6

Conclusions[1]

During the past 20 years, authoring has been part of the multimedia community's research agenda. Unfortunately, multimedia authoring has been seen as an initial enterprise that occurs before 'real' content processing takes place. This limits the options open to authors and to viewers of rich multimedia content in creating and receiving focused, highly personal media presentations. This thesis reflects on the multimedia authoring workflow and we argue that a fresh new look is required. We focused on the particular task of supporting *socially-aware multimedia authoring*, in which the relationships within particular social groups among authors and viewers can be exploited to create highly personal media experiences. Our framework is centered on empowering users in telling stories and commenting on personal media artifacts, considering the long-term social context of the user's social environment. We provided an overview of the requirements and characteristics of socially-aware multimedia authoring within the context of exploiting community content. In particular, our research involved the study of different mechanisms to allow users to explore, create, enrich and share videos based on personal relationships. Our methodology integrated knowledge from Human-Computer Interaction (e.g., focus groups/interviews for need assessment, iterative prototyping and user evaluation) and document engineering.

[1] This chapter contains extracts from the following article:

D.C.A. Bulterman, P. Cesar and R.L. Guimarães. 2013. Socially-Aware Multimedia Authoring: Past, Present and Future. ACM Transactions on Multimedia Computing, Communications and Applications (TOMCCAP), Volume 9, Issue 1s, Article 35 (October 2013), 23 pages. DOI=10.1145/2491893 http://doi.acm.org/10.1145/2491893

In this chapter we first revisit and answer the research questions of the thesis. We then reflect on the lessons learned, before concluding with a discussion of the issues that we feel can provide a fruitful basis for future multimedia authoring support. We argue that providing support for socially-aware multimedia authoring can have a profound impact on the nature and architecture of the entire multimedia information processing pipeline.

6.1 Revisiting the Research Questions

Much of the media landscape has been, and continues to be, dominated by commercially produced content. Whether image, video, audio or (to a lessor extent) text, users today have become accustomed to experiencing highly polished media messages. In spite of the dramatic impact of user contributed content sites (such as YouTube and Facebook), the amount of personal content being shared with family and friends (to say nothing of wide anonymous audiences) is minimal. A conservative estimate of media use indicates that average owners of smartphones and portable cameras capture hours of videos yearly, but that only minutes (or seconds) of content are being shared. Does this mean that user-generated content is less important? No. Personal archives have a high degree of personal value: photos of family and friends, videos of small children, audio fragments that capture the sounds of people who have played an important role in one's life. Although there may always be exceptions, it is clear, that a short video showing a child's first violin solo will not attract the same audience as, say, a slickly-produced commercial music video. This does not make the violin fragment less valuable.

In this thesis, we focused on community authoring applications, where content is contributed from many amateur sources and distributed within a relatively closed circle of viewers who have varying degrees of affinity with the content produced. We concentrated on support for situations in which both the original presentation creator and the presentation viewer play a role in determining presentation content. Given this context, we discuss and answer each of the research questions according to the work presented in the bulk of the thesis.

Question 1.1 *Can a socially-aware multimedia authoring system be defined in terms of existing social science theories and human-centered processes, and if so, which?*

In this thesis we reformulated the research problem of multimedia authoring, by investigating mechanisms and principles for togetherness and social connectivity around personal media. Our focus was on parents, family members and friends of students participating in a small-scale social event. In this scenario, parents capture recordings of their children for later viewing and possible sharing with friends and relatives. Based on a 4-year evaluation process, we specified a set of guidelines for the design and implementation of socially-aware multimedia authoring and sharing tools. We aim at nurturing strong ties and improving social connectedness by supporting *emotional intensity*, *personal effort*, and by supporting *intimacy* and enabling *reciprocity*. With these guidelines we intend to increase the feeling of connectedness, particularly among family members and friends who could not attend the social event. As shown in Chapter 2, our socially-aware multimedia authoring paradigm is aligned with the requirements needed for social communities that are not addressed by existing social media Web applications. These guidelines directly address research *Question 1.1*.

Question 1.2 *Does the functionality provided by a socially-aware multimedia authoring system provide an identifiable improvement over traditional authoring and sharing solutions? If so, how can these improvements be validated?*

To evaluate the utility and usefulness of socially-aware multimedia authoring, we realized the guidelines mentioned above in a two-phased prototype system called MyVideos. We have actively participated in the design, implementation and integration of this system, and our contributions enabled us to perform extensive field trials and these were a major part of the TA2's success[2]. Working with a test group at local high schools in two different countries (UK and the Netherlands), we investigated how focused content can be extracted from a shared repository, and how content can be enhanced and tailored to form the basis of a personalized multimedia artifact, that can be eventually transferred and shared with family and friends (each with different degrees of connectedness with the performer and his/her parents). Results from a long-term evaluation process show that all our participants (from phase 1 and 2) liked the functionality provided by our system and considered this a valid alternative to strength social interactions when apart. Therefore, using our system they would share more videos with friends

[2] The pan-European Project Together Anywhere, Together Anytime – http://ta2-project.eu.

and family. These results – complemented by more specific findings on media exploration, creation of personal memories and content enrichment (Chapters 3-5) – directly answer research *Question 1.2*.

Question 1.3 *Does a socially-aware video exploration system provide an identifiable improvement over current approaches for accessing and navigating a repository of shared media?*

While following the *emotional intensity* guideline, in Chapter 3 we discussed a two-phased design, development and experimentation of an interface for browsing a collection of user-generated videos from a shared event. Users could explore and navigate (fragments of) video clips recorded by several people based on their own personal/social interests. The design, deployment and evaluation of the system resulted in the identification of key requirements for this novel type of browsing interfaces. In particular, our approach 1) supports exploration based on the inherent event structure; 2) it makes use of contextual information to help in the navigation process; 3) it allows for flexible searches based on a combination of filters; and finally, 4) it provides a way to switch between cameras angles that might have captured different aspects of the event. Results of the evaluation process show that all participants appreciated the browsing interface and indicated that it is better than traditional tools to explore videos they care about. Therefore, they would find videos more efficiently using our system. These results clearly indicate that a socially-aware video exploration system like ours provides an improvement over current tools for accessing and navigating a repository of shared media assets, directly answering research *Question 1.3*.

Question 1.4 *Where is the balance between automatic and manual processes when authoring personalized narratives users care about?*

As for browsing a shared video collection, social relationships are key for authoring personalized stories users care about. In Chapter 4 we reported on our efforts to support the creation of personalized video stories reusing collective content. We developed a first version of an authoring system, subjected it to user testing, and then developed an improved version that follows the *personal effort* guideline of socially-aware multimedia authoring. Our initial results showed a general enthusiasm from participants, which were validated in the first evaluation phase. While the video compilations automatically produced by the initial system

were considered visually compelling, users missed the capability of personalizing those by adding their own 'imprint'. To address this limitation, we proposed a hybrid authoring approach that provides mixed support for automated creation by selecting content of personal interest and manual enhancement of personalized video stories. Based on user feedback as part of our four-year study, we have demonstrated that it is possible to satisfy casual content creators while still allowing extensive personalization to take place, if needed. These results directly answer research *Question 1.4*. We believe that the combination of automatic and manual processes provides the balance of complexity and functionality.

Question 1.5 *Does the support for timed end-user commenting within pre-authored narratives provide an identifiable improvement over current media commenting approaches?*

While concentrating in the creation process, we cannot forget that content enrichment also plays an important role in the socially-aware multimedia authoring workflow. Motivated by a survey on media watching and commenting practices, in Chapter 5 we reported on the design, implementation and user-centric evaluation of a video commenting framework that follows the *intimacy* and *reciprocity* guidelines. To realize such framework, we specified and described a set of temporal transformations for multimedia documents. Our approach allows end-users to create and share personalized timed text comments within third-party online videos. The benefit over current solutions lays in the usage of a rich commenting format – in our case SMIL [17] – which is not embedded into a specific video encoding format. The evaluation of a video commenting system that realizes our framework clearly indicates that participants appreciated our system (13/18 or 72% of the participants), and considered it helpful (100%). Our results also show that 50% of the participants considered our video commenting approach better than the one offered in YouTube and/or Facebook. These results show that our commenting framework brings a measurable increment over existing commenting systems, and directly answer research *Question 1.5*.

6.2 Reflection and Further Directions

In this thesis we provide useful insights into how a socially-aware multimedia authoring and sharing system should be designed and architected, for helping users

in recalling personal memories and in nurturing their close circle relationships. The main contribution of our work does not lie in the use of a specific technology (e.g., SMIL, NSL or Web standards) but in further understanding the fundamental trade-offs that enable better sharing of 'personal' media. Results from our evaluation process show that socially-aware multimedia authoring provides a more fruitful approach than earlier work.

Although our research has reached its aims, there are some unavoidable limitations. First, the total amount of time spent to annotate the footage (see in Chapter 2) demonstrates that this is still a very challenging problem, especially when we consider dimly lit user-generated content with different quality, encoding etc. Although these annotations are essential in our authoring framework, this thesis does not aim at solving this problem.

As to the number of subjects participating in the evaluation, we agree that 'more is more', but note: each subject needed to agree to spend some hours per evaluation (about 1h30min recording concerts plus 2h in lab studies). We found it difficult to find high school parents who would commit to this load. We are pleased that our parents – about 25% of the concert participants! – were motivated contribute this block of time. The goals of the study make it impossible to do crowdsource testing, given the focus on common personal media. Moreover, we are not aware of other studies that provide the same breadth.

Another limitation could be that we focused on a particular use case scenario. We reiterate that our participants represent a realistic sample of users: actual family members from 2 countries (NL and UK) that have been involved in the concert recordings and prototype evaluation. We agree that generalization to other events is an important problem, but before getting there we need to start somewhere. We see this as a topic for future work.

Providing support for socially-aware multimedia will significantly impact the support required for effective encoding, storage, classification, selection, transmission, protection and sharing of (potentially composite) media artifacts. The principal reason for this is that the context in which media is used will strongly determine how it is classified and accessed. Annotations and metadata will become multifaceted and dynamic, and will be determined by use rather than by design. In the following subsections we highlight some opportunities for future research in socially-aware multimedia.

6.2.1 Media Encoding and Storage

At present, media encoding is based on an agnostic view of content. This has been used to great advantage on sharing Websites and physical distribution media. The assumption has been, however, that all of the fragments related to a single story are compressed into a single fixed media object. There are usually no facilities for packaging custom versions of content from a single base encoding. Each personal version of a video (or video fragment) must be re-encoded in a new document.

One important difference required to support end-user composition is that small logical groups of media would be stored on several servers, each as individual fragments. These fragments could be mixed/matched dynamically at viewing time to support the interests of the viewer. In terms of our school concert example, this would mean that all of the individual assets captured by all of the parents could be saved in a cloud over servers. Individual presentations could then be stitched together on demand.

Having a logical media object be constructed out of dynamically combined physical fragments allows customized navigation to be supported. In YouTube (as in other commercial video sharing systems), dynamic mashups are not supported. End-users have to find suitable source material, cut it into shots, and assemble an encoded final video. While this solution does not impose hard requirements on delivery and rendering, it is limited in terms of adaptability, user interaction and seamless playback [36].

One approach to implementing such dynamic combination is supported by DASH (*Dynamic Adaptive Streaming over HTTP*), a system for HTTP-based streaming [70]. Although some efforts have investigated the use of DASH with Rich Media services [9][10], at present, it is typically used for storing pre-defined fragment encodings, nearly always based on support bitrate-adaptive resolutions (During presentation, the quality of the content can be adjusted based on environmental factors such as available bandwidth or end-user screen size). Similarly, dynamic media compositions could be achieved using a combination of HTML5 and W3C Media Fragments [21] and/or JavaScript code (e.g., Popcorn.js or Kaltura Video Platform).

Adaptivity in our work can leverage this support, but our main interests are in supporting a more abstract form of content selection: providing more trombone content to the father of the trombone player and more clarinet content to the mother of the clarinetist. This is a matter of dynamic content selection rather than (or at least in addition to) dynamic encoding selection. The selection (or generation) of

dynamic content requires more illusive criteria for content selection, such as a profile of the viewer in addition to profiles of the available content, and a content-wide temporal model that exposes logical divergence and convergence points for creating content streams. It also requires a container format that allows differential segment length to exist across candidate segments. To support this, the current model of content streaming would need to be revised: the seamless integration of individual content fragments (as opposed to encoding fragments) into a logical whole is a composition concept that most media servers and media container languages are as-of-yet ill-equipped to support.

6.2.2 Media Classification and Annotation

Personal media classification and annotation remains a challenge for supporting effective content sharing. For professional content, content is often highly segmented along the lines of established commercial distribution models. For personal content, the situation is vastly different. This shift in emphasis is new for multimedia, but there are many established examples in music, art and literature where the intentions of the composer, artist or writer are decoupled from the applications of the media itself.

At present, personal content annotation is driven by device-supplied metadata (e.g., clocks, location coordinates, file names, as well as objects and faces). For socially-aware multimedia, it is also necessary to encode relative social relationships among interested parties – plus to maintain those relationships over time. As with any large software system, the long-term maintenance costs of media will dominate the short-term development costs. This will require a new generation of iterative, socially-aware media classification tools. The analysis of content becomes then a continuing task, not an import activity. In the same vein, content recommendation needs to not only use such information, but also be sensitive to the context of use: are you watching alone, with your spouse, with your children, with your friends?

6.2.3 Customized Media Selection

Perhaps the most significant innovation in (broadcast) content selection occurred with the introduction of the video tape recorder. For the first time in history, it was the viewer that determined when content would be watched – on the precondition that it had been broadcast and recorded earlier. A next, but more minor, innovation

came with the introduction of the digital set-top box, which included an embedded program guide, providing the opportunity for more automated content selection and recording. The next logical development is to remove the TV guide altogether and to have the system itself recommend content for the family, which it found based on metadata encoded by the content providers.

One drawback of many home content systems is that a set-top box is typically not aware of who is actually watching TV. Some form of personalization is supported, but at a fairly impersonal level. At present, much research is being expended on recommender system technology. These systems depend heavily on producer-generated metadata for determining available candidate content. For socially-aware multimedia, the granularity of the metadata needs to be refocused to personal content. Another change in focus is that content selection will need to move from selecting 'programs' to selecting fragments of content. For a given viewing experience, several fragments typically will need to be dynamically combined to support end-user engagement.

6.2.4 Content-Based Navigation

One of the challenges with temporal searching along a timeline is that it is a highly unstructured activity. For instance, in a conventional YouTube interface for navigating through a video object, users can only select key frames without any higher-level narrative guidance. We note that even 1980's generation DVD technology provided more significant control through its chaptering interface. In general, the time axis provides no information on the logical structuring of the event, letting alone the performers in the concert or their relationships. Still, in the absence of any semantic structuring of content, it is often all that is available.

It will be necessary to study new mechanisms to replace timeline searching with navigation based on an overlay of structure components. One approach to provide this structure in our school concert use case is based on graphs of performers, instruments, songs or solo's. It could also be based on cinematographic classifications, such as long shots, pan shots, tight shots.

6.2.5 Ownership and Digital Rights

Reusing content brings with it questions of ownership. In printed documents, this is a solved problem: even though the base content is copyright protected, there is a

clear distinction between 'my' media and that of the original authors. For web pages and online content, the relationship is less simple.

If transparent sheets had been placed between all of the pages, we could take all of the user's comments and distribute them as separate items – all fully within current law. The content added could be further aggregated with the context created across a social network (or across the Internet), and analyzed. What are the most marked-up pages in the book? Does these represent the most interesting or most unclear sections of text? Do the markup patterns change over time? Which comments are appropriate for which users?

When annotating a piece of media – whether it is text, audio, images, or whatever – the implication has been that the annotations are of a highly personal nature. Of course, if many of these personal notes are collected and analyzed, they could provide valuable insights into the reusability of personal media assets. Even a simple density analysis of multiple media annotations could provide interesting clues for socially-aware recommender systems.

6.2.6 Security and Privacy Concerns

Content can be used or misused by various members of a user community, depending on their (possibly time-variant) relationships. Research is required to support content access and content protection that reflect time-variant social and personal relationships.

One aspect of security and privacy of socially-aware multimedia is that personal metadata will likely become too sensitive to simply place on a third-party storage system (like Facebook or *Google*): all of us will want to take back our identity and maintain our own control of our life-long information. This will require convenient interfaces. It will also probably require users to become accustomed to paying for media access and sharing services.

6.3 Closing Thoughts

Much has changed in the 'world' of multimedia. Who would have expected twenty years ago that within two decades, it would be commonplace to not only listen to music via your computer, but buy it there as well? That books would not only be written on a PC, but that the PC and its technological 'cousins' would become a handy way to read them, or to have them read aloud. That the computer would

threaten to replace not only the television, but also the movie theatre as a venue for the shared watching of content. And, perhaps more significantly in the long term, that the computer would not only render a wide range of real and artificial images, but that it would attempt to understand them as well.

In this thesis we have outlined what we mean by *socially-aware multimedia*. We have argued that the impact of supporting user-in-the-small transcends the incremental and provides a number of (fascinating) new challenges that require fundamental research results across a wide range of multimedia disciplines.

This thesis has presented the idea of socially-aware multimedia as a next step in the evolution of media authoring. By introducing the notional of a temporally-variant social content into media storage, access and sharing, we hope to stimulate a new generator of media research in which the multimedia user is given the central role that she deserves.