

# We Want More: Human-Computer Collaboration in Mobile Social Video Remixing of Music Concerts

Sami Vihavainen<sup>1</sup>, Sujeet Mate<sup>2</sup>, Lassi Seppälä<sup>1</sup>, Francesco Cricri<sup>3</sup>, Igor D.D. Curcio<sup>2</sup>

<sup>1</sup>Helsinki Institute for Information  
Technology HIIT / Aalto University  
00076 Aalto, Finland  
{sami.vihavainen,  
lassi.seppala}@hiit.fi

<sup>2</sup>Nokia Research Center  
P.O. Box 1000, 33721 Tampere,  
Finland  
{sujeet.mate,  
igor.curcio}@nokia.com

<sup>3</sup>Tampere University of  
Technology  
P.O. Box 527, FI-33101  
Tampere, Finland  
francesco.cricri@tut.fi

## ABSTRACT

Recording and publishing mobile video clips from music concerts is popular. There is a high potential to increase the concert's perceived value when producing video remixes from individual video clips and using them socially. A digital production of a video remix is an interactive process between human and computer. However, it is not clear what the collaboration implications between human and computer are.

We present a case study where we compare the processes and products of manual and automatic mobile video remixing. We provide results from the first systematic real world study of the subject. We draw our observations from a user trial where fans recorded mobile video clips during a rock concert.

The results reveal issues on heterogeneous interests of the stakeholders, unexpected uses of the raw material, the burden of editing, diverse quality requirements, motivations for remixing, the effect of understanding the logic of automation, and the collaborative use of manual and automatic remixing.

**Author Keywords:** Automation, Video, Human Factors, Music, Social, Mobile.

**ACM Classification Keywords:** H5.m. Information interfaces and presentation (I.7): Miscellaneous.

**General Terms:** Human Factors.

## INTRODUCTION

Today, more and more consumers utilize their mobile phones for recording video clips and for editing activities. Video features and recording quality of mobile phones are reaching a level where the phones are becoming a serious

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.

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.

Copyright 2011 ACM 978-1-4503-0267-8/11/05...\$10.00.

tool for on-the-move content creators. Making available a large pool of snapshot digital videos taken by the audience in the same concert can result in higher value material than individual video clips. The individual digital video clips can be remixed into compilations that potentially enhance the perceived value of the event, are useful for various stakeholders such as the artists, and the fans of the artists. Remixing can also give the fans the possibility to become creators and not just receivers, and enhance the community feeling between the artists and the fans [11].

Digital video editing is an interactive process between human and computer; depending on how the editing system is designed, the human will do some tasks and the computer will do others. A key issue with designing collaborative multi-camera video remixing systems is the level at which the editing tasks should be automated and thus allocated between human and computer. This is especially important to ensure that the editing process will not feel too troublesome for an amateur and still produce an output that adds perceived value to the stakeholders.

As we will explain in more detail, automation has a high potential to improve the performance and ease the burden of producing a video compilation. However, in a live concert type of social context, where the compilation should be able to reconstruct the collaborative experience between the band and the audience, it is not self-evident how the automatic editing algorithm should be designed.

To approach this problem, we studied the motivations for collaborative video compilations and the role of automation in the process of editing the compilations. We conducted a user-centric real world study to bring out the points of view of various stakeholders and reveal automation related socio-technical factors that we thought could be important to consider in social video remixing systems.

We formulated the following research questions:

- What are the users' motivations for collaborative video compilations?
- How do users react to manual video remixing?
- How do users react to automatic video remixing?

By publishing a video compilation, the publisher puts the content to a public space (e.g., the Internet), where the stakeholders' different interests might conflict. This is why we see that to create sustainable social video services with an excellent user experience, it is not enough to study the phenomena only from one angle, but instead we must aim for a more heterogeneous view by studying the subject from the point of view of various stakeholders. We therefore studied the two main stakeholders of a live concert event: the artist and artist's fans. Under the word "artists" we include also the artist's record company, agent, and manager. Under "fans", we include people who have great enthusiasm for the artist. We understand that the interests of the actual artist and their record company might sometimes differ in the music industry ecosystem. However, for this study, we saw that the producer-consumer distinction is sufficient. We also researched how the boundaries between consumers and producers would blur when the consumers become producers with automation as a middleman.

As a case, we compare the manual and automatic video editing processes and outcomes from the two stakeholders' perspectives. The video compilations were produced out of video clips that the audience took with camera phones in a live concert. In the manual editing process, we gave the fans of the artist the ability to use the Kaltura video editor to make their own compilations from the raw material. In the automatic editing process, we used a software prototype developed by Nokia (we will refer this as Automatic Video Remixer (AVR)). In a nutshell, this software adds context data to video clips while filming, and is afterwards able to automatically produce compilations from multi-camera video material by using the captured context data.

We begin this paper by presenting the manual and automatic video editing processes as part of our research frame. Next, we present our results on how the stakeholders reacted to the manual and automatic video remixing and how they assessed three collaborative video compilations that were produced as a part of the study. At the end of the paper, we discuss the possibilities and challenges that the automation faces in mobile social video remixing from live music events. As the main data gathering methods, we used a focus group session and individual interviews.

## RELATED RESEARCH

### Studies on video production and mobile media

Let us start by going through the research related to the artist-audience interaction process. Engström et al. [2] analyzed how dance club's video jockeys work and suggested that mobile video could enhance the interaction between the club visitors and VJs. Engström et al. [1] continued their studies and presented the SwarmCam prototype for video capture and live transmission of mobile video. Club visitors film their dancing on the dance floor and stream the video live to the VJ, who possibly broadcasts it to a mega screen. From our paper's point of view the study is interesting, given that it concentrates on the

interaction process of VJ and club visitors in often so dark music clubs. In our study the context is similar, focusing on the improvement of the interaction between the artist and audience, but in stretching the timeline of the experience with post event video remixing.

Foote et al. [3] and Kennedy et al. [8] concentrated on automation in video remixing. Foote et al. [3] concentrate on automatic analysis of audio and video material based on significant audio changes, automatic suitability analysis of the video based on camera motion and exposure. Their system then stitches audio and video material together so that high quality video clips are adjusted to match the audio segments. They also developed a semi automatic system (Hitchcock), where the user manually selects the preferred video clips and the system automatically synchronizes the video clips with the audio track. This is a very interesting study not only because of the automatic content analysis but also because it takes into account the collaboration between computer and human for easy production of personalized video compilations. Kennedy et al. [8] studied automatic video synchronization and organization of video content taken by the audience in live music concert contexts. Their results reveal practical points regarding: use of audio fingerprinting to synchronize videos taken at the same event, the effectiveness of audio synchronization under different conditions, and finding meaningful content-based metadata to retrieve and summarize video material.

Girgensohn et al. [4] describe their video remixing system Hitchcock that was also discussed in Foote et al. [3]. The system takes a semi-automatic approach on creating custom videos from raw videos taken by a basic home video camera. In their paper they concentrate in home videos and not music videos. The system analyzes video content based on the camera motion and gives the video material unsuitability scores. The work is interesting considering the focus on various levels of automation. Girgensohn et al. [5] present their user study on Hitchcock. Their results demonstrate the need for a useful balance between user control and automation. Kirk et al. [9] take a holistic user-centric view on people's practices around home videos. Their results reveal useful information about people's motivations and practices for editing home videos. One of their results is that people do not find any reason to do editing of the short video clips they had taken. Shamma et al. [14] suggested that multimedia research should shift from semantics to pragmatics by designing systems and algorithms that can usefully utilize information about how media content is being used in specific contexts. Lehmuskallio et al. [10] studied people's snapshot videography practices. They state that the models for capturing video are often taken from snapshot photography.

In addition to video editing and practices, there is related research on mobile content creation, for example Jacucci et al. [6] and Sarvas et al. [13], and on automation in mobile social applications, Vihavainen et al. [16].

Video editing, multi-camera video production, music videos, automation, live contexts, or user point of view have been discussed in several high-class studies. However, none of the earlier studies have combined those aspects in a single study. Also our objective is to take into account the viewpoints of various stakeholders that interact in the production of collaborative music video compilations.

**Automation as Part of Human Factors Research**

Automation has been studied systematically in human factors engineering. Sheridan [15] proposed an automation framework with seven levels:

- |   |
|---|
| 1. The computer offers no assistance; the human must do it all                        |
| 2. The computer suggests alternative ways to do the task                              |
| 3. The computer selects one way to do the task and asks for human approval            |
| 4. The computer allows the human a restricted time to veto before automatic execution |
| 5. The computer executes the suggestion automatically, then informs the human         |
| 6. The computer executes the suggestion automatically, and informs the human if asked |
| 7. The computer selects the method, executes the task, and ignores the human.         |

**Table 1. Levels of automation according to Sheridan [15]**

In addition, Parasuraman et al. [12] proposed a four-stage model of functions inside the system that can be automated on different levels:

1. Information acquisition	3. Decision selection
2. Information analysis	4. Action implementation

**Table 2. Stages of automation according to Parasuraman [12]**

These levels and functions make an automation design matrix that we believe might be useful in examining the human-automation collaboration in mobile video remixing.

**THE STUDY: 11 FANS IN A ROCK CONCERT**

Users from two stakeholder groups (artists and fans) were recruited in Finland to participate in the study. A focus group session and a user trial (with interviews) were conducted to discover opinions, reactions, and experiences on social video production from live music concerts.

**Users**

The artists (Artists) who represented the artist stakeholders included the three members of a Finnish rock band and four employees from their record agency (The Artists will be marked as A1...A7). Three of the artists (A1, A2, A3) had a background in movie editing, and all were professionals in the music business. All were men, age 25 to 35. The fans (Fans) were a group of eight men and four women living in southern Finland (The Fans will be marked as F1...F12). All were between the ages of 18 and 30, with one 61-year-

old man (F3). Two of the Fans (F2, F4) had an amateur background in movie editing and two had a strong background in amateur photography (F3, F12). Others were not particularly tech savvy. Most were hard-core fans of the band. They were recruited through the band’s Facebook profiles and newsletters. All were interested in going to music concerts and took photos or video during concerts to get memorabilia and share experiences with their friends.

**Data Collection**

*Focus Group Session*

The objective was to discover the Artists’ and the Fans’ views, opinions, and habits concerning video recording at live concerts and remixing the recorded video material afterwards into compilations, either automatically by a computer or manually by the users. Four participants from the Artists (A1, A3, A6, A7) and seven from the Fans (F1, F2, F3, F5, F6, F7, F12) participated in the focus group.

*User Trial*

As a user trial, we arranged a live rock concert in collaboration with the band (A3, A6, A5) and their record agency (A4). The concert was a public event and thereby an authentic situation. 11 of the Fans (F1-F11) participated in the trial. They were given Nokia N97 phones equipped with the context-recording client that was used as part of the automatic remixing process (we will present the manual and automatic remixing processes in the next chapter). The Fans were directed to film as they would normally do in the concert situation. However, they were told that from the raw material, both automatic and manual video compilations would be made afterwards. For technical reasons (mainly battery life), they were directed to film a maximum of 15 minutes, and the maximum length of a clip was five minutes. After the concert, the phones were collected and users were provided with access to the raw video material and the Kaltura video editor (Kaltura will be presented later).

*Interviews*

After the trial, six of the 11 Fans (F1-F6) who participated in the trial and three of the Artists (A1-A3) were interviewed. The interviewees were selected based on voluntariness. Before the interviews, both the manually and automatically created remixes had been made available for viewing. The interview protocol was semi-structured. During the interview, we showed the participants one manually made reference compilation that was made by us, one of the manual compilations made by F1, and one automatic remix that was made using the AVR automatic remixing system. We watched the compilations with each of the participants and let them comment on the videos. After each video, we asked whether they thought the compilation was made by a human or a machine, and asked their feelings and opinions on the compilations and the remixing processes. Six out of eight participants had not seen the compilations before the interview and did not know whether the

compilations were made manually by a human or automatically by a machine. For this paper some of the quotes from the transcriptions of the interviews and the focus group were translated into English by the first author.

### The Video Remixing Procedures

Next we address the trial procedure — including the automatic and manual video remixing procedures — that we used as a research frame in our study. Figure 1 shows the remixing process we utilized in our user trial.

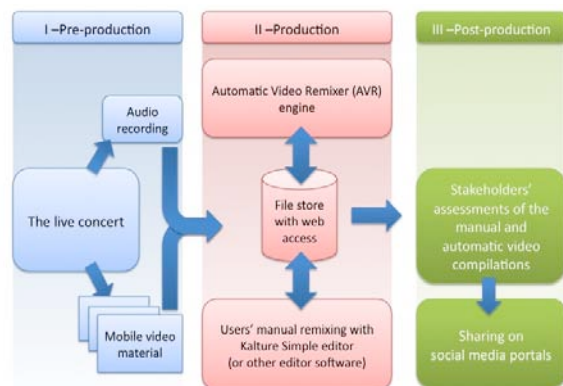


Figure 1. Remixing process

#### Automatic Remixing System

The automatic remixing system consists of the recording devices and a remixing server, and it functions in three main stages. The first is the *Pre-Production* stage, the second the *Production* stage, and the third the *Post-Production* stage.

The *Pre-Production* stage consists of recording the videos at the music event, and subsequently uploading these to the automatic remixing server. The recording devices (Nokia N97 mobile phones) were running a context recording client simultaneously with the video recorder. The context recording client captures and stores time-stamped sensor data while recording the videos. This helps in inferring the contextual information of the recorded video depending on the sensors being used. For the trial, data from the built-in electronic compass was recorded together with the video. The electronic compass gives as output the orientation of the camera-pointing direction with respect to North (for example: 90 degrees with respect to North). The information about the orientation of recording is then exploited to achieve two main goals: understanding what is the Region Of Interest (ROI) of the event for those users who are participating at the multi-user event recording, and inferring horizontal and rotational camera movements (also known as “camera panning” operations). Due to the limited network bandwidth availability with respect to the high volume of recorded data, we chose not to have direct uploading from the mobile phones. Instead, the videos and the context data were uploaded offline using a USB connection.

The *Production* stage is based on the idea of stitching together video segments cropped from the original source videos recorded by different users. The main issue consists

of understanding which video segments to consider and discovering the optimal length for each of the selected segments. The *Production* stage consists of several steps. First, all the uploaded videos are synchronized with each other such that the same events happen at corresponding timestamps among different user recorded videos (i.e., source videos) which overlap temporally. The second step consists of context data analysis: the orientation data from each user is used to infer if and when that specific user has moved the camera phone in a horizontal direction by means of a panning operation. The detected panning operations are then marked as “event points” and used in the AVR video generation as switching points between different views. Also, orientation data from all the users is used to understand if there is a specific ROI in that event. The information about the ROI is then used to assign higher priority to videos that have been recorded by pointing the camera phone within the range of orientations of interest. The video segments with higher priority will then have a higher chance of being inserted into the final AVR video. In order not to switch from one view to another too fast or too slow (causing an unpleasant viewing experience), minimum and maximum segment durations (lower and upper bound temporal thresholds) have been set in the process. Thus, if there are consecutive event points detected within a too short time interval (i.e., less than the lower-bound temporal threshold), only the first event point is used for switching the view. Instead, if there are no event points for a too long interval (depending on the higher-bound temporal threshold), then the system forces a view switch. After having stitched together all the selected video segments, the visual part of the AVR video is ready. Regarding the audio side, a combined audio track is created by using the extracted audio from the recorded videos. The approach for this combined audio track creation is based on two requirements: using the audio tracks with the best quality, and obtaining the minimum number of audio switches. The first requirement might introduce a large number of audio segments if the source content consists of frequently changing audio quality or many source videos with good audio quality of short duration. Thus, there has to be a trade-off between the requirements. At this point, the automatically generated video and audio tracks are merged to obtain the final AVR video.

The *Post-Production* stage includes the sharing and viewing of the AVR video. Due to the completely automatic nature of the AVR production, there is a possibility for inclusion of source videos that contain objectionable material (e.g. obscene gestures). When such segments are included, the related source video needs to be removed manually.

#### Manual Remixing Procedure

The manual remixing procedure roughly simulates the way a single user would preview, download, upload, and remix mobile video material that is accessible on the web. The procedure follows the three phases as described before for the automatic remixes and as shown in Figure 1, but differs

in the Production phase. In this phase the Fans were provided with a link to a website with previews of the videos as well as access to download the raw material, including a fifteen minute audio track recorded with the Nokia N97 or three specific audio tracks recorded with a video capable micro-DSLR camera. Users could watch the lower quality preview videos and then download all or part of the material in the original quality. The Kaltura Open Source video platform [7] was used to provide users with an online video editing service for creating mixes from the downloaded material. Kaltura was given only as an option and not as a mandatory requirement. All the participants were also instructed that they could use some desktop-based video editor if they desired. We wanted to offer them one tool because many of them did not have earlier experience on video remixing. With Kaltura it was easy to give technical support, if needed. Kaltura Simple Editor provides the basic video editing features needed for mixing together multiple video clips and a soundtrack; it includes timeline-based editing, trimming of and transitions between video clips, and options to use either video clip audio or an external sound file. The editor was used online on a web browser and required users to upload all the material they wanted to use in their mixes. After uploading the material, it was converted on the server to a format usable by the editor. The necessary conversion process added more waiting time on top of the time used for uploading the material. This, however, simulates the fully manual remixing process as was intended. After conversion the videos were available in the editor's Clip Library tab where the clips could be dragged onto the timeline. On the timeline the clips could be trimmed to a desired length and transitions could be added in between the clips by dragging them onto the timeline from the Transitions tab. In the Soundtrack tab the user could upload sound files into the project. One single sound file could be used to play over the video, either in looped or play-once mode. Users could then choose to mute the sounds from the videos to have them to play together with the selected soundtrack or without a separate soundtrack.

## Overview of the Data

### *Raw Material*

The trial resulted in 105 clips, totaling approximately one hour and 52 minutes of video and 2.3GB of data. Each user contributed an average of nine clips. There was a wide variation among the users on the number of clips with the minimum being four clips per user and the maximum being 20 clips per user. There was no direct correlation observed between the number of clips per user and the average clip duration. Some users were more prolific than others; the minimum total time of the clips recorded per user was 4:39 while the maximum total time of the clips recorded per user was 12:24. The maximum clip duration for each user varied between one minute and five minutes. Of these, the users chose not to include seven videos for remixing; thus, the total time of the clips was approximately one hour and 44 minutes. Most users wanted to include all of the material

they shot. Three of the four users who decided not to include some videos excluded more than one clip.

### *Video Compilations*

Out of the Fans who participated in the concert trial, only one, F1, edited her own video compilations. She made two of them. As a reference, we made a reference remix using the same process and editor as described earlier. It was done to test the process and the capabilities of the editor fully, to give an example of certain kind of editing style, and to test how it will be assessed against the other compilations. With the AVR system, we made two compilations to test what kind of compilations the system makes from the material that is taken in an authentic music concert environment.

## FINDINGS 1: GENERAL REQUIREMENTS FOR REMIXES

In the focus group and the after-trial interviews, our objective was to discover the stakeholders' requirements for the collaborative video compilation process and their views on the added value of manual and automatic social video compilations from a live music event.

### **Heterogeneous Motivations for Video Compilations**

The Fans' main motivation was to record videos as personal memorabilia of the concert. However, some people said that they share video clips with their friends, for example, on Facebook. Others noted that they would share the videos more with their friends if they could do it easily from their phones or cameras during the event. The Artists thought that video compilations are a way to promote the band image, and could be used by venue owners to publicize the band. The main motivation was to demonstrate the interaction between the band and the audience to other people who did not see the event live.

“As an agent, it is very important to me to show the customers [venue owners and consumers] how the band's live performance looks like.” – A1

Both the Fans and the Artists saw video compilations as a great way to enhance the concert experience, stretch the timeline of the concert, and promote interaction between the fans and the band.

“To get the fans to communicate with each other. That is why this is interesting.” –A1

“After a concert I often have a feeling like, too bad it's already over.” – F1

## FINDINGS 2: REACTIONS TO MANUAL REMIXING

In this section, we describe the reactions of the stakeholders to the manual remixing, including how they felt about making compilations and what kind of motivations they had to make their own.

### **Motivations for Manual Mixing**

For the Fans, the main motivation to make their own manual compilations was personal challenge, and/or publication (such as to the band's Facebook profile).

### *"I Made This!" Self-Expression and Meeting a Challenge*

The F1 who made two manual compilations was very motivated. She was proud of her compilations and excited at the prospect of them being published somewhere.

"I had this kind of proud feeling that this is my thing . . . then I can say that, hey I made this!" - F1

She felt successful after completing difficult editing of her clips to suit herself.

"Then it was like, great, I got this one thing done!" - F1

She was also motivated to make more compilations. However, she was not sure if any band would like to use these kinds of mixes because they are not very professional. One of the Fans did not consider publication important, and he felt that the only motivation was the personal challenge and that he could make something nice for himself.

"It would be like, hey I made this!" - F2

However, he would have agreed with the publishing of his compilation on the band's Facebook site.

F4 was motivated to make a remix. He even had planned what kind of a story he would tell with his remix. However, he had some problems with his Internet connection and in understanding the logic of the manual remixing process. He liked the idea of getting his remixes published.

### *"Not Interested!" Concentrating on the Concert and Not Video Editing*

For the F3 and F2 it was all about the concert experience itself. They did not have any interest in making compilations of their own. Their primary intention was just to shoot the video material in the concert and give it to other people to make the remix. They did not feel that remixing was for them.

"I shoot the material and give it to others." - F3

### **Burden of Manual Remixing**

Next, we will discuss the burdens involved the manual video compilation.

### *"It's a Tough Job But I Love it": Overcoming the Hardship and Keeping the Standards High*

The F1 who made her own compilations said that for the first one (duration of 4:24), she worked about 35 to 40 hours. This was a surprisingly long time considering that the reference compilation our researcher made was completed in about three hours. F1 said that first she used one evening to familiarize herself with the Kaltura video editor. This was evident because she did not have any video editing experience. Subsequently, after 12 to 14 hours of editing, she thought the compilation was ready. However, she was still not satisfied.

"I wanted the remix to be perfect." - F1

She also went through almost all of the raw material of one song. She made a choice to concentrate only on the video

material from one song; otherwise the work would have been too much. This demonstrates how laborious and time-consuming it can be for an amateur to manually make a satisfactory video compilation.

### *"I Was Going to Do it But Something Came Up": No Time and Technical Difficulties*

The other three of the Fans (F2, F4, F5) who had some intentions to make their own manual compilations said that they either did not have time for it, or they felt it was too much of a burden to start figuring out how the editor worked and going through the raw material. There was also a technical difficulty in getting their computers to view the material and upload the best videos to the Kaltura editor.

"It was like download clips to the computer, upload clips to the server. Edit there, put here. I was like, later!" - F2

F4's home Internet connection was so slow that he was unable to download the raw clips to his computer. He was also uncertain if he understood how to proceed with a manual compilation:

"It was a little uncertain for me, how to make the remix." - F4

### *"No Way I'm Going to Do This": Way Too Much of a Burden*

Two of the Fans (F3, F6) did not even check the manual editor and had no intention to do their own compilations. F6 said she would just forget herself while trying. She also had some problems in understanding the logic of the editing process. The M3 did not even check how Kaltura works and had no intention of making a compilation of his own.

"I just did not feel like studying it." - F3

### **FINDINGS 3: USE OF THE MANUAL EDITOR**

In this section, we will discuss how the manual editor was used. Our data comes from the F1 who made her own video compilations, and the findings are based on her experiences. We still think there are valuable results relating to how an amateur can make multi-camera live music video compilations manually, and what significant factors there are in that kind of video editing process.

### **Unexpected Uses of Raw Material**

Manual editing brought up some unexpected uses of the raw material, like use of lights, jumping audio track, and ruined video material. The F1 said that she used lights as a guide to edit the remix. Beforehand, we thought that the music would be the main factor when the user chooses when to make cuts.

"In addition to music I used lights as a guide to edit the remix. ...Because they changed in convenient points." - F1

She did not use the complete audio track that we provided. However, her idea was not to make a music video style remix with a complete audio track but something different. She thought that it was "cool" to listen to the concert from several places. She also wanted to use all the possible camera angles. Her intention was to show the viewer how

the concert looked and sounded from various places in the concert hall, and to show that the experience is different depending on the location. Because of this she tried to use every participant's material for her remix:

"I wanted to add in the element that shows how the concert sounds when you are at the balcony, near the drums or at the back of the hall." – F1

She also wanted to use a clip where the cameraman had flipped the phone by 90 degrees. She thought that would be a great special effect. The fan who had recorded that exact clip thought that the clip itself was ruined because he had accidentally held the camera with a wrong angle. We think this is a great example of how an individual video clip has much more potential when it can be remixed innovatively with other video clips.

#### **Coping With the Amount of Raw Data in Manual Mixing**

We were surprised of how fast can the raw material grow in a multi-camera recording. F1 had to mark off some material in the beginning so that going through the clips would not be too overwhelming. She decided to concentrate only on the clips from one song. Although she tried to use every participant's material for her compilation, she used her own material the most. This was mainly because she was familiar with it, and it was laborious to find good scenes from the other cameramen's raw material manually. She also used pen and paper to keep notes about which of the clips were good. It became clear that the amount of a multi-camera video material quickly becomes so large that it might be overwhelming to go through all the video material manually. As one of the Artists said:

"If one has material from 100 cameras there is no way he can go through all of them...Even if one had four cameras and a web based video editor, there are still endless ways one can go wrong and concentrate on insignificant things during the first five seconds of the editing process." – A2

Thus, for example efficient use of metadata or more strictly organized directing in the filming phase might make it easier for an amateur to handle the material and make manual video compilations.

#### **FINDINGS 4: REACTIONS TO THE MANUAL REMIXES**

In this section, we will go through the participants' reactions to the manual compilations.

##### **The Reference Compilation**

Five of the six participants guessed that humans made the reference compilation, and one was unsure. Based on the interviews, the five knew the video was human-made because it seemed somehow logical; the change of colors worked effectively; cuts were synchronized with the music; and it had "a human touch."

"It was made from the short cuts and what was special about it was the rhythm. The creator gave the viewer an understanding about what is going on [in each scene] and then it was cut to another scene and cut and another and cut..." – F3

All of the Fans thought that the reference compilation was good, and many of them were amazed how it was possible to make such a good compilation from the raw material they had been filming. The Artists had positive feeling towards the video and thought that it might be good publicity for the band. They thought that it might be the deciding factor a wavering concertgoer.

"I think and hope that based on this compilation a positive twist could happen" – A3

##### **The Manual Compilation Made by One of the Fans**

In the case of the manual compilation made by F1, the participants were a bit unsure whether it was made manually or automatically. Still, five out of six participants thought a human made it. The main reason for their uncertainty was that this compilation did not have a continuous audio track, but instead the audio was from the individual scenes' audio. This made the audio fragmented and negatively affected the viewing experience. However, the clear narration convinced them it was human-made. In the first scene, the band's lead singer welcomes people to the concert, and in the last scene, he walks away, jumping and waving. This gave the users a human impression.

#### **FINDINGS 5: REACTIONS TO AUTOMATIC REMIXING**

In this section, we will discuss how the users understood the logic of automation, reacted to automatic compilations, and what kind of use they had in mind for automatic remixing. Before the trial, the participants' first reactions to automatic remixing were mostly curious about how it would work and how the machine makes decisions. Many of the participants also thought it sounded interesting that a machine could do the editing process, but also brought up that they felt a bit skeptical about how it could be possible.

##### **The Effect of Understanding the Logic of Automation**

In the trial, the participants were not told how the automatic editor makes decisions and what information it uses in the editing process. This enabled us to study the effect of understanding the logic of the automatic editor, especially how the ignorance of the logic of automation affects how the Fans film in the concert situation. Some of the Fans thought that the automatic editor would not be able to cut the raw clips in any way. Because of this, they tried to shoot short, ready-to-use clips.

"Because it does not cut those (raw clips), because there is going to be the whole clip. That's how I assumed it works." – F4

The F3 thought it was bothersome to shoot the video clips because he did not know how the automatic systems made editing decisions.

"One should know what their algorithm is like." – F3

The Fans were thinking about the editing phase already as they shot. They tried to cooperate with the automatic editor by pre-planning what kind of material would be useful.

## Reactions to the Automatic Remix

Five participants' thought the compilation was automatically edited, and one was unsure. Thus, it was clear that the automatic compilation had special features that made it look like machine-made. Next, we will discuss what those features were and how the participants reacted to the automatic compilation before and after they came to know that it was automatically edited.

### *"This Does Not Have a Rhythm": Reactions Before Knowing it Was Made by a Machine*

Because of the low light conditions in the concert hall, there were many dark scenes in the raw material. Thus, all the compilations, manual and automatic, had dark scenes. However, in the automatic compilation, participants thought the darkness was somehow "weird" and that it flattened the feeling of the video.

"This darkness does not motivate here. It does not have any function here." – F3

None of the participants liked the automatic compilation as much as they did the first two manual ones. The automatic compilation was described as passive.

"Passivity can be a special effect to bring up dynamic stage performance. But in this there is only darkness." – F3

### *"Not So Bad After All, But No Use as a Promotional Material" – Reactions After Knowing it Was Made by a Machine*

When the Fans were told that the compilation was automatically edited, most of them then developed a different perspective. Even though it lacked the human touch, it was still viewed positively. It seemed that people did not have such high expectations for the machine-made compilation.

"It's actually good considering it's a machine made" – F4

Most of the participants saw the automatic compilation as promising and said it would be worth developing further. However, the Artists especially still saw limitations. They brought up how it is hard to get a good automatic compilation because editing is based on continuity of movement, colors, and music that enables a narration.

"It's impossible to automate drama." – A2

The Artists also thought that the image the automatic compilation projected about the atmosphere of the live show was so irrational that if it was the only reference to the band's live performance, the fans would not come to watch.

"The machine is irrational. When there is action on the stage, it shows only an illuminated exit sign." – A1

### *The Perceived Usefulness of the Automatic Compilation*

Although the automatic compilation did not get very high ratings from the participants, many of them still found situations where it could be useful, especially if it could be made more interesting. Some of the Fans thought that the automatic one could be used in personal blogs or for marketing, and could be an easy way to participate without having to make their own manual compilations.

"You can take video and then you become part of this compilation. It would be participatory activity" – F1

"Like I could watch it from YouTube." – F4

The F4 also thought the automatic remix would be great to have on his mobile phone right after the concert. This would extend the timeline of the concert experience and would be great memorabilia produced with no effort, even if it were not top quality.

### *How to Make it More Interesting*

Next, we asked the participants how to make the automatic compilation more interesting to watch and more useful. The main things were that the video should be synchronized with the music; clips should be shorter; there should be more variability in field size; and the dark scenes should be eliminated. Three participants said that the video should be more synchronized with the music.

"Now the automatic one jumped randomly somehow in a way that a human made one would not." – A3

Two participants said that the compilation should not have such long, continuous scenes from the same camera but instead it would be more interesting and lively if there were quicker changes between cameras.

"One has to remember that five seconds in a video compilation like that is a long time." – A3

Three participants mentioned that in a multi-camera compilation, there should be more systematic variation in the camera angles and field sizes. They also suggested that there should be scenes from locations other than the main concert hall, such as a bar or the men's restroom.

"Like Big Brother in a good way." – A3

One of the Artists thought that one reason the manual compilation was better is that a human can use light as a guideline to capture the feeling of the concert. Thus, if the dark scenes could be eliminated, it might improve the viewing experience of the automatic compilations.

"In a concert the light can change dramatically only for like half second. If you don't capture those tiny moments one might think that the concert was totally black." – A3

### *Need for Human Intervention*

The users often mentioned the need for human intervention as a part of the automatic editing process. A3 said that when he watched the automatic compilation, he had a feeling that he would like to make his own changes to it.

"I was like I wish I could now go there and edit that. Then one could get more out of it." – A3

He also thought that because the editing process requires a complete view of the raw material, it is difficult to leave it solely to a computer.

"There has to be a human understanding about what kind of scenes need to be one after the other because contiguous scenes affect to each other." – A3

However, one of the Artists also brought up the importance of the actual goal of the compilations:

"If it's just harsh recording, it could work without a single cut. However, if we want to do even a little bit artistic material then there needs to be a human in the process." – A2

## DISCUSSION

The study showed that in production of social video compilations from live music events it is not self-evident how to allocate the remixing tasks between human and machine.

In a fully manual setup, the user has total control of the editing process. She has the authority from the beginning to the end to sense, analyze, decide, and implement which parts of the material she uses. At best, this flexibility can give a satisfactory feeling to the users. Also, the actual video compilation can raise excitement among the other stakeholders. However, the editing process itself may prove to be troublesome. In multi-camera video editing it can be laborious to manage large amounts of material in a satisfactory way. The use of technical infrastructure for video editing can be felt as too much of a burden to even start the editing. Improvements in the manual editing interface and usability design can only reduce part of the work. The labor of going through the video clips to find the good ones and synchronizing them would still be needed.

By contrast, when the remixing process is fully automated to the point that the computer ignores the human and acts autonomously, humans do not have to do anything but film the raw material during the concert. Thus, the laborious struggle with the technology and the burden of perusing the overwhelming quantity of material disappears. However, it turned out that a machine simply cannot imitate the human touch. It was surprising that ignorance regarding the logic of the automatic editor had an effect on how the people filmed the video during the concert. People made assumptions about how the machine would edit the raw material and directed their filming based on those assumptions. We also found out that expectations for an automatically made video compilation versus the manual one were lower. Naturally, how satisfactory a video compilation is perceived to be depends on its intended use. We found that for a documentary style compilation, the artistic demands are muted. On the other hand, a compilation that has to mediate the social interaction and atmosphere of a live music concert and is also intended to be used for marketing purposes has very stringent demands in terms of artistic quality.

Our research frame considered the two most extreme options of automation: namely, fully manual and fully automatic. We note that there is a large design space between those two extremes; a space where human and machine can collaborate in various stages during the remixing process, depending on how personal and artistic one wants the compilation at one extreme or how simple and fast at the other. For this design space, we propose that designers use the levels of automation and the stages of

information processing Sheridan [15] and Parasuraman [12] have presented, and which we discussed earlier in the Related Research section.

As an example, let us next walk through a case where we set the tasks of a video remixing process into Parasuraman's four-stage model of information processing and discuss how Sheridan's automation levels could change inside the video remixing process when moving from one stage to another. Thus, we stress the use of human labor in some tasks and machine labor in others. In *the first stage*, information acquisition can cover the sensing of the raw data. If the goal is to find all scenes that have bright lighting conditions, in a fully manual model, the human must go through the material and find the proper scenes. As we have discussed, this is laborious with multi-camera material. However, if we raise the level of automation and program the machine to sense and "lock on" to all the bright scenes, this would ease the human labor. In *the second stage*, information analysis can cover the cognitive functions such as remembering which of the previously found bright scenes also include a guitar solo. In a fully manual setup, the human must manually integrate the two inputs (the brightness and guitar solo) and analyze when they both exist together. If we raise the automation level and let the computer do the analysis, we can ease human labor by letting the computer provide context-dependent summaries of the raw video data. During *the third stage*, the decision and action selection stage, some of the previously found scenes (with good lighting conditions and a guitar solo) are selected to be stitched in a particular order into a compilation. This stage could be left for human to do independently, and this is a way to raise the artistic value of the video. In *the fourth stage*, the actual stitching phase is implemented during which the selected scenes are stitched together, forming the final compilation. Human could implement this step by manually combining the previously selected scenes. If the fourth stage was automated to a higher level, a computer could execute the stitching autonomously and complete the editing process.

By this thought experiment we want to demonstrate the possibilities of dividing the video editing process into separate information-processing stages, and by that help designers and developers to systematically think of an automated process in multi-camera video remixing. When moving from one stage to another, it is possible to change the level of automation depending on the intended uses for the compilation. If the need is to get a quick documentary style compilation, greater emphasis should be put on automation. However, if the need is for an artistic compilation, higher human involvement may be required.

## CONCLUSIONS AND FUTURE RESEARCH

We compared the processes of manual and automatic video remixing from both the artists' and the fans' point of view. We studied the process from the moment when the fans shoot mobile video in a live concert to the moment when the fans and the artists assess the human or machine made video

compilations. Based on the results, automation related design decisions have several implications on the motivations and reactions the users have towards a collaborative mobile video remixing system and the actual video compilations.

*First*, it came up that the various stakeholders of a music event could have different motivations and requirements for video compilations depending on the intended use. An artist might want to emphasize the marketing value of the compilation whereas a fan might want to have it as personal memorabilia for her to share it with friends. If the compilation is wanted to have a special artistic feeling, it is hard to gain that with a fully automatic remixing. *Second*, the manual editing process can give satisfactory results but the raw material from multi-camera video shooting can quickly become cumbersome. This came up when only one of the participants made the effort to actually make video compilations. She did a lot of work for them but in the end was very satisfied about the results. *Third*, people are able to use the raw material innovatively in ways that are difficult for computers to imitate without a high-end adaptive automation. For example, video material that was accidentally shot with a 90 degrees angle was surprisingly considered as a great special effect material. *Fourth*, automatic video compilations are not assessed with the same criteria applied to human made ones. However, for this to be true, it should be evident to the viewer that a machine has generated the compilation automatically. In our study many participants thought that the automatic video compilations we showed them were quite all right only after we told them that they were remixed automatically. *Fifth*, the results revealed strong interconnections between the filming process and the editing process. Transparency and communication between both of the processes affects the automatic and manual video compilations. For example, knowledge of the logic of the automatic editing process affected the participants' filming process. People were keen to learn the logic of the automatic editing process. Before the filming stage some of the participants had made assumptions on how the automatic remixer would work and directed their filming based on those assumptions. In the manual editing stage, the effect of transparency and communication became visible when a participant ended up using her own video material more than other participants' material. The main reason was that she was more familiar with her own material than with others'. This showed that context metadata the camera phone possibly adds to the raw material, such as location, could be valuable not only in automatic remixing but also for manual remixing. For the future research we believe it is valuable to study the possibilities of automatic and user generated metadata in multi camera video remixing process and how the direction during the shooting phase affects the editing process.

#### ACKNOWLEDGMENTS

The work was part of Social Video project funded by The Finnish Funding Agency for Technology and Innovation.

We thank the participants, Anu Kankainen, and anonymous reviewers for their help and valuable comments.

#### REFERENCES

1. Engström, A., Esbjörnsson, M. and Juhlin, O. Mobile collaborative live video mixing. *Proc. MobileHCI 2008*, (2008), 157–166.
2. Engström, A., Esbjörnsson, M. and Juhlin, O. Nighttime visual media production in club environments. *Night and darkness: interaction after dark. Workshop at CHI 2008*.
3. Foote, J., Cooper, M. and Girgensohn, A. Creating music videos using automatic media analysis. *Proc. MM 2002*, ACM Press (2002), 560.
4. Girgensohn, A. et al. A semi-automatic approach to home video editing. *Proc. UIST 2000*, ACM Press (2000), 81–89.
5. Girgensohn, A. et al. Home video editing made easy—balancing automation and user control. *Proc. INTERACT 2001*, ACM Press (2001), 464–471.
6. Jacucci, G., Oulasvirta, A., Salovaara, A. and Sarvas, R. (2005). Supporting the Shared Experience of Spectators through Mobile Group Media. *Proc. GROUP 2005*, ACM Press (2005), 207-216.
7. Kaltura open source video <http://corp.kaltura.com/>.
8. Kennedy, L. and Naaman, M. Less talk, more rock: Automated organization of community-contributed collections of concert videos. *Proc. WWW 2009*, ACM Press (2009), 311–320.
9. Kirk, D. et al. Understanding videowork. *Proc. CHI 2007*, ACM (2007), 61–70.
10. Lehmuskallio, A., Sarvas, R., Snapshot Video: Everyday Photographers Taking Short Video-Clips. *Proc. NordiCHI 2008*, ACM Press (2008).
11. Lessig, L. *Remix: Making art and commerce thrive in the hybrid economy*. Penguin Pr, 2008.
12. Parasuraman, R., Sheridan, T.B. and Wickens, C.D. A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man and Cybernetics, Part A 30*, 3 (2000), 286–297.
13. Sarvas, R., Viikari, M., Pesonen, J. And Nevanlinna, H. MobShare: Controlled and Immediate Sharing of Mobile Images. *Proc. MUM 2004*. ACM (2004), 724-731.
14. Shamma, D. A., Shawn, R., Shafton, P. L., Liu, Y. “Watch what I watch: using community activity to understand content. *Proc. MIR 2007*. ACM Press (2007).
15. Sheridan, T. *Telerobotics, Automation, and Human Supervisory Control*. MIT Press, Cambridge, MA. 1992.
16. Vihavainen, S., Oulasvirta, A., Sarvas, R.: “I Can’t Lie Anymore” – The Implications of Location Automation for Mobile Social Applications. *Proc. MobiQuitous 2009*, IEEE Press (2009).