RAW: Conveying minimally-mediated impressions of everyday life with an audio-photographic tool

Joëlle Bitton, Stefan Agamanolis, Matthew Karau
Human Connectedness group, Media Lab Europe
Sugar House Lane, Bellevue, Dublin 8, IRELAND
{joelle, stefan, matt}@medialabeurope.org

ABSTRACT
This paper traces the development of RAW, a system combining a tool and a process for capturing and conveying audiovisual impressions of everyday life. The project aims to enable a relationship between the user of the tool and an audience in a different place or time with an absolute minimum of editorial mediation by a third party. The tool itself incorporates a digital camera and a binaural audio recording device that captures the minute of sound before and after a picture is taken. To inform the design process, we tested prototypes in a progression of three studies within different cultural contexts in Ireland, France, and Mali. We present the results of these experiences, in which we observed among our participants an emerging set of ways of exploiting the tool for different purposes: social glances, depictions of activities, active documentation, and intentional discourses. We also discuss more generally the advantages and pitfalls of multicultural analyses of prototype technologies like the one we undertook.

Categories & Subject Descriptors: H.5.m [Information interfaces and presentation]: Miscellaneous

General Terms: Design, Experimentation, Human Factors

Keywords: audiophotography, photography, contextual audio, binaural sound recording, everyday life, unedited media, memory, storytelling, cultural exchange, multicultural studies, Mali, Africa

INTRODUCTION

The notion of "everyday life" means different things to different people. For many the words perhaps invoke a sense of the typical routines and customs within which one is immersed: waking up, getting ready for work, eating lunch, socializing with friends, watching TV at night, and so on. We believe that recorded impressions of these kinds of activities and trivialities of everyday life can offer an incomparable perspective on the "realities" in which we live and can reveal much about rhythms of the societies and cultures that people inhabit.

In his documentary *Sans Soleil* (1982), director Chris Marker showed scenes of Japanese society one would not normally find, much less bother capturing, on film: Extended scenes of workers and travelers taking the Shinkansen, a train linking major cities, merely sleeping during the trip, for example. An observation of an "ordinary" time like this could be at first boring to watch, but it also provokes the audience to reflect upon their own memories of those forgotten, unaware moments and their possible meanings. The film was hailed as a breakthrough in its mundanely honest glimpse of what everyday life is like in that country. But even so, Marker was still an outsider to the culture, interpreting it for others via his own knowledge and background as he chose which scenes to include to paint his portrait into a time-restricted presentation form.

The RAW project began with the realization that, for many reasons, we don't always have a good sense of what everyday life is like in other places in the world, and that having this sense might be helpful in improving understanding and relations between people in different cultures. Reports and studies concerning people's pasts and presents are mediated by numerous "third parties"—researchers, directors, producers, camerapeople, distributors, censorship organizations, and so on. Further constraints arise from intrinsic restrictions of popular forms of media: at the very least, experiences must fit a certain time slice or page count to be considered palatable to a mass audience, and therefore editing must occur. Together, we feel these factors degrade the full sense of awareness and appreciation we can achieve of other peoples and places, above cultural stereotypes and clichés.

The goal of the RAW project (which relates to its potential interest in the HCI design community) is to develop a new kind of recording tool, together with a method for processing and presenting the material captured with the tool, that would enable a more direct relationship between its user and the later audience, possibly in a far away place or time. We feel such a tool should allow users themselves to reflect more directly on their everyday lives, without any task or additional structure necessarily imposed upon them,

and to control better the content that would be conveyed to an audience than if, for instance, they were interviewed by another person or followed around by a production crew. We also feel the material gathered with such a tool should remain "raw" and unedited all the way from production to archive to presentation; hence the name of the project. Finally, we feel it important that such a tool be tested and considered relevant and valuable within a plurality of cultures in order for us to declare success.

We chose the African country of Mali as a starting point for thinking about the project because we feel this country has a particularly rich and diverse culture that is not well recognized or understood within Western societies [7][11]. Western accounts of African history often start with the colonization of the continent and end with a permanent state of war and famine. The latest significant news report about Mali recounts the kidnapping of travelers in Algeria that were kept for weeks in the Malian Sahara desert. A documentary on Mali would have a specific focus and dramaturgy whether on the children of the streets or the hairdressers of the Medine market in Bamako. An anthropological or ethnographic study would observe and analyze a tradition with a theoretical distance that might make it difficult to gain a sense of its immediacy and ordinariness within daily life—the surface of its content, which we feel can be as relevant and revealing as its depth.

In the rest of this paper, we describe RAW and how its design evolved through three phases of development: the conception of a prototype and an initial test in Dublin, a workshop with children/teenagers in Paris, and a series of experimentations on a larger scale in Mali, through which we discovered among our participants an emerging set of ways of exploiting the tool for different kinds of storytelling and documentation purposes.

DESIGN OF RAW

The RAW project consists of the following basic design features, the motivations for which are described in the sections that follow.

- The RAW tool (depicted in Figure 1) is an audiovisual recording device that combines a digital still camera and a high-quality stereo audio recorder.
- The tool records the 60 seconds of sound *before* and the 60 seconds of sound *after* a picture is taken. This is possible by having the tool record continuously and by storing only the 2 minutes of audio that surround each photograph.
- Audio is recorded binaurally using high-quality miniature microphones that are placed in the user's ears, just like "earbud" headphones.
- The audiovisual material captured with the RAW tool is archived in a raw form, with no deletion or modification allowed by the user or any third party.
- Presentations of material archived from a user's session with the RAW tool must adhere to a list of "guidelines for exhibition of RAW content" (discussed later) to aid in achieving the ideal of a minimally-mediated experience for the audience.

Why sound and still image?

We believe that video and motion pictures are not the final achievement in relating sound to image. In the domain of video-making we feel that audio is typically considered secondary—that audio is seen as supplementing the visuals, not the other way around. Yet we believe audio, and especially ambient audio, holds great potential for conveying certain kinds of impressions of everyday life with a richness not possible with visual media. We wanted to innovate in the field of audiovisual expression so that sound and image could have a different and more equal dialog than they have in the domain of video. The design of the RAW system aims to consider both sound and image of the same importance. The audio provides context to the image and the image provides a context for the audio.

There have been several experiments in the domain of "audiophotography" in recent years [8][9][14]. These primarily research or artistic efforts partly trace their roots to earlier projects such as Chris Marker's *La Jetée* (1962), a short film made up entirely of still images and a soundtrack.

We are not aware of projects that have specifically addressed the notion of capturing the moments before a photograph was taken. Depending on how RAW material is presented to its audience, the availability of that minute before makes it possible to add a dimension of mystery, set up by a sense of progressive discovery—wondering how the sound being listened to will converge to the moment depicted in the image. In photography theory, a recurrent subject is the question of what could have happened just before a picture was taken. The previous minute of sound captured by RAW gives some element of an answer and reveals the picture like the chemical development process,

Figure 1: Depiction of the RAW tool, with digital camera and ear-microphones for binaural audio recording.

while the following minute of sound haunts the user who is already forgetting about their last picture.

In this sense, the RAW tool differentiates itself from the few seconds of "audio caption" available in some new digital cameras. Sound is not an option with RAW—it is a core constituent of the recording made with the tool. A composition can treat sound and image equally or it can take on different emphases. The user might desire to capture a sound that piques his attention as the primary subject. Conversely, he might illustrate a scene he likes with contextual audio, whether it's his own narration or that of a stranger talking in the street, or just ambient noise.

Why one minute before and after?

The choice to record one minute of sound before and after each photo was based on a compromise between limited storage overhead and a desire to capture as full a context of each moment as possible. If the user takes a series of photos spaced less than two minutes apart, the audio is stored continuously and covers from the minute before to the minute after each such series. Future work might include investigating the affordances and constraints presented by different audio durations.

Why binaural recording?

The RAW tool uses a binaural recording apparatus that strives for the closest possible recording and reproduction of what the user of the tool is hearing while they are taking pictures. This design decision was made in an attempt to enable the later audience to immerse themselves "into the shoes" of the person who originated the content they are experiencing, and to place greater emphasis on the subjective point of view of this original source. Another advantage of situating the microphones in the user's ears is that they are a disguise (they look like miniature headphones and therefore don't draw undue attention) and the user does not have to handle any other device besides the camera once they start using the tool. The immersive effect of binaural audio recording is best experienced when wearing headphones, which significantly dictates the form that effective presentations of RAW records can take.

Why no editing?

With home video, the producer must review the footage shot, possibly multiple times, before choosing specific moments to be edited into a final movie. This process stands in stark contrast to that designed for RAW in which no editing is allowed. In effect, knowing this constraint in advance, the user is compelled to "edit" at the same time he is capturing material. The act of taking a picture is itself a selective process that marks a specific point of interest. The immediacy of these moments is enhanced because they are chosen in real-time, not while reviewing the material at a possibly much later time. After his experience with the tool, the user obtains an immediate result that doesn't need further processing. This result intrinsically possesses a story-like quality since each photo is enclosed by two minutes of contextual audio that, together with the image, create a natural rise in suspense, a climax, and a resolution.

Why have presentation guidelines?

Because we felt there were many different ways one might want to be able to present RAW records to audiences (on the web, interactive museum installations, mobile devices, to name a few), we felt it important to create a set of guidelines that exhibitions of such records should obey in order to maintain the spirit of a minimally-mediated relationship between the user and the audience on the presentation end of the pipeline. (Of course, there is no such thing as a completely unmediated experience.)

- No modification of, deletion of, or favoritism toward any of the content is allowed by the presentation apparatus itself or by any other party acting between the capture and display of the material.
- By default, the apparatus should present the photographs together with their contextual audio in the order in which they were originally captured.
- All the photographs and audio captured in a particular user's session with the tool must be accessible to the audience. If the presentation is not interactive, then all photographs and audio that are part of a single session must be presented.
- If any portion of an image or its audio bit appears, then they must appear at some point in their entirety.
- No photograph may be presented without its corresponding audio, and no audio may be presented without its corresponding photograph.

These presentation rules were instrumental in helping to guide the design of our primary exhibition apparatus, which evolved in form through the three phases of qualitative study we undertook, described in the next sections.

PHASE 1: INITIAL PROTOTYPES IN IRELAND

Between February and May 2003, in Dublin, Ireland, we carried out a series of tests to inform the development of the technology as well as to validate the overall concept. We began by trying to develop a tool that would enable capturing impressions of "a day in the life of" the participant, and therefore we needed a technology that could sustain being used for 24 hours, without the need to recharge the batteries or empty the storage space.

Our very first prototype combined, in a small rectangular cardboard box, a miniature digital camera and a MiniDisc (MD) recorder that connected to the microphones worn in the user's ears. As a way of getting started, in this prototype, the camera was wired to trigger the MD recorder to commence recording sound only *after* a picture was taken. This early tool was tested by five people who each used it for about an hour, and they all encountered similar problems: The box didn't provide the user enough control over what they were taking the picture of, the box itself was

Figure 2: Depiction of the progressive discovery aspect of the RAW presentation prototype. While the two minutes of audio plays, the image appears first on a point of detail and slowly zooms out to its full size, and then fades away.

clumsy to handle and drew attention, and the quality of the resulting images was not felt to be satisfactory.

However, there were some positive results: Users reported feeling comfortable with the one minute audio constraint and in fact treated it more like a guideline than a hard rule because they could easily reduce or extend the duration by taking pictures more or less often. In addition, we discovered that a more time-limited experience with the tool ("an hour in the life of") could be as interesting as the 24-hour concept, especially since there seemed to be some reluctance to the idea of wearing the ear-microphones for a very long time.

Our development branched at this point—we continued working to articulate a 24-hour version of the tool, but we also quickly created an 80-minute version that simply consists of a higher-quality camera that is synchronized with an MD recorder, worn separately on the waist or in a pocket. The MD records continuously for its maximum 80 minute duration. Later, a piece of software extracts the minute of sound before and after each photograph was taken, using the image timestamps that are recorded automatically by the digital camera. Tests performed with this new version of the tool resulted in much better feedback, and all of our study participants from this point forward used this version of the tool.

Initial presentation prototypes

In this first phase of work we also prototyped and tested several different styles of exhibiting RAW records to an audience. We were in search of designs that would provide an immersive and engaging experience while also respecting the project's ideal of minimal mediation. This was a difficult task, as the idea of presentation itself somewhat implies that there is mediation or editing happening. Even in the simplest photography exhibition, the curator must make editorial decisions that will affect the perception of the content by its audience. The architectural quality of the exhibition space, the way the pictures are hung on a wall, even the wall itself all affect this perception.

We conducted a two-day workshop in which we presented 4 different presentation system prototypes to 7 participants who acted as an audience in an exhibition space. Each solution consisted of different degrees of interactivity and different representations of the media (printed and hanging on a wall or video-projected, audio played in headphones or on speakers). Most participants attended both days, and could compare all the interfaces. Others gave feedback about only one interface. The participants completed a questionnaire and we discussed their answers collectively. The questionnaire focused on their understanding of the relationship between the audio and visual content, and on their overall feelings about and interest in the experience.

This feedback helped us realize the need for some staging of the content, so the audience could better understand the rhythm underlying the capture process. Some participants wanted more control on the progression of the presentation, so they could jump forward and backward in time. This was later the main motivation for us to incorporate an interactive timeline in the interface.

The idea of a progressive discovery of the image was also inspired by their remarks and included in the presentation system from this point forward. In this design (depicted in Figure 2), when the first minute of sound begins playing, the image appears and slowly begins to zoom out from a point of detail, reaching its full size at the moment in the audio that the user actually took the picture. Because we left the audio signals of the camera turned on, the audience can hear the sound of the camera taking the picture at this precise moment. Then, while the second minute of plays, the image slowly fades away to black.

Having the image initially appear on a zoomed-in portion necessitated that we add an extra step in the capture process in which the user of the tool could select this point of detail. This weakened the ideal of all annotation occurring in "real time" during the act of capture instead of in a later review. But this weakness was seen as minimal and acceptable for the moment, as it was a quick task that didn't require much thought and something that could be easily piggy-backed into a review of the material that almost all of our participants wanted to do anyway immediately after their session using the tool.

PHASE 2: WORKSHOP IN FRANCE

We conducted a 10-day workshop in Paris in June 2003, in cooperation with an organization that helps young students with their homework. We worked on an individual basis with 2 girls and 3 boys, aged 11-13. We were hoping to gain general feedback on the potential value of the tool to a younger age group, as well as to see what new languages or

behaviors might emerge from their use of the tool in a different cultural context.

After some discussions and a test run to get acquainted with the camera, each participant was left to take the tool out for up to an hour. We chose not to show them any of the records captured in Dublin so as not to corrupt their own original thinking about what they could possibly capture or document with the tool. They were not given any specific editorial direction except that when we talked with them before their session, we tried to help them find a particular theme they might want to explore, mainly focused on the district they were living in. Upon returning, each briefly reviewed their material and performed the point-of-detail selection described in the last section while we interviewed them informally about their experience.

The results of these sessions were encouraging. Each child's record had an distinctive individual voice, an observation that pleased us as well as their mentors at the organization who said they normally do not act or express themselves outside of the mentality of their social group. One child captured impressions of things inside her home, another captured construction sites and mainly architectural features of his district, and another focused almost exclusively on random people he met on the street while walking around. Most of our participants seemed most comfortable taking the tool out together with another friend who typically walked beside and had conversations with them, all of which were captured on the audio track.

On the last evening of the workshop, we set up a public exhibition using our latest presentation apparatus, which included the progressive discovery aspect described previously, a pair of high quality headphones for listening to the binaural audio recording, and, as an experiment, a large "physical scrollbar" interface that enabled sliding backward and forward in the timeline of photos captured in one user's session. This physical slider, while impressive and fun to use by the attendees, was difficult to transport and calibrate. As an interface concept, its greatest flaw was that the slider itself could not move on its own to convey the "current position" being viewed in the presentation.

Other young people, many of them friends of the participants, attended the exhibition. We were gratified to observe several attendees, who upon entering the door appeared to be high-energy beings with the lowest possible attention spans, stop and put on the headphones for several minutes at a time, interacting and changing the image only very infrequently. Something clearly caught and held their attention longer than we were expecting. When asked informally what they liked about the experience, we received answers relating to interest in the documentation produced by one's friend, especially in the sound track, as well as interest in the sound and feeling like they were in the shoes of that friend. Although it would be wrong to generalize from such informal feedback, we felt encouraged that we were on the right track to creating the sense of immersion that we were aiming for. One boy in the audience, aged 12, asked if he could use the RAW tool as well, which he did the same evening.

This workshop prepared us to a great extent for the one we undertook in Mali a couple of months later. We gained experience in what things were easy or difficult to understand in our introduction of the general concept and the use of the tool, and also about what questions or expectations our participants might have concerning the future use of their material.

PHASE 3: THE MALI EXPERIENCE

We conducted a larger scale study over three weeks in August 2003 in three locations in Mali: Bamako, Timbuktu and Ségou. Our primary goals were similar to those we had in Paris: to observe the ways people here would use the tool for different purposes, and to assess the value and relevance the tool potentially has in this culture that is very different from the one in which it was designed.

Participants

We selected neither a focused nor an exhaustive panel of participants. Although we arrived possessing a small number of contacts obtained through prior interactions in Paris, we worked in a more opportunistic manner, finding participants through local contacts or people we got to know in each place.

We worked with a total of 23 persons from different age groups: one 10-year-old child, 2 older teenagers, 4 people in their twenties, 6 in their thirties, 7 in their forties, and 3 persons over 50. Only five of our participants were women. Originally we had hoped to have an equal number of women participate as men. The reasons for this low percentage stem not from any reluctance on the part of women to participate but rather from our opportunistic mode of operation, following chains of ad-hoc contacts that for one reason or another got us in touch with more men. Only two people we approached refused participation: one was a man who reported not having enough time, and the other a woman who didn't give any specific reasons.

Most of the participants we worked with spoke French in addition to their native Malian dialect. When it wasn't possible to communicate in French, our guides would act as translators. The occupations of our participants spanned a wide range: academics, officials, artists, musicians, craftspeople, students, shepherds, and others. We had a large number of participants involved in creative professions, a profile that does in fact reflect a larger reality in the country. Several of our participants had never before taken a picture with any kind of camera.

Method

Our strategy was inspired from the workshop we undertook in Paris. For each participant, we took the time to introduce the project in a discussion. We made a specific point of telling them we were not anthropologists coming to their

Figure 3: Participants getting familiar with the RAW tool (left) and performing the point-of-detail selection (right).

country to document aspects of their culture. Rather, we described ourselves as technology and design researchers on a field study, hoping to investigate how a variety of different kinds of people in different cultures use a new tool that we are developing.

After they agreed to participate, we demonstrated how to use the tool and helped each participant take a couple of test shots until they felt comfortable to take it out on their own for up to an hour or so. Like in Paris, we avoided showing them presentations of previous participants' RAW records so as not to affect their own original inspiration on what they might capture. Similarly, we did not provide specific tasks or themes to work around. If the participant wished to discuss possible themes, we let them suggest ideas on their own. At most, we would suggest the general theme capturing aspects of one's everyday life as a starting point.

We left each participant for a couple of hours and then returned to retrieve the device and discuss the experience with them. We would also load the images onto a laptop computer for them to review and perform the selection of the point of detail (Figure 3). We recorded each participant's contact details so that we could later send them both hardcopies of their photographs as well as a CDROM containing the digital images and audio (there are Internet cafés in most cities where these could be viewed).

Results

We ran into a variety of different problems at different times, ranging from batteries running out, to digital camera cards becoming full, to fingers appearing in pictures, to audio quality being compromised for two of our participants. However, a more serious problem was that in 8 cases, the person who used the tool was not able to do the selection of the point of detail, either because they were not available or because there was no time (once we had to leave a location in a hurry to catch transportation to our next destination). In these eight cases, in order that they could be exhibited along with the other records from Mali in the same system, we selected points of detail in the images on our own, and we specifically declare this deviation within the presentation.

However, these various difficulties were outweighed by the actual material produced by our participants which was inspirational to us in its originality and immediacy, beyond what we had experienced of the culture through other forms of media. Each one of our participants captured perspectives on their lives that we felt we clearly could not have matched in richness if we were acting on our own as photographers or documenters of some kind. The average length of use was around one hour, though one person used the tool for only 5 minutes and several others exhausted the full 80-minute available duration.

Emerging styles of use

The most significant result of our experience in Mali is that we began to observe some clear categories for how our participants exploited the RAW tool for different kinds of capture or storytelling purposes. These span a range from personal reflection to more outward styles of engagement, with either a passive or active stance toward the later audience. They emerged despite the care we took to not suggest any particular styles or themes to our participants in our initial discussions with them. The groupings, which are certainly subject to revision as more experience is gained, also relate to and build upon behaviors we observed in the Paris workshop.

Type 1: Social glances

This category represents uses of the tool that occur primarily in a social mode, or in which social contacts and spontaneous encounters are the primary content underlying the audio and visual media captured by the user. Many of our participants used the tool as a means to strike up conversations with people in their workplace, at home, or on the street. Or conversely, sometimes friends of the user would be curious about what he is doing and interrupt him during his session, resulting in a social exchange. No particular audience is addressed by the user. The relevant RAW records convey a rich impression of the social fabric and relationships that exist in the society, and they are also the records in which we hear the greatest variety of spoken languages.

Type 2: Caught in activities

Some people chose simply to perform their everyday tasks or livelihoods and capture impressions of them in more of an individual mode. Again, no specific audience is actively addressed, but the user is aware that these moments are indeed a kind of "performance" that will be experienced by an audience at a later point in time. Musicians were the most likely to share their daily experiences in this way, often capturing themselves playing in a jam or rehearsal session. One woman recorded her daily journey to obtain water from a community well in this fashion. Another woman who dyes fabrics for a living used the tool to capture glimpses of the processes employed in her profession, including final models as she depicted with photographs of other photographs from her portfolio.

Type 3: Active documentation

In this category, the user of the tool actively addresses the eventual audience of the record he is creating with spoken

narration or even live interviews with people he encounters, as a way of documenting some aspect of his everyday life or his society. There may be a specific theme, determined in advance, or a looser structure based on spontaneous encounters with interesting scenes or personalities. In some cases, the user clearly had a foreign audience in mind that would wonder what life in their Malian city is all about. Hence, the language most often used in this context was French, sometimes even English. One participant walked around Timbuktu, interviewing people along the way about the development of the city infrastructure. Another provided narration about the various aspects of a Muslim baptism ceremony while he was capturing them.

Type 4: Intentional discourses

This group refers to exploitations of the tool in which the user has a very specific message or commentary that they wish to relate to the audience. The user's speech is not intended as a narration of the audio or visual moments they are capturing, but rather it is the other way around—the pictures and sound are an augmentation of the commentary. The user may operate alone with a more "political" or "activist" perspective, and the record might take the form of a monologue or spoken contemplation. For example, one participant used the tool to convey specific thoughts concerning the development and position of academia in Malian society, which he illustrated with loosely related audio and visual impressions from his university.

Exhibition in Mali

We mounted an exhibition of the material captured by our participants in Mali on the night before we left the country. We iterated on the presentation design by adding a "timeline" at the bottom of the screen (Figure 4). The beginning and end of this line represent the start and end times of the user's session, and each photograph is represented by a small square. The distance between each square represents proportionally the length of time between each picture. The timeline is interactive and allows an audience member to click and jump from one photograph to another. A separate menu allowed the audience to change to any of the 23 complete records captured during the study.

About 50 people attended, several of whom were our participants and whose records were accessible in the exhibition. The exhibition was projected on a wall so that several people could view it at a time. We set up a pair of small external speakers so that others in the room could listen to some of the audio track while the headphones were being passed around so that everyone could have a chance to experience the binaural audio recording (Figure 5). Clearly, we would want to have multiple pairs of headphones available in future exhibitions.

The informal feedback we received from attendees of this exhibition was similar to what we received in Paris. The handful of our participants who attended and who we were able to speak with all felt satisfied with the record they

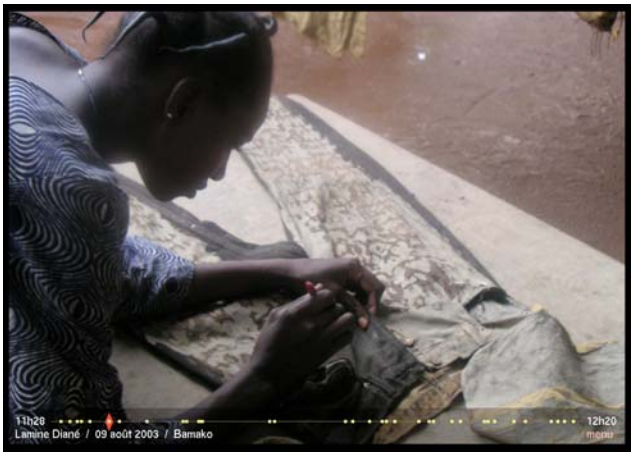

Figure 4: Screenshot of exhibition with timeline interface.

Figure 5: Attendees viewing the exhibition in Bamako.

produced. However, they had nothing else to compare it with, and they may have also been saying this because they thought it was what we wanted to hear. But our purpose was not to compare the qualities of different recording tools against that of RAW—this could be the subject of a future study. Rather we wanted to assess the relevance of our tool in a Malian context and to see what uses different people would invent for it, as a way of driving the next phases of its development. In the end, we felt the Malian participants had a good natural understanding of the audio aspect of the tool and how it can be used in coordination with the image. This perhaps relates to Mali being viewed as more of an "oral" culture, in which the spoken word takes precedence over written or visual expression, which are more a focus of the other two Western cultures we worked within.

GENERAL OBSERVATIONS

One of the biggest difficulties that we ran into in operating in different cultural contexts was that people did not always perceive the reasons for our presence in a way we would have liked. This was especially true in Mali, where, for a variety of historical reasons, assumptions were sometimes held about what the motivations were for a white person to work there. We were variously tagged as anthropologists, social workers, eco-tourists, and so on—perceptions we spent significant effort trying to break down. The importance of pre-planning and working with the highest quality of contacts and guides cannot be stressed enough, as these figures were instrumental in helping us to break the ice with the locals, to address questions about our project

and our motivations, and to understand the local customs so as to avoid tension or embarrassment.

We see the main advantage of undertaking studies in multiple cultural contexts as being the opportunity to discover behaviors and gain feedback that can potentially guide the development of a prototype such that it will be understood and valued by a much larger segment of the world population than it might have been otherwise. For RAW, this kind of study was particularly important, as we saw the tool as something that could possibly be used as a way of capturing a new kind of portrait or historical record of a culture. Everyday life happens everywhere, and interest in it comes from many directions and often grows in the years after it happens. Thus we wished to articulate a tool and process that could be considered meaningful as broadly as possible. Of course, we have only touched on three particular contexts thus far in our work, but these have already yielded insights that we felt were significant to us and perhaps to others as well.

RELATED WORK

Earlier we mentioned the existence of some other experiments in the realm of audiophotography. Frohlich, Tallyn, and Adams have highlighted the value of ambient sound recording, as opposed to "voice labeling," as a way of enhancing photographs [8][9]. Martin and Gaver took these ideas further by creating designs for potential products [14]. Balabanovic, Chu, and Wolff experiment with adding audio narration to photographs at a later time [3]. These projects, though, treat audio mainly as an "augmentation" of the photograph rather than a media of equal status.

In the realm of sound recording: Some researchers have investigated the advantages of recording "speech interactions in everyday work environments" [10]. Projects like *SpeechSkimmer* [2] develop further notions of capturing and retrieving sound in everyday situations. *Dynamic Soundscape* [12] and other projects [4][6] particularly focus on the affordances of spatialized audio. Other work, such as *Transparent Hearing* [15], investigates binaural sound recording in interactive applications.

In the realm of everyday life and mediation: Brown, Sellen, and O'Hara describe the benefits of people documenting elements of their work life on a daily basis [5]. Ananny and Strohecker's *Citizen Journalism* builds an understanding of the aspiration of people to be their own photojournalists [1]. Along similar lines, Srinivasan's *Village Voice* provides a portal for a Somalian refugee community in Boston to share with the community self-produced video documents about their cultural heritage [16]. Makkuni's *The Crossing*, another project with a cross-cultural dimension, is a reflection on the use of ICT in developing countries and how the consideration of their cultures could "shape new forms of computing technology" [13]. These projects vary greatly from RAW in the types of media employed and in the amount of editing or third party mediation incorporated into their processes.

ACKNOWLEDGMENTS

We wish to thank everyone who participated in our studies and who provided feedback on this paper. This work has been supported by partners of Media Lab Europe.

REFERENCES

1. Ananny, M. and Strohecker C. Sustained, open dialogue with citizen photojournalism. *Proc. DYD 2002*.
2. Arons, B. SpeechSkimmer: a system for interactively skimming recorded speech. *ACM Trans. on Computer-Human Interaction 4,1* (1997), 3-38.
3. Balabanovic, M., Chu, L. L. and Wolff, G. J., Storytelling with digital photographs. *Proc. CHI 2000*, 564-571.
4. Baldis J. J. Effects of spatial audio on memory, comprehension, and preference during desktop conferences. *Proc. CHI 2001*, 166-173.
5. Brown, B. A. T., Sellen A. J. and O'Hara K. P. A diary study of information capture in working life. *Proc. CHI 2000*, 438-445.
6. Burgess, D. A. Techniques for low cost spatial audio. *Proc. UIST 1992,* 53-59.
7. Colleyn, J. P. and Diawara, M. *Mali Kow*. Indigène Editions. Montpellier, 2001.
8. Frohlich, D. and Tallyn, E. Audiophotography: practice and prospects. *Ext. Abstracts CHI 1999*, ACM Press (1999), 296-297.
9. Frohlich D., Adams G. and Tallyn, E. Augmenting photographs with audio. *Personal and ubiquitous computing. 4, 4* (2000), 205-208.
10. Hindus, D., Schmandt C. and Horner C., Capturing, structuring and representing ubiquitous audio. *ACM Trans. on Information Systems 11, 4* (1993), 376-400.
11. Ki-Zerbo, J. and Niané, D. T. (dir). *Histoire générale de l'Afrique*. Vol. 4. UNESCO, Paris, 1991.
12. Kobayashi, M. and Schmandt, C. Dynamic Soundscape: mapping time to space for audio browsing. *Proc. CHI 1997*, 194-201.
13. Makkuni, R. Culture as a driver of innovation. *Proc. ICHIM 2003*.
14. Martin H. and Gaver B. Beyond the snapshot from speculation to prototypes in audiophotography. *Proc. DIS 2000*, 55-65.
15. Mueller, F. and Karau, M. Transparent Hearing. *Ext. Abstracts CHI 2002*, ACM Press (2002), 730-731.
16. Srinivasan, R. Village voice: a methodology, interface and evaluation of the collection of cultural heritage material. *Proc. ICHIM 2003.*