Challenges toward User-centric Multimedia

Janine Lachner[1], Andreas Lorenz[2], Bernhard Reiterer[1], Andreas Zimmermann[2], Hermann Hellwagner[1]

[1]Klagenfurt University, Dept. of Information Technology
Universitätsstraße 65–67, 9020 Klagenfurt, Austria
{firstname.lastname}@itec.uni-klu.ac.at

[2]Fraunhofer Institute for Applied Information Technology
Schloss Birlinghoven, 53754 Sankt Augustin, Germany
{firstname.lastname}@fit-fraunhofer.de

Abstract

Currently, much research aims at coping with the shortcomings in multimedia consumption that may exist in a user's current context, e.g., due to the absence of appropriate devices at many locations, a lack of capabilities of mobile devices, restricted access to content, or non-personalized user interfaces. Recently, solutions to specific problems have been emerging, e.g., wireless access to multimedia repositories over standardized interfaces; however, due to usability restrictions the user has to spend much effort to or is even incapable of fulfilling his/her demands. The vision of user-centric multimedia places the user in the center of multimedia services to support his/her multimedia consumption intelligently, dealing with the aforementioned issues while minimizing required work. Essential features of such a vision are comprehensive context awareness, personalized user interfaces, and multimedia content adaptation. These aspects are addressed in this paper as major challenges toward a user-centric multimedia framework.

1. Introduction

There have been considerable efforts to have multimedia content (particularly images, audio, and video), devices and applications converge into systems that end users can utilize conveniently. For instance, home environments comprise homes as *spaces* of multimedia convergence, and advanced mobile devices of nomadic users constitute *points* of convergence. The Network of Excellence INTERMEDIA, funded by the European Commission, targets to lay out a research path for user-centric multimedia content handling in extended home environments with different devices. The project seeks to progress beyond home-centric and device-centric approaches toward truly user-centric convergence of multimedia. The project vision is to place the user into the center of such multimedia systems, where services (multimedia applications) and the means for interacting with them (devices and interfaces) converge.

Human-computer interaction with user-centric multimedia applications introduces several challenges for the usability of such systems. Among the most important issues are, according to the authors' experience, context awareness and the personalization of multimedia content and of the user interface. Additionally, despite a human-centered design process, the final deployment of multimedia applications may require instant adaptation of content and functionality due to increasing situational dynamics. Therefore, the INTERMEDIA project targets a software framework that provides developers with means for improving the usability of user-centric multimedia software at both construction and deployment time. This software framework is to empower developers to easily and flexibly create applications that closely meet their respective requirements and to adopt the human-centered computing paradigm [5] to diverse application scenarios involving multimedia content.

This paper primarily focuses on the challenges involved in the creation of a conceptual framework as a basis of the software framework elaborated in a subsequent step. It provides considerations about requirements for the design, development and functioning of user-centric multimedia software that need to be reflected by such a framework. The following section outlines important research and technology fields pertaining to the concept of user-centric multimedia. In the main sections, the paper describes the user-centric view and the most challenging technical aspects relevant to the creation of a user-centric multimedia environment.

2. Scenario and vision of user-centric media

The concept of user-centric multimedia provides freedom to a modern nomadic person to consume multimedia content without the requirement to carry a range of mobile devices, by providing personalized access to media regardless of device type. It will extend personal media spaces for nomadic life by removing spatial constraints in our daily activities.

Imagine a typical user-centric scenario where a person (let us call her *Nina*) is able to seamlessly enjoy her favorite movie while cooking and eating, then relaxing in the living room, and later going out. A context aware system continuously monitors her current environment and accordingly performs adaptation actions. For instance, the sound is turned down if it is late in the evening and a complaining neighbor has just arrived; or, the movie is migrated onto the HDTV screen in the living room from the small kitchen TV after Nina has finished dinner and has moved to relax on a comfortable couch. It is also possible for her to go out and continue watching the movie on her PDA. Coincidentally, she meets a friend in the subway who joins her watching the movie. As the friends are in public space now, the content is automatically adapted to omit age restricted scenes of the movie. Later on, when strolling through the city center's streets, they decide to switch to inspect the product offerings of the fashion shops they pass by, including sample videos of the products' looks (since the shops are closed at that time). Meanwhile, the movie is being recorded on Nina's home server. The user interface of the PDA presentation changes to present Nina's avatar for trying the fashion products as well as her favorite method of payment. The media stream on the PDA is continuously adapted according to the friends' current environment, i.e., the volume is turned up because it is noisy on the streets, and the frame rate is adapted according to the bandwidth currently available while they are on the move. In addition to the automatic adaptation based on the monitored context, Nina can use her personalized device to make further settings that improve her media experience even more, using simple and natural interaction, e.g., voice controls.

This scenario illustrates the general vision of user-centric multimedia, which is that a user should have access to multimedia applications and services

- offered by the surrounding environment
- and/or providing personal/personalized content
- in an easy-to-use and intuitive way
- regardless of device type and physical position
- seamlessly across various networks
- through a personalized interface
- according to his/her commands, gestures, behavior
- as sensed by the environment.

Key to this vision is the idea that users can interact using personalized interfaces and consume their personalized content regardless of their location, the particular set of physical devices used for interaction (on the body or in the environment), and other environmental influences, rather than forcing users to organize their lives around the computational technologies required for multimedia access.

To realize a vision like this, significant advances in a number of research and technology fields need to be made. For example, content annotation and an efficient and interoperable content description mechanism guarantee the user that she gets the content of interest. Content discovery methods and the process of selecting content have to take into account the user's preferences and other usage environment factors. Other areas that have to be considered are digital rights management and privacy and security issues (i.e., authentication and authorization). Regarding the delivery of multimedia data over dynamic networks, it is important to address robust transmission, quality of service issues, the mobility of users, as well as energy efficient delivery. Furthermore, the design of devices, software, and interfaces has to be considered for the consumption of the content. However, these areas are important for a vast amount of scenarios regarding multimedia delivery, not only for a typical user-centric approach, where they have a rather supporting role. In this paper we want to concentrate on three areas that are especially significant for user-centric scenarios: personalized interfaces, context awareness, and content adaptation, as illustrated in Figure 1.

In the next sections, technologies and possibilities of the three areas we consider most important for a user-centric multimedia approach are discussed in more detail.

3. Personalization and context management

In order to employ user-centric multimedia consumption, it is crucial that the user can communicate with the system using a natural, personalized interface to issue commands and express preferences. For example, a user might want to use her media player with a personalized skin to choose some of her favorite video clips for playback.

Jameson [6] defines systems automatically performing adaptation to the individual user in a nontrivial way as *user-adaptive systems*. Personalization allows users to obtain information that is adapted to their needs, knowledge, interests, or other characteristics delivered by user models [9]. Therefore, the user is placed in the center of Figure 1. Several generic user modelling servers support adaptation by

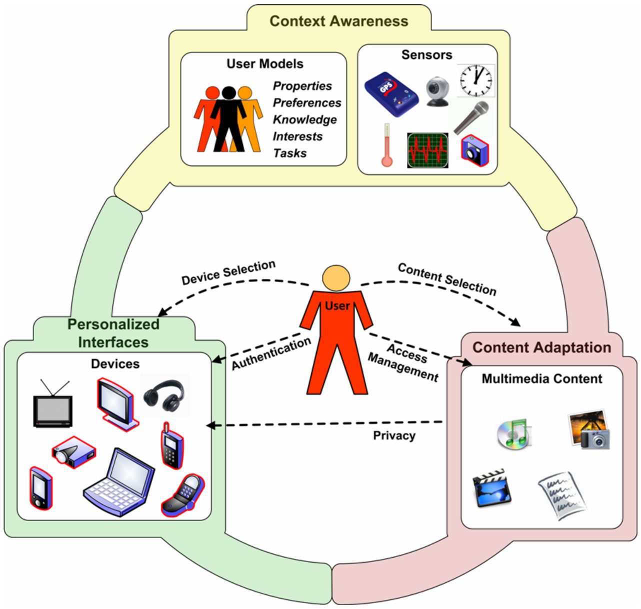

Figure 1. The user in the center of a multimedia environment

providing user modelling services to application systems [8]. In this context, Fink [2] presents research on requirements, design, and evaluation of user modelling servers. Since users might form groups to share content with others or to interact with each other in a shared session, we also consider user groups in addition to a single user.

3.1. Adaptation targets

The major aims of personalized adaptive systems are improvements in effectiveness and efficiency of user interaction together with higher user satisfaction. User-aware applications consider five properties that can be tailored to the context and the user:

Human-computer interaction covers the modality of entering commands or data, and receiving information and services.

Information presentation regards methods and coding required for receiving and displaying content (front-end).

Functionality addresses the features needed to perform tasks (back-end) and the ability of the application to solve a single task or a set of tasks.

Information and service selection covers the content, density, and depth of information as well as the functionality and complexity of necessary services.

Knowledge base discusses the collection, organization, and retrieval of the knowledge about the user, the context, and the application.

3.2. Features for adaptation

In user-oriented systems the adaptation of the system behavior to its current user is the main focus of interest. It is common to base the adaptation process on a user model representing the demographical data and other personal characteristics of the user or of shared attributes of the group members, possibly enhanced by the goal (task) model. Additionally, in order to provide adaptive services in mobile and ubiquitous computing a context model has to be added.

The application scenario described in Section 2 demonstrates examples of changing conditions that might trigger an adaptation. These characteristics can be categorized into inter-individual, intra-individual, and environmental differences.

Inter-individual differences address the varieties among several users in multiple dimensions. The characteristics that are primarily user-related and hence relevant for multimedia applications are described in the following:

User preferences: The opportunity to define personal preferences is one of the most popular features of adaptable systems. Without much doubt, multimedia applications will significantly fall down in their acceptance by users without the opportunity to adjust preferences like language, color schemes, menu options, security properties, or modality of presentation and interaction, and numberless other personal favorite preferences.

User interests: Awareness of the user's interests and disinterests can be applied to configure filters for the selection of multimedia content presented to the user. On the one hand, the user can define her interests herself, for example to create personal playlists. On the other hand, genre and categories of selected contents can be used to recommend content the user might be interested in.

Knowledge and expertise: The user's level of factual, general and domain-specific knowledge is a valuable source for adaptive operations. Possible indicators for knowledge and expertise are the level of detail of information the user requests as well as the frequency of usage. Knowledge determines both the content and the interface: the more experience the user acquires, the more she is able to use more complex user interfaces with advanced features and to consume more detailed information.

Intra-individual differences consider the evolution of a single user over time, as well as variable requirements for a certain task at different times. User requirements may change over a period of time, since activities and goals of the user evolve. Thus, the system has to be continuously adapted to conform to the increasing experience and changing interests of the user. For example, a user might be overstrained when first using a system, but lack some possibilities as soon as her expertise is increasing. In the same manner, the need for highly flexible and adaptable systems is driven by changing requirements for the tasks that are to be accomplished with the system.

Environmental differences result from the mobility of computing devices, applications, and people, which leads to highly dynamic computing environments and the integration of context aware functionality, e.g., location information, currently available network bandwidth, capabilities of the current media renderer, or natural environment characteristics. Thus, the user is continuously monitored to guarantee personalized media consumption with the best possible media quality.

The context aware functionality of a user-centric multimedia application needs to be conclusive in order to make the adaptation decisions of the application accessible to the user and to allow an overriding of the behavior. For the user of a converged multimedia application, the lack of conclusiveness bears the risk of getting into situations in which the context aware behavior implemented by the developer is inappropriate or undesired. If the user becomes subordinate to automatic mechanisms, she will abandon useful context aware functionality of the multimedia application [1] [3].

Derived from [16], the following context attributes are important in a user-centric multimedia application:

Places and locations: The characteristics of the user's environment, the presence and the arrangement of objects, people and communication partners, and the technological characteristics of available devices are primarily determined by and strongly depending on the spatial and temporal context of the user.

Environment characteristics: Locations carry context information such as data on temperature, humidity, brightness, and sound levels, acquired by sensors in the environment.

People: Information about people describes people currently in the environment (e.g., identity and user profiles for access control or service personalization).

Technology characteristics: The adaptation to characteristics of the utilized technology reflects the increasing variety of available devices with specific properties, capabilities, software specifications, operating systems, or available media players. The content selection, presentation, and interaction settings have to be adapted depending on the screen or display size, the bandwidth and reliability of the accessible network connection. In particular, for the adaptation of output medium, modality, or device, the observation of the technology available for use plays an important role.

Time: The user's time context consists of all available information about temporal coherences like point in time, length of time, or existing limits in time.

Services: Technically, a service can be anything ranging from a Web service, a Web (or client-server) application, or a functionality of a networked appliance (such as a projector). Sessions contain information about the status of a service instance, including user preferences and choices (e.g., navigation history).

3.3. Context management

In order to develop user-centric applications the considered attributes have to be gathered and managed in a conclusive way [17]. Different approaches for context management have been identified [15]. Context widgets are building blocks aiming to hide the complexity of gathering and managing context information from the application by encapsulating complex sensor functionality. The one-dimensional context information is communicated over the network by messages and callbacks. In service based approaches, on the other hand, clients have to discover and establish a connection to each service providing relevant context information. This architecture lacks a global context management and is thus more flexible, however, each component has to manage an increased amount of functionality, e.g., network connections. Another approach is the blackboard model that is implemented as a shared message board where sensor components provide the context information for the applications. The main part of this architecture is a centralized server that is responsible for routing the messages from sensors to processing components.

4. Content adaptation

Ubiquitous use of media content requires the adaptation of content to the current context. For example, a user wants to avoid watching violent scenes in an action movie, or a visually impaired user receives better audio quality instead of the video stream. Media adaptation is currently a very active research topic. MPEG-21 Digital Item Adaptation (DIA) [4] defines standardized structures for describing factors relevant for adaptation, collectively known as Usage Environment Descriptions (UEDs). UEDs deal with characteristics from the areas covered in Section 3, e.g., terminal capabilities, location and movement, network conditions, and users' perception impairments. Using such descriptions, a variety of adaptation operations can be performed as outlined in the following subsections.

4.1 Types of adaptation operations

Adaptation may involve various kinds of manipulations resulting in the fulfillment of the user's needs. The spectrum of such manipulations has been subdivided in various ways, and so different flavors have been given to the concept of adaptation itself. The existing views on adaptation do either include or exclude certain subproblems. Steps to be applied during content adaptation may belong to one or several of the following categories [10]:

Transcoding changes the coding (compression or container) format of multimedia content.

Scaling produces alternative variations of multimedia content with reduction along one or multiple dimensions, e.g., frame size, frame rate, or color space. Scaling is sometimes subsumed in transcoding [10], especially as scalable coding formats [14] are emerging that allow the application of scaling operations in the binary domain, i.e., without decoding and re-encoding.

Selection, removal and merging operations change the structure and amount of essence contained in a multimedia data entity. Examples of such operations are the removal of additional audio tracks, filtering of violent scenes, and multiplexing a video stream and an audio stream into a composite audiovisual stream.

Summarization and modality change usually result in alternative contents more deviating from the original than this is the case for other types of operations. A typical summarization approach is to extract keyframes from a video and to deliver them as a sequence of images [10]. Modality change produces content in a modality different from the original one, e.g., converting text to speech, or rendering a vector graphics-based animation for a device only capable of playing videos.

Besides these categories, any multimedia editing operation could make sense in certain adaptation scenarios.

In [11], the impact of adaptation on the experience a user gets by consuming content is emphasized. While many adaptations of certain content are possible, the entertaining value or the transport of knowledge could be reduced so far below an acceptable rate that it may be worth considering canceling the content's delivery under the given circumstances.

4.2. Utility-based multimedia adaptation

The approach of utility-based multimedia adaptation [12] [13] aims at maximizing a user's multimedia experience. For every user, a comprehensive utility model is constructed. It is derived not only from usual adaptation constraints such as terminal capabilities and network conditions, but includes also factors such as intuitive rules, the user's judgements, perceptual impairments, demographic features, and favorite content types. An example of a utility-related hypothesis validated in the course of this research is that for different genres, various quality features have different priorities. Adaptation decision taking is based upon values that were statistically obtained from experiments with a number of test users. For finding the optimum adaptation decision, different algorithms were implemented and compared. A heuristic hill-climbing method performed best.

A prototype of the system has been integrated into an HTTP server with a Web GUI. The user can configure her profile either via the GUI or by transmitting an MPEG-21 DIA UED. Adapted versions of the desired content can then be either viewed in the browser or requested as an HTTP download for use with an arbitrary application. This demo system can, e.g., process files that are available on the Web, but it is even possible to adapt DVB streams in real-time.

4.3. Knowledge-based media adaptation

Adaptation is more and more shifting towards user centricity. In the long term, the thus growing set of user characteristics and preferences to be supported requires open and intelligent solutions.

The knowledge-based multimedia adaptation (koMMa) framework [7] [10] can be used to set up intelligent adaptation servers. The idea underlying the framework resulted from the observation that there is a growing number of possible constraints in content adaptation and personalization. Besides well-known transcoding problems such as changing coding formats or spatial and temporal resolution, the user's constraints and preferences may require more complex processes, e.g., various kinds of splitting and merging content in order to select the relevant parts for the user. The conclusion drawn here is that monolithic approaches to adaptation cannot cope with the full spectrum of possible requests, especially as the world of multimedia keeps evolving. Thus, the koMMa framework is designed to dynamically integrate simple, well-defined adaptation operations into an adaptation server. A multimedia content adaptation request is transformed into an AI planning problem, composing complex adaptation tasks from the simple operations. Solutions to such a problem are arrangements of the available operations that can then be executed by the server. The output of this execution will be a version of the multimedia content that suits the user's needs.

Operations can be added to a running adaptation server without any programming effort. Instead, declarative descriptions of operations are required for the planning process. The initial state for the planning problem is automatically derived from MPEG-7 descriptions of the source content. Similarly, goals to be fulfilled by the plan are constructed from MPEG-21 DIA UEDs.

A prototype of the framework is implemented in Java, using SWI-Prolog for the planning process of the Adaptation Decision Taking Engine. Java methods and constructors are described as planner actions. The prototype fulfills MPEG-21 DIA's specification of a DIA Engine: a source Digital Item (consisting of both content data and metadata) is adapted to a destination Digital Item with respect to a given UED.

5. Conclusion

In this paper, personalized user interfaces, context awareness, and multimedia content adaptation were discussed as the most important areas to achieve user-centric multimedia. One of the objectives of the NoE INTERMEDIA is to develop solutions and provide a software framework for the problems addressed earlier in order to make progress toward the vision of user-centric multimedia applications and services.

The previous sections indicated important requirements for the envisioned user-centric multimedia framework. Besides its basic functionality of comfortable sharing and distribution of content to any suitable device, adaptation w.r.t. a multitude of factors must be incorporated. A basic set of such factors considered in many adaptation scenarios are terminal and network capabilities. However, the INTERMEDIA project is focussing on user-centric multimedia convergence and thus attempts to cover in particular the properties and context of the user.

In the field of user-centric multimedia, adaptation and other tasks are subject to a large number of constraints and preferences that are not yet fully identified. Because of this, an extensible knowledge-based approach (as presented in Subsection 4.3) to various decision taking processes is suggested. Further investigation may disclose a subset of factors to be of high importance while others may be almost insignificant. These results might lead to further design considerations regarding more efficient procedural techniques (similar to the utility-based approach presented in Subsection 4.2) or hybrid approaches. Furthermore, AI techniques may enable the system to not only maximize the quality of any requested action but also to recognize a certain type of situation in which a proactive operation could increase the users' perception of a highly user-centric and personalized system.

Acknowledgment

This work is supported by the NoE INTERMEDIA funded by the European Commission (NoE 038419).

References

[1] A. Dey, D. Salber, and G. Abowd. A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. *Human-Computer Interaction*, 16:97–166, 2001.

[2] J. Fink. *User Modeling Servers - Requirements, Design, and Evaluation*, volume 289. Akademische Verlagsgesellschaft Aka GmbH, Berlin, 2004.

[3] K. Henricksen and J. Indulska. Developing context-aware pervasive computing applications: Models and approach. *Pervasive and Mobile Computing*, 2, 2005.

[4] ISO/IEC 21000-7:2004. Information Technology – Multimedia Framework – Part 7: Digital Item Adaptation, 2004.

[5] A. Jaimes, D. Gatica-Perez, N. Sebe, and T. S. Huang. Human-Centered Computing: Toward a Human Revolution. *IEEE Computer*, 40(5):30–34, May 2007.

[6] A. Jameson. Systems that adapt to their users: An integrative overview. In *Tutorial presented at 9th International Conference on User Modelling*, Johnstown, PA, USA, 2003.

[7] D. Jannach and K. Leopold. Knowledge-based multimedia adaptation for ubiquitous multimedia consumption. *J. Netw. Comput. Appl.*, 30(3):958–982, 2007.

[8] A. Kobsa. Generic user modeling systems. *User Modeling and User-Adapted Interaction*, 11(1-2):49–63, 2001.

[9] A. Kobsa and W. Wahlster, editors. *User models in dialog systems*. Springer, New York, NY, USA, 1989.

[10] K. Leopold, D. Jannach, and H. Hellwagner. A Knowledge and Component Based Multimedia Adaptation Framework. In *Proc. IEEE Sixth International Symposium on Multimedia Software Engineering (ISMSE'04)*, pages 10–17, 2004.

[11] F. Pereira and I. Burnett. Universal Multimedia Experiences for Tomorrow. *IEEE Signal Proc. Mag.*, 20:63–73, 2003.

[12] M. Prangl and H. Hellwagner. A Framework for Personalized Utility-aware IP-based Multimedia Consumption. *IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WOWMOM)*, 2007.

[13] M. Prangl, T. Szkaliczki, and H. Hellwagner. A Framework for Utility-Based Multimedia Adaptation. *IEEE Trans. Circuits Syst. Video Techn.*, 17(6):719–728, 2007.

[14] H. Schwarz, D. Marpe, and T. Wiegand. Overview of the Scalable H.264/MPEG4-AVC Extension. In *ICIP06*, pages 161–164, 2006.

[15] T. Winograd. Architectures for context. *Human-Computer-Interaction, Special Issue on Context-Aware Computing*, 16(2-4), 2001.

[16] A. Zimmermann, A. Lorenz, and R. Oppermann. An operational definition of context. In *Proc. 6th International and Interdisciplinary Conference on Modeling and Using Context (CONTEXT '07)*, pages 558–571, 2007.

[17] A. Zimmermann, M. Specht, and A. Lorenz. Personalization and Context Management. *User Modeling and User-Adapted Interaction*, 15(3-4):275–302, August 2005.