Mobile Collaborative Live Video Mixing

Engström, A., Esbjörnsson, M., Juhlin, O.
Mobility, Interactive Institute
P.O. Box 1197, SE-164 26 Kista, Sweden
+46 (0)8 751 72 30

{arvide;mattia;oskarj}@tii.se

ABSTRACT

We report on design research investigating a possible combination of mobile collaborative live video production and VJing. In an attempt to better understand future forms of collaborative live media production, we study how VJs produce and mix visuals live. In the practice of producing visuals through interaction with both music and visitors, VJing embodies interesting properties that could inform the design of emerging mobile services. As a first step to examine a generation of new applications, we tease out some characteristics of VJ production and live performance. We then decide on the requirements both for how visitors could capture and transmit live video using their mobile phones and how this new medium could be integrated within VJ aesthetics and interaction. Finally, we present the SwarmCam application, which has been implemented to investigate these requirements.

Categories and Subject Descriptors

H.5.1 Multimedia Information Systems: video

General Terms

Design, Human Factors

Keywords

VJ, mobile video, production, collaborative, club, public displays, hybrid media, real time editing, socializing, nighttime, visualization.

1. INTRODUCTION

In recent years a number of trends on user content creation on the Internet, such as photo- and video-sharing and collaborative editing have emerged. This phenomenon has gained interest from the HCI research community, where a number of projects have been presented, exploring mobile aspects of user content creation. Tazaki [26] displays a fictitious, however thoroughly elaborated, concept of turning passive media consumers into collaborative media producers. Jacucci et al [12] and Sarvas et al [21] explore the potential in camera phones when co-experiencing events in groups. Kirk et al [13] investigate what people do with video when it comes to aspects of recording, editing and sharing.

Nevertheless, the question remains: How should new and innovative services, supporting mobile user content creation, be designed? In the following we suggest that mobile and collaborative user created content will be indispensable in the production of future hybrid media [7]. To understand, as well as to design for, these emerging forms of mobile and collaborative media production [8], we bring forward VJing, with its unique characteristics, as a favorable use case. VJing is a predecessor when it comes to experimentation with form, content and presentation in visual media production and display. Clubs and nightlife remain the main platform, where they use video, film projections and lights to accompany a DJ's music set and to interact with their audience [5].

In recent years HCI research has found its way into the nightclubs. Several studies report on gadget-oriented projects [27; 22; 1; 24; 2; 14] supporting live media production, which fits well with the experimental characteristics of the VJs in their utilization and adaptation of new technologies. There are also a number of empirical studies revealing the interactional practices between performer and spectators [20; 10; 23], as well as a few texts discussing the history and culture of the club environment [17; 5]. Despite that the related studies concern the investigation, or introduction, of new technologies supporting the relation between the crowd of spectators, and the performer in club settings the majority of them concern the DJs and their audio-centric performances. Hence, we now return to the club environment to gain a deeper understanding of visual media production in performances, and investigate the possibilities for designing services which invite the club visitors to take an active part in the real-time creation of visuals, thus transforming it to a collective effort.

Our topic and methods are framed within the area of design research [29]. It is concerned with the integration of theoretical knowledge, technical opportunities as well as explorations of naturalistic everyday life studies. Through an associative process, design researchers redefine and reorient available resources to accomplish a meaningful change of states [9]. It results in new theories, technologies and better understanding of social life.

Initially we created a design space [19] spanning new mobile technologies and existing applications supporting video capture and editing, the existence of collaborative video production practices, as well as spatially distributed social practices [8]. In order to further investigate the topic we conducted ethnographic studies of use contexts, and generated new prototype applications. The selection of a naturalistic case study is an important step to inform the design research process. VJing was chosen because it includes real-time editing of visuals, elements of social interaction and visitors who often utilize mobile equipment. We have in this

case particularly focused on the perspective and practices of the VJs. We wanted to learn how their media production was fitted into various forms of social interaction, as well as to understand the potential of adding new forms of live media production into this context. In our analysis we identify the characteristics in their aesthetical preferences; the interaction pursed during a VJ set, as well as their interdependencies. By juxtaposing the concept of mobile collaborative video production and our analysis of VJing, a number of design implications emerged. These have played a major role in the design and implementation of the SwarmCam prototype, a mobile video production support for VJing, enabling real-time collaboration.

The paper is outlined as follows. We begin by giving a brief account of VJing, followed by a section on related research. The paper continues with a description of the method and setting for our research. The empirical material is analyzed, and is succeeded with a section summing up the important findings relevant for the design. Finally we present the implantation and description of the SwarmCam prototype, and conclude the paper.

2. VJing

The term VJ, short for video jockey, can be traced back to early music television and to night clubs in the late 1970's and early 1980's [5; 28]. The analogy between VJ and DJ (disk jockey) best explains the VJ's work. While the DJ selects, combines and mixes music in real time, the VJ uses similar tools to dynamically design the visual appearance of the space where she performs, usually in conjunction to music [15]. Thus, the VJ is responsible for the creation, playback and live mixing of moving visual art in the space of the performance [28]. VJing was originally influenced by contemporaneous visual arts, but gained its independent form in co-evolution with electronic music [5]. Performers who are more affiliated with the art world than the electronic music scene may label themselves visualists or live cinema artists rather than VJs. Even today's club VJs are expanding out of the electronic music scene into audio-visual and performance art, cultural institutions, concerts and mainstream business [28]. Since the majority of the performers in this study work mainly in the club space, we will continue to use the term VJ here.

VJing comprises a variety of media, reflecting the background of the performer. The origin of the content may be any combination of graphics, video, 3D animation, photography, film, and the aesthetics displayed by performers are as diverse. Again analogous to electronic music, a VJ set is typically pieced together from short, looped clips of different source media [15]. The logic of using loops as building blocks allows for instant adaptation to the beat of the music and makes it possible to create virtually unlimited variations of image sequences from the selected material [28]. The short looped sequences stand in contrast to traditional cinema's linear edited sequences, and reflect the need for modularity and adaptation to music, which is also built largely on repetition. Virkhaus [28] elaborates on the role of loops, and distinguishes between editing and mixing in VJing: *"A concert visual is a dynamic piece that may be composed of pre-produced sets of visual content used as building blocks, which are then timed to live music. It is not created through editing, but instead through live mixing. Sets of video loops may be constructed for each part of a song – introduction, verse, chorus, bridges and so on – which are triggered on-the-fly during a live performance"* [28]. This shows how the effort of producing a VJ performance is divided in time into production of pieces of content in advance, and mixing compositions live at the time of the performance. In the live mixing of media, VJs combine clips using some basic approaches for image montage brought from traditional film making and digital postproduction. For our design discussion on incorporating user generated video in VJing later in this paper, it may be useful to outline these basics approaches:

Content mapping is the action of mixing clips or loops based on their significant content. This approach relies on the power of human association and creation of meaning. The meaning is consequently used to either associate or contrast familiar elements within the images.

Mixing clips by *visual expression* is similar to content mapping, but associates or contrasts clips by making use of their visual properties such as tonality and contrast. These properties can be manipulated in production, in the studio and during the live performance, using effects and mixing modes available in VJ software [3].

Temporal and *spatial* montage are the two main cinematic dimensions of montage [16]. Temporal montage is what we call editing in film; producing sequences of separate shots over time. Spatial montage can be divided into several categories, as discussed by Manovich [16], among them montage using layers, as enabled by digital compositing techniques, and two-dimensional movement of layers in relation to the display surface, as in for instance a video mosaic.

3. RELATED RESEARCH

As indicated in the introduction we refer to two important areas of related research, where the *first* concerns mobile content creation. Jacucci et al [12] investigates how camera phones are used to enhance a shared spectator experience. They argue that in this context, mobile devices can be considered beyond person-to-person messaging and beyond passive consumption of multimedia content. They emphasize how spectators co-experience events in groups, and how mobile imaging can be a participative practice enhancing the spectator experience on-site, rather than merely documenting it. Sarvas et al [21] present a study with some similarities, however where the sharing, discussion and storytelling aspects of mobile pictures were done in retrospect. In a more recent study on 'Understanding Videowork', Kirk et al [13] investigate video recording among teenagers, and how they use their own mobile phones. This ownership of the devices affected what was recorded, seeing that traditional video cameras were only brought to events, while mobile phones are used whenever, for whatever, emphasizing spontaneity in capture. The spontaneity was also visible in the sharing of videos, which was done locally immediately after recording. They did not see the point of manipulating the clips, seeing that these were short snippets of action, and the clip title gave enough information. Tazaki [26] presents her conceptual idea of the InstantSharecam, which enables this spontaneity in capture, but even more important is how she emphasizes the collaborative process. She envisions a group of users, each with a video camera, simultaneously shooting and co-directing coverage of an event in real time.

The *second* area of research derive from the club environments, and covers gadget-oriented approaches, introducing prototypes adding new dimensions, new interfaces, restoring 'old-time'

interfaces, and removing tasks to ease the cognitive load of the DJ, and empirical studies revealing interactional practices between performer and spectators. The prototypes contribute in various ways by adding features and assisting the DJ in his performance. The AR/DJ-system [24] provides a 3D interface by which the DJ can 'place' his music and effects. Accordingly, with the positioning of sound on the dance floor Stampfl [24] adds a new dimension, in excess of rhythm and light, to create an appealing club atmosphere. When it comes to new interfaces Beamish et al [2], Lippit [14] and Slayden et al [22] provide physical interfaces to the interaction with digital music. In the systems D'Groove [2], Lupa and Audile [14], they attempt to restore some of the qualities of the 'old-time'-types of interaction, where the DJs used vinyl only. D'Groove clearly resembles qualities of physical turntables, which is also something provided in several commercial products. The DJammer-system [22] contains a handheld device, equipped with an accelerometer which *"..allows DJs to interact more with their audience, but it cannot come at the expense of control over music.."* [22]. This characteristic is also visible in commercial products such as Pacemaker® [18], which is a professional pocket sized DJ-system. Rhytmism [27] is a VJ performance system with maracas based devices, which also emphasize *"..the freedom of the user's physical movement and to realize the big attraction that the performance itself has..".* Accordingly they argue for the manipulation as an important part of the performance as well. A last category of prototypes concerns the automation of the important task of beat matching. Andersen [1] claims that *"..beat matching is a task where technical skill can be shown off, but little artistic and creative expression is possible..",* and consequently this could easily be removed.

Several of the prototypes mentioned above are, according to the articles, informed by studies of DJs and the club environment. However, none of them clearly refer to how the design is supported by the empirical findings. It is also the case that some of the design elements presented are contradictory to results from studies with a stronger empirical focus. Gates et al [10] deliver the critique that several technology-driven projects have sidestepped the DJs role in the process of connecting awareness information to the music. From their empirical research they found that DJs interact with the visitors in complex ways, and that the DJs are adept at reading the visitors, despite the demands of selecting and mixing music [10]. They conclude in their design guidelines that, among other things, new services should not add to much cognitive load to the DJs, in terms of introducing complicated new equipment. However, somewhat contradictory, they also argue that new technology should be used for information gathering rather than automation. The DJ should be in control. Neither technology, nor club visitors should command the show. They also emphasize the interaction between the DJ and the visitors. This is facilitated by the physical position of the DJ, most often very close to the dance floor [10], as opposed to how Sheridan et al [23] describes that the DJs *"..are more often than not tucked away, obscured from view..".* By highlighting the interaction between the DJ and the audience Gates et al [10] emphasize the importance of the performance, and it is not only in terms of amplified effects [20]. The highly visible position of the DJ, and the design of physical interfaces, put forward the importance of amplified manipulations [20], i.e. expressive interfaces come to play an important role in the club setting.

4. METHOD AND SETTING

We have interviewed nine VJs from five VJ acts from Sweden, Finland, France and Spain. As part of the research we have also visited several VJ performances in club and festival settings in Sweden, Finland and Japan. The interviews were recorded and transcribed. This empirical focus obviously leaves out a study of e.g. club visitors and DJs. Our analysis draws upon the interviews, but also on our own participation and experiences of VJing in night clubs and new media festivals. When analyzing the ethnographic data, we aimed at producing a general understanding of being a VJ, as well as specific knowledge on social interaction in this environment. The study also influenced several design activities focused on associating the conceptual idea with the findings from the interviews. Here we organized design workshops, which included VJs, DJs, as well as participants from the mobile industry. Finally, we settled on a design concept that was implemented to illustrate, and eventually test, how mobile technologies for collaborative real time editing match to hedonistic club environments.

5. ANALYSIS

In the following we identify a set of themes of relevance for the design space, based on the interviews as well as our own interpretation of VJ performances. We identify characteristics in their aesthetical preferences; the interaction pursed during a VJ set, as well their interdependencies.

5.1 Music matching

The activity of producing a tight rhythmic connection between visuals and music, i.e. beat matching, is a defining feature of a good VJ set, according to our interviewees. Most of them name rhythm as central to their expression. A lack of connection between visuals and music is seen as a failure and a loss of an essential quality of VJing as an art form. *"You can play as beautiful material as ever but (...) if it doesn't go with the beat then it's...it doesn't really matter."* (HR) Another VJ states that: *"most of our clips are based on rhythm. There is something pulsating in them, which in turn is rather easy to tap in to BPM."* (DA). Thus, he identifies a salient visual element with a rhythmic potential within the source video and then selects it for editing. The clip is then looped during the live performance. The VJ controls playback speed and manually aligns the rhythmic element to the rhythm of the music, which is measured in beats per minute (BPM).

Hectic vs. fluid tempo expression: There is an easily distinguishable tempo in dance music. The tempo of the visuals is related but not equal to the tempo of the music. Visuals have their own pace determined by the VJ's aesthetic preference and decisions in the live performance. The VJ most often strives towards matching the visuals to the beat. To accomplish this he may go for a *hectic* mixing style, driving up the pace of the visuals, or a slower more *fluid* pace.

Manual vs. automatic mixing: The matching of video to the music is predominantly a *manual* activity. VJs use special hardware mixers and software mixers running on laptops to physically fit the visuals to the beat. Improvisation through manual mixing is regarded as a central part of their skill set. But parts of the task can also be done automatically, and some of them had been experimenting with it. Audio analyzers enable control over

playback speed as a function of the audio output. But these tools introduce delays, which make some VJs hesitant to using such equipment. For the same reason, beat matching is hard to conjoin with use of external input from sensors and cameras. *"It can be very annoying. But if you work with it the whole night you learn how the delay works and you get it right"* (YS). Thus, VJs get to know the special constraints of their setup and adapt to it. A less common approach was to let the music follow the visuals. VJs had been experimenting with: *"AV-connections, more samplings from the VJ side. More audio from video clips and animation, bringing it in, and sending it back to the music mixer, to kind of merge it all together."* (MW) and *"artificial intelligence. Improvisation is augmented in real time, the music controls the visuals and vice versa."* (BG). This illustrates how loosely controlled live input from music and the performance space is regarded as a means of creative experimentation as long as it does not disrupt the VJ's creative control of the core functions of mixing.

Consequently, none of the VJs expressed any interest in automating the basic tasks of their work performing live, which disaffirms some earlier attempts in HCI research in this direction [1].

5.2 Aesthetic preferences

Mashup vs. uniform media combinations: VJs who express that they strive to adapt to the music in a reactive way, can combine media, such as graphics, video, 3D-animation, photography and film in different ways to achieve this. As shown, VJs can combine clips to associate – or contrast – them in order to produce a VJ set, using some basic approaches from traditional film making; *temporal* and *spatial* montage, *context mapping* and by visual *expression*. These approaches are supported in the basic set of tools in most or all VJ software, and form the technical framework the VJ operates within.

Aesthetically, the VJ can adapt one of two main approaches to producing a set. First, he can create highly associative *mashup* compositions (see fig. 1), comprising many and diverse media types, techniques and aesthetics: *"Mashup. Everything at once. (…) not necessarily a lot of effects, but a lot of layers on top of each other"* (DA) Mashup aesthetics seem to work well with the modular logic of a VJ performance, which according to another VJ *"gets enhanced the more you add to the mix. It's the same with adding more objects"* (YS).

Figure 1: Mashup composition using still photography, animation and typography

Second, they can apply a *uniform* aesthetic, in terms of source input (see fig. 2). The composition is held together by limiting the use of techniques and media types to one or two, and using reoccurring graphic patterns and colors to create a more coherent expression over time.
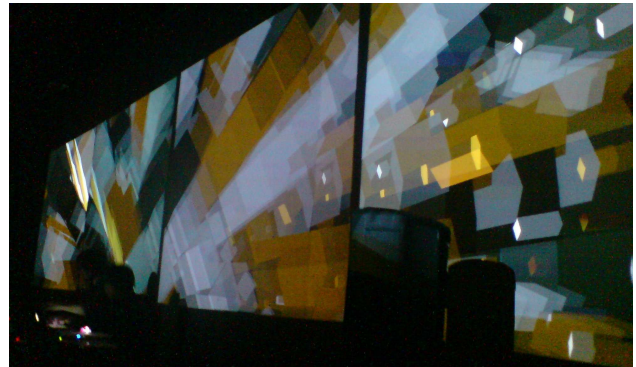
Figure 2: A uniform composition

Rough vs. smooth editing: A VJ set is typically pieced together from short, looped clips of video or animation. The live mixing is preceded by editing and rendering these building blocks in pre-production. These elements can then be looped and repeated to produce a rhythm within the sequence. We discern between two main editing styles. First, the cuts can be *rough* and marked, which is an effective way of creating building blocks for rhythm in visuals. A VJ described their typical clips as *"Things that fill up space, draw attention. Distinct cuts, I would say."* (DA) Small and simple units can then be combined to build complexity in a live set. A rough edit would imply less precision in regular video editing, but in VJing rough cuts can be effective in creating a beat in the cut itself, since the clip is looped and matched to the beat of the music. Second, *smooth* editing is achieved through isolating a single action in the source video and work with the beginning and end of the clip to make the transition as even as possible. *"Often enough you find like a seamless loop from a linear video, like you know, one action."* (HR). This process requires more work in preproduction and renders an accordingly smoother result. Here, rhythm is created through looping content elements *within* the clip, while the cut itself is seamless.

Balancing light and darkness: The dim lighting of the club is an important prerequisite for the display of visuals. They are purposefully designed for dark spaces and heterogeneous screen setups. Just like in a movie theatre, darkness has an important function in providing contrast and enhancing the visuals. Knowing how screens are positioned and controlling ambient lighting, the VJ has more freedom to design for a visual experience in the club space as a whole. This goes beyond the individual screens as the visuals become a dynamic part of the lighting. VJs working outside of the club space and with irregular projection surfaces described how they would adapt their content to resemble *"(…) moving stencils, or living stencils"* (SS) using black to effectively mask out the empty space toward a dark background.

Visuals erase the gloom they depend on. On the other hand, the lighting conditions put restrains on the performance. The amount of light projected has to be balanced. Bright images may emphasize the screen surface and have other undesirable effects. *" (…) doing material with white backgrounds or very bright clips is*

often a bit counter productive in club environments since you only light up the room instead of getting that club feel" (MW). The VJs had differing opinions on this balance. Another VJ stated that "brightness can be very effective. Lots of light in the images, and flashes" (FC). Darkness also affects image quality from input devices. Using live cameras and projectors in unconventional environments like clubs and in the streets at night presents its challenges. "You really need a camera which is able to do good night shots (...) vice versa the camera needs a lot of light to get good images but the beamers are prepared for dark places in the evenings" (SS).

5.2 Interface preferences

Size vs. quantity of the interface thumbnails: The video bank, a matrix where video clips, animations and graphics are organized visually as thumbnails, is the commonly used interface metaphor for media content in VJ and live performance software. The video bank is typically integrated in the VJ mixer interface (see fig. 3), displaying a given number of source clips at a time, each represented by their first frame. There is a trade-off between a visible size of the thumbnails and the need for storing a large number of clips, up to 300-400 for a show in the case of two of our interviewed VJs.

Figure 3. A section of a typical VJ interface, showing part of the video bank to the right (Modul8).

The video bank typically accommodates this trade-off by arranging clips in pages where the VJ can arrange the rest of the clips in lower levels. The interface is designed to provide the most visibility and best access to the videos, given the restricted space of the video bank.

5.3 Social interaction

Ambient interaction with the audience: The interaction is very limited in between VJ and audience, although they are standing very close to each other. The audience does not necessarily associate the visual effects to the people on stage who is putting it together. It becomes obvious for example on those occasions when they do talk to each other. A VJ says that *"...often we're mistaken for the DJ. Drunk chicks who ask if we can play some R&B"* (FH) This does not mean that there is no interaction. VJs interpret the activities of the audience as feedback on what they do. In that sense they are successful if the audience is dancing. They are also successful if they take a visual interest in their graphic production. However, if the audience just looks at what they are doing, this could be negative feedback: *"[W]e don't want people to just stand and gaze at our graphics, we want people to dance"* (MW). They look for the subtle combination of people dancing and looking at their visuals: *"[I]f someone is dancing looking at a screen, then we understand that it contributes to the atmosphere"* (MW). Hence the interaction with the audience is

highly subtle, or ambient, and occurs partly mediated through the dance per se.

Orientation to the DJ: The VJ's interaction with the DJ is not only mediated through the beat of the music. First, VJs and DJs talk to each other during the set, even though it occurs rather rarely. It is also possible to follow the DJ by looking at what he or she is doing, as stated by an interviewee: *"[L]ooking at the DJ is enough. You can see when he's changing records and so"* (FH). Furthermore, the interaction could be rehearsed and preplanned, which gives the VJ and DJ better opportunity to coordinate their performances. Accordingly, some of the VJs preferred working with DJs they knew. When a tight interaction between VJ and DJ is missing, it is most often a result of bad planning. It is seen as a lack of insight of the importance of pairing musical and visual expressions on the part of arrangers who often view VJing as mere eye-candy. The result is perceived by the VJs to be less interesting.

But failures in beat matching can also be caused by a stressed DJ who does not orient to the VJ. *"If there is a DJ or somebody who plays and they're stressed and only want to take it to the next level without any kind of mature framing, I get stressed too (...) It takes time to build. Then I get stuck and it's perceived as out of synch"* (FH). Ability to predict the DJs actions ahead of time will prevent this, which can be achieved if the DJ and VJ have collaborated before, or if the DJ provides orienting gestures to the VJ or by cues in the mixing of the music.

At the same time, pre-planned collaboration is not essential for a successful set. In cases where the DJ is only slightly known to the VJ, the VJ can select and prepare customized material in advance, and then rely upon live mixing skills and improvisation to produce a unified set. The VJs exemplifies with an upcoming show where they plan their performance with very scant knowledge of the DJs plans for the evening: *"What we're going to do is pick up a lot of pictures and pre-edit and pre-animate them, some stuff with his name... This is like the basic structure for that, we can relate to the DJ already, to the artist"* (SS).

Interaction between VJs: A VJ can also be involved in live-collaborations with other VJs. They can either work in pairs, or in larger groups. A favored setup is to work on parallel workstations and parallel screens. There is more space for preparation and elaboration when there is one online composition and one offline composition. It is perceived as an advantage by some of the VJs: *"Preferably, you want to concentrate fully on what is being shown and use the MIDI-controllers to sort of do rhythms and such (...) when you're done with it you can just mix to your buddy and be more concentrated on one comp."* (MV). Here, they use functional separation, where the editing and beat matching is constantly done by two different persons. Two of the groups have developed the collaboration even further. They have a person with a specific mix master role for some performances. It lifts the overall direction task of the set off the other VJs who can then focus on the creation of their individual compositions without worrying about the critical timing of transitions between segments.

As VJs co-produce material and collaborate live over time, they orient their performance styles toward each other. They rely on the knowledge of the other person's material and personal expression. *"We've been playing for a long time so it's kind of easy to predict what is happening in the head. There are phrases*

or sentence structures that have already been said earlier. It's anyway a dialogue; he says something and I reply somehow and the story kind of evolves from that" (SS). It implies a common visual grammar, which is mediated through the imagery.

5.4 Interdependencies

In this section we will tease out some of the ways in which there are interdependencies between aesthetical preferences and interactional characteristics. The identification is predominantly done by the authors by juxtaposing the VJs' comments on individual characteristics in their practice.

Media composition and tempo expression: The ways in which the VJ create visual tempo interdepend with the media compositions they select. We have identified a specific pattern among most our interviewed VJs which combine *mash up* media compositions with a *hectic* tempo. We suggest that mash up material provides a wide range of possibilities for creating rhythm through loops and mixing and quickly adapting compositions to the hectic beat of the music. It also works well for pushing out large amounts of visuals at a fast pace to "compete" with the music for the club visitors' attention. The mash-up style, with hectic tempo expression, lends itself to both collaboration and friendly competition with the music.

The combination of *uniform* compositions, with *fluid* tempo expressions, is rarer. Only one of the interviewees favored such an approach. It creates a visual accompaniment to a DJ set by establishing a rhythm within the image and staying with a main theme throughout several songs. Gradual variations of elements within the image replace hectic mixing as the means to mark rhythm and make the visuals evolve over time. This particular media consumption fits well with flowing and soft visuals. It would be hard to fit a uniform approach to hectic expressions, since the composition as a whole is hard to adapt with single elements or beats in the music as quickly as mash-up techniques.

Of special interest for the investigation of our design space is the way in which VJs, within those two interdependent approaches, add *live improvisations to their pre-edited material.* VJs that created mash up and hectic performances were more prone to using live cameras and other live sensor input from the club space. A VJ gave us an example: *"[T]he sky was a perfect blue for an hour or so, so we could key out the audience with one of those really ugly analog keyers. That was great."* (FH). *"It's like aikido. You're using other peoples' energy to produce your own material. (...) it took a while before they recognized themselves. I had effectively only made silhouettes out of them"* (FC). Both quotes illustrate how the VJs improvise and use live input to enhance the live aspect of their set. These VJs also experimented with ways to interact with club visitors through devices such as dance mats and other sensors in the live space. The use of such live input means that the VJ let go of some of their control of the visual expressions, to instead gain unpredictable and sometimes interesting effects from interacting with club visitors. Our interpretation is that a mash up style, with multiple image sources and few set rules as to how to interact with the material, is more forgiving to such unpredictable input than for instance a style based on abstract 3D animation. It is a bit unclear how well, e.g. live video input of a dancing crowd, fits with the demand for hectic visual expressions. Perhaps such a view provides requested pulse, or it could that it is not as necessary as their attachment to mash up compositions.

The VJ in our study who favored uniform and fluid presentations also utilized means to add live improvisation to pre-rendered material. He occasionally used audiovisual connections in the form of automatic manipulation by pre-programmed software, like in *Grains and Pixels* live cinema shows using artificial intelligence, [11] than manual mixing. Although there are exceptions like the highly physical and rhythmic generative graphics of *SanchTV* [6], the genre of uniform visuals most often conveys an interpretation of a mood in the music rather than distinct beat matching. Bad examples of this genre has led critics to liken VJing to computer screen savers or "visual wallpaper" [3], with seemingly no connection to the beat of the music.

Tempo expression, media compositions and collaboration: The tempo expression affects the VJs' modes of collaboration. It is difficult to collaborate when the tempo is hectic: *"[I]t's harder to do stuff that is as hectic, because if you suddenly start combining five layers from VJ A with five layers from VJ B, it can get really chaotic."* (MW) However, the most mashup-oriented VJs viewed this as less of a restraint, since the lack of overview is seen as a positive resource for improvisation. It allows them to produce new combinations within the material. We argue that in the cases where two or more VJs play together, controlled collaboration in producing a well held together expression is easier accomplished when the tempo expression is fluid. This gives all members the time needed to manage the separate tasks involved such as; live manipulation based on pre-recorded material; view and adapt to the other members' compositions; as well as take part in the directing of the set as a whole. In both cases, however, the large amount of work in interacting with other members and multiple sets of material is evident in that the groups prefer to assign the task of mixing between compositions and thus control the directing of the set to one single person

The influence of interface configuration on aesthetics: The structuring logic of video clips, graphics and other material is built into the interface of the software used for VJ-performance. Depending on its logic and layout, the interface may or may not support uniform media compositions of material that may be helpful for a more structured, uniform style. VJs who favor uniform compositions may structure their media content on the interface according to a planned sequentiality of a set. On the opposite end of the scale are mashup aesthetics. The VJs in favor of this approach tended to be less descriptive about how they produced and organized their material, and some even utilized a conscious disarrangement of material prior to a set as a method to build in a potential for interesting superpositions and associations between sequences as testified by two VJs: *"That is what playing live is about, mixing clips."* (YS) and *"(...)finding new stuff in old clips too. So we usually throw it all around. We just keep everything in a mess, because then you can't find what you're looking for"* (FH). This practice of overriding the structure proposed by the VJ software supports Gates point that in designing tools for performance, building in automation may actually be counter productive [10]. Structuring should at least allow for explicitly non-structured storage of content. On the other hand, a VJ software interface could support associative mixing by design, a dimension up to this point unexplored by available software.

6. DESIGN IMPLICATIONS

In this section we will juxtapose the concept of mobile collaborative video production and VJing. First, we will discuss how visitor generated video would fit with current video mixing practices. Second, but much more scantly, we will turn our attention to the club visitors and their role as camerapersons. In general, the VJ seeks to provide distinctive visuals with rhythm, which is matched to the music the DJ is playing. This visual beat is marked in various ways such as in the content (e.g. a door that opens and shuts), cuts in between the loops, thematic presentations, as well as by the use of effects. A big part of the work in VJing goes into preparing the visual presentations before the set. Introduction of resources for live video in night clubs, produced by the visitors, delimit the possibilities to do so. The VJ can only work with parts of the material that will be transmitted in real-time. In the following we will discuss the consequences of this shift of work organization and how this could influence VJing that includes live video feeds.
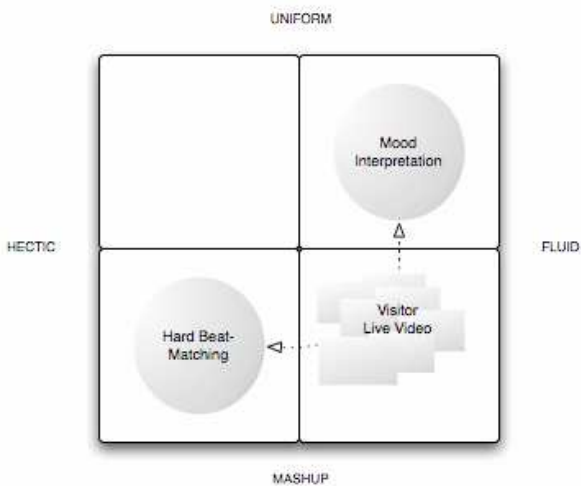

Figure 4: Mapping visitor live video to current VJ formats

We have previously discussed how VJs combine on the one hand hectic and mashup styles, and on the other hand fluid and uniform expressions. Hectic beat matching fits better with a mashup, fractured style (see low and left corner in figure 4). The rhythm of the visuals gets to dominate over the coherence. On the other hand, a slow or fluid representation allows the visuals to take on a parallel aesthetic to the music. It becomes more of an alternative but consistent experience (top right corner in figure 4). Allowing club visitors to contribute to the visual performance seems to move the style into a somewhat anomalous combination (see low right corner in figure 4). Raw live material will in many ways lack in strong beat matching, and be slower than e.g. looped video sequences, seeing as hectic material demands more preproduction work. At the same time, it will be unpredictable and thus sit well together with mashups. These properties suggest that collaboratively produced video would fall into a style category outside of the common VJ practices. This needs not to be a problem, since the practice is already ambiguous and open to new expressions. However, we suggest that the mixer should be designed to support adaptation of this material to current VJ aesthetics.

6.1 Mobile video input as media compositions

Uniformity and mashups are important aesthetic dimensions. Since the VJ can not control the visitors' video production it will be difficult to provide a *uniform* VJ set where the content of the video maps to other graphical element. Not surprisingly, we noted in our study that the VJs who had previously experimented with live video favored *mashup* aesthetics. However, it might be possible to predict the types of videos that the visitors would be generating and then adapt the presentations to make the set more coherent.

6.2 Real time beat matching with visitors' video input

VJs draw upon hectic and fluid aesthetics. *Fluid* aesthetics are marked by slowly altering visuals, with low or limited beat matching. It is unlikely that club visitors by themselves generate video which matches the beat. Not even captures of dancing visitors will have this match, since there will always be some time delay in the transmission. However, the VJ can adapt the video to a fluid tempo expression by avoiding cuts and going for long sequences, conjoining the video within themes that express fluidity in various ways, as well as by the use of effects on the video per se. A *hectic* expression with strong beat matching is favored by VJs, but normally depends on preparation of video material e.g. in loops crafted to produce a distinct visual rhythm. Control of frame rate, a basic tool for increasing tempo, is out of the question when working with live video, since it requires material ahead of the current frame. This narrows the options considerably when designing support for beat matching.

6.3 Mobile video input alters light conditions

The VJs orient their performances to *lights* in the rather *dark* club setting. The gloom is both a perquisite for the generation of a good experience, and a hampering factor. The darkness gives a scene to their expressions i.e. contrast to their works, but it also requires that they keep their performance on a well balanced level, since their performance competes with the darkness. Visitor generated video will face a similar balancing problem. We envision that the visitors would like to capture activities that are not the VJ performance, i.e. the projections, and thus occur in dimness of the club. But then we face a problem with acquiring sufficient technical quality. Using the lamp on the mobile video cameras would increase the chances of capturing the images, but would then risk lighting up the darkness in unwanted ways. Thus, the balance between light and darkness, which is now a challenge in the VJ presentation, will occur also in the production.

6.4 Visitors' video sets new demands on mixer interface

Unpredictable live content has another implication on the design of the mixer interface in that it sets new demands on preview. While static thumbnail images of clips in the video bank is sufficient for a VJ's recognition and identification of old and prerecorded material, it would not be sufficient for live video streams.

6.5 Visitors' video input influence the current organization

The demand to handle uncontrolled generation and live feeds will also affect the *organization* of the VJs. The VJ seeks to follow the

music and thus the DJ. The efforts being made to organize DJs and VJs more tightly will hardly fit with visitors' live video input. It will be hard to establish a tight connection to the DJ when the visitors require to be taken into account as well. If this concept is introduced, the VJ must in some sense not only follow the DJ but also the visitors. This will lead to a higher workload for the VJs. It may be necessary to separate the VJing tasks in between several persons to keep up with such increased ambitions. As the study shows, this division of labor, for instance between image manipulation and mixing between compositions, is already done in VJ groups in order to give each member more time and less tasks to handle.

6.6 Support for the club visitor as cameraman in a collaborative production

The focus in this paper has been to investigate how VJing is done and how it could be associated with the idea of mobile collaborative video production. Although we currently have done no specific studies on visitors, we need to make some initial accounts of how the concept maps to their use practice. Ethnographic studies of professional collaborative TV-production points to the intricate interaction between an individual camera man and the technical director, who mixes the many sources into one production [4]. The technical director talks to the camera man over the intercom and lights up a red diode on his camera, to mark when he is on air. This is done to make him hold on to a selected and favoured camera angle, and refrain from the mode in which he swirls around looking for an interesting topic. Symmetrically, the camera man proposes input by holding the camera steady. In our case, the club visitor who captures video on a mobile phone could be seen as a camera man. Similarly, the VJ shares the work tasks of real time editing of many sources, with the professional image engineer who mixes TV-input from various sources. Thus, we need to support ways in which the visitor can make propositions, as well as forms in which the VJ can indicate that the propositions are accepted, and then get the visitor camera man to hold on to a selected shot. But the relation between the club visitor and the camera man also differs. The camera man will continuously suggest shots to the engineer even though he is never, or very rarely, selected, since that is part of his job. The club visitor needs to be motivated by other means. We suggest that the design of the interaction support between the visitor and the VJ also account for the need to motivate the previous, which include both understanding that the VJ receives his material as well as when he find it interesting.

7. IMPLEMENTATION OF THE SWARMCAM

In this section we will present a concrete implementation and discuss the ways in which it draws on our previous association of design concept and our study of the club environment. The purpose of the application is to explore the possibilities with our suggested concepts, as well as provide a tool to investigate the VJing qualities further.

7.1 Use scenario

The SwarmCam application is intended to be used in the following manner. A visitor spends the evening at a night club which presents a DJ set supported by VJing. The visitor engages in dancing on the dance floor, making his moves publicly visible to the fellow party people. During the clubbing, she brings out her mobile phone and starts the SwarmCam application. She can now capture her friends cool dancing skills on her phone, and it is directly streamed to the VJ. A VJ looks at it, and thinks that it is pretty decent. He cuts it in, and combines it with some effects, which merge it nicely into the overall VJ performance. As her video recordings are selected she gets a notice on her mobile screen indicating that she is "on air".

7.2 System Architecture

The live streaming of video is managed by Movino, an open source program for streaming video from a mobile phone to a web interface, which we have implemented in parts and modified to fit the prototype. Movino has a Symbian S60 client which captures the video stream from the built-in camera and streams it over TCP/IP over the 3G network [25].

A Movino video server receives the incoming video streams. The server has decoding and encoding functionalities for encoding the data into the open OGG format, using the Theora codec. A data buffer is set to provide data for the codec, the buffer size set to a minimal in order to avoid delays in the live video transfer. The server has an archiving function which makes the video content available for recycling clips during a set.

A VJ mixer program, built in max/msp/jitter, runs on a laptop and is displayed on a screen or a setup of multiple screens., The data stream is transferred to the mixer over HTTP (Hypertext Transfer Protocol). The mixer program reads the video streams through a video player component that converts the data stream to readable frames of video.

The mixer consists of a GUI displaying preview windows of the incoming streams and controls for a basic set of mixing functionalities and effects such as brightness, contrast and hue controls, and tools for spatial montage. The interface also contains an output window, equivalent to a program monitor, which can be set to full screen mode or output to a separate screen.

An external hardware mixer serves as a last step before the screen output, allowing the VJ to mix between the SwarmCam composition and the output of a piece of regular VJ software within the same station.
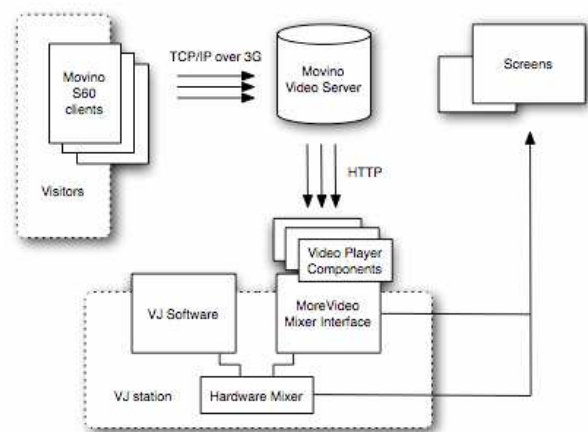

Figure 5: SwarmCam system architecture

A collaborative production with the SwarmCam prototype typically proceeds in the following manner:

1. The users start the application independently on their mobile phones and begin recording. The source YUV video is encoded to MPEG1 by the FFmpeg encoder within the S60 client and transmitted over TCP/IP to the video server.

2. The MPEG streams are then transferred over TCP/IP over the 3G network to the Movino server, where each stream is coded into the OGG format, given a unique URL.

3. As soon as a full frame of video is ready to be read by the mixer, the VJ receives an alert by the system. He then opens each stream through the SwarmCam mixer, which reads it to RAM memory on the dedicated laptop computer. From RAM, the SwarmCam mixer displays each incoming video stream in a preview window within the interface, visible only to the VJ. As long as the users keep filming, the VJ can now apply basic effects on the streams individually, using live image processing tools in the mixer interface.

4. The streams are then mixed or composited together in the output window. The output image can in turn be mixed and composited with other material from the VJ station using a hardware mixer, much in the same way a DJ mixes between two record players.

5. The output from the hardware mixer is displayed on a screen or a setup of multiple screens in the performance space.
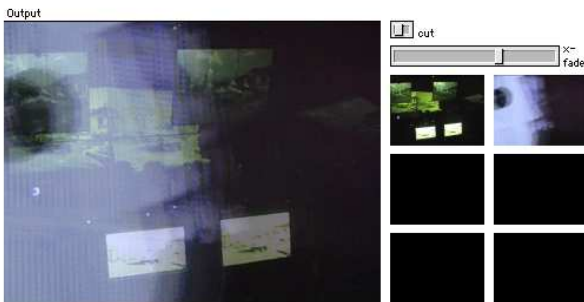

Figure 6. Screen capture of mixer with crossfade functionality.

7.3 How the application meets requirements

7.3.1 Allowing visitor video in media compositions
There is no possibility to pre-edit visitors' content. The VJs will loose some of the control over content, but they can still utilize standard features for VJ mixing which allows transformations of the video feeds in various directions. Those features, which were discussed in section two, include:

- Content mapping i.e. juxtapositions and temporal unfolding of material dependent of the meaning of the visuals, provided by traditional editing tools, cross fade and access to standard video bank.

- Spatial formatting on projection surface through tools for scaling, splicing and mosaics.

- Expressions, i.e. transformations of content through filters altering brightness, contrast, color and pixelation.

The VJ can through those tools adapt content to fit within preferred media composition style such as mashup or uniform. The interviews indicate that the type of content may fit better with mashup compositions.

In the same manner, beat matching through content will be hard to produce, but hectic matching can still be done with the available tools above, e.g by layering live content with pre-produced material and effects.

7.3.2 Dynamic thumbnails
The interface must in some way inform the VJ on the content. For familiar video material, a still image representing each clip is sufficient. Since live video is less predictable, the thumbnails in the SwarmCam interface instead show the live video stream from each camera.

7.3.3 Fixed spatial interface
The structure of the mixer interface is fixed. Thus, we provide now way of reorienting thumbnails to support mashup VJing. We suggest that the introduction of live video feeds introduces enough unpredictability in itself to support such media compositions.

7.3.4 Visitor-VJ interaction
The visitor needs a way to know if he is selected by the VJ. This helps her to continue to provide video material until she is deselected. We suggest that the projection screen will not be enough and provide cues on the screen of the video phone as well.

7.3.5 VJ organisation
Live video mixing will put an extra effort on top of the already constrained VJs. We suggest therefore that these types of material are handled among VJs working with dedicated mix masters. The mix master controls the physical slider between the SwarmCam mixer and the standard mixer. It could further be needed to have a specific VJ handling the new media as well, dependent on how much of it that is generated and its importance for the VJ set.

8. CONCLUSION
The design-oriented research in this paper presents the SwarmCam prototype, an innovative system for video capture and live transmission on mobile phones. The system fits with VJing and night clubbing, opening up the VJ performance, making it a collective effort where club visitors can contribute with live video.

We have teased out the detailed characteristics of how visuals are produced in real time in night club environments. This knowledge has influenced our design. We suggest that there is a fit on many levels. In general, the VJ's main role is to create an appealing club atmosphere for the club visitors. There is already an orientation to experiment with input from the club visitors of various sorts. The VJs also often open for experimentation, which is more than just an "open" attitude. The interest in new combinations is a basic part of most VJs mashup media compositions. New mobile technology has also features, such as video cameras and high bandwidth networks, which make it suitable for collaborative video production. At the same time, there are characteristics in VJing that speaks against such a combination. Today, their interaction with visitors is only mediated through the visuals, and they make serious efforts just to follow the music and the DJ. It is

not evident that it is possible for them to engage in direct interaction with club visitors. Thus, it is not enough to ground a design principle in a detailed study of VJing. Obviously, the next step is to bring the implemented concept into a club setting to investigate the characteristics of the technical implementation as well as our analysis of VJing. For the latter reasons, future work also includes adding new features into the application. We would like to provide support both for real-time production of hectic loops, as well as support for uniform media compositions. We would also like to enable access between camera functionalities, such as the light, and the VJ mixer.

The general motivation for the design research presented in this paper is to investigate how mobile collaborative video production should be designed in detail, as well as to investigate possible use contexts. Night clubbing was selected partly because it is a social event, including VJ's real time video production. The system has been developed for VJing in specific but contains general qualities, and contributes to the area of mobile user content creation. Thus, we will also have to move the concept out of the night clubs to investigate its generality to other use contexts.

9. ACKNOWLEDGMENTS

10. REFERENCES

[1] Andersen, T. H. (2005). In the Mixxx: Novel Digital DJ Interfaces. Proceedings of CHI 2005. ACM, 1136-1137.

[2] Beamish, T., Maclean, K. and Fels, S. (2004). Manipulating Music: Multimodal Interaction for DJs. Proceedings of CHI 2004. ACM. 327-334.

[3] Bernard, D. (2006). Visual Wallpaper, VJ Theory.net Published 12/10/06.

[4] Broth, M. (2004). The Production of a live TV–interview through mediated interaction. Proceedings of Logic and Methodology 2004. SISWO

[5] Crevits, B. (2006). The roots of VJing – A historical overview. VJ audio-visual art + VJ culture (ed. Mi-chael Faulkner/D-Fuse). Laurence King Publishing, 14-19.

[6] Dessens, D. (2008). SanchTV. Available at: http://www.sanchtv.com/ (Accessed 3rd, Feb, 2008).

[7] Engström, A., Esbjörnsson, M., Juhlin, O. and Norlin, C. (2007). More TV! - Support for local and collaborative production and consumption of mobile TV. Adjunct Proceedings of EuroITV 2007, 173-177.

[8] Engström, A., Esbjörnsson, M. and Juhlin, O. (2008). Nighttime visual media production in club environments. To be presented at the Night and darkness: Interaction after dark - Workshop, at CHI 2008.

[9] Esbjörnsson, M., Juhlin, O. and Östergren, M. (2004). Traffic Encounters and Hocman - Associating Motorcycle Ethnography with Design. Personal and Ubiquitous Computing, Springer Verlag, vol. 8, no. 2, 92-99.

[10] Gates, C., Subramanian, S. and Gutwin, C. (2006). DJ's Perspectives on Interaction and Awareness in Nightclubs. Proceedings of DIS 2006. ACM, 70-79.

[11] Gondouin, B. (2004). Grains and Pixels. Available at: http://www.bertrandgondouin.net/post/Grains-Pixels (Accessed 3rd, Feb, 2008).

[12] Jacucci, G., Oulasvirta, A., Salovaara, A. and Sarvas, R. (2005). Supporting the Shared Experience of Spectators through Mobile Group Media. Proceedings of Group 2005. ACM, 207-216.

[13] Kirk, D., Sellen, A., Harper, R., and Wood, K. (2007). Understanding videowork. Proceedings of CHI 2007. ACM, 61-70.

[14] Lippit, T. M. (2006). Turntable music in the digital era: designing alternative tools for new turntable expression. Proceedings of NIME 2006. IRCAM, 71-74.

[15] Lund, H. (2008). Philosophy and cultural theory of vjing and realtime interactive performance. Available at: http://www.vjtheory.net/book.htm (Accessed 3rd, Feb, 2008).

[16] Manovich, L. (2001). The Language of New Media. MIT Press, 155-157.

[17] Miller, P. D. (2004). Rhythm Science. MIT Press.

[18] Pacemaker. (2007). Available at: http://www.pacemaker.net/. (Accessed 6th, Dec, 2007).

[19] Redström, J. (2001). Designing Everyday Computational Things (doctoral thesis). Gothenburg Studies in Informatics, no. 20, Dept. of Informatics, Göteborg University.

[20] Reeves, S., Benford, S., O'Malley, C. and Fraser, M. (2005). Designing the Spectator Experience. Proceedings of CHI 2005. ACM, 741-750.

[21] Sarvas, R., Viikari, M., Pesonen, J. And Nevanlinna, H. (2004). MobShare: Controlled and Immediate Sharing of Mobile Images. Proceedings of MUM 2004. ACM, 724-731.

[22] Slayden, A., Spasojevic, M., Hans, M. and Smith, M. (2005). The DJammer: "Air-Scratching" and Freeing the DJ to Join the Party. Proceedings of CHI 2005. ACM, 1789-1792.

[23] Sheridan, J. G., Dix, A., Lock, S. and Bayliss, A. (2004). Understanding Interaction in Ubiquitous Guerilla Performances in Playful Arenas. Proceedings of HCI 2004. Springer Verlag, 3-18.

[24] Stampfl, P. (2003). Augmented Reality Disk Jockey – AR/DJ. Proceedings of SIGGRAPH 2003. ACM Press, 1.

[25] Storsjö, M. (2007) The design of Movino - S60 phone client, OS X components and video server. http://www.movino.org/ (Accessed 3rd, Feb, 2008).

[26] Tazaki, A. (2006). InstantShareCam: Turning Users From Passive Media Consumers to Active Media Producers. Presented at the Workshop Investigating new user experience challenges in iTV: mobility & sociability, held at CHI'2006.

[27] Tokuhisa, S., Iwata, Y. and Inakage, M. (2007). rhythmism: A VJ Performance System with Maracas based Devices. Proceedings of ACE 2007. ACM, 204-207.

[28] Virkhaus, V. (2006). Concert visuals. VJ audio-visual art + VJ culture (ed. Mi-chael Faulkner/D-Fuse). Laurence King Publishing, 14-19.

[29] Zimmerman J., Forlizzi, J., and Evenson, S., Research Through Design as a Method for Interaction Design Research in HCI. Proceedings of CHI 2007, April 28–May 3, 2007, San Jose, California, USA.