

Automated Creation of Mobile Video Remixes: User Trial in Three Event Contexts

Jarno Ojala
Tampere University of Technology
Korkeakoulunkatu 1
33720 Tampere Finland
jarno.ojala@tut.fi

Sujeet Mate, Igor D.D. Curcio,
Arto Lehtiniemi
Nokia Technologies,
Tampere, Finland
firstname.lastname@nokia.com

Kaisa Väänänen-Vainio-Mattila
Tampere University of Technology
Korkeakoulunkatu 1
33720 Tampere Finland
kaisa.vaananen-vainio-
mattila@tut.fi

ABSTRACT

This paper describes a user evaluation study of automated creation of mobile video remixes in three different event contexts. The evaluation contributes to the design process of the Automatic Video Remixing System, deepening knowledge to wider usage context. The study was completed with 30 users in three different contexts: a sports event, a music concert and a doctoral dissertation. It was discovered that users are motivated to provide their material to the service when knowing they get an automatically created remix containing many capturers' content in return. Automatic video remixing was stated to ease the task of editing videos and to improve the quality of amateur videos. The study reveals requirements for pleasurable remix creation in different event contexts and details the user experience factors related to the content capturing, sharing, and viewing of captured content and the remixes. The results provide insights into media creation in small event-based groups.

Author Keywords

Mobile videos; collaborative systems; user study; video remix

ACM Classification Keywords

H.5.1 multimedia information systems: video; H.5.3 group and organization interfaces: collaborative computing;

General Terms

Human factors; Design; Experimentation; Theory

1. INTRODUCTION

Most of us have been to a concert, a sports event, or similar, where numerous people in the crowd held a mobile phone to capture a memento of the event. It is rather common to see part of the crowd holding their mobile devices above their heads capturing the event. The habit of spontaneously capturing videos at any chosen event is becoming more common. What happens to these video clips after they are captured is an interesting area to develop new solutions. A major part of the social media use and personal content management nowadays happens with mobile devices such as smart phones, tablets, and other hand-held

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

*MUM '14, November 25 - 28 2014, Melbourne, VIC, Australia
Copyright 2014 ACM 978-1-4503-3304-7/14/11...\$15.00*

devices. The habit of amateur mobile video creation is a growing phenomenon [8, 9]. Online entertainment relies increasingly on user-generated content in social networking services (SNS) and social media. SNS such as YouTube, Facebook, Vine, and Vimeo rely on the video and photo content captured and shared by the users. Mobile video capturing, however, poses problems, as users are struggling with the growing amount of video content they have captured. In a study by Lehmuskallio et al. [14], editing these snapshot videos is a prominent problem that the users face. Eventually this content may be left on the devices, even though the original intention would have been to share it.

This paper presents findings from a user trial of a concept for collective creation of automated mobile video remixes. The concept is called "Automatic Video Remixing System" (AVRS). AVRS is a fully automatic, collaborative video remix creation system. AVRS uses the multiple videos captured by multiple users in an event to create an automatic video remix. The automatically generated remix utilizes multiple perspectives captured by the users' recordings at the event. The remix and the related collaborating group are created by the system in relation to an identified common event like a music concert, a sports event, or a party.

AVRS was originally introduced in [23], where the study compares the product and processes of automatic and manual remix creation. According to the study, although the amateur manual remix performed better in terms of subjective viewing quality, the users were shown to reduce their expectations if they knew beforehand that a remix was generated automatically. Subsequently, the AVRS was used to study the effectiveness of an automatically generated video remix as memorabilia [22]. In the second study, automatic remixes were seen to be fairly equal in acceptability as digital memorabilia of an event. The first two studies were about concert events, these studies did not address user experience aspects that may be of significance when using AVRS in a wider context. Studies did not investigate the design requirements of the front-end of AVRS system or the users' motivations or habits of capturing the videos in the first place.

Different types of events vary by the captured content, audience, and parameters for salient features. For example, a sport event may constrain the user to record from a fixed location whereas recording in a party event can be unconstrained. The audience in a concert may not know each other but have gathered for watching the same band perform. The salient features of a sport event (e.g., a goal or audience reaction) are different compared to a music concert (e.g. a popular song or a speech from the band) or a party (the host and the guests). Consequently, the authors found it

essential to investigate the issues and requirements for collective automatic remix creation in different event contexts.

The goal of this study is to understand four areas which our previous studies of AVRS system did not address. Firstly, it aims to understand motivations and requirements for capturing and contributing video content for automatic video remix creation in different event contexts. Secondly, it identifies automatic video remixing requirements from different types of events. Thirdly, it studies how the collectively created remix is perceived by the users. Fourthly, it identifies features that are desirable to users in a collective video remix system and presents them as a guideline.

This work contributes to the understanding of requirements of an automatic remix and collective video creation in different contexts by event-based small groups. Additionally, the work contributes to the topic of social user experiences [24] by identifying factors that motivate users to share or contribute their video contents to an automatic video remixing service. Our approach of studying the automatic collaborative remix requirements in different event contexts is novel, which helps confirm some previous findings and bring up some results which indicate the need for further study.

2. RELATED WORK

A large number of studies have addressed the habits and patterns of photo sharing and experiences related to mobile photos (e.g., [11, 18, 21]). As mobile videos are increasingly becoming easier to capture and share, the photo-sharing knowledge needs to be extended by the special characteristics of the video content, as videos differ from photos in their temporal dimension. While a number of studies have addressed the areas of collaborative creation and content sharing (e.g., [18, 20, 21]) and collaborative video creation [2, 3, 4] the requirements of different events and the group formation remains a less studied area. Users face problems with their video content editing, especially in the situations where multiple streams of content are available. Automation in video editing can therefore drastically reduce users' time and make the process of video creation more enjoyable.

2.1 Automatic Video Remix Creation and Collaborative Video Creation

Many systems have been studied that utilize a semi-automatic approach to video editing in a collaborative setting for different scenarios, but its development and usage in a collaborative setting are still not completely understood. Engström et al. [3] investigated collaborative video production in a live video setting. The system uses a human-mediated approach for decisions about the choice of what is included from the content received from multiple users. In our study, we also explore the effect of automation in a collaborative video reproduction setting. Girgensohn et al. [5] used a semi-automatic approach for creating home videos, which required assistance from automation in analysis regarding the video motion's characteristics. In contrast with the above-mentioned approaches, the fully automatic approach presents new findings regarding the effectiveness and advantages of such an approach.

Systems using a fully automatic approach for music events have also been studied. Kennedy and Naaman [10] exploited the audio fingerprints from concert videos to organize the content. This approach depends on the number of overlaps to determine what is interesting enough to create an event representation. Shrestha et

al. [26] presents an automatic mash-up creation approach that uses content from multiple users who were recording a music event. Our study investigates the human aspects related to user content contribution, collaboration, and effectiveness of automatic remixes in music and non-music events.

A prototype solution for collaborative video production, called *Caleido* is presented in the work by de Sa et al [2]. *Caleido* offers support for capturing the videos collaboratively, coordinating video capturing. Another approach by Bao et al. [1] utilizes mobile devices as sensors for recording and sensing the environment for creating event highlights. The work mainly focused on significant event detection and its effectiveness, it did not cover the larger user-related experiences regarding content contribution and collaboration. In a system proposed by Zsombori et al. [28], a narrative specification-based approach is used to create video compilations that utilize semi-automatically annotated content; the narrative is chosen by the viewer or derived from viewers' preferences. In the system proposed by Jansen et al. [7], the work by Zsombori et al. [28] is used as a dynamic video compilation.

None of these previous works have provided a detailed study of human aspects about the collaboration motivation and effectiveness of fully automatic system being used for different event types. Collaboration in video creation requires learning, which is addressed in the work by Weilenmann et al [25]. The learning can happen playfully by imitating the professionals, as the work by Juhlin et al [8] suggests. Whereas the presented systems utilize collaboration in the video creation, AVRS aims for collective video creation, since the collaboration is not needed on the video capturing moment. Instead, remixes are created from the collectively captured and shared videos. Interaction with the system in the moment of capturing is kept minimal.

Vihavainen et al. [23] studied use of AVRS at a large-scale festival. The study results suggest that remixes were assessed as important memorabilia equal to the manual remixes from the same event. In the study, users trusted the service and willingly handed over their video clips, even though they stated that they did not want to get acknowledged if their content ended up in the remix. Monroy-Hernandez et al. [16] divide acknowledgement in the content to "attribution" (automatic and computer generated) and "credit" (by other users). How interesting the content is to a user depends on its freshness, the person's relation to the content, the personal nature of the content, and whether the content is actually targeted to the receiver, as previous work suggests [15]. From this, it can be seen that preferences with regards to attribution and credits, as well as the audience [15], may vary depending on how personal the content is to the user.

2.2 Small Groups and Spontaneously Formed Groups

Previous studies suggest that people are willing to share personal content in private circles such as family or close friends [7, 17, 18]. Close-knit groups have needs for demonstrating their group identity and for collectively managing the content [17, 20]. This work extends the idea of small groups to the spontaneous groups that relate to a certain event and thus have a relatively short lifespan. Sharing with the people who were present in the capturing moment is referred to as "reminiscing" [20], and "storytelling" is telling those who were not present about the event [13]. In previous studies relating to small group sharing [18], it was found that small groups have problems in sharing the

picture content from many devices within the group and that people have suspicions over sharing the data on social media. One of the solutions that support small group sharing is the social camera [18]. These studies imply the value of a collective online folder for the photo experience for small groups, especially after meaningful events.

However, targeted sharing to a small group poses problems, as the group formation may vary drastically at different events. SNSs generally face problems with the balance in user-generated content, with massive consumption but little creation [19]. However, creativity can be motivated by giving users a sense of social interaction and connectedness and by lowering the threshold of sharing as the work on social user experience suggests [24]. Social networking services can add collective value to the content by facilitating the sharing of personal media, thus offering a sense of community [12]. Captured and shared content facilitates social interaction and collaboration related to content, and both enrich the content and can lead to new content types and entities. The social user experience happens in a social context, where users and their presence define the actual interaction.

Väänänen-Vainio-Mattila et al. [24] defined factors for social user experience. Curiosity, learning, self-expression, suitability of content and functionalities, completeness of networks, and competition were identified as the motivational drivers for social user experience, which was extended in [18, 24]. The findings presented in this paper contribute to the understanding of content-mediated social user experience with individually recorded video content contribution for automatic video remix creation as well as automatic video remix sharing in small event-based groups in three different event contexts.

3. THE STUDY SETUP

Our research approach is that of constructive design research [27], in which the phenomenon is approached by giving a designed artefact to the study subjects. By the behavior and the feedback of the study subjects the artefact is developed further. In this study, the back end of AVRS, namely the remixing feature, was utilized as the artefact. Artefact was developed further based on the findings of the study. The study was part of user-centered design process aiming to understand the usage patterns of collaborative mobile video remixing and additionally to collect knowledge of the user behaviour in the video capturing events for building the AVRS client, front end of the system. More specifically, this study aims to solve the following research goals:

- 1) What are the motivations of capturing and sharing mobile video content for collaborative remixes?
- 2) What type of requirements do the different events bring to capturing and remixes?
- 3) How do the users perceive the collaboration after seeing the end-product, namely the remixes?
- 4) What type of features should be implemented in the AVRS client application?

Methodologically, the study was organized partly as observed field trial and partly as a qualitative interview study. Observation was done by the researchers in the video capturing events to identify the habits of video capturing that the client application has to support. A total of 30 participants were selected for the study. Fourteen of the participants took part in the video capturing events and sixteen participated as video viewers. All of the 30

participants watched the videos and were interviewed in the final sessions.

The Automatic Video Remixing System (AVRS)

The automatic video remixing system (AVRS) is a fully automatic, collaborative video remix creation system. It was introduced in [23], where it showed that it can be an invaluable tool in reducing the burden of generating video remixes, compared to manual remix creation. This becomes even more prominent in a collaborative environment in which content from multiple users needs to be processed. The quantity of content increases, resulting in the increase of time required for making manual remixes [22]. Figure 1 introduces the four logical phases of the collective video remix creation.

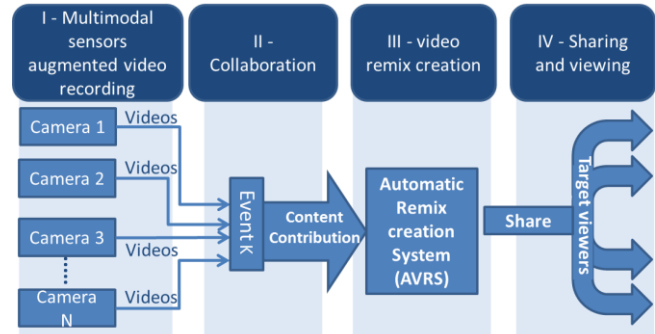


Figure 1: Process of creating video remixes from a user's viewpoint

The automatic remix creation consists of essentially four logical steps (Figure 1). The first step is the multimodal sensor augmented video recording. This phase consists of recording videos that are augmented with multimodal sensor information (compass, accelerometer, and GPS). The second step consists of collaboration for generating the remix. This phase requires collaboration by multiple users who recorded content at the event and contributed their content for making a video remix. The collaboration mechanism consists of creating an “event” in the AVRS system. The “event” acts as an identifier for collecting the content contributions, which are envisaged to be used as input for generating the video remix. The event identifier is used as a logical common repository for all the related content contributions. The other users at the event need to join the created event. Subsequently, the users select the content from the list of recorded content to be uploaded to the AVRS system.

The video remix is generated automatically by the AVRS system after the predefined minimum content availability threshold is fulfilled. The third step is video remix creation. This consists of generating the automatic remix from the contributed content. The AVRS system has been improved compared to the previous version [23]. The improvements relate to the use of the best quality content from the available content in the video remix, the inclusion of relevant views from all of the available views, and changes of views depending on the audio rhythm, aiming for a more interesting remixes. The final step is sharing and viewing the video remix generated in the previous step. This step signifies the fruition of all the effort that the multiple users have invested in making a video remix. Sharing of the video remix with the audience of interest is an important step for user satisfaction, since it enables viewing of the video remix by the intended audience. Sharing the content to the server can be handled either instantly or after the event.

The AVRS system enables people to collaborate by allowing them to form spontaneous groups based on a certain event. As an addition to the small group sharing in the previous work, this work gives a perspective about the event-based small groups. For public events, anyone attending the event can join this collective effort, even if the users do not have each other's contact information or know each other. The remix includes multiple video views over a common audio background track, which represents the common audio scene at the event. The audio source is selected based on the quality of audio. The rhythm of switches between views is in accordance with the audio tempo to allow new views in the video remix.

3.1 Participants and Method

We recruited 30 people living in (removed for blind review) for this study, 13 males and 17 females, with ages varying from 20 years to 50 years. The average age was 28.6 years. Among these 30 people, 21 were students, while the remaining nine worked in different fields. Fifteen participants worked or studied in an ICT-related field. Nine had some previous experience in video editing. All participants received a small reward.

This study had two phases. The first was a video capturing session that included a briefing about this study. In the video capturing sessions, researchers took part and observed the events and how participants captured the videos. The second phase was the final interview session with a debriefing. Fourteen participants were involved in both phases, as video takers at the events and as interviewees. Sixteen participants took part only in the interview sessions as video material viewers. This selection was made to reflect the real users' situations and that of the interest groups, in which only some of the video remix viewers had actually attended the related event.

The AVRS concept in full was presented to the users as a storyboard that explained the functionality in a real-use scenario (Figure 2).

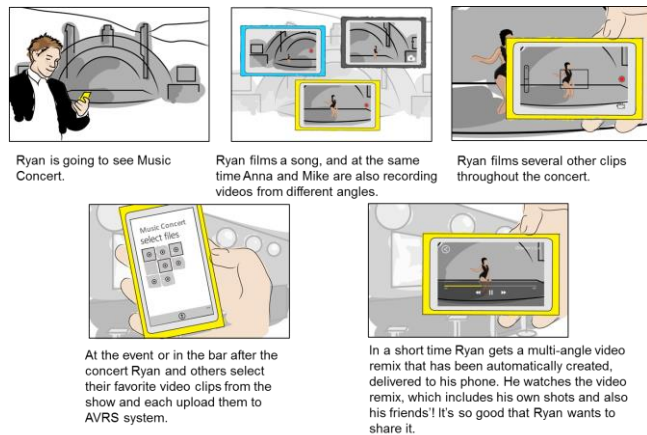


Figure 2: Concept of AVRS. The concept slides were presented to the participants in the interview sessions.

Figure 2 shows the concept slides that were shown to the users in the interview session to describe the functionalities of the concept. This phase of study was organized for collecting feedback for the AVRS client implementation in order to discover the features the application has to offer on the time of video capturing. The actual concept was introduced to participants using a concept slides, as the actual concept requires minimal interaction during the video capturing in the events (Figure 2). Similar approach, but using

low-fidelity prototypes has been introduced in work by de Sa et al [2]. The interview evaluation was complemented by a user experience questionnaire. Comparison of the automatic and manual remixes is beyond the scope of this paper, since it will require detailed treatment to present the results and discuss the user experience implications on the system requirements.

In the final interview sessions, semi-structured interviews were carried out, consisting of individual session and group sessions (of two to three people). In these interview sessions, the participants watched and evaluated three video clips. The first clip was a randomly selected raw video clip from the event that was not edited in any way. The second clip was a manual remix made from the raw video clips recorded by the trial participants in phase 1 (made by one of the authors). The third clip was an automatic video remix clip of the raw video clips recorded by the trial participants in phase 1. Remixes 2 and 3 were shown in a random order. The users watched one raw clip, two video remixes, and the AVRS as a concept to get the idea of the remixing functionality and its capabilities.

The final interviews were audio recorded, resulting in a total of almost 90 hours of raw interview data. Users were not informed beforehand about how the remixes were made. After the remixes were shown and the user experience surveys and interview questions were answered, we revealed that one of the two video remixes was automatically created. For each interview, the responsible researcher wrote notes. Data was then analyzed by using the Affinity Diagram approach electronically [6].

3.2 Video Capturing Events

Trial participants were divided into groups of video capturers and viewers. Three different events, each belong to a different event type, were organized for this study. The different event types were sport event (an Ice Hockey match), a music concert and a formal event (a doctoral dissertation defense and dinner party). Each of these events brings wide variation in the content capture situation (Ice Hockey event was in big stadium, the music concert was in a small club and the doctoral event in a more private venue), the composition of the audience and the parameters for determining the salient features to be included in a video remix.

The chosen events represent a diverse contextual situation, and hence it was considered a good choice for discovering new user requirements. While the video capturers attended the organized events as well as the interviews, viewers took part only in the interview sessions. The choice of the specific events was influenced also by practical considerations like ability to recruit users who may be actually interested in recording in the events and also have interest in the content. These practical constraints in user recruitment did not allow including niche events like exhibitions, museums, trade fairs, etc., The study was designed in that manner in order to simulate the real usage of the AVRS concept, whereby only some of the users capture videos at the events. This assumption is also valid for user-generated content consumption in general. Users who record or create content share it with others, and in many cases, the viewer group is more numerous than that of people who record or create content. This design gave us the possibility to study the differences in the ratings between the groups.

In the video capturing events, users were instructed to capture videos at specified times using all the devices together and to capture more at will. After the events, the smartphones were collected from the users and the material was uploaded to a server.

In this study, the users did not complete the uploading part. Instead, they saw the end result remix in the final interview sessions. In all of the following events, users captured videos with three Nokia Pureview N808 smartphones and additional N8 smartphones.

Event 1: Finnish national league ice hockey game at Hakametsä ice hall in Tampere. At the event, six participants and one of the researchers shot videos. Three of the participants knew each other beforehand. In addition to the capturers, six viewers watched the material during the interview sessions.

Event 2: Music concert held at a local venue, called YO-talo. One of the researchers' band performed at the event. Five participants and one researcher captured video material at the event and seven viewers took part in the final interview sessions.

Event 3: Doctoral defense held at a local university. Three participants shot the videos at the event. All of them were part of the same project group and knew each other well. They also knew the doctoral candidate. Three viewers, who knew the doctoral candidate, took part in the interviews.

4. RESULTS

The user study findings are presented in a similar order than the processes described in Figure 1. First, the factors that motivated people to capture and share video content at the specified events are presented. Second, the requirements that different events bring to the remixes are discussed, and the benefits of using the system are assembled. Third, the factors that affect the ownership of videos and collaboration are presented. In the end, user needs for the remixes are discussed and finally complemented with the requirements for the implementation.

Generally, users who attended the events and captured videos saw the concept as handier than those who only watched the videos. The concept idea was described as fun and easy in the interviews. The majority of the users stated that the automatic end result, the video remix, was of better quality than they would have been able to make by themselves with manual editing. Nine of the participants had some experience in video editing and even these participants saw value in the automatic remixing. The ease in producing the remixes from many video sources was appreciated as well as the quality of the automatic remix. *"I have masses of photos and videos that are only on my phone, but whenever I happen to see them, they evoke memories!"* (P26). *"It was quite exciting. I could not believe that computer could end up with such a good result."* (P17).

4.1 Self-Expression and Connectedness: Motivations for Capturing and Sharing Raw Video Material

The study investigated the motivations to capture and share video material in the events. The following section describes the findings related to content capturing and sharing. The AVRS concept was intended to add reciprocity and a feeling of social presence and awareness [24] to the video capturing. When you contribute to the collective video remix, you get others' content in return. This also motivates users by providing extra material and viewpoints in addition to one's own recorded content, which is obviously captured from the same spot where the capturer experienced the event, thus adding a feeling of connectedness with other capturers [18]. Others' materials can enhance their own captured material. *"Single capturer cannot take all the angles and*

in some events move at all. It can be more interesting with the multiple cameras. It raises the watching experience" (P21).

Capturing and sharing videos was stated to add a social dimension to the events and interaction mediated by the content afterwards. Social dimension motivates users to participate and contribute content [15, 18] Ease of creating the remixes was stated to be the main benefit of using the service. Videos tend to be left on the personal devices, even though the intention was to share them. Automatic remix creation provides a channel to the content. *"Videos and photos are shared in FB in a closed group or by e-mails. It might take two years in some events."* (P16).

Figure 3 gives an overview of the social user experience with the service being studied. Figure 3 shows that statements related to the sociability of the AVRS got high ratings regardless of the event. Users saw the remix as a social effort and they were mainly willing to be social with the other users of the service.

Statements on a Likert scale (1=totally disagree to 7= totally agree) Average (standard deviation in brackets)	Ice Hockey N=12	Concert N=12	Dissertation N=6	Total N=30
This concept idea would make it easy to share videos with the people who attended the event.	6.50 (0.52)	6.33 (0.78)	6.00 (0.89)	6.33 (0.71)
This concept idea would make it easy to share videos with the people who did not attend the event.	6.42 (0.79)	5.50 (1.00)	5.50 (1.22)	5.87 (1.04)
I'm interested in knowing whose video clips I'm watching.	4.75 (1.48)	5.00 (1.71)	5.33 (0.82)	4.97 (1.45)
It is fun to see videos including content captured by other users	6.17 (0.58)	5.75 (0.87)	5.83 (1.17)	5.93 (0.83)
Overall grade for the concept idea that was presented?	5.83 (0.52):	5.75 (0.62)	5.83 (0.72)	5.80 (0.66)

Figure 3: Sociability of User Experience with the AVRS

As Figure 3 shows, the overall grade for the concept (N=30) was an average of 5.83 on the Likert scale of 1 to 7. Figure 3 gives an overview of the ratings that the different groups gave to the concept. Participants stated that collective video remixing and knowing that there will be captured content from others allows them to be creative and express themselves. Self-expression and creating users' own identity is also a driver of the social user experience in the previous work [24]. Figure 3 expresses the difference in the nature of the events. The ice hockey match was seen as a mass event, which could be of interest to those who had not participated. Higher rating on the "This concept idea would make it easy to share videos with the people who did not attend the event" statement suggests that concept was seen more convenient for events like that. On the other hand, dissertation was more intimate event for a smaller group, which can be seen in

the statement “I’m interested in knowing whose video clips I’m watching.”, where the dissertation event got higher ratings.

AVRS concept can help users in being creative in their video capturing. Knowing that the main focus will be shot by multiple capturers allows users to freely express themselves and capture the unexpected and interesting things happening in the background. *“An option is to personalize the stuff for yourself, shoot everything where other cameras do not point. You can see, for example, what your own friendly group or celebrities in the concert did in the audience!”* (P26).

Being a part of the video collective was stated to be a motivating by many of the participants. Content from many capturers was stated to result in a better end-product, if the remixing was handled automatically. Surprisingly, automatic remixes were stated to be artistic and varied. The automatically created remixes from many sources can raise the quality of YouTube live videos. *“It can give very diverse remixes, by combining the stuff from many shooters. Professionals can do it, but to hobby shooter it can really be supportive.”* (P21). *“Your own material will be better when others’ material is automatically added”* (P7).

Fundamentally, capturing videos at special events can shift the focus from the enjoyment and experience of the event. Current design of AVRS aims for minimal interaction with the client during the video capturing. *“I don’t usually like to shoot videos. It takes something away from the enjoyment of the gig”* (P21).

4.2 Requirements of the Different Kinds of Events

The study investigated the requirements that different event types bring to the AVRS concept. The AVRS concept was stated to be effective in offering additional amateur video content to be mixed with the professionally captured content and thus adding new angles to the experience. Different kinds of events where the videos are captured by event attendees and amateur capturers impose various requirements for the video remixing system. The following section describes the requirements for the different event contexts.

Sporting Events: Requirements for a sporting event, such as a hockey game, are built around the earlier habit of watching games on television. Earlier experiences dominate the perception and anything different can feel wrong at the beginning. Sporting event broadcasts follow certain conventions that must also be followed in the remix. For example, conventions do not allow 180-degree turnovers during the game. Also certain highlights such as goals and player information are familiar and their absence lowers the perception of the remix quality. Users wanted to have relatively long periods without any switches and smooth camera changes, even though the pace of the sport may be fast. Reactions and the feeling of an audience presence is important in sporting events that fundamentally rely on spectators. *“If somebody manages to capture something special, for example, in the audience, the audio track can still follow the game at same time”* (P25).

Music Concerts: The concert setting was the event type that was also covered in the earlier work by Vihavainen et al. [23]. In a concert setting, automatic video remixing can bring extra value to the classic mobile video shots. In the interviews, the users very clearly indicated that expectations of the live videos shot with mobiles were relatively low, and users can easily be surprised positively by using material from multiple sources. Users can get the viewpoint of others in the audience. *“Since your own seat may*

be fixed and cannot move freely, it will be interesting to see content from other viewpoint” (P12).

It was clearly important to have the overall atmosphere included in the video, namely the audience and the venue. Concert settings give freedom to the camera changes, but there are still parts that can raise frustration, if they are accidentally cut out of the video remix. *“It was pleasant to watch. The camera changes were smooth, and it didn’t feel like randomly shooting around. The end was still stupid, because it cut away the part where singer was about to give a speech. If that happened to the video of a band that I’m a fan of, it would be irritating!”* (P19).

Formal Events: Formal events such as big celebrations and work- or study-related events have different requirements for the video capturing and what viewers expect of the video remix. Events like a dissertation presentation include a lot of speech, during which the speakers are sitting still and comprehension of what is being said is important. *“Sub-titling should be included if the audio is not good. Audio and spoken words are so important”* (P28).

Formal events pose problems for the video capturers. Video capturing must not disturb the flow of the event and has to be unobtrusive. Balancing between the formality and informal parts is important regarding the audience for the remix. At the formal events, it is important that the main persons are in their main roles in the remix. An absence of the main persons lowers the feel of quality of the remix. *“It would have been possible, if the camera was on a tripod or remotely controlled, to avoid making a lot of fuss. That would not have disturbed us that much”* (P27). *“If the whole dissertation is remixed, it must include dialogue between the candidate and the opponents and the presentation. It has to be formal at that point”* (P25).

All of the studied event types shared certain similarities in the requirements of the video remix. In all of the studied events, camera changes needed to happen for a reason or to support storytelling to get the best experience. The reasons and the way of storytelling are different in the various contexts, and the storytelling has to follow the conventions of the event type. For example, the camera should not change to a long shot or bird’s-eye view when something is happening. Users stated that the concept would be useful at events that are not captured professionally. In the events where audience has a significant role in building the atmosphere, the audience should be audible and visible in the remix also. *“I would like to hear the sound of supporters. You expected that the audience would explode into screams when goals comes. (P1)”*.

Additionally, at the events where there are lots of things happening at the same time and people are scattered around, it would be useful for mediating the events to those that are on different locations. *“For the events that are not recorded in other ways. Junior league football matches or special events like the one where <removed for the blind review> United Supporters team got a promotion to the fifth division!”* (P11). *“Festivals are relevant. Things happen in various stages, so you want to see what happens elsewhere. You shoot one gig and get other in return!”* (P12).

4.3 Collaboration and Discovery: Ownership of the Remixes and Videos

The study focused thirdly on the factors that affect the ownership of videos and how the automatic remixing enables collaboration. This section describes how the collective remix can enable

collaboration. In automated video remixing the users do not actually collaborate in the remix making, but instead they collaborate as the content creators when they allow their videos to be used in the resulting remix. Videos are captured collectively and therefore the system differs from the previous collaborative video systems. This creates fundamental difference, since the collaborative work is mostly automatic.

Understanding the audience the remix can reach and possibility to limit it were important for the users. Even though content such as large-scale festival videos can be public, all of the shared content is not perceived as widely public content. Small group sharing and limiting the audience are important. Even at the mass events, some people were interested in seeing the “viewpoint of my friends” or a similar limited edition of the video remix, consisting of recorded content only from a subset of the event participants. Participants saw many possible uses for the automatic remixes. The end product would be useful as a way to combine material from social and family events. The end product could be handed out as a gift to friends and relatives, or as a bigger group memorabilia. “I would take videos and photos of my godchild and then make the remix on the first birthday” (P19). “I got invited to an event where I see people I haven’t seen in three years. This could give the whole group a memorabilia of the event!” (P18).

Figure 4 describes the ratings related to the video content sharing from the user experience questionnaire. Figure 4 shows users gave relatively high ratings to the statements related to willingness to share their video content to the service. Answers are divided into capturers and viewers at all events. The ratings were relatively high, regardless of if they participated in the video capturing or just watched the videos.

Statements on a Likert scale (1=totally disagree to 7=totally agree) Average (standard deviation in brackets)	Capturers N=14	Viewers N=16	All N=30
I would allow my personal video clips to be used on the remixes.	5.65 (1.62)	5.31 (1.25)	5.50 (1.46)
I would allow others to edit raw video I have captured.	5.76 (1.64)	5.46 (1.05)	5.63 (1.40)
I would like to do the video remixes between more private or closed group (group of my friends or family for example).	6.24 (1.09)	5.92 (0.95)	6.10 (1.03)
I would give the videos I have captured to use in the system for making the video remixes.	5.06 (1.82)	5.38 (1.26)	5.20 (1.58)
I’m interested in seeing in which remixes my clips end up into.	6.12 (1.45)	6.46 (0.66)	6.27 (1.17)
I’m interested in seeing who sees my video clips in the remixes.	5.35 (1.50)	5.31 (1.38)	5.33 (1.42)

Figure 4: Content sharing related statements on the AVRS user experience questionnaire. Comparison is made between the capturers and viewers of the content.

Ownership of the videos was not important for the participants, but getting recognition for what they had made was, as high ratings in the statement “I’m interested in seeing in which remixes

my clips end up into.” suggest (Figure 4). In the user study, the participants were willing to hand over the video material for this kind of service. Since the video material was shot at the request of the users, obviously it makes it more impersonal, and the case will be different in real life. Occasionally, concerns regarding the ownership of the recorded content were raised. The concerns were about the presence of copyrighted content in the recorded material or if the video remixes were used for commercial purposes. Unlike the previous studies by Vihavainen et al. [22, 23], some users wanted credit for their material in the remix. However, it was also stated that there are other channels of creating and sharing the video if you want to make it your own work of art. “The shooters name or tag should be visible in the video” (P11).

Figure 5 shows how the answers to statements related to video sharing differed between the different events.

Statements on a Likert scale (1=totally disagree to 7= totally agree) Average (standard deviation in brackets)	Ice Hockey N=12	Concert N=12	Dissertation N=6	Total N=30
I would allow my personal video clips to be used on the remixes.	5,83 (1,03)	5,17 (1,31)	5,50 (1,86)	5.50 (1.46)
I would allow others to edit raw video I have captured.	5,67 (1,61)	5,67 (0,90)	5,50 (1,96)	5.63 (1.40)
I would like to do the video remixes between more private or closed group (group of my friends or family for example).	5,50 (1,17)	6,42 (0,67)	6,67 (0,83)	6.10 (1.03)
I would give the videos I have captured to use in the system for making the video remixes.	5,08 (1,83)	5,00 (1,08)	5,83 (0,79)	5.20 (1.58)
I’m interested in seeing in which remixes my clips end up into.	6,75 (0,45)	5,75 (0,72)	6,33 (0,49)	6.27 (1.17)
I’m interested in seeing who sees my video clips in the remixes.	5,42 (1,38)	5,17 (1,17)	5,50 (0,81)	5.33 (1.42)
Other users’ video clips were interesting	5,75 (0,75)	5,08 (1,44)	6,00 (0,91)	5.53 (1,28)

Figure 5: Content sharing related statements on the AVRS user experience questionnaire. Comparison is made between the different event groups.

As figure 5 suggests, participants were less willing to share their video clips in the concert setting. Partly because they were not familiar with the bands performing, which also show in the lower ratings to the statement “Other users’ video clips were interesting”. Sharing the video remixes between smaller target group was more important in the concert and dissertation setting,

as ratings to statement “I would like to do the video remixes between more private or closed group” suggest. Mass sports event are fundamentally open and broadcasted events. However, in such mass events, users are particularly interested in seeing if their own video clips reach the remix, as high ratings from ice hockey event group in statement “I’m interested in seeing in which remixes my clips end up into” suggests.

As the users were willing to share their video clips, they at the same time felt connection to the material they had captured. High ratings on the statements “I’m interested in seeing in which remixes my clips end up into” and “I’m interested in seeing who sees my video clips in the remixes” suggest, that users were willing to know how their own video clips were used in the remixes.

As the content is uploaded to the server for the remix purposes, it offers an opportunity to find and store content afterwards, thus adding the possibilities of content discovery [18, 19]. *“I could add social dimension to the concert if the whole group of friends would shoot videos and share. Even more if another friend has been on the same gig” (P25).*

Participants shared the fundamental idea that the contributors owned the remix all together, even though their own clip did not end up in the final remix. The experience of creating the video remix was stated to add collaboration to the user experience, even though the creators may not know each other [18, 19]. They wanted the service to be responsible for the legality of the material in the end. The copyrights should be owned by all the users, for example, if the remix goes public in news services. *“In a way, the shooters own it together, but I’m not sure if they really have the license or copyright to the artwork. You cannot expect that basic users take care of the copyrights” (P8).* *“The videos are shot everywhere, but it is kind of mixed up situation with the copyrights. If the remix is made from a commercial concert, it would be good if the service could take care of the legal stuff” (P16).*

4.4 Design Implications for Collective Mobile Video Creation

Finally, study gathered a list of requirements of the system from the user feedback. These findings were analyzed and elaborated as design implications for similar solutions. They are presented in the following section.

Automatic remix creation, the pro-activity of the concept and level of user control raised concerns amongst participants in this study. The level of user control is previously addressed in the work by Vihavainen et al [23]. The first concern raised by the users was that interesting parts will be left out of the video remix, and the second concern was that something would be published unintentionally. Combining automation and user control was stated as the most efficient way to end up with a sufficient remix result. Finding the right balance between automation and user control and user efforts determines how useful and pleasurable to use the AVRS solution is.

As an approach to control the content in remixes, two prominent methods were discussed in the interviews: automatically detecting the important parts and detecting them with the help of user feedback. Two important factors define the need for the annotations: identifying important clips and the clips that can be left out of the remix. For making the annotations, there are two possible ways to add the information to the video content. The

first possibility is when the raw videos are watched. The second possibility is at the time of watching the video remix. Annotations that are made at the time the videos are shot, using simple interactions, were said to be most time efficient. Making annotations afterwards is hard and time consuming, as well as non-motivating. Making annotations must not disturb the video capturing at the events. *“Users should be able to mark the interesting moments of the event when capturing the videos. Users should be allowed to be lazy” (P1).* *“It may be that you have only one hand free for the video shooting, so it has to be that easy” (P24)* *“Maybe with simple interactions where you select interesting moment and want to see more: more camera angles. Here’s a concept from skate boarding: you capture hours of shots and when you get the perfect shot, the cameraman puts hand over the lens and then you can see the mark when you watch the clips” (P4).*

The number of video capturers at a specific time and to a specific direction offers a data to detect the most important and interesting moments, Detecting the moments that gathered collective interest is an interesting development area to research further. *“If there is something important shot from different angles, it will most likely be important. (P3)”*

Current design of AVRS aims for minimal interaction with the client during the video capturing. Participants gave ideas on how the system could give help in the moment of capturing the videos. They wanted a system that could work as a real-time director of the multiple cameras. For example, it could tell how many cameras are recording certain view. *“If the picture is low quality due to light or shaky, it would help.” (P21).* *“Give each camera certain roles. If someone is covering one of the important things, the other people can cover other things” (P28).*

Detecting the most interesting parts automatically by gestures, laughter, and funny faces was suggested to be a promising approach to detect the interesting parts of the clips. Additionally, automatic selection of the close-ups and detailed shots could make the remixes better and more interesting. Additionally, the system could exploit face recognition in order to make sure that there is video from all the important persons from the event. *The shooters could have a common sense of who are the main persons. Maybe first by tagging the faces and then the system could tell that at the moment no one is shooting the doctoral candidate for example and show the red light” (P28).*

Participants wanted someone to be in charge of the final remix in situations like weddings and formal celebrations. For example, in the dissertation, one of the contributors could be nominated as a director. Content must be previewed by the concerned people for privacy and emotional reasons in such events. Making selections and annotations with the help of crowdsourcing was seen as a promising approach, as well as democratic principles to decide on the remix publishing. Users wanted to give different parameters to create personal and iterative remixes. The motivation behind this was to be creative and test different combinations. Users additionally stated that they like the idea that remix creation can introduce randomness to the remixes intentionally. *“If you could mark the stuff on the process and the remix could evolve every time. (P19).* *“Allow easy way for people to be creative. If they have a chance to influence the result they feel more related to the remix. Implement it like a lottery machine and varying video remixes come out. You may have few options: funny, intense more meditative etc. try different things with the system and see the results.” (P27)*

A Fundamental problem was stated to be the formation of a group in the events, where capturers do not know each other. Collaborative video creation solution has to offer features for initiating the video collective in the spot or include features for pro-actively initiating it. AVRS allows users to form “collectives” related to events based on the spontaneity. Spontaneity itself is a corner case in the video creation, since the events can also be planned beforehand, e.g. formal events.

5. DISCUSSION

This study addresses issues related to capturing and sharing video content within event-based groups. The results show that different kind of events and group formations require different functions from the remix. Sports events require following the conventions and including the audience in the remix. Concerts require multiple views of the performers and views showing the venue and atmosphere, whereas formal events require including the main persons and full comprehension of their speeches. Users were satisfied with the quality of the remixes in the music and dissertation events. At the ice hockey event, AVRS did not support the user needs as well.

In terms of the social user experience, the findings of the study relate closely to following categories: self-expression, connectedness, collaboration, and discovery, which are identified in previous work [18, 19, 24]. AVRS supports connectedness by offering a feeling of being related to others who took part and captured videos at the event and shared memorabilia with them. AVRS supports self-expression by allowing a user to give one’s own content to the remix and thus be a content creator. It supports collaboration by creating a group memento from the event in collaboration with the group. And finally, it supports discovery by enabling finding and seeing new videos and thus new viewpoint of the events.

Getting acknowledgement if the contributed content is visible in the remix brought out variance in the current study in comparison to the previous study of AVRS in the large-scale music festival scenario [22]. Participants were clearly interested in seeing whose content has reached the remix. The previous study included users who did not want any acknowledgement in the final remix. However, in this field study, users were interested in seeing if their video ended up in the final remix and to know who contributed the other video clips. One reason for this could be that some participants knew each other before the study, and consequently they were interested in the contributions of others but also to know whose contributions were included in the final video remix. Juhlin et al. [9] have introduced a research agenda for video interaction and in their work one of the goals is to understand the value and utility to the users. Results are promising in a sense that automatic remixing is obviously needed for the collaborative videos.

An obvious limitation in the study setting was the actual spontaneity of the groups. All the participants were invited to the study and explicitly instructed to capture video content. As the study setting defined the group to share the video content with the situation is fundamentally different to a real situation where the group should form spontaneously or even needs activity from a certain user in initiating the group. Pro-active features in the application can however ease the group formation by initiating the group based on the location and the event. The implementation of a client in any collective video creation solution needs to address this issue.

Findings of this study suggest that users are willing to hand their video material to create automated remixes, even with the strangers. A group formation on the events with the strangers however is an area that needs further studies, since in this study the groups were instructed to capture videos. Thus the actual spontaneity can be criticized. The study was completed with a population of 30 users, so the validity should be validated with a broader population to have more statistically robust results and to further investigate the differences between the groups. For the future research and development, tools for iterating the remixes would allow more flexibility in the end remix creation. It was the most desired new feature in the study.

6. CONCLUSION

We have presented a user study on a concept that enables user groups to create automatic mobile remixes in different event types. The most prominent findings of the study imply that people are motivated to use such a service as well as contribute to the service by sharing their personally captured video content. The automatic video remix creation was seen effective in giving a good presentation of what happened at the event, and resulting interesting remixes. Users were motivated to capture and share their content because they wanted to access others’ material and an interesting final remix in return. Taking part in the community of the event video capturers motivated the users since they felt connected and related to other users recording and sharing videos in the event. Automatic video remixing was stated to ease the pain of editing videos. AVRS was stated as giving a channel to share the videos. Sharing the collectively created remixes was stated as offering an easy and efficient way to have memorabilia of the events. Evidently, the group formation in the event is a challenge that AVRS aims to solve. AVRS aims to offer pro-active platform for enabling the spontaneous video capturing and utilizing the video clips in a collective video remix.

Of the three event types, the AVRS system was considered to support users, especially at music concerts, followed by formal party events and sports events. As an addition to the previous findings in similar solution in festival event setting [22, 23], the results suggest that AVRS as a solution can be expanded to other event types as well. Combining automatic approaches of selecting the most interesting and high-quality sections of the video with user annotations was seen as an ideal way to make best possible remix that would remove the shortfalls in the current system.

7. REFERENCES

1. Bao, X. and Choudhury, R.. 2010. Movi: mobile phone based video highlights via collaborative sensing. In *Proceedings of Mobile systems, applications, and services 2010*, ACM Press (2010). 357–370.
2. de Sa, M., Shamma, D. A., Churchill, E.F. 2014. Live mobile collaboration for video production: design, guidelines, and requirements. *Personal and Ubiquitous Computing*, 2014, 18. 693-707.
3. Engström, A., Esbjörnsson, M. and Juhlin, O. 2008. Mobile collaborative live video mixing. In *Proceedings of MobileHCI 2008*, ACM Press (2008), 157–166.
4. Engström, A., Perry, M., and Juhlin, O. 2012. Amateur vision and recreational orientation: creating live video together. In *Proceedings of CSCW 2012*, ACM Press, 2012. 651-660.

5. Girgensohn, A. and Lee, A. 2001. Home video editing made easy—balancing automation and user control. In *Proceedings of INTERACT 2001*, ACM Press (2001), 464–471.
6. Holtzblatt, K. Wendell, J. B., and Wood, S. 2004. *Rapid Contextual Design*. Morgan Kaufmann, 2004.
7. Jansen, J., Cesar, P., Guimaraes, R., and Bulterman, D.C.A. 2012. Just-in-time personalized video presentations. In *Proceedings of DocEng 2012*, ACM Press (2012). 59-68
8. Juhlin, O.; Engström, A., and Örnevall, E. 2014. Long tail revisited: from ordinary camera phone use to pro-am video production. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI 2014, ACM Press (2014), 1325-1334.
9. Juhlin, O., Zoric, G., Engström, A., and Reponen, E. 2014. Video interaction: a research agenda, *Personal and Ubiquitous Computing* (2014), 18: 685-692.
10. Kennedy, L. and Naaman, M. 2009. Less talk, more rock: Automated organization of community-contributed collections of concert videos. In *Proceedings of WWW 2009*, ACM Press (2009), 311–320.
11. Kirk, D., Sellen, A., Harper, R., and Wood, K. 2007. Understanding videowork. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI 2007, ACM Press (2007), 61–70.
12. Lampe, C., Walsh, R., Velasquez, A., and Ozkaya, E. 2010. Motivations to Participate in Online Communities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI 2010, ACM Press (2010). 1927-1936.
13. Kindberg, T., Spasojevic, M., Fleck, and R., Sellen, A. 2005. The Ubiquitous Camera: An In-Depth Study of Camera Phone Use. *IEEE Pervasive Computing*. (2005). 4(2), 42–50
14. Lehmuskallio, A., and Sarvas, R. 2008. Snapshot Video: Everyday Photographers Taking Short Video-Clips. In *Proceedings of NordiCHI 2008*, ACM Press (2008), 257-265.
15. Malinen, S. and Ojala, J. 2012. Maintaining the instant connection - Social media practices of smartphone users. In *Proceedings of Design of Cooperative Systems 2012*, Springer Press (2012). 197-211.
16. Monroy-Hernández, A., Hill, B.M. & Gonzalez-Rivero, J., Boyd, D. 2011. Computers can't give credit: how automatic attribution falls short in an online remixing community. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI 2011*. ACM Press, (2011). 3421-3430
17. Odom, W., Sellen, A., Harper, R., Thereska, E. 2012. Lost in translation: understanding the possession of digital things in the cloud. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI 2012, ACM Press (2012). 781-790.
18. Ojala, J. Väänänen-Vainio-Mattila, K., and Lehtiniemi, A. 2013. Six Enablers of Instant Photo Sharing Experiences in Small Groups Based on the Field Trial of Social Camera, In *Proceedings of Advances in Computer Entertainment 2013*, LNCS (2013). 344-355.
19. Ojala, J. 2013. Personal Content in Online Sports Communities: Motivations to Capture and Share Personal Exercise Data. *International Journal of Social and Humanistic Computing* 2(1-2), 68–85 (2013), doi:10.1504/IJSHC.2013.053267
20. Olsson, T. 2009. Understanding Collective Content: Purposes, Characteristics and Collaborative Practices. In *Proceedings of C&T 2009*, ACM Press (2009) 21-30.
21. Van House, N. 2009. Collocated photo sharing, story-telling and the performance of self. *International Journal of Human-Computer Studies*, 67, 12, (2009) 1073-1086.
22. Vihavainen, S., Mate, S., Liikkanen, L., and Curcio, I. 2012. Video as Memorabilia: User Needs for Collaborative Automatic Mobile Video Production. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI 2012, ACM Press. 651-654.
23. Vihavainen, S., Mate, S., Seppälä, L., Cricri, F., and Curcio, I. 2011. We Want More: Human-Computer Collaboration in Mobile Social Video Remixing of Music Concerts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI 2011, ACM Press (2011). 287-296.
24. Väänänen-Vainio-Mattila, K., Wäljas, M., Ojala, J., and Segerstahl, K. 2010. Identifying Drivers and Hindrances of Social User Experience in Web Services. In *Extended abstracts in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI 2010*, ACM Press (2010), 2499–2502.
25. Weilenmann, E. Säljö, R., and Engström, A. 2014. Mobile video literacy: negotiating the use of a new visual technology. *Personal and Ubiquitous Computing*, 2014, 18. 727-752.
26. Shrestha, P., de With, P. H. N., Weda, H., Barbieri, M. and Aarts, E. H. L. 2010. Automatic mashup generation from multiple-camera concert recordings. In *Proceedings of Multimedia 2010*, ACM Press (2010). 541-550.
27. Zimmerman, J., Forlizzi, J. and Evenson, S. 2007. Research through Design as a Method for Interaction Design Research in HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, ACM Press (2007), 493-502.
28. Zsombori, V., Frantzis, M., Guimaraes, R.L., Ursu, M.F., Cesar, P., Kegel, I., Craigie, R., and Bulterman, D.C.A. 2011. Automatic generation of video narratives from shared UGC. In *Proceedings of Hypertext and Hypermedia 2011*, ACM Press (2011). 325-33