The Mobile Vision Mixer: A mobile network based live video broadcasting system in your mobile phone

Arvid Engström[1], Goranka Zoric[2], Oskar Juhlin[3]
Mobile Life @[1,3]Stockholm University, [2]Interactive Institute
Box 1197, SE-16426 Kista, Sweden
{arvid, goga, oskar}@mobilelifecentre.org

Ramin Toussi
Ericsson
Torshamnsg. 23, SE-16483 Kista
ramin.toussi@ericsson.com

ABSTRACT
Mobile broadcasting services, allowing people to stream live video from their cameraphones to viewers online, are becoming widely used as tools for user-generated content. The next generation of these services enables collaboration in teams of camera operators and a director producing an edited broadcast. This paper contributes to this research area by exploring the possibility for the director to join the camera team on location, performing mixing and broadcasting on a mobile device. The Mobile Vision Mixer prototype embodies a technical solution for connecting four camera streams and displaying them in a mixer interface for the director to select from, under the bandwidth constraints of mobile networks. Based on field trials with amateur users, we discuss technical challenges as well as advantages of enabling the director to be present on location, in visual proximity of the camera team.

Categories and Subject Descriptors
H.5.1 [Information interfaces and presentation]: Multimedia Information Systems (Video).; H.4 [Information Systems Applications]:Miscellaneous

General Terms
Performance, Design, Human Factors

Keywords
live, broadcast, mobile network, bandwidth, vision mixer, mobility, collaboration, real-time mixing, video, production, webcasting.

1. INTRODUCTION
Online video is growing rapidly, and video streaming tools for mobile phones are some of the most interesting new services in that space. A first generation of mobile live streaming services, such as Bambuser and Ustream, is widely used in a range of contexts from citizen journalism to social media. Recent work has explored the possibility of adding *collaboration* to these service concepts, allowing a group of users to act as a camera team and produce an edited live broadcast for remote viewers [1,2]. This addition is of great value as it allows broadcasters to cover live events using multiple camera angles and editing, which are key tools for storytelling in traditional TV productions. However, systems to date are either partly stationary or restricted in network connectivity, and thus do not take full advantage of the mobility offered by using cameraphones.

In this paper, we explore the possibility of bringing the entire group of people, including the director mixing and controlling the broadcast, out into the location of the filmed event. We present a technical solution for enabling an all-mobile live video production, and discuss its potential advantages as compared to available systems. We present an example of the next generation of live video streaming services, the Mobile Vision Mixer (MVM) system. The system lets a group of people produce a live broadcast together, through connecting their camera phones and mixing between their video feeds. The design essentially takes the simplest functionality of a studio mixer, mixing, and transfers it to a mobile phone, omitting the use of cables and heavy broadcast equipment, and enabling users to broadcast from anywhere within the mobile network.

Mobile networks are chosen for their wide coverage, as mobility is set as a key requirement for the MVM system. By comparison, wireless networks such as MANETs (Mobile Ad-hoc Networks), WLANs (Wireless Local Area Networks), and satellite networks all satisfy specific service needs in terms of coverage, latency, bandwidth or cost, but are severely restricted in reach, or unavailable for consumer services. 3G cellular networks offer high mobility but low bandwidth. In order to reduce the needed bandwidth, the MVM combines multiple streams into one, while retaining the same functionality on the mobile mixing device.

This research is of relevance for the growing field of video interaction research in HCI (Human-Computer Interaction). Mobile broadcasting services enable a unique combination of mobility and live streaming in a consumer device. But these new technologies and practices also present new interactional, experiential and production-related challenges that have acquired some attention in research. Juhlin et al. [3] provided a qualitative content analysis of mobile broadcasting. Reponen [4] conducted a field experiment to investigate how it supports group communication. Potential uses ranging from rescue operations [5] and journalism [6] to nightclubs [7] and concerts [8,9] have been explored. Challenges for the wider appropriation of new mobile services involving live video are still largely unexplored, and include topics such as technical quality, consumption, interactivity and video literacy [10]. This paper builds on the existing body of work on collaborative live video, and discusses advantages of bringing increased mobility and co-presence to teams of people creating live video broadcasts together.

2. RELATED WORK
ComVu Pocket Caster, launched in 2005 [4], later renamed Livecast [12], is the pioneer in mobile live broadcasting. It was followed by several other services like Qik [13], Bambuser [11], and Ustream [14]. These services all allow

MUM '12, December 04 - 06 2012, Ulm, Germany

instant sharing of individual users' videos, filmed with mobile phones, with an audience on the Internet. Yet none provides any features for collaborative production or support for multi-camera filming. Current research literature addresses collaboration in mobile video from various perspectives. The Mobicast [15] system enables collaboration between multiple users streaming the same event from their mobile phones. Incoming video streams are stitched together and a panoramic view is constructed in real-time thus enabling the enhanced viewing experience of mobile-casted events. Shamma et al. [8] propose a system that relies on ad-hoc sharing, offering users the ability to select between available mobile devices for viewing after the event. Vihavainen et. al [] compare manual and automated remix compilations of co-created mobile video content. Engström et al. have presented the Instant Broadcasting System (IBS) [7], enabling people to collaboratively produce live video by using their mobile phones and a computer, connected through mobile networks. Although this system provides users with basic tools for mixing and broadcasting in real time, it is partly desktop based. The CollabraCam [16] application enables shooting and simultaneous editing of live video from multiple cameras on-the-fly using iOS devices over local WiFi. Though similar in mixing functionality, the WiFi as compared to 3G severely restricts the mobility of both the camerapersons and the director. With live streaming relying on the 3G networks, the Mobile Vision Mixer system provides a combination of mobility and collaboration, beyond what existing systems offer.

3. THE MOBILE VISION MIXER SYSTEM

The Mobile Vision Mixer provides a minimalistic collaborative production environment, where the mixer resources are provided on a mobile handset. It enables a group of users to co-produce and broadcast live footage.

Figure 1: Video mixing with MVM on Nokia N86

The MVM is intended to be used by amateurs, giving them more elaborated tools for live storytelling than the available single user webcasting services. It intended for a wide variety of events and gatherings, where people on location can broadcast the event to a non-present audience. Taking an example from our field study setting, a group of five friends, who are spending afternoon at the skateboarding park, decide on the spot that it would be fun to broadcast others skateboarding, and share it with the friends who couldn't join. They bring out their mobile phones and decide on the fly who is going to be camerapersons, and where they should be situated to provide complementary views of the field. They plan to have the set-up in which two of them cover detailed shots of skateboarders, the third provides a wider overview, and the fourth camera covers the environment. The five friends also decide who should be the director and mix in between the views of the action. During shooting, the director is able to see live previews of the above four cameras and can cut between them at any moment, (see Figure 1). Their task is then to decide, on a moment-by-moment basis, which camera to select for the live broadcast. The video created in this way is the coverage of the event through a spontaneous collaboration that is then made publicly visible on a website, where it is also recorded. Thus, non-present friends get to see how the afternoon in the park unfolds. The tedious job of editing the videos is avoided as the video is mixed instantly. In the same way, the system could be used in other settings, from amateur news or sport reporting to weddings and local music festivals.

3.1 System architecture

The MVM system components - mobile cameras, a mobile mixer application, a local MVM server and Bambuser, as well as communication details in between them, are show in Figure 2. Video from mobile cameras (phones) is streamed over the 3G network to Bambuser. The local MVM server fetches these four camera feeds from Bambuser, combines them into one and presents them to the mobile mixer application through Bambuser. The director can select any of the four camera feeds to be broadcast. Mobile cameras and the mobile mixer application are connected through 3G with other system components, while the communication between the local MVM server and Bambuser goes over TCP/IP. Our initial experiments showed that transmitting four live video streams together to a mobile mixer over 3G network would require more bandwidth than it is possible to provide during typical mobile network conditions leading to the unsatisfactory system performance. We address this problem by the specific design of the local MVM server. The main system components are described next.

3.1.1 Mobile cameras

The MVM uses mobile phones equipped with built-in cameras and the Bambuser application for live video streaming. For practical reasons when displaying video streams, the system currently supports up to four camera phones.

3.1.2 Bambuser

Bambuser [11] is a live streaming online service that lets its users broadcast from their mobile phones or desktops and share broadcasts instantly with viewers on the Internet. In the MVM system, Bambuser is used for broadcasting of camerapersons' individual streams and the combined mixer stream, as well for outputting the finalized broadcast. Bambuser was chosen due to its high immediacy and short delays in broadcasting of images. It streams live video over TCP/IP channels using the FLV (Flash Video) container format.

3.1.3 Local MVM server

The local MVM server runs processes for live stream combining and switching. The video combining process requests and fetches the current broadcasts from Bambuser, stitches them together in a cross view and broadcast the composite image back to Bambuser as one stream (from there it is streamed to the mobile mixer application). The video switching process listens to the mobile mixer application and performs requests for switching between the sources. The resulting broadcast is fed back to Bambuser where it is made available to viewers.

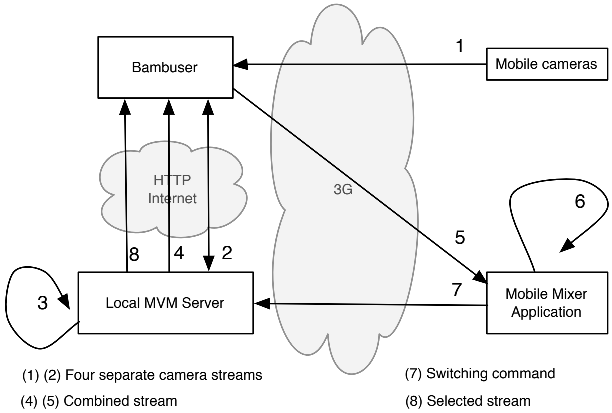

Figure 2: MVM architecture and data flow

3.1.4 *Mobile mixer application*

The mobile mixer application allows a director to see a quadruple live preview of four camera streams (see Figure 3), and to select one for broadcast, using the number keys 1-4 on the mobile phone. A red frame labeled "On Air" overlaid on the image marks the camera stream currently selected for broadcast. Complementary information of each camera, including camera index, the title of the video, the start time of filming, are displayed, as well as the label "Ready" for each of the incoming video feeds. The application is implemented as an Adobe Flash Lite object, suitable for mobile interactive applications. This object can run on any mobile device with support for Flash.

Figure 3: Four-camera view of the mobile mixer interface

4. METHOD AND SETTING

The study is an ethnographic field trial where participants were video recorded and observed while using the MVM system. Data collection took place on two occasions in two public places; a skateboard park and a science center, during the summer of 2010. The studies were performed altogether in seven sessions with participants of ages 11 to 17. At the beginning of each session participants were given an introduction of the system and its use. To conclude the sessions, participants were brought together and interviewed in groups. The broadcasting sessions and following interviews were video recorded and analyzed after the event.

In each session, participants were divided into four camerapersons and one director. The director or camera operators were also allowed to choose an assistant, when available. Two of the directors worked with an assistant, the other two on their own. Each group was asked to provide footage of their respective events, in the best understandable way, to remote viewers. Their briefing included basic descriptions of the roles of director and camera operators. Study participants during the first occasion (skateboard park) were free to choose their topic of filming while those in the second (science center) were instructed to cover an exhibition object of their choice.

5. ANALYSIS

The analysis of the field trials is focused on two specific aspects of the use of the MVM prototype, representing its conceived contribution to currently available systems. First, it is concerned with the director's experience of using the proposed technical solution for viewing and mixing between four parallel camera streams. Second, it presents observations and feedback on the practice of mixing on a mobile device in-situ, in visual proximity of the camera team. The analysis is based on empirical data collected during the field trials; interviews with participants, observations, and the recorded videos of their activities.

5.1.1 *Managing the task of mixing video streams*

The most appreciated mobile mixer feature was the split-screen live preview of what all camerapersons were filming (Figure 3). There were no reported problems with viewing and understanding the layout of the combined stream. The image quality was generally perceived as sufficient for performing the mixing task, and all directors understood the red bounding box indicating the selected video stream. Both camerapersons and directors frequently chose to have an assistant or a co-director to collaborate better with the other team members. By doing so, the cameraperson could focus on the task while the assistant would coordinate by talking to the director or propose topics for filming. In the two cases of directors working in pairs, the co-director maintained an overview of the positions of the team and the live action, and communicated this verbally to the director, who could then focus more on their tasks. *"All of them seemed very spread out, so the thing is to sort of keep track of where everybody is, to make it easier (…) we are the only ones keeping track of everybody",* one co-director commented. This task separation between director and co-director was visible only in the pairs where both were present on the filming location.

No in-situ performance tests of technical aspects of the system were performed at this stage. Instead we focus on the directors' experience of the delay, image quality and layout of the mixer, as they are part of the experience of mixing on a mobile device. Delays in the video feeds were clearly a significant technical problem that needs to be addressed in future development. Subsequent lab test have shown that the delay between the camera and its corresponding presentation in the mobile mixer application is on the minimum value of 11.120 seconds, which is enough to severely affect the director's performance, since he/she instructs the camerapersons and makes mixing decisions based on his/her view in the mixer console. The final output delay had an average value of 13.20 seconds, but is less critical as a remote viewer cannot perceive the delay as long as it is constant.

5.1.2 *Taking advantage of mobility while mixing*

One of the director/co-director teams chose to sit in a fixed position outside of the filmed action, while the other three stayed close to the action and in visual proximity of the camera team. We observed how these directors used both the mobile interface and the live action as resources for mixing, in what could be called "*in view mixing*", distinctly different from the practice of a remote mixer. The directors were mixing while looking at the actual scene, and viewing it through the mobile mixer application, interchangeably. They were able to do this because they were physically present at the location of the filming, and could make out the relative positions of the camerapersons at the scene of the event. Being present on location and mixing "in view" was a way to practically manage two problems. First, it helped directors manage the delay in displayed video streams, which becomes more noticeable when in-situ as the direct comparison between live action and video feed is made available. *"It was lagging a bit too, so it was a little tricky – you saw this person doing something. Then it came up afterwards on the camera so you didn't exactly know how to… (…) You had to wait and see when it showed up on the camera"*, one director commented. This makes evident how they were attending to both the live action and the mixer interface while mixing. The comment also confirms a general observation; that although the delay was clearly a problem, the directors also managed to use it as a resource, in that they used the live action as a "*preview*" of events to act on when they appeared in the interface.

Second, visual proximity helped the person mixing direct the camerapersons. Less experienced camerapersons tended to stand closer together while filming, allowing them to communicate and compare shots. This behavior presents a problem to the director, as she has fewer perspectives to select images from. Users more familiar with filming (in the skate park setting) were more mobile and individualistic. One of them commented while filming: *"Ok, but now everybody is in the same spot so I'm going to spread out a little",* acknowledging the usefulness of multiple camera angles. Directors reported that they were trying to verbally direct the camera operators on location to make sure that they provided interesting and complementing material.

6. DISCUSSION AND FUTURE WORK

The first version of the MVM system, presented in this paper, was designed and developed to explore collaborative video production where all participants in the production are mobile and co-present. It is achieved by bringing live mixing to a mobile handset, and thus letting the director out with the filming camerapersons. Our first trials show that mixing in-situ, on a mobile device is feasible. The proposed technical solution made it possible to stream over the 3G network and to view parallel streams in sufficient image quality on the mobile phone screen. However, although the aim here was not the performance evaluation of the system, the initial user studies showed obvious delay problems. Delay between actual event and its display on the mobile mixer was especially problematic, and will require further technical development for future versions of the system to be truly useful in production. This is in progress, and our plan is to do performance tests in realistic settings as delays become more manageable. Use of 4G will increase the performance greatly, but effective use of bandwidth and processing on mobiles will still be needed. The MVM system emphasizes one novel feature, and its potential advantages. Many of the other production features included in more elaborated systems, e.g. a feedback channel, editing and special effects, were not explored. Future work will involve the addition of a communication channel in between the director and camerapersons, adapted to fit lightweight collaborative live video production. Despite these early technical and functional limitations, our field trials indicate that enabling the director to be mobile and in visual proximity of the camera team has advantages for amateur users. It helps the director maintain an overview of the cameras and live action, select the most useful camera, and direct the operators on location. "*In view mixing*" gives the directors an additional visual resource. It also lets them use predictable delays, and use live action as a "*preview*" for mixing and broadcasting decisions. As some delay can be expected even in more optimized future systems due to properties of mobile networks, this indicates advantages for systems that allow the director to be co-present on location during the broadcast.

7. ACKNOWLEDGMENTS

The research was partly made possible by a grant from the Swedish Governmental Agency for Innovation Systems to the Mobile Life VinnExcellence Center, in partnership with Ericsson, Microsoft Research, Nokia Research, TeliaSonera and the City of Stockholm.

8. REFERENCES

[1] Engström, A. Perry, M., Juhlin, O. 2012. Amateur vision and recreational orientation: creating live video together. In *Proc. of* CSCW '12. ACM, New York, NY, USA, 651-660.

[2] Bentley, F., Groble, M. 2009. TuVista: meeting the multimedia needs of mobile sports fans. In *Proc. Of* MM '09.

[3] Juhlin, O., Engström, A., and Reponen, E., 2010. Mobile broadcasting: the whats and hows of live video as a social medium. In *Proc. of MobileHCI '10*.

[4] Reponen, E., Huuskonen, P., and Mihalic, K., 2008. Primary and secondary context in mobile video communication. In *J. of Personal Ubiquitous Computing Volume 12, Issue 4*.

[5] Bergstrand, F., and Landgren, J., 2009. Information sharing using live video in emergency response work. In *Proceedings of the 6th ISCRAM Conference*.

[6] Dougherty, A. 2011. Live-streaming mobile video: production as civic engagement. In *Proc. of MobileHCI '11*

[7] Engström, A., Esbjörnsson, M., and Juhlin. O., 2008. Mobile collaborative live video mixing, In *Proc. of MobileHCI '08*.

[8] Shamma, D. A., de Sá, M., and Churchill, E. F., 2011. Capturing moments together: Collaborative mobile video production. In *Workshop on Video interaction, CHI'11,* 2011.

[9] Vihavainen, S., Mate, S., et al. (2011) We want more: Human-computer collaboration in mobile social video remixing of music concerts. Proc. ACM CHI, 287-294.

[10] Juhlin, O., Reponen, E., Bentley, F., and Kirk, D., 2011 Video interaction - making broadcasting a successful social media. In PART 2 *Proceedings of CHI EA '11*.

[11] Bambuser. http://www.bambuser.com/ (23.12.2011.).

[12] Livecast. http://www.livecast.com/ (05.01.2012.).

[13] Qik. http://qik.com/ (05.01.2012.).

[14] Ustream. http://www.ustream.tv/ (05.01.2012.).

[15] Kaheel, A., El-Saban, M., Refaat, M., and Ezz, M., 2009 Mobicast: a system for collaborative event casting using mobile phones. In *Proc. of MUM '09* (2009).

[16] Collabracam. http://collabracam.com/ (05.01.2012)