigned P3 as the controller with P1 and P2 following. They were unable to enact this strategy with the baseline. P1 started with a countdown so they could all press Play together. They watched the video together at first, but once somebody decided to pause or scrub the timeline, synchronization broke down. Toward the end of the task, each user browsed different parts of the video. The same pattern is apparent for Group 2 (Figure 8b). When using our system, P4 and P5 explored the video separately until they found something interesting. Then they synchronized playback, either manually using the timeline and gaze visualization as guidance, or through the view sharing tools. In the baseline, they made a small effort to synchronize playback, but mostly watched separately, occasionally sharing something interesting.

The *View similarity* × *Task time* charts show how long each group aligned their views (Figure 8). It is clear that CollaVR better supported view alignment than the baseline. The *view similarity* between P1–P2, P2–P3, P1–P3, and P4–P5 in our system are: 51.1%, 27.6%, 27.9%, 19.8%, respectively, and in the baseline are: 14.0%, 9.8%, 10.5%, 7.1%.

Communication patterns

Each participant conducted their review by finding and discussing details about the editing of the video. When using CollaVR, groups discussed more than in the baseline (Table 1). We also counted the number of implicit and explicit spatial references used (Figure 9). Participants of both groups used more implicit references when referring video elements in CollaVR than the baseline, suggesting that our system enables participants to use the same natural verbal cues as in face-to-face interactions such as deixis and gaze [7].

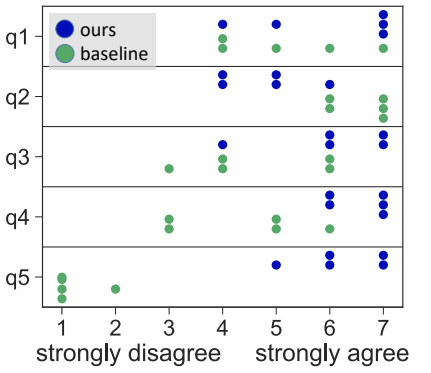
In contrast to CollaVR, participants using the baseline required much more effort to establish common context. P5 mentioned in the post-study interview that "I can't just tell somebody to look at the upper left corner, because there is no corner, you can look anywhere." Moreover, using video elements for reference can lead to misunderstandings. In Group 1, P2 told P1 to look at the left side of a biker, but because the biker was facing the camera, her left was P1's right. P1 misunderstood and looked to the wrong side. P3 actively discussed the video when using CollaVR but said that when using the baseline, he "mostly watched quietly, allowing others to talk." In the baseline condition of Group 2, we counted 8 instances when a person initiated a discussion without any responses from their peers. These results are consistent with previous studies on social VR video watching [25, 34]. Without common conversational context, users' communication suffers.

Subjective feedback

Figure 10a shows participants' responses about perceived comfort (Q2) and collaboration of each system (Q1, Q3, Q4, and Q5). None of the participants reported any symptoms of motion sickness after the study. However, Q2 ratings for our system were slightly lower than the baseline. Q2 does not distinguish between physical comfort and ease-of-use however. P1 and P4 rated Q2 neutral (4) and said they felt a bit overwhelmed to learn all of CollaVR's tools, but they loved the familiarity of the 2D GUI design and thought it would not take long to master. On all four collaboration questions, participants rated CollaVR higher than the baseline.

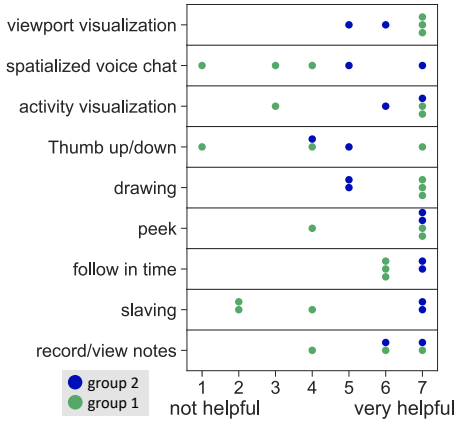
We also asked users to rate each features' helpfulness in supporting collaboration (Figure 10b). Most features received high ratings. Users named their favorite features as: drawing (P1, P2, P3, P5), viewport visualization (P1, P3, P5), follow in time (P1, P3, P5), slaving (P1, P4, P5), peek (P4, P5), and record/view notes (P1, P2, P4, P5).

Although all participants appreciated voice chat, opinions differed about spatialized audio. Group 1 rated it lower, and P1 and P2 said they did not pay much attention to it. Figure 8a shows they mostly aligned views, so the spatial cues may have not been useful. In contrast, Group 2 favored this feature. P4 explained it helped him find collaborators without having to think about it. P5 found an unexpected creative use: he sometimes left the Peek panel open, so he could monitor the video through the collaborator's view and infer its direction from the spatialized audio. Such monitoring is common in group collaboration [35].



Q1: I felt my collaborators and I worked well together
Q2: I felt comfortable using the system
Q3: It was easy to understand what other people are talking about
Q4: It was easy to refer to items in the video
Q5: It was easy to record feedback

(a) Post-task questionnaire



(b) Subjective feedback on our system

Figure 10: Results of the post-task questionnaire (a) and subjective feedback (b) on individual features of our system. Each dot (●) is a rating of participant on a 7-point Likert scale.

Slaving also received diverging ratings. Group 1 gave low scores. P1 and P2 slaved to P3 initially, but since both were active in the discussion, they quickly switched to follow-in-time so they could look around freely. Group 2 rated slaving highly. As a cinematographer, P4 mentioned he could use it to point out both scene details and abstract content like flow or story. P5 praised the visualization, saying that he was surprised it did not make him sick. As a director, P5 deemed this feature as very beneficial as he could have editors locked to his view during review.

Finally, participants found feedback recording very helpful, though they used it differently. In Group 1, P3 recorded the entire session for the whole team, while P2 recorded important details. In Group 2, each participant recorded feedback independently. Participants emphasized that this feature could also be very useful for offline (asynchronous or remote) collaboration. Participants using the baseline stayed in-headset the majority of the time and did not take notes on pen and paper.

DISCUSSION & LIMITATIONS

Our study shows that CollaVR's features enable expert users to collaboratively review VR video relatively unhindered. Compared to the baseline system without any collaboration support, participants using our system spent more time with aligned views and were able to discuss with more deictic references. Watching video together, gesturing, and talking are all natural in-person interactions. Our results suggest the awareness visualization and view sharing tools of CollaVR help users establish common context to discuss videos in-headset, akin to face-to-face collaboration. This is important because understanding VR video requires people to experience it in-headset, and our system helps filmmakers share that experience and exchange ideas within the medium.

The additional collaboration support of CollaVR might have aided participants' reviewing performance. Compared to baseline, they spent more time in the task and initiated more group discussion about VR video editing techniques. Participants also engaged in collaborative notetaking. They used our feedback recording tool to coordinate the notetaking task, or share notes with each other, all in the headset. Although our preliminary study is limited in scale, these results are encouraging and motivate more explorations of collaboration techniques to support in-headset reviewing.

Our study also confirms that multimodal recording is promising to capture interactions in VR video, consistent with previous research on collaborative review [29, 37]. Participants highlighted its potential not only in notetaking, but also in capturing the entire collaboration session and in asynchronous review. Exploring filmmakers' needs for recording and visualizing feedback is interesting future work [2, 22].

CollaVR could be extended to other VR applications. Our features could be integrated with other in-headset production tools such as Vremiere [27] to simultaneously support reviewing and editing, similar to how editors discuss and try alternatives when reviewing regular video [29]. They can also work with other video domains such as training or surveillance. The view sharing and feedback tools can be applied to general 3D environment in VR. For example, they can be used to share lessons in a VR drawing application. Since our system is relatively light-weight, it's possible to use on other hardware platforms. One interesting direction for future work is to explore collaboration in asymmetric hardware setups [13].

In terms of scalability, our system is currently designed with small-group collaboration in mind, which is typical in most studios. For larger groups, such as in a director-team meeting, our view sharing and notetaking tools would still be applicable, but our awareness visualizations may create visual clutter, and spatial voice chat can be less effective. One solution is to let users select a subset of users to interact with or watch, based on their roles in the group. There might be other issues that require further investigation [14].

We conducted our exploratory study in a lab. To better understand CollaVR's real performance requires deploying it in actual studios. Rather than using local video files, we could stream video from an editing suite, enabling editors to dis-

cuss drafts more easily. Better audio compression techniques are needed to improve voice chat performance. Stereoscopic VR video is common and requires special treatment of UI elements [18]. To be more widely useful, we could support low-end headsets such as Google Cardboard and Samsung GearVR as well. The choice of input device is also an open area for research. Our system currently uses mouse and keyboard as input because of their efficiency and their familiarity in professional video workflow. Although none of the participants reported any issues, some users may have difficulties retrieving these devices because they cannot see them.

CONCLUSION

We present CollaVR, an in-headset interface for collaborative VR video review. Our core contribution is a set of techniques that support multiple users to watch VR video, exchange feedback, and take notes together without being hindered by VR headsets, by reproducing the benefits of natural face-to-face interactions. Our preliminary expert review study showed that filmmakers are positive about the potential to review VR videos in CollaVR over a baseline interface. These results highlight the potential of VR video as a collaboration space that we have taken only one step in exploring.

ACKNOWLEDGEMENT

We thank the professionals who participated in our interviews and evaluation for their time and feedback. Figure 1, 2, 5, and 6 use images from YouTube users EU2016NL under a Creative Commons license. This work was supported in part by NSF IIS-1321119.

REFERENCES

1. Ellen Baker, John Geirland, Tom Fisher, and Annmarie Chandler. 1999. Media Production: Towards Creative Collaboration Using Communication Networks. *Computer Supported Cooperative Work (CSCW)* 8, 4 (dec 1999), 303–332. DOI: http://dx.doi.org/10.1023/A:1008616002814

2. Paulo Bala, Mara Dionisio, Valentina Nisi, and Nuno Nunes. 2016. IVRUX: A tool for analyzing immersive narratives in virtual reality. In *Interactive Storytelling*, Ulrike Spierling and Nicolas Szilas (Eds.). Lecture Notes in Computer Science, Vol. 5334. Springer Berlin Heidelberg, 3–11. DOI: http://dx.doi.org/10.1007/978-3-319-48279-8_1

3. Jessica J. Baldis. 2001. Effects of spatial audio on memory, comprehension, and preference during desktop conferences. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 166–173. DOI: http://dx.doi.org/10.1145/365024.365092

4. Steve Benford, Chris Greenhalgh, Tom Rodden, and James Pycock. 2001. Collaborative Virtual Environments. *Commun. ACM* 44, 7 (July 2001), 79–85. DOI: http://dx.doi.org/10.1145/379300.379322

5. M. L. Brown, S. L. Newsome, and E. P. Glinert. 1989. An experiment into the use of auditory cues to reduce visual workload. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 339–346. DOI: http://dx.doi.org/10.1145/67449.67515

6. Gerd Bruder, Frank Steinicke, Phil Wieland, and Markus Lappe. 2012. Tuning self-motion perception in virtual reality with visual illusions. *IEEE Transactions on Visualization and Computer Graphics* 18, 7 (July 2012), 1068–1078. DOI: http://dx.doi.org/10.1109/TVCG.2011.274

7. Mauro Cherubini, Marc-Antoine Nüssli, and Pierre Dillenbourg. 2008. Deixis and gaze in collaborative work at a distance (over a shared map): a computational model to detect misunderstandings. In *Proceedings of the Symposium on Eye Tracking Research & Applications*. 8. DOI:http://dx.doi.org/10.1145/1344471.1344515

8. cineSync. 2005. Retrieved 2017-04-03 from https://cospective.com/cinesync/

9. Maxime Cordeil, Tim Dwyer, Karsten Klein, Bireswar Laha, Kim Marriott, and Bruce H. Thomas. 2017. Immersive collaborative analysis of network connectivity: CAVE-style or head-mounted display? *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (Jan. 2017), 441–450. DOI: http://dx.doi.org/10.1109/TVCG.2016.2599107

10. Paul Dourish and Victoria Bellotti. 1992. Awareness and coordination in shared workspaces. In *Proceedings of the ACM conference on Computer-supported cooperative work*. 107–114. DOI: http://dx.doi.org/10.1145/143457.143468

11. Ajoy S. Fernandes and Steven K. Feiner. 2016. Combating VR sickness through subtle dynamic field-of-view modification. In *Proceedings of the IEEE Symposium on 3D User Interfaces*. 201–210. DOI: http://dx.doi.org/10.1109/3DUI.2016.7460053

12. Mike Fraser, Steve Benford, Jon Hindmarsh, and Christian Heath. 1999. Supporting awareness and interaction through collaborative virtual interfaces. In *Proceedings of the ACM symposium on User interface software and technology*, Vol. 1. 27–36. DOI: http://dx.doi.org/10.1145/320719.322580

13. Jan Gugenheimer, Evgeny Stemasov, Julian Frommel, and Enrico Rukzio. 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 Press, 4021–4033. DOI: http://dx.doi.org/10.1145/3025453.3025683

14. Carl Gutwin and Saul Greenberg. Design for individuals, design for groups: tradeoffs between power and workspace awareness. In *Proceedings of the ACM conference on Computer supported cooperative work*. 207–216. DOI:http://dx.doi.org/10.1145/289444.289495

15. Jefferson Han and Brian Smith. 1996. CU-SeeMe VR immersive desktop teleconferencing. In *Proceedings of the ACM international conference on Multimedia*. 199–207. DOI:http://dx.doi.org/10.1145/244130.244199

16. Jeffrey Heer and Maneesh Agrawala. 2008. Design considerations for collaborative visual analytics. *Information Visualization* 7, 1 (Jan. 2008), 49–62. DOI: http://dx.doi.org/10.1057/palgrave.ivs.9500167

17. Rorik Henrikson, Bruno De Araujo, Fanny Chevalier, Karan Singh, and Ravin Balakrishnan. 2016a. Multi-device storyboards for cinematic narratives in VR. *Proceedings of the ACM Symposium on User interface software and technology* (2016), 787–796. http://doi.acm.org/10.1145/2984511.2984539

18. Rorik Henrikson, Bruno De Araujo, Fanny Chevalier, Karan Singh, and Ravin Balakrishnan. 2016b. Storeoboard: sketching stereoscopic storyboards. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 4587–4598. DOI: http://dx.doi.org/10.1145/2858036.2858079

19. Eugenia M. Kolasinski. 1995. *Simulator sickness in virtual environments*. Technical Report ARI-TR-1027. Army Research Institute for the Behavioral and Social Sciences. DOI:http://dx.doi.org/10.1121/1.404501

20. Hao Li, Laura Trutoiu, Kyle Olszewski, Lingyu Wei, Tristan Trutna, Pei-Lun Hsieh, Aaron Nicholls, and Chongyang Ma. 2015. Facial performance sensing head-mounted display. *ACM Transactions on Graphics (Proceedings of SIGGRAPH)* 34, 4 (July 2015).

21. LookAt. 2017. Retrieved 2017-04-03 from https://www.lookat.io/

22. Thomas Löwe, Michael Stengel, Förster Emmy-Charlotte, Steve Grogorick, and Marcus Magnor. 2015. Visualization and analysis of head movement and gaze data for immersive video in head-mounted displays. *Proceedigns of the Workshop on Eye Tracking and Visualization* 1 (2015), 1–5.

23. Andrew MacQuarrie and Anthony Steed. 2017. Cinematic virtual reality: Evaluating the effect of display type on the viewing experience for panoramic video. In *2017 IEEE Virtual Reality (VR)*. IEEE, 45–54. DOI: http://dx.doi.org/10.1109/VR.2017.7892230

24. Charles Malleson, Maggie Kosek, Martin Klaudiny, Ivan Huerta, Jean-Charles Bazin, Alexander Sorkine-Hornung, Mark Mine, and Kenny Mitchell. 2017. Rapid one-shot acquisition of dynamic VR avatars. In *2017 IEEE Virtual Reality (VR)*. 131–140. DOI: http://dx.doi.org/10.1109/VR.2017.7892240

25. Mark McGill, John H Williamson, and Stephen Brewster. 2016. Examining the role of smart TVs and VR HMDs in synchronous at-a-distance media consumption. *ACM Transactions on Computer-Human Interaction* 23, 5 (Nov. 2016), 1–57. DOI:http://dx.doi.org/10.1145/2983530

26. Meredith Ringel Morris and Eric Horvitz. 2007. SearchTogether: an interface for collaborative web search. In *Proceedings of the ACM symposium on User interface software and technology*. 3. DOI: http://dx.doi.org/10.1145/1294211.1294215

27. Cuong Nguyen, Stephen DiVerdi, Aaron Hertzmann, and Feng Liu. 2017. Vremiere: In-Headset Virtual Reality Video Editing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. DOI:http://dx.doi.org/10.1145/3025453.3025675

28. Costas Panagiotakis and George Tziritas. 2005. A speech/music discriminator based on RMS and zero-crossings. *IEEE Transactions on Multimedia* 7, 1 (Feb. 2005), 155–166. DOI: http://dx.doi.org/10.1109/TMM.2004.840604

29. Amy Pavel, Dan B Goldman, Björn Hartmann, and Maneesh Agrawala. 2016. VidCrit: video-based asynchronous video review. In *Proceedings of the Symposium on User interface software and technology*. 517–528. DOI: http://dx.doi.org/10.1145/2984511.2984552

30. Julien Phalip, Ernest a. Edmonds, and David Jean. 2009. Supporting remote creative collaboration in film scoring. In *Proceeding of the ACM conference on Creativity and cognition*. 211. DOI: http://dx.doi.org/10.1145/1640233.1640266

31. Road to VR. 2016. Facebook social VR demo. Retrieved 2017-04-03 from https://www.youtube.com/watch?v=YuIgyKLPt3s

32. Anne Sèdes, Pierre Guillot, and Eliott Paris. 2014. The HOA library, review and prospects. In *International Computer Music Conference | Sound and Music Computing*. 855 –860. https://hal.archives-ouvertes.fr/hal-01196453

33. Dave Snowdon, Elizabeth F Churchill, and Alan J Munro. 2000. Collaborative virtual environments: digital spaces and places for CSCW. *Collaborative Virtual Environments* (2000), 1–34. DOI: http://dx.doi.org/10.1.1.114.9226

34. Anthony Tang and Omid Fakourfar. 2017. Watching 360° videos together. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. DOI:http://dx.doi.org/10.1145/3025453.3025519

35. James R. Wallace, Stacey D. Scott, Eugene Lai, and Deon Jajalla. 2011. Investigating the role of a large, shared display in multi-display environments. *Computer Supported Cooperative Work* 20, 6 (Dec. 2011), 529–561. DOI:http://dx.doi.org/10.1007/s10606-011-9149-8

36. Robert Xiao and Hrvoje Benko. 2016. Augmenting the field-of-view of head-mounted displays with sparse peripheral displays. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 1221–1232. DOI: http://dx.doi.org/10.1145/2858036.2858212

37. Dongwook Yoon, Nicholas Chen, François Guimbretière, and Abigail Sellen. 2014. RichReview: blending ink, speech, and gesture to support collaborative document review. In *Proceedings of the ACM symposium on User interface software and technology*. 481–490. DOI: http://dx.doi.org/10.1145/2642918.2647390