

Issues in Designing Novel Applications for Emerging Multimedia Technologies

Hyowon Lee

CLARITY: Centre for Sensor Web Technologies
School of Computing
Dublin City University, Ireland
hlee@computing.dcu.ie

Abstract. Emerging computational multimedia tools and techniques promise powerful ways to organise, search and browse our ever-increasing multimedia contents by automating annotation and indexing, augmenting meta-data, understanding media contents, linking related pieces of information amongst them, and providing intriguing visualisation and exploration front-ends. Identifying real-world scenarios and designing interactive applications that leverage these developing multimedia technology is certainly an important research topic in itself but poses a number of challenges: the currently practiced methodologies and tools in the field of Human-Computer Interaction and Interaction Design seem to work better when the target users and usage requirements have been clearly identified and understood in advance whereas much of what emerging multimedia technology could offer is expected to create completely new user activities and usage that we are not aware of; immature multimedia tools currently being researched are not good enough to be the core engines of real-world applications today, making realistic user studies through deployment difficult; our future interaction platforms will be more than just desktop PC, Web, or mobile devices but many other forms of tangible, embedded, physical appliances which we expect the currently developing multimedia technology would be coupled with. In this paper, these challenges and the insights into how we could get over them are explored based on the author's decade-long experience in designing novel interactive applications for multimedia technology.

Keywords: Human-Computer Interaction, Multimedia, Interaction Design, Innovation, Design Methodology, Novel Technology

1 Introduction

With so much R&D in multimedia research in developing various novel techniques and algorithms promising effective and efficient access to our growing media archives and collections, thinking about how to channel those developments into usable and feasible usage scenarios and actual, practical, innovative and high-impacting real-world applications is certainly a significant issue today.

The “demonstration systems” often featured in many multimedia research papers are in a sense such an effort to bring the developing computational tools in multimedia into real-world applications, even though they tend to be short-sighted, technically-oriented, poorly-designed and usually lack an understanding of end-users and contexts in which such a system is to be used.

On the one hand, most multimedia research groups’ lack of expertise in Interaction Design and Human-Computer Interaction (HCI) knowledge and perhaps their lack of collaboration with neighbouring HCI groups within their departments may be to blame. But on the other hand, the HCI community itself does not seem to offer more prescriptive methods or procedures to help develop novel scenarios and applications in such a way as to leverage the technologies developed by these multimedia researchers. Monitoring and interviewing end-users and coming up with requirements to develop a better version of an existing application is one thing, but trying to come up with a usable application not based on end-user data but based on emerging technology itself is another.

While acknowledging the power of User-Centred Design and how the strong emphasis on end-users and establishing comprehensive user requirements promotes a product that satisfies existing user needs and thus fitting the system well to their work context, an attempt to develop a completely novel application that does not have existing user base or current practice of use tends to face a lack of methodological support due to the HCI tools and procedures that seem to be focused on supporting today’s tools and activities rather than tomorrow’s.

The culprit of the plethora of complex technology demos with poor usability in multimedia witnessed today is not only the lack of inter-disciplinarity or lack of collaboration between multimedia and HCI, but the lack of general understanding of how a novel application could and should be developed where the starting point is technological possibility instead of an identified, unmet user needs in a specific domain.

The focus of this paper is on “emerging” or “novel” multimedia techniques rather than well-studied or already-robust techniques since the latter group of tools, once matured enough, become the realm of “conventional” technologies that any software house or companies can readily adopt in developing a highly usable multimedia application with the well-established software development and usability assessment processes available today. With this in mind, this paper will describe three interweaving issues that make it challenging in developing future interactive applications that incorporate emerging multimedia tools and techniques, and in doing so draw on the author’s hands-on experience of designing a variety of novel multimedia applications in a large multimedia research group over the past ten years.

To name but a few of these novel applications, a Web-based video retrieval system where a user can search a video collection by semantic features such as the existence of faces/people, indoor/outdoor, and cityscape/landscape [3]; an object-based photo archive where a user can search for objects in photos (e.g. a vase, a car or a ship) in terms of the object characteristics rather than whole photos [25]; a security video search system where CCTV video footage in a uni-

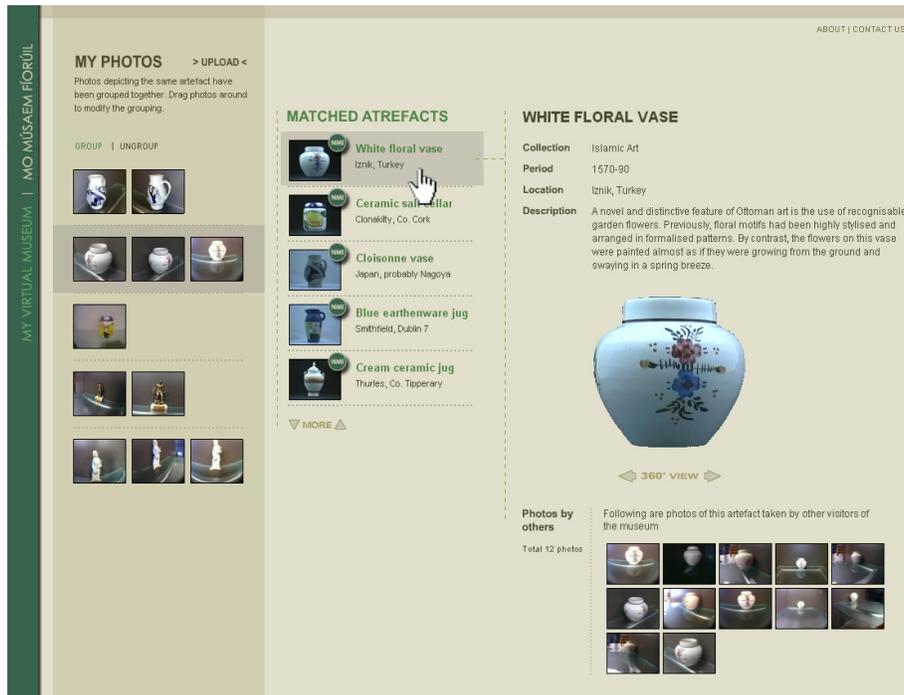


Fig. 1. Photos taken from a museum are automatically grouped by individual artefacts (left panel), and matched to the museum’s authoritative photo collection (middle panel), then linked to detailed information about that artefact (right panel)

iversity campus is captured and indexed by people’s contours and later a security staff can trace whereabouts of a suspicious person appearing in multiple camera locations [14]; a museum artefact explorer where a user can upload the photos taken at the museum to match and group the same artefacts even if the photos show different angles or sides, and display information on them (see Figure 1 for a screen shot) [2]; an online photo organiser where the uploaded personal photos are automatically annotated by the people appearing in the photos [23]; a Lifelogging browser where a large number of passively captured photos from a wearable camera are automatically structured by individual events, their relative importance identified and presented in an intriguing comic-book style montage as to help the user review their day [15]; a route finder where a collection of in-vehicle video footage is automatically indexed and seamlessly interweaved with the vehicle routes on a geographic map [19], and many more. All of these applications incorporated one or more combinations of multimedia techniques that were on-going research topics at the time (and many of them still are), and were more than just “technology demos” in the sense that an extensive Interaction Design effort was expended in developing them and specially designed to support some form of novel activities that conventional applications do not support. Go-

ing one extra mile to package the developing multimedia tools to come up with feasible and usable applications and studying the usage of such applications is not commonly witnessed in multimedia or any other technology-focused research communities today.

Serving as a “design consultant” and “usability expert” in the highly technically-oriented but extremely talented multimedia group that have been generating such a great variety of exciting computational tools, the author’s unique position as the HCI and Interaction Design researcher-practitioner contributing to “all things end-users” in the 40+ members of technology researchers over the past decade brought many exciting outcomes. This paper tries to highlight the issues by reflecting on the author’s own practice of communicating with his multimedia colleagues, of dealing with multimedia ways of thinking, of monitoring the aggressive and dynamic formation of new ideas, and of the design effort in applying various HCI tools to developing novel applications inspired by these technical ideas in multimedia.

2 Designing Interactive Multimedia Applications and the HCI Stance

One might attempt bringing in the well-structured HCI methodology in the design of a novel interactive multimedia application. Many of the HCI tools and procedures currently practiced (and with desirable effects), however, are geared towards those applications that we currently use, with relatively well-understood domains and with an existing user base. Involving the target end-users at various stages of the development process in order to reflect the users’ wishes and needs into the design is the central premise of the User-Centred Design (UCD) approach. Currently popular ethnographic methods such as Cultural Probes [7] and Contextual Inquiry [9] as well as other traditional interviews, questionnaire, interaction logging and eye-tracking, are the examples of these methods where the purpose is to obtain the information about people’s current practice of certain activities in order to incorporate them into the design.

Some of the multimedia technologies do promise an enhancement of existing tools in existing practices. For example, *face detection and recognition* technique could be plugged in to an existing personal photo or video management service such as Flickr and YouTube, to help reduce the user’s manual annotation burden by automatically tagging the photos or video clips by the appearance of individual persons’ names. *Advertisement detection* in a video stream could be plugged in to our VCR/DVD recorders at home to help save the storage and skip the annoying ad breaks in the middle of a movie, by automatically identifying them and removing or at least tagging them at the time of recording. These scenarios are certainly a useful target for the computational multimedia tools to be applied to and be made useful in the real-world, in the way that makes people’s life easier and more convenient by automating those elements that so far had assumed human labour and intervention.

There are, however, ways in which technologies could support very new activities, or rather, *create* new activities that people have not done before: online chatting was not a known activity that anybody did until the infrastructure of the Internet and the software that supported it appeared; the activity of blogging did not exist only 5 years ago until web blogging services appeared and people started using them; video sharing/voting and twitting are also the examples of recently invented activities whereby those technological applications appeared first then the uptake of the usage happened afterwards. Designing for these novel applications (albeit the lack of any technological innovations in some of the above examples) that will support new activities is fundamentally different from designing, for example, a next generation of a word processor or a better version of a library management system.

We expect the tools and techniques researched in many multimedia R&D groups today will in time create many such new activities in the form of “novel interactive applications.” Because the design for the new activities expected to be invented cannot rely solely on existing user data or precedent design examples, the starting point of designing such an application is on shaky grounds (Section 3.1). While the novel multimedia tools themselves might not provide robust and accurate performance today thus making any decent user-experiment or deployment effort difficult (Section 3.2), we expect by the time these tools become better understood and more mature in terms of their performance, the kinds of interaction platforms that we will be interacting with on a day-to-day basis will be much more diverse (Section 3.3). The next section will address these issues in more detail.

3 The Issues

3.1 Shaky Starting Point

An interaction designer typically gets input from (1) engineers, who provide technologies; (2) anthropologists, who provide field data; (3) behavioural scientists, who provide models and theories to support the design of artefacts [31]. While the modern HCI practice strongly advocates the User-Centred Design approach where the emphasis is on establishing requirements by understanding the users and the environment where the system in concern is to be located and used, designing novel multimedia applications often do not have the “field data” or the information on the existing users’ practice of the application area, as that application scenario itself is something new. Also, many successful designers rely on their past experiences of designing similar artefacts [11], and many aspects of design activity itself are based on building on successful precedences. Designing novel applications, as there is no such initial exemplar or successful products, has a very shaky starting point.

Some inspirations do come from observing people’s current lives and their activities. Trying to use ethnographic studies to explore new possible scenarios and innovations [22] and studying a particular practice to learn about the underlying motivation for a more grounded innovation [17] have been suggested.

These approaches try to start by studying end-users in developing technological innovations. Also, we could first study the usage of existing applications of a similar nature (probably a manually-intensive equivalent that the new application is to supersede) and hope that they could help inform the design of the new. For example, in order to design a novel application that uses a *video summarisation* technique, we could study how people browse their DVD collections using a chapter selection feature provided in most DVD movies, or use an Electronic Program Guide to quickly get the gist of the contents, and identify how the current experiences in these activities could be enhanced with the automatic video summarisation. Collecting usage data from the “proxy users” in this way is one of the ways to ground the new technological development to the real-world. In a recent article in interactions [18], Norman warns against the tendency to rely on ethnography in technological inventions by explaining how major breakthroughs and innovations in history came from technologists who had very little understanding of users, and how most often “technology will come first, the products second, and then the needs will slowly appear.”

Because it is often impossible to predict whether a newly-supported activity is something that people will want to do, or rather, impossible to predict how people will accommodate and assimilate a novel activity afforded by the new application into their lives, it is more important to quickly develop a robust application first then get people to start using it. In this sense, the design decisions for novel multimedia applications should be made as to come up with an artefact that is open-ended in terms of its eventual usage or purpose (because we don’t know what they would be at the time of design) but that strictly adheres to usability principles [26]. Thus, quickly prototyping an application and conducting a usage study with it is one sure way to go about developing a novel multimedia application, without spending too much resources on trying to establish initial requirements or understanding the usage context at the beginning. As more and more people have the access to the Web, developing a Web-based application has the great benefit of being able to easily deploy it for a number of users to use although this can be a problem in limiting the usage scenarios for near future (see Interaction Platform Issues Section below).

The shaky starting point tends to result in longer design time to produce an initial application scenario and the user-interface, as having no previous examples to follow and having no good understanding of usage makes it difficult to structure and streamline the design process in any way. For example, it took the author well over full-time 4 months to design a video clip searching application where a clip-to-clip content similarity measure is used to support a “find more like this” type of query and interactive refinement of the query and retrieval results in a highly efficient way [20]. Having a query panel on the left where example video clips and their associated transcripts are added and retrieval results on the right showing matched shots’ keyframes as well as a few preceding and following shot keyframes in different sizes, had no such precedences and everything had to be designed from scratch. For this, a series of iterative sketches and brainstorming sessions with the technical members were undertaken as the main

process over a 4-month period. After the system was fully implemented and a user experiment conducted, we had much better understanding of the application in terms of people's opinions about such a tool and how it might be further refined or re-branded into a product. It has been noticed that a number of other multimedia groups designed similar user-interfaces in the subsequent years, and now designing a similar interface can be a matter of days (precedences and practical know-how/experience makes such a big difference in design).

The design of a lifelogging photo browser where a large number of a person's lifelog photos from a SenseCam can be reviewed [15] took more than a year during which the strategies to connect between the supporting back-end techniques and their possible front-end manifestation and interaction schemes gradually developed - there was no precedence of such an application and there still is no usage of such an application as it will take many more years for the passive capture device such as SenseCam to be used by the general public. Many of the novel applications incorporating multimedia techniques developed in our group took months for each of them to form any concrete user-interface, because of the lack of understanding in the application areas and the lack of examples to narrow down the design space.

While in engineering fields the lack of initial understanding or requirements generally poses challenges in formulating the questions and problems to work on, design studies show that the inherent quality of designers is the ability to work on an ill-defined problem space and quickly reach an initial design solution by framing the problem space in a creative way, then going back and forth between the problem and solution space over time [5]. When it comes to designing novel multimedia applications, such quality will be very much needed and it may be an important clue in any attempt to make explicit the procedure for innovative design that lacks user/usage information.

3.2 Imperfect Back-end Performance and Implications

Developing effective video *shot-boundary detection* and *keyframe extraction* techniques have been a very active sub-topic in the multimedia field since the late 90s and early 2000s and their accuracy level is said to have reached above 95-98% today (for straightforward hard-cut transitions anyway) [29]. We call this a "solved problem" and these well-understood techniques are today featured in many video editing tools to help quickly browse the video contents. *Face detection* has been, as its potential value for such a technique is huge, an another active research topic in computer vision and now many digital cameras feature real-time face detection and highlighting in the camera viewfinder to help the user focus the camera to those face areas. Some of these older topics in the field are having fruition today by being featured in these commercial products. There are a variety of different techniques in multimedia that are currently being studied and sooner or later to come out of the laboratories into real-world applications.

One of the problems of making a working prototype of a novel multimedia application is that the multimedia technique that the system is to demonstrate

is, almost by definition, an imperfect technology that multimedia researchers are working very hard today to improve in their laboratories. As the technology research community, it is important to design novel applications *now* even though their back-end multimedia techniques themselves may not yet be good enough to be used in the real-world, in order to envision and shape the near-future when such techniques will have become sufficiently matured.

Having an immature technology as a back-end of a system not accurate or robust enough to be used un-supervised in a real-world application is a big handicap as an interactive application: users will notice the inaccurate results on the front-end and this lowers the perceived value and attraction of the application enormously, even if all other usability aspects are crafted to high standards. It can be compared to, say, an online shopping mall site where product information is occasionally incorrect or some times wrong photos are displayed for a product - whatever the reason, it is simply unacceptable. This means the precious comments, feedback and usage of the test users that otherwise could have shown valuable information about possible new usage or wishes will be clouded by the obvious functionality flaws.

In order to remove such complaints on the performance of the system (and to get useful comments on all other aspects), we need to manually correct the results of the imperfect algorithms and show the perfect outcome to the users.

For example, we had an online video retrieval system we called Físchlár (“Fís” means *video*, and “chlár” means *programme* in the Irish language) which was deployed within the university campus for over 6 years (1999-2006) during which more than 2,000 students and staff users registered and actively used the system. The system ingested video streams from broadcast TV signal on our users’ requests and processed, indexed, and structured the incoming video stream for browsing, searching and playback. One of its latest variations incorporated an automatic *news story segmentation* technique and processed daily 9pm news from the national TV channel RTÉ, and presented a news story-based searching and browsing interface to the users. A lot of effort was put into capturing the real usage of this system with interaction logging, diaries and questionnaire [16]. While the system had to be made robust and accessible at all times (our users used the system in the morning before they start working, at lunch breaks, between lab sessions and in late evenings), the most problematic aspect was that the story segmentation engine itself did not produce 100% accurate results, thus our users would frequently see incorrect story units in their browsing. In order to prevent this, the newly added news videos were manually checked every morning to correct wrong segmentation with a simple editing tool before our users started accessing the site in the morning. This early-morning manual intervention continued every day for more than 5 months during which this particular study was conducted. Costs associated with deploying an application, especially when its back-end uses an experimental multimedia techniques that is not mature, can be so high that it may seem almost a luxury to be able to conduct this kind of long-term deployment study with such an application.

In this particular study, our users thought the system fully-automatically indexed the news stories without any human intervention, but the more illuminating fact for us was that they did not really *care* whether it was automatically done or otherwise: all they wanted was to be able to locate the news stories they were interested in and watch them to kill some time, as if it were a regular online news service such as Irish Times Online or BBC News Online. They instead complained about the system not presenting the news stories by overall categories such as “International”, “Politics”, “Entertainment” and “Sports”, which did provide us with future desired features of the application.

An alternative strategy is to design the user-interface in such a way that whenever a user notices such imperfect output of the back-end processing, he/she could very easily fix or correct it while doing the task. The corrected facts can then be fed back to the back-end processing and propagate to the rest of the data, overall improving the accuracy of the system. Encouraging or motivating the user to provide annotation during the use [27] is probably a good strategy for providing a service with imperfect performance as long as the task of annotation is not lengthy or laborious. For example, in a personal photo management system we developed in 2007 [21], uploaded photos were automatically annotated with the name of the people present in each of the photos using face recognition augmented with a body-patch algorithm. As the accuracy of these algorithms was not 100% accurate, sometimes the users noticed incorrect names being labeled for some faces in the photos. Using a simple mouse-over action over the photo and the selection of a predicted alternative names pushed to the user by the system, he/she corrected the annotation with a minimal effort whenever such error was noticed and the system re-calculated remaining photo collection with the revised certainty based on that manual correction, just as is done in current systems like Apple iPhoto.

Another stream of strategy is that when the back-end performance is imperfect, we could develop usage scenarios and the user interaction in such a way that such imperfection would not be so blatantly noticeable or crucial in the user’s task. For example, our online Movie Browser application [1] used movie *scene segmentation and classification* techniques to present movie content by chunks of Exciting, Dialogue, or Montage scenes (see Figure 2 for a screen shot). While the state-of-the-art scene segmentation tools are still not mature enough to ensure reliable results, our user-interface arranged the browsing mechanism with a large timeline where the segmented scene blocks are highlighted in different colours depending on its identified scene types. When a user clicks any of the highlighted blocks on this timeline, the series of scenes around that time region is presented and the user can browse movie events around that time region. In our deployment study of this system with over 260 students in a Media Study course for a full semester, the students used the system freely to help write their movie analysis assignments. Interviews, focus group and questionnaire were used to obtain their usage and opinions, and the inaccurate scene boundaries and classification appearing on the interface was not an issue partly because the interface naturally reduced any negative effects by showing neighbouring scenes but also

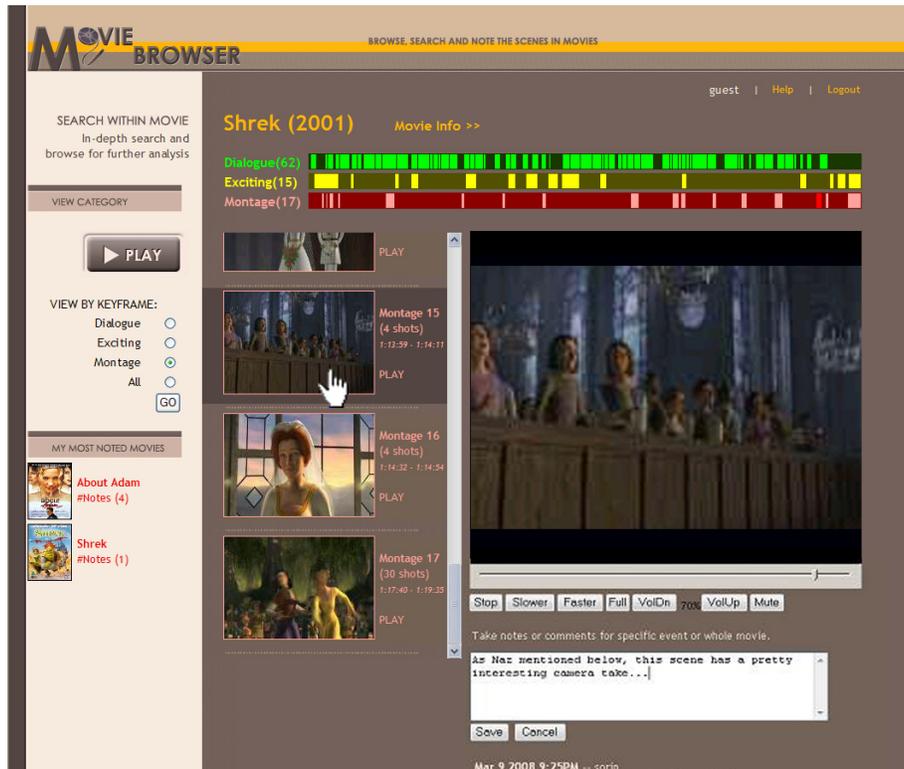


Fig. 2. Three-band timeline shows where dialogue, exciting, and montage scenes are in the movie, and clicking any area on the timeline will show the scene keyframes around that area below

partly because the concept of scene boundary itself cannot be objectively defined or agreed upon by everybody: as long as the users were able to roughly reach a point of their interest quickly and navigate around, they were happy and did not complain about the segmentation inaccuracy.

The choice of these strategies will depend on the domain and the way the usage scenario is drawn for that application. Thus, it will probably be more difficult to design the above-mentioned news story browsing interface in such a way that the incorrect story segmentation results will not be too obvious or matter, because news stories are generally more well-defined and are often the isolated unit of contents that news consumers will want. Even so, there might be some novel scenarios that we could develop where the blending of adjacent news stories is the main feature, perhaps a visualisation of an aggregation of daily news stories over a long period.

3.3 Interaction Platform Issue

Consider some of the multimedia tools and techniques we are currently developing are packaged into an intriguing application running on an Apple iPad, with which a user can use a simple finger touch to easily command the functions, search and browse with the powerful automatic indexing and visualisation techniques in a way that conventional applications cannot. Are there particular multimedia tools that might be better suited for this “Multimedia iPad” scenario than others?

So far, whenever somebody showed a “demo” of a multimedia research project, it was shown on a desktop PC or a laptop displaying a stand-alone or Web-based user-interface in some way. Large databases of complex and colourful multimedia information in image, video and text may be nicely visualised on a large monitor with a mouse and keyboard, and help explore the data, adjust sophisticated parameters, pinpoint the area of interest, draw a region, and navigate a deep hierarchy of menus relatively efficiently. More recently some demos were shown on mobile devices, with academic events specifically dealing with mobile multimedia appearing such as the International Workshop on Mobile Multimedia Processing (ICPR 2010), the International Conference on Mobile and Ubiquitous Multimedia (MUM 2010) and the ACM Multimedia Workshop on Mobile Video Delivery (MoViD 2010).

The great thing about Web-based and mobile multimedia applications is that they can be packaged to be more or less readily accessible for anybody who has access to the Internet or has an appropriate mobile device. Deploying a Web-based prototype application or an iPhone app is methodologically very effective in terms of getting people to use it to get feedback and monitor the emerging usage of such applications.

However, *by the time* many of the emerging multimedia tools studied today become mature enough to be used outside the lab and into real-world applications (say in 5-10 years as a ballpark figure for a cycle in which a newly proposed computational tool becomes a commercialisable piece of technology), we will be interacting in our daily lives with more variety of devices and platforms, most likely far *more frequently* than with desktop PCs or mobile platforms: interactive tabletops, with such a ubiquity of physical tables at home, restaurants, cafes, airplanes and schools, will probably become one of the most commonly encountered interaction platforms in the near future; interactive TV, augmenting conventional TV boxes with social connectivity, storage and processing power, will become the future way of watching TV; large multi-touch walls will dominate our streets, at bus stops and shopping malls providing novel public spaces with multi-user interactivity; other embedded appliances will be around us such as electronic magazines, sophisticated digital picture frames, touch-screen in-home displays (perhaps a variation of the next generation of iPad), door panels with media alerts and weather forecasts, etc.

Thus we envisage that it will be more constructive and forward-thinking to try to couple the emerging multimedia techniques that we are developing today with these more novel interaction platforms, than just demoing on a desktop PC

or a laptop. As different interaction platforms assume different physical and contextual characteristics and consequently different appropriate modes and styles of interaction, we should consider that some types of multimedia tools currently being developed might be more suitable to manifest on particular types of platforms than others in order to maximise or best leverage their potential power and benefits.

For example, an earlier project involved developing an automatic news story segmentation tool from daily TV broadcast news videos and recommended individual news stories to mobile users based on user preference, viewing history and collaborative filtering [8] - the mobile platform with small screen and awkward input mechanism makes the content analysis and structuring techniques coupled with automatic recommendation of meaningful units of information quite ideal for providing the kind of user interaction where the interaction effort is minimised by the system intelligently digesting the information, structuring it, summarising it, then selectively presenting the most useful piece of information to the user on the mobile screen. Thus, multimedia techniques that result in elaborate visualisation or explorative interaction is less of a value to mobile platforms but summarising, structuring, and selective pushing types of techniques seem more promising for mobile applications. For example, the sports summarisation technique we have developed [24] analyses any field-sports video content and identifies those segments that contain high probability of important events happening in the game. Stitching up those identified segments can result in a 3-minute video summary of important events from a 90-minute football match. Such a summarised form of a playable video could very well support a mobile entertainment service where the user's interaction with the device is to simply watch a 3-minute summary video rather than typing in text, selecting options, dragging panels, and exploring a sophisticated visualisation scheme which all cause interaction burden to the mobile user.

Down the road, we can imagine an ideal mobile device where its back-end mechanism is so advanced that as soon as the user turns on the device he/she will see only those information and interface that the user wanted to use without any menu navigation selection of options requiring visual attention, as a result of the combination of accurate information processing, intelligent inference and adaptive interface.

Exploring the suitability of various interaction platforms for emerging multimedia techniques should be one very important part of the interactive multimedia application design. The obvious challenge in trying to develop such applications is that the developed system (whether it being on a tabletop, iTV, Multitouch Wall, or any other platforms) is difficult to deploy due to the current unavailability of such platforms to the test users, unlike Web-based systems or iPhone apps that could relatively easily be deployed as many users have access to these platforms anyway.

In 2005, our group developed an interactive tabletop application that incorporates a number of video retrieval techniques to allow multiple users sitting around the table to collaboratively search for video clips [28]. While the back-

end multimedia tools were used to segment video streams to manageable chunks and to allow similarity-based searching amongst the chunked clips, the front-end table interface tested how the interaction can be arranged in such a way that the collaboration amongst the users could be encouraged and conducted smoothly. The tabletop is still an on-going field of study in the HCI community, and coupling multimedia techniques on an interactive table gives very interesting research opportunities to this line of study. Because regular users do not have an interactive tabletop and in most cases have not even used such a platform, studying people's use of the application required for us to arrange appointments with the test users in the room where we showed it and got them to do training tasks, in a typical laboratory user testing style.

In 2008, we developed a "Multimedia TV" application which featured a number of multimedia techniques (including shot/scene/story boundary detection, sports highlight detection, face detection, clip similarity calculation, etc.) in a TV set and a viewer with a conventional remote control can use a small number of buttons to use these capabilities [12]. *Lean-back interaction* that characterises a TV interaction requires extremely simplistic screen element design and the mechanism to invoke the functions while watching the TV. During the design stage of the TV widgets and screen elements, we brought in test users to our TV set and informally observed their behaviour and discussed with them. Even though realistic usage can only be tested in people's own homes, deploying such a TV to a home is currently impossible as nobody has such a TV.

It is not easy to see any effort in experimenting the emerging multimedia techniques on novel interaction platforms, as most multimedia researchers are busy with their algorithms and techniques to seriously think about applying them to these platforms and the HCI researchers often have little understanding of what emerging multimedia tools offer.

While quite a lot of interaction design knowledge, experience and skill set are today available for the desktop PC/Web platform, relatively little is understood or practiced for the mobile platform, and far less for other more novel platforms. The design knowledge for each of these platforms are, however, slowly growing today as more and more experimental applications appear in the labs and the trial-and-error process starts identifying what makes good interaction strategies for each of these platforms. Growing design knowledge for each of these platforms is highlighted and summarised in [13]. Leveraging the increasing amount of design knowledge for various novel platforms will be one of the keys to successfully deploying usable interactive multimedia applications of the future.

4 Conclusion

The way we can interact with multimedia data depends on how the data is organised, indexed and presented, so there is certainly a very important link between multimedia technology and interaction design [10]. A competent interaction designer for multimedia applications will be someone who is equipped with HCI

tools, design thinking and skills while at the same time understands the nature of multimedia technologies and the trend of their progress over time.

In highlighting the issues, this paper did not emphasise the lack of design thinking by the multimedia researchers although this is most likely one of the factors slowing down the development of highly usable and attractive multimedia applications. Multimedia researchers without practical design skills usually like to argue a multimedia application's user-interface with such HCI terms and concepts as "click distance", "screen real-estate", "consistency" as it seems they find these concepts most comfortable to understand in their scientific and engineering frame of mind. The fact of the matter is that a competent interaction designer usually does not constrain himself to such terms in designing or assessing the interface. For designers these terms are largely a hindsight, post-design analysis or post-design rationale. He tries to find a wholesome solution that can elegantly solve a range of issues in a single, coherent and unifying theme (sometimes called "primary generator" [11] or "first principle" [4]), even at the cost of a sub-optimal level of, say, click distance or screen real-estate, more drawn from his/her experiences and intuition, frequently ignoring an elaborate requirements analysis that multimedia researchers in the science and engineering tradition prefer to start with. Certainly design decisions are based on a series of good judgements of the designer [30] who can juggle with many unknowns and known factors, and the success of a design is at the designed outcome, not the way design choices were argued and rationalised [6]. It is understandable how multimedia researchers struggle to understand and often become frustrated when their interaction design colleagues immediately start working on the final solution in a sketch when given a description of the technologies to be incorporated.

The way forward would be on the one hand for the multimedia community to embrace these "designerly ways of knowing" [5] and leverage their ability in developing the new generation of interactive applications that incorporate the emerging technologies the multimedia researchers are working so hard on today. On the other hand, more competent interaction designers and HCI practitioners should start working directly with the multimedia technology researchers, picking up those multimedia tools and algorithms with huge potential to design novel scenarios and applications that will forever change the way people work and play.

Acknowledgments. This work is supported by Science Foundation Ireland under grant 07/CE/I1147 and by Dublin City University Research Fellowship Scheme.

References

1. N. M. Ali, A. F. Smeaton, H. Lee, and P. Brereton. Developing, deploying and assessing the usage of a movie archive system. In *HCI Intl. '09 - 13th International Conference on Human-Computer Interaction*, 2009.
2. M. Blighe. Mo msaem foril: A web-based search and information service for museum visitors. In *ICIAR 2008 - Intl. Conf. Image Analysis and Recognition*, 2008.

3. P. Browne, C. Czirjek, C. Gurrin, R. Jarina, H. Lee, S. Marlow, K. M. Donald, N. Murphy, N. O'Connor, A. F. Smeaton, and J. Ye. Dublin City University Video Track Experiments for TREC 2002. In *TREC 2002 - Text REtrieval Conference*, MD, USA, 2002. National Institute of Standards and Technology.
4. N. Cross. Creative cognition in design: processes of exceptional designers. In *C&C '02: Proceedings of the 4th conference on Creativity & cognition*, pages 14–19, New York, NY, USA, 2002. ACM.
5. N. Cross. *Designerly Ways of Knowing*. Birkhäuser, Basel, Switzerland, 2006.
6. N. Cross. Creative thinking in design: an introduction. In *SoD '07: Proceedings of the 2007 Symposium on Science of Design*, pages 2–3, New York, NY, USA, 2007. ACM.
7. B. Gaver, T. Dunne, and E. Pacenti. Design: Cultural probes. *interactions*, 6(1):21–29, 1999.
8. C. Gurrin, A. F. Smeaton, H. Lee, K. McDonald, N. Murphy, N. O'Connor, and S. Marlow. Mobile access to the Físchlár-News Archive. In *Springer LNCS*, volume 2954, pages 124–142, Berlin / Heidelberg, Germany, 2004. Springer.
9. K. Holtzblatt and S. Jones. *Contextual Inquiry: A Participatory Technique for System Design*, pages 180–193. Lawrence Erlbaum, Hillsdale, NJ, 1993.
10. A. Jaimes, N. Sebe, and D. Gatica-Perez. Human-centered computing: a multimedia perspective. In *MULTIMEDIA '06: Proceedings of the 14th annual ACM international conference on Multimedia*, pages 855–864, New York, NY, USA, 2006. ACM.
11. B. Lawson. *How designers think: the design process demystified (4th Ed)*. Architectural Press, 2006.
12. H. Lee, P. Ferguson, C. Gurrin, A. F. Smeaton, N. E. O'Connor, and H. Park. Balancing the power of multimedia information retrieval and usability in designing interactive TV. In *UXTV '08: Proc. Intl. Conf. Designing interactive user experiences for TV and video*, pages 105–114, New York, NY, 2008. ACM.
13. H. Lee and A. F. Smeaton. Interaction platform-orientated perspective in designing novel applications. In *Create 2009 - Creative Inventions and Innovations for Everyday HCI*, 2009.
14. H. Lee, A. F. Smeaton, N. O'Connor, and N. Murphy. User-Interface to a CCTV Video Search System. In *ICDP 2005 - IEE International Symposium on Imaging for Crime Detection and Prevention*, pages 39–43, 2005.
15. H. Lee, A. F. Smeaton, N. E. O'Connor, G. J. Jones, M. Blighe, D. Byrne, A. Doherty, and C. Gurrin. Constructing a SenseCam visual diary as a media process. *Multimedia Sys. J.*, 14(6):341–349, 2008.
16. H. Lee, A. F. Smeaton, N. E. O'Connor, and B. Smyth. User evaluation of Físchlár-News: An automatic broadcast news delivery system. *ACM Transactions on Information Systems*, 24(2):145–189, 2006.
17. S. Ljungblad and L. E. Holmquist. Transfer scenarios: grounding innovation with marginal practices. In *CHI'07*, pages 737–746, New York, NY, USA, 2007. ACM.
18. D. Norman. Technology first, needs last: The research-product gulf. *interactions: Exploring Aspects of Design Thinking*, XVII(2), 2010.
19. N. O'Connor, T. Duffy, P. Ferguson, C. Gurrin, H. Lee, D. Sadlier, A. F. Smeaton, and K. Zhang. A content-based retrieval system for UAV-like video and associated metadata. In *SPIE Defence and Security Conference 2008*, 2008.
20. N. O'Connor, H. Lee, A. F. Smeaton, G. Jones, E. Cooke, H. le Borgne, and C. Gurrin. Físchlár-TRECVid2004: Combined text- and image-based searching of video archives. In *ISCAS 2006 - IEEE International Symposium on Circuits and Systems*, 2006.

21. N. O'Hare and A. F. Smeaton. Context-aware person identification in personal photo collections. *IEEE Transactions on Multimedia, Special Issue on Integration of Context and Content for Multimedia Management*, 2009.
22. Y. Rogers and V. Bellotti. Grounding blue-sky research: how can ethnography help? *interactions*, 4(3):58–63, 1997.
23. D. Sadlier, H. Lee, C. Gurrin, A. F. Smeaton, and N. O'Connor. User-feedback on a feature-rich photo organiser. In *WIAMIS 2008 - 9th Intl. Workshop. Image Analysis for Multimedia Interactive Services*, 2008.
24. D. Sadlier and N. O'Connor. Event detection in field sports video using audio-visual features and a support vector machine. *IEEE Transactions on Circuits and Systems for Video Technology (Eds. F. Pereira, P. van Beek, A.C.Kot, and Ostermann J.)*, 15(10):1225–1233, 2005.
25. S. Sav, G. Jones, H. Lee, N. O'Connor, and A. F. Smeaton. Interactive experiments in object-based retrieval. In H. Sundaram, M. Naphade, J. R. Smith, and Y. Rui, editors, *CIVR2006 - 5th International Conference on Image and Video Retrieval. Springer Lecture Notes in Computer Science Vol. 4071*, volume 4071 / 2006, pages 1–10, Berlin/Heidelberg, Germany, 2006. Springer.
26. P. Sengers and B. Gaver. Staying open to interpretation: engaging multiple meanings in design and evaluation. In *DIS'06*, pages 99–108, New York, NY, 2006. ACM.
27. B. Shneiderman, C. Plaisant, M. Cohen, and S. Jacob. *Designing the user interface: strategies for effective Human-Computer Interaction (5th ed.) cf. 3.4 Development methodologies; 3.5 Ethnographic observation, and 3.7 Scenario development*. Addison-Wesley, 2009.
28. A. F. Smeaton, H. Lee, C. Foley, and S. McGivney. Collaborative video searching on a tabletop. *Multimedia Systems Journal*, 12(4):375–391, 2006.
29. A. F. Smeaton, P. Over, and A. Doherty. Video shot boundary detection: Seven years of trevid activity. *Computer Vision and Image Understanding*, 2009.
30. T. V. Wolf, J. A. Rode, J. Sussman, and W. A. Kellogg. Dispelling “design” as the black art of CHI. In *CHI'06*, pages 521–530, New York, NY, 2006. ACM.
31. J. Zimmerman, J. Forlizzi, and S. Evenson. Research through design as a method for interaction design research in hci. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 493–502, New York, NY, USA, 2007. ACM.