International Journal of Digital Multimedia Broadcasting

Personalization of Mobile Multimedia Broadcasting

Guest Editors: Harald Kosch, László Böszörményi, Günther Hölbling, David Coquil, and Jörg Heuer

Personalization of Mobile Multimedia Broadcasting

International Journal of Digital Multimedia Broadcasting

Personalization of Mobile Multimedia Broadcasting

Guest Editors: Harald Kosch, László Böszörményi, Günther Hölbling, David Coquil, and Jörg Heuer

Contents

Hindawi Publishing Corporation
International Journal of Digital Multimedia Broadcasting
Volume 2008, Article ID 238073, 6 pages
doi:10.1155/2008/238073

Editorial
Personalization of Mobile Multimedia Broadcasting

Harald Kosch,[1] László Böszörményi,[2] Günther Hölbling,[1] David Coquil,[1] and Jörg Heuer[3]

[1] *Department of Informatics and Mathematics, University of Passau, 94030 Passau, Germany*
[2] *Institute of Information Technology, University Klagenfurt, 9020 Klagenfurt, Austria*
[3] *Siemens AG, 80333 Munich, Germany*

Correspondence should be addressed to Harald Kosch, harald.kosch@uni-passau.de

Received 18 August 2008; Accepted 18 August 2008

1. INTRODUCTION

This special issue is devoted to a well-focused subject: personalization of mobile multimedia broadcasting. Nevertheless, the topics of the papers published here demonstrate an amazing diversity. This phenomenon suggests that our subject is both highly relevant and experiencing a period of rapid change. Until recently broadcasting has been a well-established, relatively stable technology. However, new usage scenarios, mobile consumers together with mobile devices, and the desire for personalized content are providing new challenges. We currently have many more questions than answers.

We are confronted with a range of subtly different techniques, such as digital TV, IPTV, video-on-demand, Web-TV, live casts, mobile TV, peer-to-peer TV, and video-portals, which use different encoding/decoding standards, transmission protocols, streaming methods, quality-of-service levels, and interactivity features. In addition, they often require different bandwidth and different infrastructure.

In view of this diversity, it is sensible to take a fresh look at the basic concepts. The rest of this special issue is dedicated to presenting a nice selection of timely, ongoing research. Therefore, this editorial introduction starts (Sections 2–4) with a contextual overview authored by László Böszörményi. This overview concentrates on "the past and the future of this topic"—leaving the present, together with all of its unsolved questions, to be the subject of the rest of the papers.

An overview of the different contributions in this special issue closes this editorial introduction in Section 5.

At first glance, broadcasting and personalization seem to contradict one another. The idea of broadcasting is to transmit a message from an authority to everybody; the idea of personalization is to exchange messages between individuals. Broadcasting offers a high degree of sharing and a low level of privacy. Personalization, on the other hand, usually offers the opposite: privacy increases, but sharing decreases. There are a number of basic issues requiring very different, often contradictory treatment, and strategies. May be the most important of these issues are (1) authenticity and popularity, (2) personalization and privacy, (3) sharing, (4) interactivity, and (5) rights management.

2. A VIEW AT THE PAST

2.1. A bit of ancient (mainly european) history

The idea of broadcasting might have its roots—as almost everything—in the attitude of the ancient Greeks, interpreting the thunderbolt as an expression of the anger of Zeus. This message was *authentic*—coming from the main god directly—and everybody could perceive it—actually *had to* perceive it. There was no way not to listen to Zeus's "word" and thus, it did not leave much room for *privacy*. There was, however, room for many *different interpretations*.

Zeus used the air as a common, *shared* medium, making this kind of communication very efficient and reliable. If Zeus repeatedly created thunder, this only emphasized his anger. Thus he used a combination of acoustic and visual signals, ensuring that it was impossible not to listen to the acoustic expression of his anger, even with the eyes closed. The combination of these two modalities still plays an important role today.

The first sign of *personalization* lies in the diversity of Gods, thus allowing at least a choice among them. In the great war between Greeks and Trojans, the Greeks, especially Odysseus, followed the goddess Pallas Athena, whereas the Trojans were advised by Aphrodite. Communication

channels were shared, but with a limited *radius*. Privacy was higher and realized though the relative freedom of selecting which god to follow.

In the ancient Jewish religion, God speaks often personally and in secret to selected persons. He can be heard but cannot be seen. He speaks a more sophisticated and understandable language than that of thunderbolts. Even *interaction* is often possible. A choice among gods is, however, not allowed as it is a monotheistic religion. Greek mythology also permitted some people to *interact* personally with gods. Interestingly, in these cases the corresponding god took the figure of a human. For example, Odysseus meets Pallas Athena in the form of a young swineherd and his son Telemachus meets the same goddess in the shape of King Menelaus. Incidentally, they both recognize the presence of Pallas Athena by the phenomenon that their partner appeared in a supernatural beauty. This suggests that a personal conversation with a god was seen as something *beautiful*, whereas an impersonal message, such as thunderbolt, was *frightening*. Regardless of this, the message was still very authentic, although personal. Although *sharing* of communication channels disappeared for the sake of *interactivity*, a certain level of sharing was still available, as some gods, such as Pallas Athena, had the admirable capability to appear at two different physical places at the same time—we would say a kind of *virtual replication*.

Greek gods were omnipresent, and therefore *mobility* was not a problem. Greek people were extremely mobile and could listen to their favorite god or goddess everywhere as, unlike some later people, they did not need a special church for this. Communication seemed to work without difficulties also among people. The Greeks (consisting of many small groups of people) and the Trojans were at war, but they never experienced difficulty in speaking with or understanding one another. Unfortunately, we do not know in which language they communicated.

2.1.1. Some medieval history

The second major step in the history of broadcasting was presumably the invention of printing by Gutenberg in the fifteenth century. Previously, visual material on paper (or clay, etc.) had to be physically replicated and transported in order to be broadcasted. This was extremely expensive. Copying a book manually could take a small group of monks a year or more and bringing it to a different monastery often took several weeks. The printed book and especially the invention of the newspaper was a revolution in the technology of broadcasting. Compared to the ancient Greeks, we can observe a number of changes. Authenticity starts to decline. Although it was originally only the Holy Bible that was replicated, fairly soon a large number of publishers with different levels of authenticity appeared on the scene. Authenticity was step by step replaced by *popularity*. A thunderbolt had to be perceived regardless of whether people liked it or not. A book or a newspaper must be bought; it must be liked or "popular." Whether they are still "true"—authentic—is another question. This obviously led to a certain degree of contradiction and competition. At the same time, the level of personalization starts to grow. People have a rich selection of choices. They also have the opportunity to become publishers, a process that is definitely easier than it was for Greeks to become a god (or at least a half-god). At the same time, the issue of right management emerges: authors of books and newspapers want to have some control over their publications. Previously they remained even anonymous—their only reward was in the eternity of god. This changed radically in the new age.

2.1.2. Radio, TV, telephony

The next revolution in broadcasting was the appearance of analogous radio at the beginning of the 20th century and that of television somewhat later. These "classical" broadcast media resemble in many senses the communication paradigm of the Greek gods. They transmit a central, *authentic* message, essentially and continuously; the only "escape" for listeners is to change the channel or to switch off the receiver. Beyond channel selection, no interactivity is provided. Each channel has fixed bandwidth, a fragment of the spectrum. This channel is *shared* by all listeners of the same channel, making broadcasting highly efficient. The senders themselves share the "air" as a common medium, but try to avoid any kind of further sharing. Even though they often transmit more or less the "same" information, they try to do this in a different form—as competition and profit are the basic principles keeping them alive.

The first steps in switching from analog to digital technology tried to maintain the traditional view of authenticity, based on a limited number of highly trustable senders. As for sharing, new competitors have emerged, especially the Internet. Interestingly, for a while, wireless broadcasting was considered "old-fashioned" technology, as the new technology was wired. This situation has changed once again.

A further, very important aspect was the appearance of telephony roughly at the same time as analogous radio, the most important technological step in the development of personal communication. Railway, beginning with its modern form in the middle of the 19th century, can be regarded as similarly important; however, transportation is not primarily devoted to communication. Telephony allows people to communicate with each other synchronized in time while being released from the constraints of space. Analog telephony relies on circuit switching, giving their customers the illusion of having *private* connections while at the same time intensively *sharing* the same cabling system. Privacy is principally provided, but at that time everybody is aware of the fact that even private conversations may have uninvited listeners, not even necessarily on purpose but rather due to usual errors. Telephone conferences and broadcasting remain rather rare applications. It is interesting to note that the Internet has been strongly connected with "plain, old" telephony from the beginning through its use of the telephone system as a transportation medium.

Furthermore, extremely significant step is the appearance of wireless technology both in computer networking and telephony. The idea of "ubiquitous" computing emerged in the eighties, strongly related with pervasiveness and mobility.

This is—very roughly sketched—the situation in which ongoing research finds itself. A large number of papers have been published in recent years, addressing a lot of the issues of this big picture. Instead of trying to reflect on this diversity, we ask the following question: what is the future of combined broadcasting, personalization and mobility?

3. AN ATTEMPTED GLIMPSE INTO THE FUTURE: THE "FRENCH-REVOLUTION" OF BROADCASTING

We assume that in the future virtually everybody may broadcast any kind of message at any time and can of course also receive any such message at any time, at any place, equipped with any kind of device. We could say that broadcasting will become *democratic*. That shatters the fundamentals of broadcasting, as broadcasting is—as we tried to show—basically *undemocratic*. Therefore, the development of personalization, mobilization, and enhanced flexibility is not just an option—but a necessity. In the near future, we can expect radically new usage patterns to arise, characterized by the following main features.

(1) Digital multimedia will be produced by many sources and injected into a fully distributed and multimodal environment at many different locations.

(2) A huge "web" of multimedia data will be produced and consumed with various aims and requirements.

(3) Beside entertainment, professional use of digital multimedia will grow considerably.

(4) Production and consumption of multimedia data will be better integrated into the computing environment than is the case today.

In such a world, production, search, access, delivery, processing, and presentation of multimedia data must become much more flexible; in many cases it must become "spontaneous." While spontaneity is the enrichment in everyday life, it is extremely hard to apply to technology. Thus research will be confronted with new challenges. Let us consider a few example scenarios of this new multimedia world.

3.1. Some usage scenarios

3.1.1. Live event with professional and amateur producers and consumers

A first example is a live event—such as the "Iron Man" competition, where a few thousand athletes are competing in swimming, biking, and running on a large but limited geographical area for several hours—followed by tens of thousands fans. A huge number of still and moving pictures are created by a variety of sources, including some professional camera teams, static surveillance cameras, and a great number of private people equipped with very heterogeneous photographic abilities. In addition, people with a wide range of interests would like to consume these pictures. Many of the consumers are watching just for fun, some others in order to track a certain participant and yet others, such as the event organizers, are watching to obtain a global view of the whole competition. How can these users find easily and exactly what they need, without being bothered by long sequences that they are not interested in? How can they get the required content without substantial delay and in exactly that quality they require (neither better nor worse)? Currently, there is no system that is able to cope with (or even approximately cope with) such a complex and spontaneous world.

3.1.2. Public motorway equipped with sensors and cameras

A company operating public motorways equipped with thousands of sensors and hundreds of cameras actually produces more broadcast material than a number of TV channels together. This material is obviously not of a trivial or entertaining nature; nobody wishes to watch traffic on motorways for days or even hours. What is needed, is a system which automatically identifies interesting events and offers them to the users (typically professional staff of the company, may be police and ambulance officers, or even public users planning their routes) to observe and evaluate. In many cases, the pictures of one single camera do not suffice, a group of cameras and related sensors should be identified, delivering relevant data for a certain, important event (traffic jams, accidents, etc.) or for enabling a global view on a major section (e.g., traffic in a certain area is quiet, whereas hectic at another, connected area). Current systems are still very far from providing such complex services.

3.2. Popularity management, as a compromise between sharing and personalization

We have known for many years that popularity of videos essentially follows the laws of Zipf and Pareto. This means—albeit overly simplified—that roughly 20% of all videos, stored somewhere accessible over the Internet, will be downloaded or streamed more than once. The remaining 80% will remain essentially unused. What remains unreported is that the same laws hold for the scenes inside videos. That is, only portions amounting to 20% of all downloaded videos will be watched. Putting these two observations together, we come to the result that ca. 4% of all video material is watched by a second person (beyond the author), the rest is just there. However, in the resource management, this issue is hardly considered and even if, then at most on the level of the popularity of entire videos, but hardly on the level of individual scenes. Efforts are typically made to provide good resource management for the entire 100%—although instead what we need is effective management of the relevant 4%. Even if the resource management takes popularity into consideration on the level of entire videos and even if techniques, such as partial caching, do consider popularity on the level of scenes to a certain extent, a huge potential for savings, up to two orders of magnitude, still remains. The difficulty is of course obvious as follows: we usually do not know which 4% should be supported. Therefore, we need a new model of video delivery. Instead of viewing videos as sequential streams of data (resembling the video-tape

paradigm), they should rather be regarded as direct-access media. Direct accessibility in several dimensions include that users should get exactly (1) *what* they need, (2) *when* they need, and (3) *how* (in what quality) they need.

This implies a transition to a flexible management in heterogeneous environments. The above observation may open a new chapter in combining *sharing* and *personalization*. Even though classical broadcasting (sending the same content to everybody) cannot work in future, even "democratic" systems can efficiently share resources by carefully tracking popularity. To explore this, let us take a look on such a possible, future delivery system. We make the following basic assumptions for a new model of video delivery.

3.2.1. Nonlinear video delivery

We assume that videos are rarely watched sequentially. In many usage scenarios, and especially professional situations, people want to quickly find certain scenes and avoid watching long sequences they are not interested in.

3.2.2. Two-phase delivery

We distinguish between video *offering* and video *delivery*. Offering should be fast, interactive and should provide information about the videos available within a certain context. During the offering stage, the underlying resource management system should be able to make preparations for an efficient delivery. We could use a restaurant as a metaphor. In a good restaurant, guests are served essentially in two main phases. In the first—the offering phase, they receive the menu and some appetizer very quickly. This enables them to make their favorite choices comfortably and also leaves time for the kitchen to be prepared which represents the second, the delivery phase—the main dish. In a video delivery system, the whole issue is much more complicated. We might have many "cooks" (video providers) and many guests and they may even change their roles (and places). Offering and delivering are overlapped activities all the time. Offering is push-based, that is, the service provider more or less "aggressively" announces meta-information about the available content. This must be very fast, as studies show that it is better to present to the clients something they did not explicitly require than to present nothing (or a rotating "hourglass"). "Real" content can be delivered pull-based (or in a hybrid way).

3.2.3. Video composition/decomposition

Data should be decomposed into units of "meaningful" size (how large meaningful is depends on the actual context) and can be composed under quality-of-service (QoS) constraints into continuous "movies." For performance reasons, the decomposition may be performed in a lazy way, in order to avoid decomposing data which is never used or is only used for "traditional" streaming in its entirety

A traditional, long, continuous video is defined as a "special case" as a sequential composition of data units, under certain QoS constraints (e.g., 25 frame/sec, jitter <10 msec). The interesting point is that we may compose any data units in any order under arbitrary QoS constraints. The user becomes thus from a passive consumer to an active "composer." This does not mean that the interactive human user always has to take the burden of the composition: predefined profiles of user-classes may serve as composition patterns. What is essential is that the system basically supports free and flexible composition. In real usage scenarios, full freedom must of course be reasonably restricted.

Decomposition and composition obviously come at a price. Using them only to support traditional usage patterns is hardly a good idea. However, if we assume that videos are rarely watched sequentially from the beginning to the end, rather certain important or "popular" parts are watched often, then we are confronted with new optimization possibilities. Popular parts may be replicated much more intensively than others. Moreover, the same data might be replicated in different qualities, following different replication strategies. Optimal data management is a most challenging research question.

3.3. Self-organizing delivery

The delivery system should strive for self-organization. Each node of the delivery system (no matter whether a server, a proxy, or a P2P client) can follow a simple, local *goal-function*. A goal is a state of affairs where an optimal utility value is reached and stabilized over a certain period of time. For example, a proxy could have the local goal of maximizing of its own throughput by building groups with other proxies sharing the same kinds of video segments. Even data units might have goals (e.g., to be replicated somewhere). The system has a required global behavior, expressed as a global goal. In an *ideal* self-organizing system, this global goal emerges as a result of the local behaviors. In practice, some parts of the global behavior might be controlled in a non-self-organizing manner, leading to a *nonideally* self-organizing system.

3.4. Trust management as a compromise between authenticity and popularity

Who will broadcast in the future? The answer is simple: virtually everybody. This leads immediately to the question of *authenticity*. How will we be able to decide on the value of the received information, if the sender is not necessarily trustable, may be not even known? This dilemma is of course already very well known, as demonstrated, for example, by discussions on the value of Wikipedia entries. This becomes more difficult if the information changes rapidly, as is the case in live events. This already occurs in some extreme cases, for example, in the case of natural catastrophes, where pictures and reports of eye witnesses are of high value, even though the technical or artistic quality is low. If pictures are taken at such an event, then they usually reach a trusted broadcaster via more or less "private" communication, who subsequently checks them as far as possible, before publishing them.

It would be, however, much better, if future broadcasting systems would offer well-defined services (1) to submit input messages "spontaneously," (2) to check them for authenticity and to assign a certain level of trust, and even (3) to offer a way of rewarding the providers of such input. Authenticity or trust management must become an integral part of future services.

3.5. Metadata management as a compromise between sending everybody the same versus sending everybody something else

What will be broadcasted in the future? The *answer* is once again simple: virtually everything. The content will be multimodal including continuous data. Moreover, as the previous considerations show, it is not enough to deliver pure data; we need additional information, generally called metadata. Level of trust is—for this aspect—an example of metadata. Current scientific literature on multimedia delivery concentrates almost exclusively on the delivery of "real data." If metadata is required, its availability is simply "assumed." (A good example for this is the MPEG-7 metadata standard, leaving the delivery of metadata simply out of scope.) However, in dynamic scenarios, as described above, users have no chance to get the data they need without sophisticated metadata management. Much more than a simple electronic program guide (EPG) is needed. As long as one has the choice between two public TV channels, the selection is relatively easy. If a user has to choose among 200 channels, then the decision is harder. If a user has to choose between thousands of sources, some of which cannot be properly identified but useful, then a radically new technology is required. In Sections 3.2.2, we introduced the idea of a two-phase delivery, consisting of an offering and a delivery phase. This is a possible, partial solution for the general issue is that the metadata management must be an integral part of any future broadcasting system.

Also digital right management (DRM) belongs to the same category. The MPEG-21 standard offers the necessary tools for interoperable DRM. Why its acceptance is lagging behind the expectations is one of the questions which are harder to answer. Nevertheless, in the long term, we can assume with certainty that a business model will be generally accepted that enables consumers to access digital content as freely as possible and producers not to starve. There seems to be no alternative. Even if everybody has the possibility of becoming a broadcaster, no one is likely to agree with starvation.

4. CONCLUSIONS

Broadcasting is indeed in the state of a revolution. Our well-known and well-understood concepts have to be revisited.

(1) The idea of a kind of "divine" authority and authenticity will be replaced by the "democratic" notion of popularity, tightly coupled with integrated trust management.

(2) The traditional view of personalization based on the free selection between a small number of channels is definitely outdated. The user must get what she needs, when she needs it and how (in which quality) she needs it. Especially she should *not* be presented with content that she does *not* need. The traditional view of privacy, of being encapsulated in a kind of sand-box, will also disappear in the future. Future broadcast systems will need to be able to switch dynamically between private and public data and contexts.

(3) The sharing of resources based on sending everybody the same content is outdated. This can be efficiently replaced by a delivery model that shares information on the *popularity* of data and that subsequently favors popular data. This promises a good compromise between share-everything and share-nothing standpoints.

(4) Interactivity will become a central issue. Not only in the sense that consumers must receive very detailed metadata, which serves as a basis for making qualified selections, but also in the sense that everybody may change from being a consumer into a producer and vice versa.

(5) Rights management is probably not a workable concept and should be replaced by business model. Valid business models, enabling the highly flexible scenarios as described previously, without hurting the interests and rights of either producers or consumers must emerge soon.

5. OVERVIEW OF THE CONTRIBUTIONS IN THIS SPECIAL ISSUE

This special issue presents a selection of state-of-the-art research works in the domain of mobile multimedia broadcasting (MMB) with a focus on personalization.

In the first paper "Acceptance threshold: a bidimensional research method for user-oriented quality evaluation studies," S. Jumisko-Pyykkö et al. present a survey of state-of-the-art methods of acceptation assessment based on subjective user feedback, and study their validity in the context of mobile television. Personalized multimedia applications need to make use of multimedia adaptation methods. Two papers of the special issue present contributions in this domain.

In the second paper "Adapting content delivery to limited resources and inferred user interest," C. Plesca et al. present adaptation policies specifically designed for highly dynamic and partially or fully observable contexts typical of mobile environments with an application to film browsing service.

In the third paper "Efficient execution of service composition for content adaptation in pervasive computing," Y. Fawaz et al. propose a method for executing multimedia documents adaptation plans based on composition of services.

In the fourth paper "Two-level automatic adaptation of a distributed user profile for personalized news content delivery," the authors present a work that pertains to two major issues of the domain. The first one is the implementation of personalization features in the specific concept of mobility. The second one is the collecting and usage of user feedback in order to offer a better personalized service, which in this case is implemented using machine learning techniques. An important application domain for MMB services is the home

multimedia environment, in which Universal Plug and Play Audio Visual (UPnP-AV) devices are often used.

In the fifth paper "Context-aware UPnP-AV services for adaptive home multimedia systems," M. Papadogiorgaki1 et al. propose an enhancement of UPnP-AV that enables the adaptation of multimedia content based on contextual information. In order to offer optimal personalization features to their users, new MMB applications need to go beyond traditional adaptation methods based on parameters such as image size, color scale, bitrates, and so forth. by implementing finer-grained adaptation features. Two examples of such applications are presented in this issue.

In the sixth paper "Region-based watermarking of biometric images: case study in fingerprint images," K. Zebbiche et al. propose a personalization method of biometric images using region-based watermarking. In the last paper "Extracting moods from songs and BBC programs based on emotional context," M. K. Petersen and A. Butkus make an initial contribution toward the goal of emotion-based personalization by showing how moods can automatically be extracted from songs.

Harald Kosch
László Böszörményi
Günther Hölbling
David Coquil
Jörg Heuer

Hindawi Publishing Corporation
International Journal of Digital Multimedia Broadcasting
Volume 2008, Article ID 712380, 20 pages
doi:10.1155/2008/712380

Research Article

Acceptance Threshold: A Bidimensional Research Method for User-Oriented Quality Evaluation Studies

S. Jumisko-Pyykkö,[1] V. K. Malamal Vadakital,[2] and M. M. Hannuksela[2]

[1] *Tampere University of Technology, Human-Centered Technology, P.O. Box 553, 33101 Tampere, Finland*
[2] *Nokia Research Center, P.O. Box 1000, 33721 Tampere, Finland*

Correspondence should be addressed to S. Jumisko-Pyykkö, satu.jumisko-pyykko@tut.fi

Received 5 March 2008; Accepted 17 July 2008

Recommended by Harald Kosch

Subjective quality evaluation is widely used to optimize system performance as a part of end-products. It is often desirable to know whether a certain system performance is acceptable, that is, whether the system reaches the minimum level to satisfy user expectations and needs. The goal of this paper is to examine research methods for assessing overall acceptance of quality in subjective quality evaluation methods. We conducted three experiments to develop our methodology and test its validity under heterogeneous stimuli in the context of mobile television. The first experiment examined the possibilities of using a simplified continuous assessment method for assessing overall acceptability. The second experiment explored the boundary between acceptable and unacceptable quality when the stimuli had clearly detectable differences. The third experiment compared the perceived quality impacts of small differences between the stimuli close to the threshold of acceptability. On the basis of our results, we recommend using a bidimensional retrospective measure combining acceptance and satisfaction in consumer-/user-oriented quality evaluation experiments.

1. INTRODUCTION

Consumer acceptance is a critical factor in the adoption of new mobile multimedia products and services. Acceptance is defined as the minimum level of user requirements that fulfills user expectations and needs as a part of user experience [1, 2]. User experience as a broad concept refers to a consequence of user's internal state, characteristics of designed system, and the context within interaction occurs [3]. Modern mobile services are collective results of several product elements and combine the effort of several players in a field from content owners, producers, and service providers to platform developers [4]. In the product development process, the quality of critical components is adjusted or optimized separately from the end-product or prior to the completion of the end-product. For example, in streamed mobile multimedia, the quality of network connection may represent one of these elements. To ensure that qualities of components developed in isolation are not barriers to the adoption of end-products, their acceptability should be studied in their optimization process.

In the development of signal or system quality as product components, subjective quality evaluation experiments are conducted. Subjective quality evaluation, also called perceptual, affective, or experienced quality evaluation, or even more broadly referred to as sensorial studies, is based on human judgments of various aspects of experienced material based on perceptual processes [5–7]. For the consumer-oriented critical product component assessment, an overall quality evaluation approach is appropriate. It is suitable for the evaluation of multimodal and heterogeneous stimuli [5, 7], and also assumes that human knowledge, expectations, emotions, and attitudes are integrated into quality perception [5, 7]. The overall evaluation approach has been applied in subjective quality evaluations of mobile television to study different codecs, audio-video compression parameters such as frame rates, bitrates, and screen sizes [8–10].

Subjective overall quality is mainly measured as an affective degree-of-liking, whereas only little attention has been paid to acceptance of quality. Subjective quality is usually measured as one-dimensional satisfaction based on

the methodological recommendations of the International Telecommunication Union [11]. Recently, the Quality of Perception (QoP) model has been proposed to combine two dimensions, namely, satisfaction and cognitive information assimilation, into one measure of subjective quality [12, 13]. However, these methods have not paid any attention to acceptance of quality. There are only few studies in which measures of acceptance have been reported [14]. However, no extensive theoretical background has been presented. Furthermore, these methods are applicable only when the quality is close to the acceptance threshold, and are not discriminative above or below the acceptance threshold, that is, the methods cannot be applied for the comparison of good qualities. These approaches necessitate changing the data-collection method for the duration of quality evolution. In sum, there is a clear need to develop an overall quality evaluation method of acceptance to ensure fulfillment of user minimum quality requirements in quality optimization and to provide comparability between studies independently of levels of quality under continuous technical development.

The aim of this paper is to develop research methods for assessing overall acceptance of quality. We present a literature review of acceptability and research methods as a basis for development in Sections 2 and 3. We conduct three experiments to develop and test validity under heterogeneous stimuli in the context of mobile television. The first experiment examines the possibilities of using a simplified continuous assessment method for assessing overall acceptability. The second experiment explores the perceived boundary between acceptable and unacceptable quality in four error rates having clearly detectable differences between stimuli. The third experiment compares the impacts of four different error control methods on perceived quality close to the threshold of acceptability with small differences between the stimuli. Finally, we present a discussion on all the experiments, provide recommendations for use of the methods, and conclude the study in Section 7.

2. MULTIMEDIA QUALITY

Multimedia quality is a combination of produced and perceived quality. Produced quality describes the technical factors of multimedia which can be categorized into three different abstraction levels, called network, media, and content [15, 16]. Perceived quality represents user's or consumer's side of multimedia quality, which is characterized by active perceptual processes, including low-level sensorial and high-level cognitive processes. A typical problem in multimedia quality studies is to optimize quality factors produced under strict technical constraints or resources with as little negative perceptual effects as possible.

2.1. Produced quality

Huge amounts of data, limited bandwidth, vulnerable transmission channel, and constraints of receiving devices set specific requirements for multimedia produced quality. Network-level quality factors describe data communication over a network and are often characterized by loss, delay, jitter, and bandwidth [15, 17, 18]. Network-level quality factors are discussed in greater detail in the subsequent paragraphs as they have a central role in this paper. Media-level issues include media coding for transport over the network and rendering on receiving terminals [15]. Recent studies on media-level quality factors have addressed the compression capability of codecs [19, 20], temporal factors in terms of video frame rates [13, 19], spatial resolution [9, 10], bitrates, spatial factors (e.g., monophonic or stereophonic sound), and temporal parameters of audio, such as sampling rate [20]. Increasing interest has been expressed in the topic of audio-video factors, like skew between audio and video streams [21] and shared resources between the streams, like bitrates [8, 9, 19, 20], and audiovisual transmission error control methods [22, 23]. The content level quality factors concern the communication of information from content production to viewers [15]. The topics studied include impacts of content manipulations [24], content comparisons (e.g., [8, 10, 13]), and text size [25]. High level of optimization, especially in the network and media levels, can cause noticeable degradation in perceived quality.

Network-level quality factors relate closely to imperfections of transmission channels. In fact, erroneous transmission of data may occasionally occur in any transmission channel. The causes of errors depend on the transmission channel and its characteristics. For example, in many wired-line networks, the main causes of errors are queue overflows at network nodes, while in a wireless network, the main cause of data corruption is due to the physical characteristics of the radio channel. Furthermore, the statistical characteristics of errors may also vary. They may be either isolated individual errors, burst errors, or a combination of both. Therefore, any methods to resolve errors in a transmission channel must take into consideration the cause of error as well as the nature of error that corrupts the data.

In wireless channels, the radio channel properties, such as interference from other cochannel signals, multipath propagation due to signal reflection from different natural, and man-made structures in the vicinity of the receivers, together with fading are the major causes of errors. If the receiver is a mobile terminal, errors may also occur due to the Doppler effect caused by the speed of the receiver. These errors typically occur as bursts rather than isolated individual errors [26, 27]. The nature, frequency, and duration of errors may vary regardless of the cause of errors.

Broadcast services typically fix transmission errors with forward error correction (FEC) coding, such as Reed-Solomon FEC codes [28]. FEC repair symbols are appended to the actual data such that when errors are encountered, the combination of the data and the FEC repair symbols can be used to obtain the correct data. The correction capability of FEC codes is limited, however, and once the number of transmission errors exceeds the correction capability of the FEC code, typically no lost data can be recovered. Consequently, the use of FEC codes causes an abrupt threshold between produced quality free of network-level errors and severely impaired quality due to transmission errors.

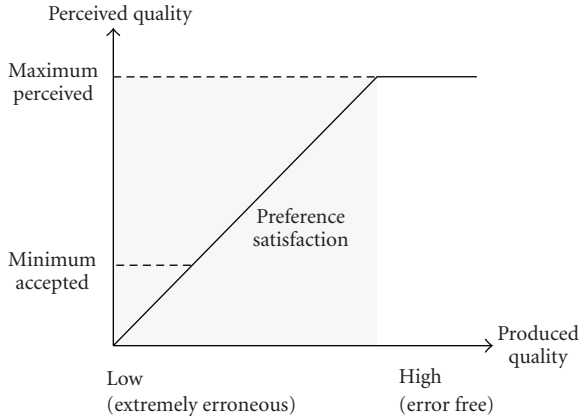


Figure 1: The levels of produced and perceived quality.

2.2. Perceived quality

Quality perception is constructed in an active process. Early sensory processing extracts relevant features from the incoming sensory information. In vision, brightness, form, color, stereoscopic, and motion information are distinguished in the early perceptual process while pitch, loudness, timbre, and location are attributes of auditory processing [29, 30]. However, the final quality judgment is always a combination of low-level sensorial and high-level cognitive processing. In cognitive processing, stimuli are interpreted through their personal meaning and relevance to human goal-oriented actions. This process involves individual emotions, knowledge, expectations, and schemas representing reality, which affect the importance of each sensory attribute and more broadly enable human contextual behavior and active quality interpretation [31–33]. For example, quality evaluations are not restricted to the characteristics of interpreted stimuli. The assessment of usefulness or fitness to purpose of use is included in human evaluations of quality [34].

2.3. Levels of produced and perceived quality

Multimedia quality can be presented as a relation between produced and perceived quality. We present this relation by applying basic conventions of psychophysics (originating from Fechner 1860 overview, e.g., [7, 35]), but widening the view to actual user quality requirements. The quality produced may have a wide range from low and extremely erroneous to extremely high fidelity and error-free presentation (Figure 1). However, the human perceptual processes cannot detect all levels of produced quality. In addition, the whole quality range is not appropriate for the consumer products.

When the produced quality is extremely high, the threshold of maximum perceived quality is reached. This means that an increase in produced quality does not improve the perceived quality since the differences in produced quality become undetectable and impossible to recognize. In psychophysics, this is called terminal threshold [7]. In consumer products, top-end multichannel audio or high-definition visual presentations may reach these thresholds under certain rendering constraints in the near future.

Below the maximum perceived quality, the levels of produced quality can be organized into orders of preference if the difference threshold between the stimuli is reached. Perceived quality at this stage represents satisfaction or pleasantness. Preferences can be compared until the stage, at which the decrease of produced quality no longer decreases, perceived quality. The lower edge of detection and recognition threshold is reached [7]. Produced quality that is close to lower thresholds is not appropriate for studying consumer products or services.

Discrimination testing is used to gather data on conventional thresholds. There are different types of discrimination tests and their further applications, such as method of limit, constant stimuli, and adjustment. Common to all of these methods is the binary data collection form. Either there is sensation or there is not "no sensation or yes, I perceive something" [7, 35] .

We assume that there are also other types of meaningful thresholds between those located at the extremes of perceived quality. When the produced quality approaches the level of very poor and erroneous presentation, there is the area of minimum acceptable quality within the perceptual preferences. The concept of minimum accepted quality can be expected to be relevant in system quality assessments for consumer electronics as an indicator of useful level of produced quality and as an anchor for user requirements. A more detailed conceptual presentation for acceptability is given in Section 3 from the perspectives of acceptance as technology adaptation and acceptance as sensorial experience.

3. ACCEPTANCE AND QUALITY EVALUATION METHODS

3.1. Technology acceptance—the wide audience approach

In the broadest sense, acceptability refers to the market decision whether to accept or reject products or services characterized by willingness to acquire the technology, use it, and pay for it [36, 37]. This approach is popular in the fields of consumer studies and human-computer interaction. In one of the most widespread theories, called the Technology Acceptance Model (TAM), factors predicting the intention to use information system and adoption behavior are formed [38, 39]. TAM was originally developed to measure the acceptance to use information systems for mandatory usage conditions, but later, it was adapted and modified for consumer products and mobile services (e.g., [40–42]).

In TAM, the main predictors of behavioral intention to use the tested technology are usefulness and ease of use. Usefulness refers to the degree to which a person believes that a certain system will help perform a certain task. Ease of use is defined as a belief that the use of the system will be relatively effortless. Low produced quality may be one of the obstacles in the acceptance of technology [38, 39]. In the context

of mobile multimedia, failures of produced quality factors, such as screen size and capacity, interface characteristics of mobile devices, wireless network coverage, as well as capabilities and efficiency of data transform [40, 42–44], may have indirect effects on usage intentions or behavior by affecting perceived usefulness and ease of use [38, 39]. From the broad viewpoint of acceptability, subjective quality evaluation experiments on certain techniques should ensure that perceptually minimum accepted quality level is reached for the developed information systems or services to be an enabler of wide audience technology adaptation.

3.2. Quality evaluation methods

Subjective quality evaluation experiments are conducted for signal or system development purposes. Information about these studies is used in the optimization of a system, like network or media parameters, or in the development of objective metrics. In the literature perceptual, hedonic, or experienced quality evaluation are typically used as synonyms for these measures depending on the different emphases [5–7]. These studies are conducted in a controlled environment to ensure a high-level of control over the tested variables and repeatability of measures. For consumer-oriented quality evaluation, overall quality judgments are used. Evaluations of excellence of stimuli are based on human perceptual processes. As the evaluations are based on human perception of the excellence of stimuli, knowledge, expectations, emotions, and attitudes are integrated into the final quality perception of stimuli [5, 7]. The overall quality evaluation can be used to evaluate heterogeneous stimuli material (e.g., multimedia) because it is not restricted to the assessment of a certain quality attribute, such as brightness, but rather based on a holistic view of quality [5].

There are three main approaches to evaluate subjective perceived overall quality which can be applied in the measures of relatively low produced multimedia quality. A summary of the essential properties of the methods is given in Table 1. The International Telecommunication Union Recommendation [11] provides a reliable research method called Absolute Category Rating (ACR), which is applicable for performance or system evaluations with a wide quality range [11]. In ACR, also known as the single-stimulus method, test sequences are presented one at a time and rated independently and retrospectively. The short stimuli materials and mean opinion score (MOS) using labeled scales to set the evaluations into order of preference in ACR. One of the ultimate aims of method development has been to create a very reliable subjective method providing comparable data for the construction of objective or instrumental multimedia quality evaluation metrics [11]. It is maybe not surprising that the method is especially widespread in engineering.

Quality of Perception (QoP) is a user-oriented concept and evaluation method combining different aspects of subjective quality introduced by Ghinea and Thomas [12, 13]. QoP is a sum of information assimilation and satisfaction formulated from dimensions of enjoyment and subjective, but content-independent objective quality (e.g., sharpness). Information assimilation is measured with questions on audio, video, or text in different content and in the analysis right answers are transformed into the ratio of right answers per number of questions. Both satisfaction factors are assessed on a scale 0–5. Final QoP is the sum of information assimilation and satisfaction that sets the stimuli into order of preference. Both ARC and QoP result in subjective evaluations in the form of a preference order and can be applied in studies on low produced quality, but they are not restricted to it. However, these methods do not indicate any threshold of acceptance among these preferences.

McCarthy et al. [14] tackle the problem of quality acceptability on the basis of the classic Fechner psychophysical method of limit. The threshold of acceptance is achieved by gradually decreasing or increasing the intensity of the stimulus in discrete steps every 30 seconds. At the beginning of the test sequence, participants are asked if the quality is acceptable or unacceptable. While watching, participants evaluate quality continuously. They report the point of acceptable quality when quality of stimuli is increasing or the point of unacceptable quality when quality is decreasing. In the analysis, binary acceptance ratings are transformed into a ratio calculating the proportion of time during each 30-second period that quality was rated as acceptable. The results are expressed as acceptance percentage of time. This method is powerful when studying variables around the threshold but not those clearly below or above it [7].

The duration of stimuli differs between the three overall quality evaluation methods. The ACR recommends to use short stimuli (10 seconds). This approach pays attention to the constraints of the human working memory, which is about 20 seconds in duration and has limited capacity for units [45, 46], also, it assumes that it is possible to remember all impairments of a stimulus when assessing quality. In contrast, QoP and the method of limit use longer-lasting stimuli materials. They focus more on user and aim to maximize the ecological validity of the viewing task in the experiments and therefore stress less about an ability to remember each of the imperfections the stimulus had [12–14]. It is also worth mentioning that the use of short-stimuli material might be constrained by the measured phenomena, for example, they might fit for measuring compression, but not for transmission quality factors.

In contrast to the overall quality evaluation methods presented, there has been interest in studying instantaneous changes of real-time variation in quality. Originally, the method was developed to go beyond the limitations of the working memory and to enable the use of long material, even up to the duration of a full television program, for testing of time-varying image quality [49–51]. In continuous assessment, participants express their quality evaluation moving the slider on a graphical 5-point labeled MOS scale while watching the content. It has been used to assess the excellence of video and audiovisual quality [50–52]. Similarly to ACR and QoP, the acceptance threshold is hard to locate on this scale. Later, continuous monitoring has been reported to be too demanding evaluation task, especially for multimedia quality evaluation [52]. It may also impact on the natural strategy of human information processing [53].

Table 1: Overview to overall quality evaluation methods.

Method	ACR	QoP	Method of limit
Presentation	Single stimulus, Independently	Single stimulus, Independently	Continuous, gradually decreasing or increasing the intensity of the stimulus
Duration of stimuli	≤10 s	App. 30 s	210 s, quality changes every 30 s
Scales	5/9/11-point scales, MOS	Satisfaction (0–5) Enjoyment Objective quality Information assimilation: ratio of right answers	Binary acceptable/unacceptable
Applied	Audio-video bitrates, codecs, resolution, packet loss [8, 19, 20, 22, 23, 47]	Framerate, delay, jitter, devices, [12, 13, 17]	Framerate, quantization, audio-video bitrate, resolution, text quality, [9, 10, 14, 25, 48]

3.3. Acceptance evaluation

In most consumer-oriented quality evaluation or sensorial studies, acceptance represented refers to affective measurements and represents degree of liking. These measures are used to gather the subjective responses of potential customers or users to a product, product idea, or specific product characteristics [35]. Typically, acceptance is measured on an ordinal scale of overall preference of product or specific preference for a certain sensory attribute [35]. For example, in the context of video or audiovisual quality studies, Apteker et al. [54] and Wijesekera et al. [55] both used ordinal acceptance scales to study framerates whereas Steinmetz [56] studied acceptance of media synchronization on a nominal scale (acceptable, dislike, and annoying). When measuring acceptance as a degree of liking, it lacks of the same detail of threshold of acceptance as quality preferences derived from ACR methods. In contrast to the preference approach, there are only few studies by McCarthy et al. [14] and later Knoche et al. [9, 10, 25, 48] in which acceptance has been seen as a binary phenomenon representing the nature of conventional thresholds (Table 1). Apart from these few studies, acceptability has not typically been measured in the quality assessment of mobile multimedia.

Recent studies have assessed preferences of low produced qualities to optimize the quality of service parameters for mobile devices and networks. Most of the studies compare compression parameters, like low framerates, bitrates or audio-video bitrate share, modern codecs, small display size, and their interactions [8, 10, 19, 20, 57]. Impacts of transmission errors on perceived quality is less reported [58]or the studies focus on one media at a time [59, 60]. Independently of the source of impairments in produced quality, some of these studies compare extremely poor qualities [8, 9, 20] and, therefore, their feasibility can be questioned as follows. How relevant are comparisons of poorness of quality when evaluations are clearly targeted at consumer services? Where is the threshold of minimum accepted level in these preference evaluations?

This leads to the connections between acceptability and preference. As Jumisko-Pyykkö [8] has concluded earlier that "*to improve the connections between the quality preferences or pleasances to the real usage, the anchor of binary acceptability is necessary to…set parallel to quality preferences.*" This is important in quality evaluation studies comparing several parameters, media, and their interaction at the same time. Further, it becomes even more significant when studying the novel optimization problems derived from technology totally lacking previous knowledge about perceptual impacts of parameters. "*This (acceptability) would show the useful quality levels…and target the focus in this field to the meaningful and necessary parameter comparisons*" [8]. In the long term, the goal is to ensure that the produced quality is set in a way that constitutes no obstacle to the wide audience acceptance of a product or service.

For the sake of clarity, we call degree-of-liking or ordinal measured preference of quality satisfaction in this paper. Acceptance of quality refers to the binary measure to locate the threshold of minimum acceptable quality that fulfills user quality expectations and needs for a certain application or system.

4. EXPERIMENT 1

The first experiment had two goals. Firstly, the aim was to develop a new subjective quality evaluation method. Our main focus was on an assessment method for the overall evaluation of acceptance and satisfaction. We also wanted to develop a simplified continuous assessment method for instantaneous quality evaluations which would avoid the previously reported problems of conventional methods being too demanding [52, 53]. Secondly, we wanted to study the impact of simplified continuous assessment on retrospective evaluations between two samples.

4.1. Research method—test set-up

4.1.1. Participants

Two samples, each with 15 participants (equally stratified by age between 18–45 years and gender) conducted a study in a controlled laboratory environment. The samples contained mostly (80%) naïve or untrained participants. They had no previous experience of quality evaluation experiments, they were not experts in technical implementation, and they were not studying, working, or, otherwise, engaged in information technology or multimedia processing [11, 61]. In addition, they did not belong to any group of innovators and early adopters regarding their attitudes to technology [62].

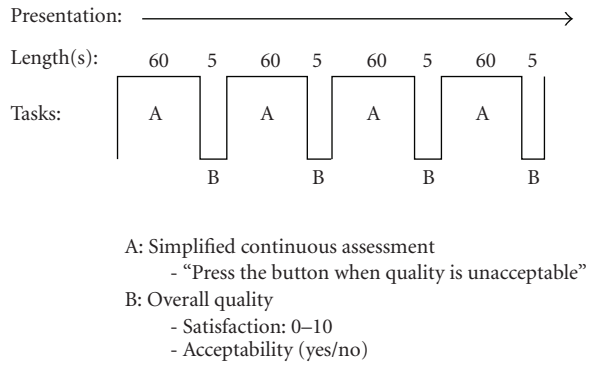


Figure 2: Experimental setup: simplified continuous assessment and retrospective ratings of quality and acceptance.

4.1.2. Test procedure

The test procedure was divided into pre-test, test, and post-test sessions. In the pre-test session, vision and hearing tests with demographic data collection took a place. All participants had normal or corrected-to-normal visual acuity (20/40) as well as normal color vision and hearing. In the combined training and anchoring, participants were shown the extremes of the sample qualities as examples of the quality scale and they became familiar with the contents and the evaluation task.

In the test, the test group evaluated quality with simplified continuous assessment parallel to retrospective ratings (Figure 2: Tasks A + B). The control group used only retrospective ratings (Figure 2: Task B). The sample material was shown using the Absolute Category Rating method where clips are viewed one by one and rated independently [11]. During the clip presentation, the test group used a simplified continuous assessment method in which instantaneous unacceptable quality was indicated by pressing a button on a game controller while viewing the content. After each clip, participants marked retrospectively the overall quality satisfaction score of a clip on an answer sheet using a discrete, unlabeled scale from 0 to 10 and the acceptance of quality (yes/no choice). 9 and 11-point scales are recommended over narrower scales because they compromise the end-avoidance-effect and problems of labeled scales [7]. The widely used labeled MOS scale was not used because it has been criticized for having unequal distances between the labels [49] and the meaning of these labels are not the same between cultures [63, 64]. Acceptance was measured on a binary scale imitating the measures of thresholds [7, 35].

The instructions for the quality evaluation tasks were as follows. For gathering the quality satisfaction score, the participants were asked to assess the overall quality of the presented clip. The measure of acceptance of quality was instructed by asking whether the participants would accept the overall quality presented if they were watching mobile television. No other evaluation criteria were given.

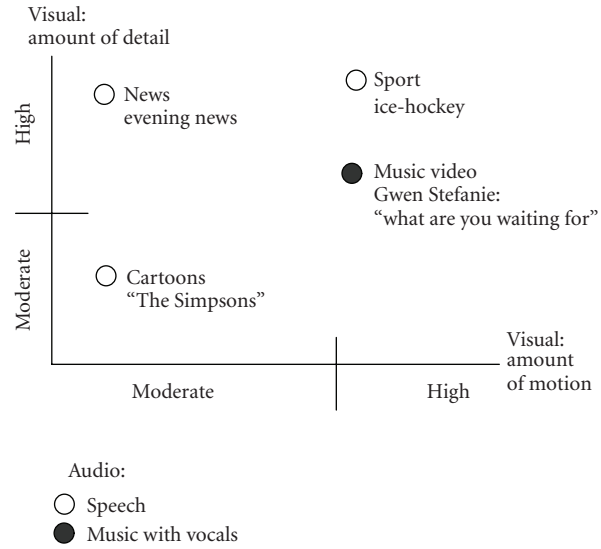


Figure 3: Genre of stimuli, contents, and their audiovisual characteristics.

The post-test session gathered qualitative data on experiences of erroneous streams. One test session lasted for about 1.5 hours.

4.1.3. Selection of test material

Four types of content, news, sport, music video, and animation were selected for test clips according to their potential for mobile television [48, 65, 66], popularity, and audiovisual characteristics (Figure 3). Each clip contained a meaningful segment of a TV program without cutting the start or end of a sentence, some textual information, several shots with different distances and angles to be representative of mobile television content.

The length of stimuli was approximately 60 seconds (61–63 seconds). The chosen duration enabled at least one impairment to appear with the lowest error rate. The use of shorter stimuli is recommended due to the limitations of human-working memory [45, 46], but with the chosen impairment rate, shorter stimuli would have been meaningless.

4.1.4. Network-level characteristics of mobile television

The target application for which the test was setup was mobile television. One of the most prominent standards for mobile television is the Digital Video Broadcasting-Handheld (DVB-H) standard [67], the characteristics of which are briefly reviewed in this section. DVB-H uses Internet Protocol (IP) packet encapsulation for datacasting. These IP packets are further encapsulated into User Datagram Protocol (UDP) packets, Real-Time Protocol (RTP) packets, and lastly Multi-Protocol Encapsulation (MPE) sections before being segmented into 188 byte (inclusive of 4 byte header) transport stream (TS) packets. DVB-H uses time-slicing for reducing power usage in receivers. The error

Table 2: Number of errors, mean durations, and standard deviation (in seconds) of burst errors for error patterns in different error rates.

Error rate		Error rate 1.7%		Error rate 6.9%	
Content		N	Mean(SD)	N	Mean(SD)
Cartoon	Audio	0–3	0.33(0.28)	3–6	0.37(0.20)
	Video	1	1.57(0.51)	3-4	1.06(0.54)
Music video	Audio	0–3	0.27(0.38)	3–7	0.70(0.17)
	Video	1	1.65(0.38)	2-3	1.21(0.43)
News	Audio	2	0.33(0.29)	2–6	0.38(0.20)
	Video	1	1.94(0.45)	2–4	1.08(0.35)
Sport	Audio	0–3	0.34(0.28)	4–6	0.34(0.21)
	Video	1-2	1.10(0.34)	2–4	1.06(0.44)
Error rate		Error rate 13.8%		Error rate 20.7%	
Cartoon	Audio	11–14	0.32(0.19)	9–22	0.30(0.15)
	Video	7-8	1.61(0.97)	13–15	1.31(0.75)
Music video	Audio	11–14	0.31(0.19)	9–22	0.31(0.19)
	Video	7–9	1.27(0.74)	12–15	1.27(0.75)
News	Audio	11–14	0.32(0.19)	9–22	0.30(0.15)
	Video	7–9	1.41(1.00)	11–13	1.40(0.99)
Sport	Audio	12–15	0.30(0.18)	13–22	0.30(0.14)
	Video	7-8	1.61(0.81)	11–14	1.50(0.90)

correction system of DVB-H, known as MPE-FEC, is based on Reed-Solomon FEC codes computed over the IP packets of a time-sliced burst of data [68].

4.1.5. *Production of test materials—transmission error simulations*

The test setup simulated DVB-H reception. The goal of the error simulations was to produce four detectable different transmission error rates with varying number, length, and location of errors. To achieve this goal, 6 pilot experiments were conducted to make a final decision about the final error rates. The simulation of the DVB-H channel was done with a Gilbert-Elliot model that was trained according to a field trial carried out in an urban setting with an operable DVB-H system. Four rates (1.7%, 6.9%, 13.8%, 20.7%) for erroneous time-sliced bursts after FEC decoding (known as MPE-FEC frame error ratio, MFER) were chosen for the simulations. It is noted that these residual error rates do not represent typical DVB-H reception but rather are examples of extremely harsh radio conditions. Such severe radio conditions were selected for the test to discover the threshold between acceptable and unacceptable quality.

The selected test materials were encoded using recommended codecs for IP datacasting over DVB-H [67]. Visual content was encoded using a baseline H.264/AVC encoder with the quarter common interchange format (QCIF), a bitrate of 128 kbps, and a frame rate of 12.5 frame per second [8, 19, 67, 69]. For audio encoding, Advanced Audio Coding (AAC) was used with a bitrate of 32 kbps and sampling rate of 16 kHz as monoaural. An Instantaneous Decoder Refresh (IDR) frame was inserted per time-sliced transmission burst to minimize tune-in delay to new receivers tuning in to the channel and to provide better error resilience under DVB-H channel error conditions. The protocol stack of DVB-H was applied conventionally. The length of transmission burst interval was set at approximately 1.5 seconds, and a code rate of 3/4 was used for MPE-FEC [70].

At the receiver, simple error concealment procedures were used. When a picture of video was lost, all subsequent pictures were replaced by the last correctly received picture in presentation order until the arrival of the next IDR picture. Thus, errors in video produced discontinuous motion. Similarly, the lost audio frames were replaced by silence, resulting in gaps during playback. The error characteristics are presented in Table 2.

4.1.6. *Presentation of test materials*

The experiments were conducted in a controlled laboratory environment [71]. The stimuli materials were viewed on a Nokia 6630 handset with a Nokia player. During the viewing, the device was enclosed in a stand and adjusted to eye level with a viewing distance of 44 cm [8]. For audio playback, headphones were used and the level of audio loudness was adjusted to 75 dBA.

A game controller (Logitech Dual Action gamepad) was used to instantaneously mark unacceptability in the simplified continuous evaluation. A logging program was run on a laptop (Fujitsu Simens Lifebook Pentium 3, Windows 2000) to collect the user input. The logging program run on Python 2.3.5 and used PyGame 1.6 module for accessing the game controller button events. When the button of the game controller was pressed, the program saved the number of seconds elapsing from the reference time at the beginning of the presentation. All clips were played three times in random order and the positions of the transmission errors varied in each repetition.

4.1.7. Method of analysis

Acceptance

To compare the acceptance ratings between the samples, we used Chi-square test, which is applicable to measure the differences of categorical data in independent measures [72].

Satisfaction

To compare the differences in satisfaction ratings between the samples, we used the Mann-Whitney U test as a nonparametric method (Kolmogorov-Smirnow: $P < .05$). The Mann-Whitney U test to measure differences between ordinal measured two independent samples [72]. A significance level of $P < .05$ was adopted in this study.

4.2. Results

We examined the effect of simplified continuous assessment on retrospective overall quality evaluation of acceptance and satisfaction. We compared the retrospective evaluations between the test group and the control group.

Acceptance

When the effects in all combined evaluations of acceptance were compared, the effect was not significant ($\chi^2 = .803$, df = 1, $P > .05$, nor was there any significant effect in the comparison samples in different error ratios ($P > .05$). Moreover, in the comparisons between the samples in each content and error ratio, there was no significant effect of continuous assessment on evaluation of acceptance in 15/16 cases ($P > .05$). The only exception appeared in the sport clip with error ratio 20.7 ($\chi^2 = 4.05$, df = 1, $P < .05$).

Satisfaction

There was no significant difference in the retrospective overall quality assessment of satisfaction. There was no significant effect in the comparison of all given evaluations ($U = 246999$ $P > .05$, ($P = .12$), nanoseconds), nor was the effect significant in the comparison of all error ratios ($P > .05$) or in the comparisons of each content in each error ratio between the two research methods ($P > .05$).

4.3. Discussion

The results showed that the simplified continuous assessment method did not affect the evaluations of retrospective acceptance and satisfaction between the studied samples. Earlier continuous assessment methods have been criticized for requiring a high level of involvement on the part of the evaluator and for possibly changing the way of information processing while evaluating quality [52, 53]. It is known that the difficulty, similarity, and practicing of tasks are the basic factors affecting performance of dual tasks [73]. Our study indicates that the simplified continuous assessment task developed is easy enough to be used parallel to retrospective evaluations without negative impact. Our results are also supported by Reiter and Jumisko-Pyykkö [74]. They concluded that while viewing the content, simple parallel tasks like pressing the button or catching the object, did not impact on the requirements of quality in audiovisual applications. Based on these results, we will use simplified continuous assessment in parallel with other methods to evaluate overall quality in different transmission simulations.

5. EXPERIMENT 2

To apply the developed overall quality evaluation methods, we used them to measure the impact of transmission errors. As in experiment 1, we assumed a mobile television usage scenario using the DVB-H standard. The goal of the experiment was to study the effect of four clearly detectably different residual transmission error rates on perceived quality. We aimed to locate the threshold between acceptable and unacceptable quality, examine the quality satisfaction, and also express acceptance percentage of time. In addition, we examined the relations between the results of these three different methods to evaluate their reliability.

5.1. Research method—test setup

5.1.1. Participants

30 participants, recruited according to the same criteria and meeting the same sensory requirements as in experiment 1, participated in the experiment.

5.1.2. Test procedure

The test procedure was identical to the test sample procedure in experiment 1 (Figure 2: Tasks A + B). The simplified continuous assessment was used parallel to retrospective ratings of acceptance and satisfaction.

Test materials, Test material production—transmission error simulations, and material presentation were identical to those in the experiment 1.

5.1.3. Method of analysis

Acceptance

McNemar's test was applied for the nominal retrospective acceptance evaluations to test the differences between two categories in the related data [72].

Satisfaction

Satisfaction data were analyzed using Friedman's test and Wilcoxon matched-pair signed-ranks test because the presumption of parametric methods (normality) was not met (Kolmogorov-Smirnow $P < .05$) [72]. Friedman's test is applicable to measure differences between several and Wilcoxon's test between two related and ordinal datasets [72].

Acceptance percentage of time

To formulate the data of simplified continuous assessment in the form of overall Acceptance percentage of time, nominal data was converted to a scale variable using the conversion introduced by McCarthy et al. [14].

$$(1 - (\text{unacceptable pressings/length of the clip})) * 100 \quad (1)$$

After the conversion, each of the stimuli was given a score showing the percentage of acceptable quality of stimuli presentation. Friedman and Wilcoxon's tests were then used in the actual analysis.

Relations between different measures

To analyze the connections between the different overall quality evaluation measures, Spearman's correlation as a nonparametric method for ordinal data was used and the Chi-square test of independence evaluated independence between distributions of two variables measured on a categorical scale [72].

5.2. Results

5.2.1. Acceptance

The results of acceptance measurements showed that error rates 1.7% and 6.9% of uncorrectable time-slices were experienced as giving acceptable subjective quality, while error rates of 13.8% and 20.7% were perceived as unacceptable. The differences between the error ratios were significant (All comparisons $P < .001$; Animation: 13.8% versus 20.7% $P < .05$) except the difference between the error rates 13.8% and 20.7% in the news, music video, and sport clips evaluations (Figure 4 $P > .05$, nanoseconds).

5.2.2. Satisfaction

In terms of satisfaction, the order of preference in all combined evaluations of error ratios was 1.7%, 6.9%, 13.8%, 20.7%. Error rates had a significant effect on quality scores ($F_R = 437.6$, df = 3, $P < .001$) and the differences between the error rates were significant ($P < .001$).

The preferred order of satisfaction was the same in the content-by-content examination but there were some variations in the pairwise comparisons of the highest error rates (Figure 5). Error rates had significant effect on all satisfaction evaluations in all contents (Animation: $F_R = 183.3$, df = 3, $P < .001$, Music video: $F_R = 145.2$, df = 3, $P < .001$, News: $F_R = 183.4$, df = 3, $P < .001$, Sport: $F_R = 203.6$, df = 3, $P < .001$). The evaluations differed significantly between all error rates in animation ($P < .001$), sport ($P < .001$), and music video content presentations (between 13.8% and 20.7% $P < .01$; all others $P < .001$). In the presentation of news content, the differences were significant ($P < .001$) excluding the ratios 13.8% and 20.7% ($P > .05$, nanoseconds).

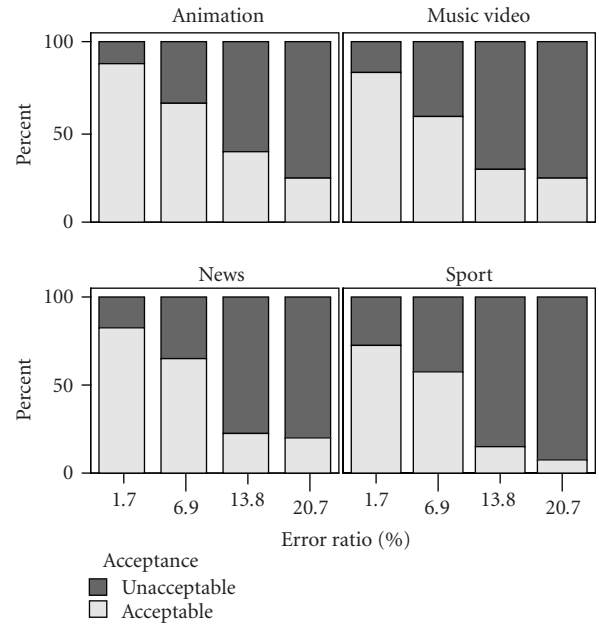


Figure 4: Acceptance of different error rates for all contents.

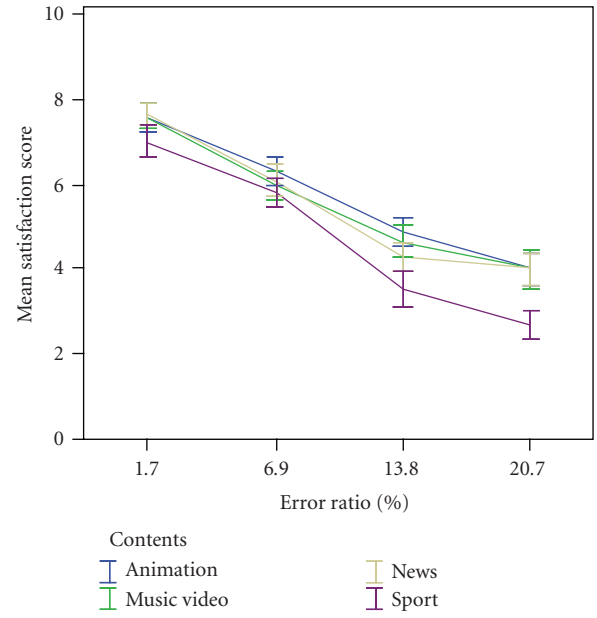


Figure 5: Mean satisfaction scores for all contents. Error bars show 95% CI of mean.

5.2.3. Acceptance percentage of time

Three outliers were removed from the data because they either expressed unacceptable quality very rarely during the presentation or they expressed it infinitely. Similar personal variation has also been expressed in the use of conventional continuous assessment [51].

The acceptance results based on a combination of continuous assessment were similar to the results of retrospective ratings.

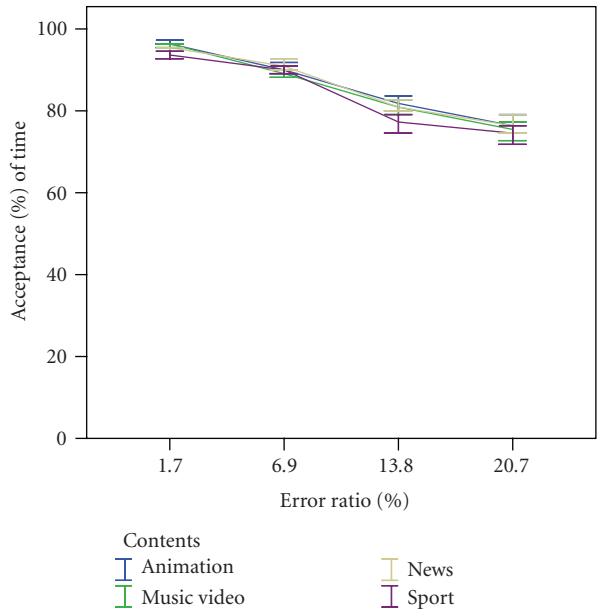


Figure 6: Acceptance percentage of time for all contents. Error bars show 95% CI of mean.

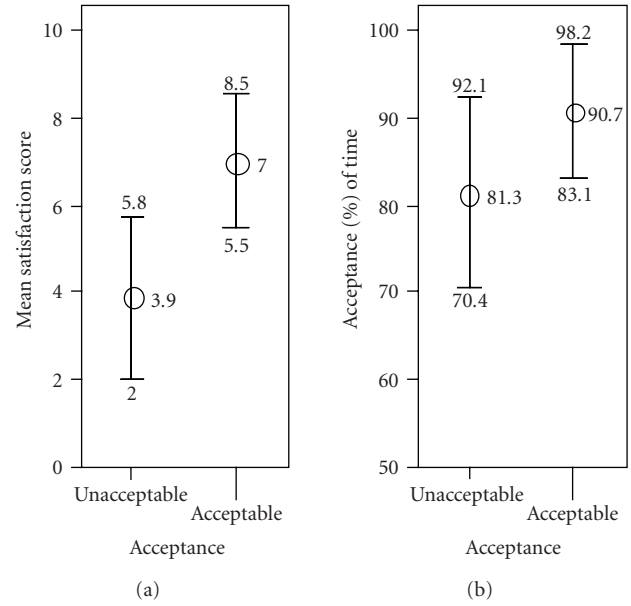


Figure 7: Relations on the scale between retrospective acceptance and satisfaction; and retrospective acceptance and acceptance based on continuous assessment. Bars show mean and standard deviation.

The lowest error rate 1.7% gave acceptable viewing experience for approximately 95% of the time whereas the highest error rate gave the acceptable experience only approximately 75% of the time (Figure 6). The acceptance evaluations were significantly affected by the error rates (F_R = 774.4, df = 3, $P < .001$) and the evaluations differed significantly between all tested error rates ($P < .001$). The effects of different error rates were similar to the combined evaluations in content-by-content examination. In the animation (F_R = 210.9, df = 3, $P < .001$), music video (F_R = 190.5, df = 3, $P < .001$), and news (F_R = 176.5, df = 3, $P < .001$) content evaluations differed significantly between all error rates ($P < .001$). In the sport content evaluation (F_R = 208.1, df = 3, $P < .001$), the differences between the evaluations varied significantly between error rates ($P < .001$; and 13.8% and 20.9% $P < .01$).

5.2.4. *Relations between the overall quality evaluation methods*

All quality evaluations based on three different evaluation methods were related to each other. Retrospective acceptance was discriminative on a scale of satisfaction, but not on the acceptance based on simplified continuous assessment. Related or correlated measures indicate that results measured on one scale can be used to interpret the results in another scale. Discrimination between the scales, such as the independence of the acceptable and unacceptable ratings from the satisfaction scales, can be examined in a further analysis for locating the threshold of acceptability. The idea resembles the classical Thurstonian scaling, aiming to construct nonoverlapping concepts with equal intervals on the attitude scale (e.g., [7]) .

Acceptable quality was expressed between scores of 5.5 and 8.5 (Mean = 7.0, SD = 1.5; Figure 7) and unacceptable quality was located between scores of 2.0–5.8 (Mean = 3.9, SD = 1.9). The distribution between acceptable and unacceptable ratings on the satisfaction scale differed significantly ($\chi^2(10) = 683.2$, $P < .001$). In relation to evaluations based on continuous assessment, acceptable quality was located between 83% and 98% (M = 90.7% of time SD = 7.6; Figure 7) of total acceptances of time, overlapping with unacceptable quality evaluations (M = 81.3% of time SD = 10.8). The distributions between acceptable and unacceptable ratings on a scale of acceptance % of time likewise differed significantly ($\chi^2(36) = 319.1$, $P < .001$). The retrospectively rated satisfaction and acceptance based on continuous assessment were positively and linearly related (Spearman: $r = .725$, $P < .001$). In practice, the acceptance threshold is located in the range of 5.5–5.8 on the satisfaction scale in this experiment. It is not justifiable to draw a similar conclusion for the measures of acceptance percentage of time because the threshold is located between 83.1 and 92.1 and the confidence intervals of unacceptable and acceptable percentage of time overlap to a great extent.

5.3. *Discussion*

The perceived preference order in all measured scales for error rates was 1.7%, 6.9%, 13.8%, and 20.7%, respectively, indicating clearly detectable differences between stimuli. Acceptance ratings give a quality anchor for this preference order showing that the threshold between acceptable and unacceptable quality lies between error rates of 6.9% and 13.8% and this result is not dependent on content. In practice, acceptable quality can be reached when approximately 4/60 seconds are corrupted, resulting altogether in a maximum 10 detectable errors [59, 60] in different media.

In the literature, an error rate of 5% is the conventionally used limit value of operative quality of restitution (QoR) for mobile reception [68] but our result showed a slightly higher tolerance of errors.

The order of preference for different error rates collected using different methods was similar in all contents with few exceptions. Exceptions were found especially in the comparisons of acceptance ratings of the highest error rates. In these error rates, the produced quality is relatively modest. The evaluation criterion of acceptance may be much tighter compared to the task of evaluating quality satisfaction or it may be hard to accept any such erroneous presentations as the goal of viewing can no longer be achieved [34]. In addition, a binary acceptance scale may be useful only in the identification of the threshold, not in detailed comparisons of preferences regarding low qualities. In summary, the assessment results were closely related between all three measures indicating good reliability, and they had good discriminative capability when differences between stimuli were distinguishable and the stimuli not extremely erroneous.

6. EXPERIMENT 3

For further estimation of the reliability and discriminative ability of the overall quality evaluation methods presented, we continued the work with heterogeneous error characteristics, realistic in multimedia broadcasts. The third experiment aims to compare two different error rates on both sides of acceptability by pre- or postprocessing them with four different error control methods. This combination was assumed to produce detectable, but relatively small differences between stimuli.

Few studies have reported comparisons of error control methods related to DVB-H to improve experienced quality. Hannuksela et al. [23] have compared unequal and equal error protection methods with two different error rates. Unequal error protection (UEP) method uses priority-based segmentation of media streams in which audio and the most important coded video pictures have the best protection under harsh channel transmission conditions. By contrast, all media data are of equal importance in the conventional equal error protection method (EEP). The experiment compared these methods with error rates of 6.7% and 13.8% and concluded that in the highest error rate UEP improved the subjective quality. Further, Hannuksela et al. [22] also compared audio redundancy coding and conventional error protection methods with two different error rates (6.7% and 13.8%). Audio redundancy coding (ARC) aims to ensure audio continuity in very erroneous channel conditions and their results showed it to improve perceived quality, especially with the harshest error rate. Earlier studies have shown that error control methods can provide some quality improvements depending on error rate, but no extensive study of different error control methods and error rates has been published.

The aim of the experiment is to compare the interactions of four different error control methods and error rates close to the threshold of acceptability with small differences between the stimuli. In addition to measuring overall satisfaction of quality and acceptance percentage of time, we are interested to ascertain if the boundary of acceptability can be affected by error control methods. To evaluate reliability, we also examine the relations between results of three different methods.

6.1. Research method—test setup

6.1.1. Participants

Our participants were 45 participants, recruited according to same criteria as in experiments 1 and 2.

6.1.2. Test procedure

The test procedure was identical to that of experiment 2. The total duration of the experiment was approximately 2 hours.

6.1.3. Selection of test material

Test materials were identical to experiment 2.

6.1.4. Material production process—transmission error simulations

The aim of the error simulations was to produce stimuli material with relatively small, but detectable differences between stimuli in various forms. As a base for error simulations, two different error rates known to be perceived around a boundary between acceptable and unacceptable (experiment 2) qualities were selected and further four different error concealment methods were applied to these. The simulated error rates produced a varying number, length, and location of errors, and error concealment methods caused different audiovisual appearance form for these errors (Table 3).

Four different error resiliency methods were tested. While one of the error resiliency methods gave more importance to audio, another gave video error resiliency more importance. The remaining one used channel-assisted error resiliency based on unequal error protection. These methods are described in greater detail below.

The first method, called conventional transport with picture freeze (CT-PF), did not use any kind of additional error resiliency measures apart from the protection provided by DVB-H MPE-FEC. The method was used as a base for comparing other error resiliency methods tested. It assumed a compliant audiovisual decoder, albeit with no intelligence. In this method, when the decoder encountered errors in a video stream, it stopped decoding any subsequent pictures until an Intra Decoder Refresh (IDR) picture arrives. IDR pictures use no other pictures as a prediction reference and therefore provide a resynchronization point in an erroneous bit stream. During the period when the decoder stopped decoding, it presented the last uncorrupted decoded picture. Subjectively, when this method was used, an error was perceived as jerky motion in visual streams. The duration of these jerks in visual streams depended on the IDR interval

Table 3: Number of errors, mean durations and standard deviation (in seconds) of burst errors for error patterns in different error rates and error control methods.

Concealment content		N	Mean(SD)	N	Mean(SD)
CT-PF			Error rate 6.9%		Error rate 13.8%
Cartoon	Audio	7–9	0.15(0.08)	18	0.18(0.09)
	Video	3	1.2(0.58)	5-6	1.5(0.71)
Music video	Audio	7–9	0.15(0.08)	18	0.17(0.09)
	Video	3–5	0.92(0.55)	5	1.72(1.01)
News	Audio	8	0.15(0.07)	18	0.17(0.09)
	Video	3-4	1.53(1.13)	5–7	1.63(1.04)
Sport	Audio	8	0.15(0.07)	18	0.17(0.07)
	Video	2–4	1.22(0.72)	6	1.29(0.07)
SAR-PF			Error rate 6.9%		Error rate 13.8%
Cartoon	Audio	2	0.11(0.06)	7–11	0.11(0.04)
	Video	2-3	2.41(1.26)	4-5	1.90(1.01)
Music video	Audio	2	0.11(0.06)	7–11	0.12(0.04)
	Video	2–4	2.11(1.52)	5	1.82(0.80)
News	Audio	2	0.11(0.06)	7–11	0.11(0.03)
	Video	1–3	2.28(1.53)	1–3	6.03(3.60)
Sport	Audio	2	0.11(0.06)	7–11	0.12(0.03)
	Video	1–4	2.30(2.00)	3	3.04(1.47)
CT-EC			Error rate 6.9%		Error rate 13.8%
Cartoon	Audio	7–9	0.15(0.08)	18	0.18(0.09)
	Video	7–9	0.18(0.06)	18	0.18(0.09)
Music video	Audio	7–9	0.15(0.08)	18	0.17(0.09)
	Video	7–9	0.17(0.07)	15–18	0.19(0.09)
News	Audio	7–9	0.15(0.08)	18	0.18(0.09)
	Video	7–9	0.18(0.07)	17–19	0.18(0.10)
Sport	Audio	8	0.15(0.07)	18	0.17(0.07)
	Video	7-8	0.20(0.08)	17–19	0.19(0.11)
UEP-PF			Error rate 6.9%		Error rate 13.8%
Cartoon	Audio	4-5	0.29(0.14)	11	0.34(0.19)
	Video	7–12	0.43(0.65)	14-15	0.54(0.69)
Music video	Audio	3–5	0.27(0.16)	10	0.38(0.24)
	Video	8–12	0.32(0.42)	11	0.72(1.21)
News	Audio	3-4	0.32(0.18)	9–11	0.36(0.22)
	Video	8–10	0.34(0.44)	13–17	0.44(0.49)
Sport	Audio	3	0.35(0.17)	9-10	0.39(0.22)
	Video	6–12	0.34(0.47)	9–12	0.60(0.92)

and the position of the error between two IDR intervals. The audio compression scheme used in the tests encoded 1024 samples of every audio channel as frames. These frames were all independent of each other and a loss of any one frame of the bit stream did not affect any other subsequent frames of an audio channel. When an audio frame was lost, it was replaced with a null frame perceived as silence by the listener. Subjectively, audio frame losses were perceived as discontinuous audio.

The second method used audio redundancy coding to achieve better audio reception in heavy DVB-H channel error conditions and is therefore called Synchronised Audio Redundancy coding with picture freeze (SAR-PF). When MPE-FEC frames were constructed with audiovisual data as input, audio packets that constitute the next MPE-FEC frame in transmission were replicated and sent in the current MPE-FEC frame. The audio decoder expected two copies of every coded audio frame. However, when errors destroyed an audio frame, the decoder looked for the second copy of the same audio frame and if received correctly, used this copy instead. This redundancy of audio packet coupled with their transmission in different time-sliced bursts greatly reduced the probability of any audio frame being completely lost. Video error concealment was identical to what was done in the CT-PF method described above. However, to account for the additional bit rate overhead incurred due to redundant audio packets, the video bit rate was dropped such that the overall bit rate was the same as the other error resiliency

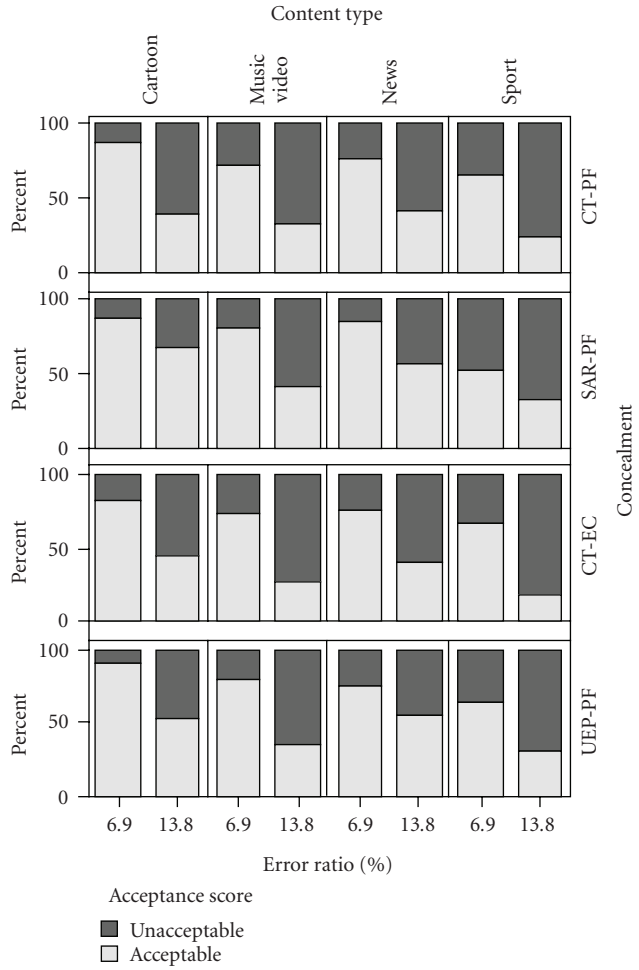


Figure 8: Retrospective acceptance of different error rates and concealment methods for all contents.

methods. In other words, the media-level-produced quality of the coded video was poorer than in the CT-PF method. More details of the SAR-PF method are available in [23].

The third error concealment method, called Conventional Transport with Error Concealment (CT-EC) used a very simple decoder-based visual error concealment method for concealing lost parts of the video sequence. When a picture of the sequence was lost, the decoded picture buffer (DPB) replicated the last correctly received picture (in presentation order) and used it instead of the lost picture. The reason for this replacement was the assumption that spatial video redundancy can be fairly high (depending on the video sequence) and the replaced picture is a good enough estimate of the lost picture. However, since the replaced picture was not the exact representation of the lost picture, motion compensation errors occured in pictures using the replaced picture as reference, and these errors propagated until an Intra picture and/or IDR picture arrived. For audio, the error concealment was similar to what was used in the CT-PF method, where the last audio frames were replaced with a silent frame.

The fourth method of error resiliency is called Unequal Error Protection with Picture Freeze (UEP-PF). First, the media datagrams covering a certain period of playback time were assigned priorities. In the tests, two priorities were used. Audio packets, video reference pictures (both IDR and reference predicted pictures) were assigned priority 1 (the highest), and nonreference pictures were assigned priority 2 (the lowest). The priority-assigned datagrams were grouped together such that all datagrams in a group had the same priority. The protection of the priorities was chosen such that priority 1 datagrams were protected with a 3/4 MPE-FEC-code-rate while the priority 2 datagrams were completely unprotected. These grouped and protected MPE-FEC matrices (called peer MPE-FEC matrices) were then sent back to back without any delay between these MPE-FEC frames. More details on the UEP-PF method are available in [23, 75]. The first and last five seconds of presentation were left error-free to avoid memory effect (primacy and recency) in evaluation of long test materials [49, 53].

6.1.5. Presentation of test materials

The presentation of the test materials was similar to that in the previous experiments. All clips were played twice in random order and the positions of the transmission errors varied in both repetitions.

6.1.6. Data-analysis methods

Selection of data-analysis methods followed the methods described for experiment 2.

6.2. Results

6.2.1. Acceptance

Between error rates

Lower error rate (6.7%) provided mostly acceptable and higher error rate (13.8%) unacceptable quality with significant difference between them in all studied concealment methods and contents ($P < .01$; Figure 8).

Between error concealments

All concealment methods were evaluated equally acceptable in error rate 6.9% ($P > .05$). In contrast, in error rate 13.8%, SAR-PF and UEP-PF ($P > .05$) were evaluated equally and more acceptable CT-PF and CT-EC ($P < .001$) which were in same level as well ($P > .05$).

In error rate 6.7%, mostly all error concealment methods were evaluated into same level, but there were some content-dependant variations. There were not differences between the concealment methods in animation and music video presentation ($P > .05$). News content, concealed with SAR-PF, was evaluated more acceptable than other methods (SAR-PF versus others $P < .05$; all other comparisons $P > .05$). In contrast, SAR-PF provided the most modest quality for sport presentation ($P < .05$; all other comparisons

$P > .05$), approaching the boundary between acceptable and unacceptable quality.

In error rate 13.8%, SAR-PF and UEP-PF provided more acceptable quality than CT-PF and CT-EC. For animation and news presentation, most of the participants considered SAR-PF and UEP-PF as equally acceptable ($P > .05$) and CT-PF and CT-EC as equally unacceptable ($P > .05$) with significant differences between them (SAR-PF versus CT-PC, CT-EC $P < .05$; UEP-PF, and CT-PF $P < .01$). In music, SAR-PF was significantly better than CT-EC ($P < .05$), while all other methods were in same level ($P > .05$). For sport presentation, SAR-PF and UEP-PF were rated as the most acceptable ($P > .05$) with significant difference to other methods ($P < .05$). Error rate 13.8% is in general evaluated as unacceptable, but in the case of cartoon and news with concealment method, SAR-PF quality can become acceptable or, with method UEP-PF, reach the boundary of acceptable and unacceptable ratings.

6.2.2. Satisfaction

Between error rates

Similar to the results for acceptance, error ratio 6.7% was reported more satisfying than error ratio 13.8% in all contents and error control methods ($P < .001$; Figure 9).

Between error concealments

Error ratios and error concealment methods affected satisfaction evaluations ($F_R = 982.1$, df = 7, $P < .001$) and error concealment strategies had a significant effect on evaluations within both error rates (6.9%: $F_R = 17.252$, df = 3, $P < .01$, 13.8%: $F_R = 94.381$, df = 3, $P < .001$).

In terms of satisfaction, CT-EC provided the lowest quality in comparison to other concealment methods ($P < .05$), which were equally evaluated ($P > .05$) for error rate 6.9%. In error rate 13.8%, the most satisfying quality was given by SAR-PF, followed by UEP-PF and the lowest quality by equally rated CT-PF and CT-EC ($P > .05$) with significant differences between all ($P < .01$).

There were also content-dependent preferences between the concealment methods in different error rates. For the lower error rate of 6.9% for animation content, all concealments were evaluated at the same level ($P > .05$). UEP-PF and SAR-PF were evaluated equally, giving the most satisfying quality in music video ($P > .05$), but only differences between UEP-PF and others were significant ($P < .001$). SAR-PF was evaluated as the most satisfying for news content compared to other methods ($P < .01$). In sports, CT-PF and UEP-PF were found equally good ($P > .05$) and significantly better than SAR-PF ($P < .05$).

In error rate 13.8%, error concealments SAR-PF and UEP-PEF were among the most satisfying methods in all contents. For animation presentation, SAR-PF and UEP-PF were evaluated equally being more satisfying ($P > .05$) than other methods ($P < .001$). In music video, SAR-PF, UEP-PF, and CT-PF ($P > .05$) were more satisfying than the concealment method called CT-EC ($P < .05$). The SAR-PF and UEP-PF were equally evaluated ($P > .05$) in news presentation in which SAR-PF was significantly better than CT-PF and CT-EC ($P < .01$) and UEP-PF significantly better than CT-PF ($P < .05$). For sport content, SAR-PF, and UEP-PF ($P > .05$) were more satisfying than the others with SAR-PF significantly outperforming both CT-PF and CT-EC ($P < .001$).

6.2.3. Acceptance percentage of time

Between error rates

Lower error rate (6.7%) was reported to give a higher acceptance rate percentage of time compared to higher error rate (13.8%) ($P > .001$; Figure 10). An exception was found in news presentation with error rate 6.7%, methods CT-EC and UEP-PF were evaluated at the same level with error rate 13.8% concealed with SAR-PF ($P > 0.05$, ns).

Between error concealments

Error ratios and error concealment methods affected acceptance evaluations based on simplified continuous assessment ($F_R = 1335.0$, df = 7, $P < .001$). The error concealment strategies also had a significant effect on within error examination (6.9%: $F_R = 48.5$, df = 3, $P < .001$, 13.8%: $F_R = 223.0$ df = 3, $P < .001$). In error rate 6.9%, SAR-PF yielded the highest acceptance percentage of time with significant difference ($P < .01$) to others being on the same level ($P > .05$). Similarly, SAR-PF yielded the highest acceptance % of time in error rate 13.8% ($P < .001$), followed by UEP-PF and CT-PF ($P > .05$) and UEP-PF and CT-EC ($P > .05$).

There were also some content-dependent variations between the concealment methods with the lower error rate of 6.9%. For presenting cartoons, the longest acceptable presentation for cartoon content was given by SAR-PF outperforming the others ($P < .05$), followed by UEP-PF (difference from others $P > .05$). In music video, SAR-PF and UEP-PF were evaluated at the same level ($P > .05$, difference from others $P < .05$). The concealment SAR-PF also provided the highest quality ($P < .001$) for news content with significant difference from other methods which were evaluated equally ($P > .05$). In sport content, there were no differences between the methods ($P > .05$) except the UEP-PF, which yielded the lowest quality ($P < .001$).

In the higher error rate (13.8%), CT-PF, SAR-PF, and CT-EC ($P > .05$) were more satisfying than the most modestly assessed UEP-PF ($P < .001$) for cartoon content. In music video, SAR-PF is the highest quality with a significant difference from the others ($P < .001$), UEP-PF is the second highest ($P < .05$), and the other methods were evaluated at the same level ($P > .05$). For the news, the concealment called SAR-PF yielded the highest quality ($P < .001$) and all other methods were on the same level ($P > .05$). As in news content, SAR-PF yielded the highest quality for sport content with significant difference from the others ($P < .001$), CT-PF and CT-PC the second highest ($P > .05$), and UEP-PF the most modest ($P < .05$).

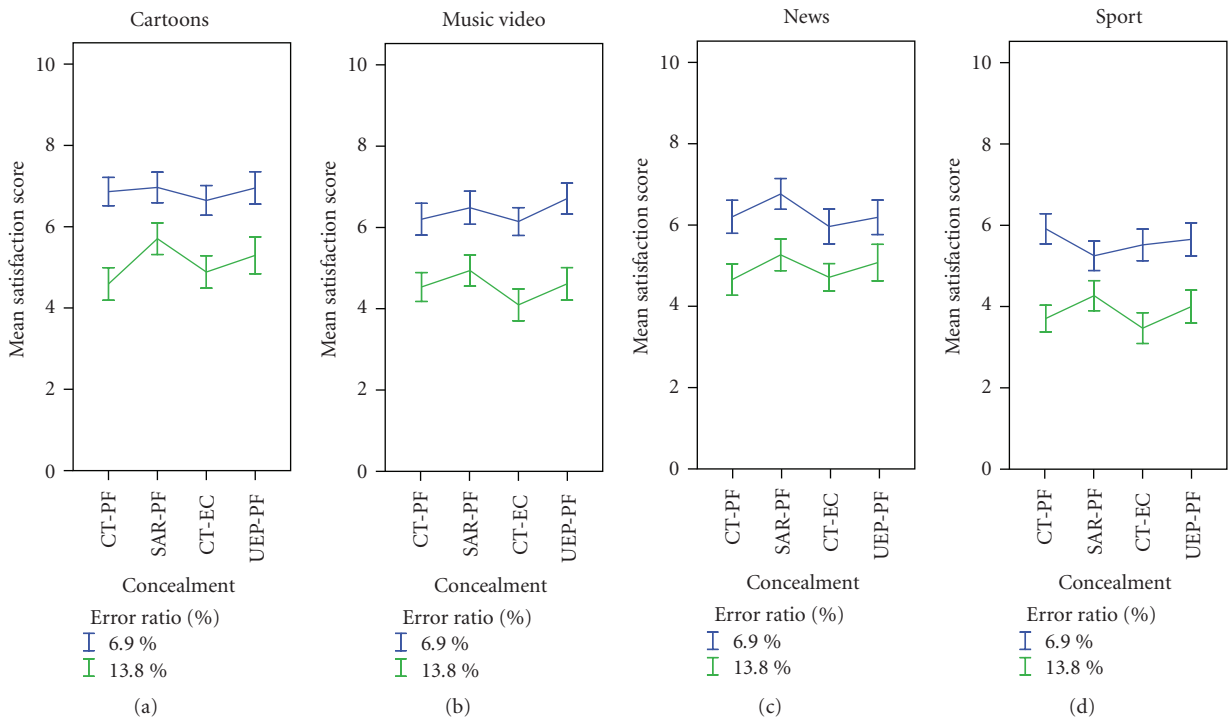


Figure 9: Retrospective satisfaction of different error rates and concealment methods for all contents. Error bars show 95% CI of mean.

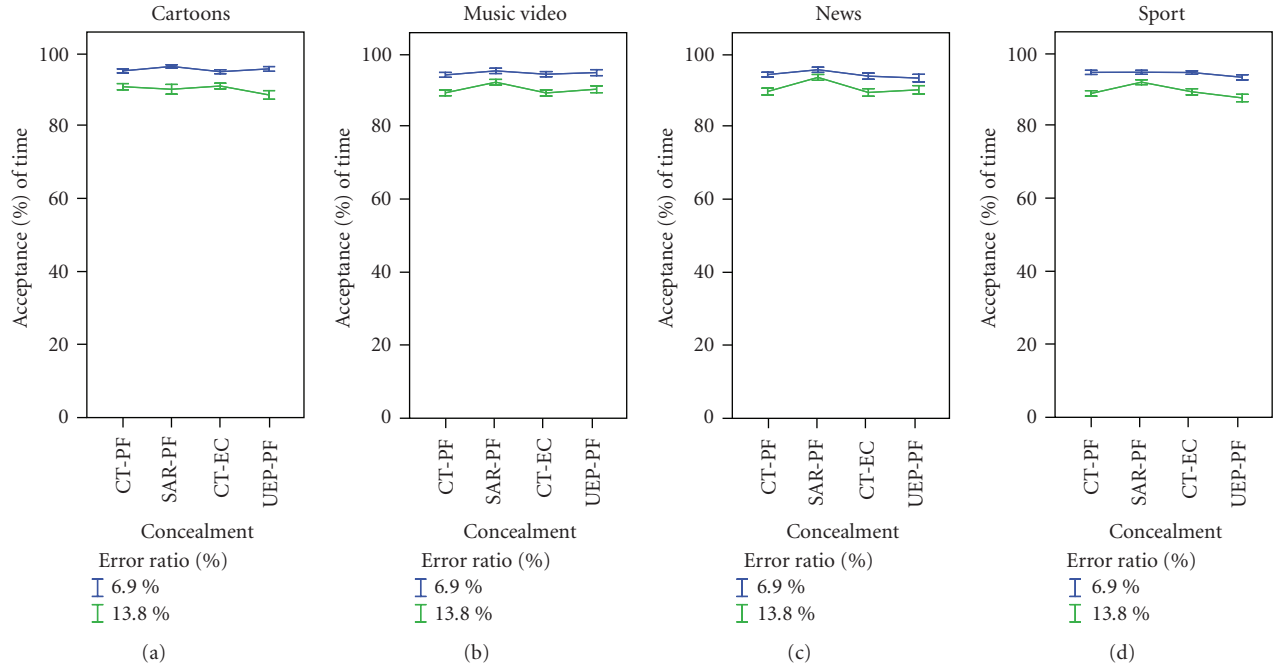


Figure 10: Acceptance percentage of time of different error rates and concealment methods for all contents. Error bars show 95% CI of mean.

6.2.4. *Relations between the overall quality evaluation methods*

As in experiment 2, the three different evaluation methods were related to each other. Acceptable and unacceptable quality was clearly detectable on a scale of satisfaction, but not on a scale of acceptance percentage of time. Acceptable quality was connected to scores between 5.2 and 8.1 (Mean = 6.6, SD = 1.45; Figure 11) on a satisfaction scale and unacceptable quality to scores between scores of 2.1–5.4

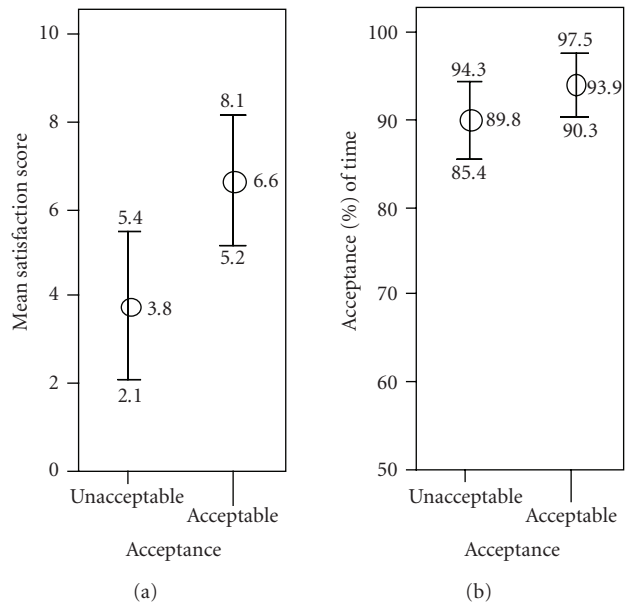


Figure 11: Relations on the scale between retrospective acceptance and satisfaction; and retrospective acceptance and acceptance based on continuous assessment. Bars show mean and standard deviation.

(Mean = 3.8, SD = 1.67). In the examination of the relation between acceptance and acceptance percentage of time, acceptable quality was located between 90 and 97% (M = 93.9% of time, SD = 3.6) on a scale of acceptance percentage of time with widely overlapping unacceptable quality range (M = 89.8% of time, SD = 4.4). As in the previous experiment, both the distributions of retrospectively rated satisfaction and acceptance ($\chi^2(10) = 1370.3$, $P < .001$) and the distributions between the retrospectively rated acceptance and acceptance based on continuous assessment ($\chi^2(49) = 632.0$, $P < .001$) differed. The retrospectively-rated satisfaction and acceptance based on continuous assessment were also positively and linearly related (Spearman: r = .542, $P < .001$). In practice, according to this experiment, the threshold between acceptable and unacceptable ratings is between the scores 5.2 and 5.4 on the satisfaction scale. The threshold on a scale of acceptance percentage of time is between 90.3 and 94.3 in which the overlapping of the confidence intervals constrains the interpretation of results.

6.3. Discussion

All the evaluation methods were able to detect the differences in the level of error rates confirming the results of experiment 2. Higher error rate was experienced giving poorer quality compared to lower error rate in all methods measured. In the measures of acceptance percentage of time, only one exception appeared in which the poorest quality of lowest error rate was evaluated equal with the highest quality of most erroneous error rate.

When error control methods were compared, variations were found in the results gathered using retrospective and continuous methods. In error rate 6.9%, the requirements for different error control methods varied content dependently. For example, in news content, SAR-PF outperformed the other methods in all measures, whereas all methods were equally retrospectively evaluated for cartoons. CT-PF and UEP-PF were among the methods that provided highest quality for sport content in the retrospective measures, whereas UEP-PF was the poorest method according to acceptance percentage of time measures. In high error rate, retrospective methods had excellent agreement in acceptance and satisfaction revealing that SAR-PF and UEP-PEF were among the most satisfying methods in all contents. These error control methods even enabled cartoons and news to reach the 50% acceptance threshold. In contrast, according to simplified continuous assessment, SAR-PF provided the highest acceptance percentage of time while UEP-PF did not produce the highest quality in any of the cases. In all of the cases measured with continuous assessment, SAR-PF was among the methods producing the highest acceptance percentage of time.

From the viewpoint of research methods, there are two main conclusions. Firstly, good agreement between the retrospective methods indicates that detailed analysis is not needed for both of the measures. Both of the methods are needed in data collection, but different emphasis is given in the analysis. As quality satisfaction is measured using an ordinal scale and therefore providing a chance to use sophisticated and efficient methods of analysis [72], it should be used as a primary data source for analysis. Data on acceptance of quality may only be analyzed to locate a certain threshold of acceptance and these thresholds can be used as references in the interpretation of the results of quality satisfaction. Secondly, simplified continuous assessment may not be a reliable method for overall quality evaluation to discriminate stimuli having small noticeable differences. The results of simplified continuous assessment differed from the results of retrospective measures when the differences between the stimuli were small.

There are two main conclusions about the error rates and error control methods we studied. Error rate seems to be a more important factor in perceived quality than an error control method. Further research may focus on error rates and more detail examination of different impacting error characteristics, such as duration, location, and modality within these error rates. In addition, the results of the comparisons of error rates and error control methods also reflected the relation between content dependency and level of quality. In the low error rates, some dependant preferences appeared. For example, the error control methods improving audio quality was emphasized in news presentation while improvements in visual quality were highlighted in sport content. By contrast, extremely erroneous quality seems to hide the content-dependent preferences highlighting the importance of audio quality in all contents. These results are supported by an earlier study comparing several audio-video bitrates. These authors concluded that relations between optimal audio-video bitrates are content dependent, but in low qualities audio qualities is emphasized [8].

7. DISCUSSION AND CONCLUSIONS

In this paper, we examined research methods for assessing overall acceptance of quality in three experiments. At first, we explored the possibilities of using simplified continuous assessment in the evaluation of overall acceptance parallel to retrospective measures. Secondly, we studied the boundary between acceptable and unacceptable quality using clearly detectable differences between stimuli. Finally, we studied the acceptance threshold with small differences between stimuli under heterogeneous conditions. We conducted these studies in the context of mobile television with varying error rates and error control methods with several television programs in a controlled environment. Our results showed that instantaneous and retrospective evaluation methods can be used in parallel in quality evaluation without causing changes to human information processing. All measures were discriminative and correlated when clearly detectable differences between stimuli were studied. By contrast, when small differences between stimuli were examined, the results of retrospective measures correlated but differed from the results based on the evaluation of instantaneous changes. In this section, we discuss the main results and make recommendations for the use of these methods.

7.1. Bidimensional method for retrospective evaluations of overall acceptance and satisfaction

As the main result of this study, we recommend parallel use of retrospective measures of acceptance and satisfaction in quality evaluation experiments. Acceptance, representing the first dimension, is needed to ensure that test variables reach the predefined thresholds depending on the goal of the study (e.g., 50%, 80%). However, the nature of measuring a threshold has some constraints. Firstly, the measure is discriminative when studying variables close to the threshold, but is not clearly below or above it. Secondly, as acceptability is measured on a binary scale, it imposes limitations on the use of efficient methods of analysis which are needed in careful pairwise comparisons [7]. To go beyond these constrains and broaden the use of the method to the other quality ranges, we recommend studying satisfaction of quality parallel to acceptability. Satisfaction, the second dimension, as a degree-of-liking is most commonly measured on a 9- or 11-point ordinal scale which enables the use of efficient methods of analysis [7]. In addition, it allows using same data-collection method for the duration of continuous quality evolution.

Data-collection and analysis using a combination of acceptance and satisfaction methods are summarized in Figure 12. We recommend a separate analysis for both of the measured dimensions, but as a starting point, the relation between the measures needs to be considered to ensure the reliability. There are two options for extracting the desired threshold. Firstly, the tested parameters can be dissected from the frequencies of acceptance data. In the second option, the value range of the threshold between acceptable and unacceptable scores can be identified on satisfaction scale in the case the measures are not strongly overlapping. Further, the located threshold can be used in the interpretation of results of a detailed analysis of preferences derived from the satisfaction data.

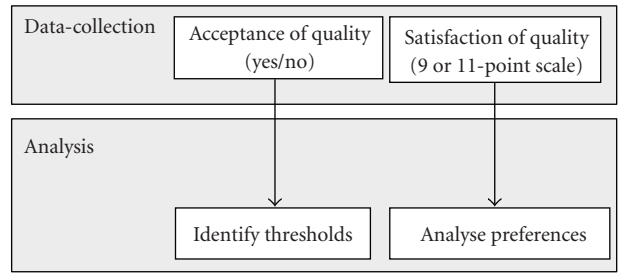


Figure 12: Data-collection and analysis for bidimensional measure combining retrospective overall acceptance and satisfaction.

This work focused on presenting a bidimesional research method, but it did not aim to model bidimensionality in the level of analysis. Our studies showed evidence that the location of acceptance threshold on satisfaction scale is relative to the measured phenomena. To name the constant values for the threshold on a satisfaction scale might be impossible and it might restrict the use of method for measuring different quality ranges. However, our study was limited to the evaluations of clearly detectable and small differences around the threshold. Further work needs to explore the behavior of these measures on the high or low levels of produced qualities for modeling the actual usage of the different scales. In addition, to validate the bidimensional method, further studies need to apply it for studying all multimedia abstraction layers and their interaction. This study targeted only the network and media levels while less attention was paid to the content layer. Finally, to broaden the presented method, there is a need to explore acceptability evaluations in relation to other user-oriented assessment tasks, like examination of goals of viewing.

7.2. Overall acceptance based on evaluation of instantaneous changes

As a minor result, a simplified continuous assessment task to evaluate instantaneous quality changes can be used in parallel with retrospective evaluation methods in quality assessment. This data-collection method can offer insights for the annoying factors in time-varying quality [76] without changing human information processing which has been the shortcoming of the previous methods [53]. When talking about constructing overall evaluations based on instantaneous assessments, there are some challenges. Our results showed that the overall acceptance scores of continuous assessment were relatively high all the time and were not very well distinguishable in the terms of retrospective acceptance. Moreover, their ability to differentiate small differences between stimuli was limited. All of these aspects might have been impacted by an additive approach for constructing the overall evaluations we used. In this trail, further work needs to examine other possibilities for the use of instantaneously

recorded data to predict overall quality by weighting the certain segments of evaluations, like peaks and ends [77]. This perspective can also reveal something new from the fundamental problem of relation between parts and whole in the human information processing. On the other hand, this approach does not necessarily erase the need of measuring a retrospective acceptance anchor. In the current phase, we recommend using a simplified continuous assessment method for tracking the acceptability of instantaneous changes (e.g., [76]) in parallel to retrospective methods, but not as the only method for evaluating overall acceptance.

7.3. Conclusions

This study presented an evaluation method of acceptance representing the minimum level of user requirements in which user expectations and needs are fulfilled. The proposed bidimensional evaluation method combining acceptance and satisfaction can be extended or integrated into any consumer- or user-oriented sensory studies to ensure the level of minimum quality of a relevant component. For example, in the context of multimedia quality, it can be added to an existing QoP model targeting the measurement of quality preference and goals of viewing [12, 13]. The method can also help system developers to test meaningful parameter combinations when testing a novel set of parameters, parameter combinations or several modalities (e.g., audio-video parameter combinations for mobile 3D television).

However, acceptability measurement is just one of the first steps on the way to understanding consumer- or user-oriented experienced multimedia quality. Our long-term aim is not only to focus on acceptance evaluation as method to ensure the quality of a critical system component, but also to understand the effect of user characteristics, system design, and the actual context of use on experienced quality.

ACKNOWLEDGMENTS

This study was funded by Radio- ja televisotekniikan tutkimus Oy (RTT). RTT is a nonprofit research company specialized in digital television datacasting and rich media development. Satu Jumisko-Pyykkö's work was funded by the UCIT graduate school and this article was supported by the HPY research foundation and Finnish Cultural Foundation. The authors wish to thank Hannu Alamäki and Kati Nevalainen about their work in the project.

REFERENCES

[1] R. J. Abbott, *An Integrated Approach to Software Development*, John Wiley & Sons, New York, NY, USA, 1986.

[2] V. Roto, *Web browsing on mobile phones—characteristics of user experience*, Ph.D. dissertation, Helsinki University of Technology, Helsinki, Finland, 2006.

[3] M. Hassenzahl and N. Tractinsky, "User experience—a research agenda," *Behaviour & Information Technology*, vol. 25, no. 2, pp. 91–97, 2006.

[4] S. J. Barnes, "The mobile commerce value chain: analysis and future developments," *International Journal of Information Management*, vol. 22, no. 2, pp. 91–108, 2002.

[5] S. Bech and N. Zacharov, *Perceptual Audio Evaluation: Theory, Method and Application*, John Wiley & Sons, New York, NY, USA, 2006.

[6] P. G. Engeldrum, *Psychometric Scaling: A Toolkit for Imaging Systems Development*, Imcotek Press, Winchester, Mass, USA, 2000.

[7] H. T. Lawless and H. Heyman, *Sensory Evaluation of Food: Principles and Practices*, Chapman & Hall/CRC, New York, NY, USA, 1998.

[8] S. Jumisko-Pyykkö, ""I would like to see the subtitles and the face or at least hear the voice": effects of picture ratio and audio-video bitrate ratio on perception of quality in mobile television," *Multimedia Tools and Applications*, vol. 36, no. 1-2, pp. 167–184, 2008.

[9] H. Knoche, J. D. McCarthy, and M. A. Sasse, "Can small be beautiful? Assessing image resolution requirements for mobile TV," in *Proceedings of the 13th Annual ACM International Conference on Multimedia (MULTIMEDIA '05)*, pp. 829–838, Singapore, November 2005.

[10] H. Knoche, J. McCarthy, and M. A. Sasse, "How low can you go? The effect of low resolutions on shot types in mobile TV," *Multimedia Tools and Applications*, vol. 36, no. 1-2, pp. 145–166, 2008.

[11] ITU-T P.911 Recommendation P.911, "Subjective audiovisual quality assessment methods for multimedia application," International Telecommunication Union - Telecommunication sector, 1998.

[12] G. Ghinea and J. P. Thomas, "QoS impact on user perception and understanding of multimedia video clips," in *Proceedings of the 6th ACM International Conference on Multimedia (MULTIMEDIA '98)*, pp. 49–54, Bristol, UK, September 1998.

[13] S. R. Gulliver, T. Serif, and G. Ghinea, "Pervasive and standalone computing: the perceptual effects of variable multimedia quality," *International Journal of Human Computer-Studies*, vol. 60, no. 5-6, pp. 640–665, 2004.

[14] J. D. McCarthy, M. A. Sasse, and D. Miras, "Sharp or smooth?: comparing the effects of quantization vs. frame rate for streamed video," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '04)*, pp. 535–542, Vienna, Austria, April 2004.

[15] K. Nahrstedt and R. Steinmetz, "Resource management in networked multimedia systems," *Computer*, vol. 28, no. 5, pp. 52–63, 1995.

[16] G. Wikstrand, *Improving user comprehension and entertainment in wireless streaming media: introducing cognitive quality of service*, Ph.D. thesis, Department of Computer Science, Umeå University, Umeå, Sweden, 2003.

[17] S. R. Gulliver and G. Ghinea, "Defining user perception of distributed multimedia quality," *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 2, no. 4, pp. 241–257, 2006.

[18] S. Jumisko-Pyykkö, M. V. Vinod Kumar, M. Liinasuo, and M. M. Hannuksela, "Acceptance of audiovisual quality in erroneous television sequences over a DVB-H channel," in *Proceedings of the 2nd International Workshop in Video Processing and Quality Metrics for Consumer Electronics*, Scottsdale, Ariz, USA, January 2006.

[19] S. Jumisko-Pyykkö and J. H"akkinen, "Evaluation of subjective video quality of mobile devices," in *Proceedings of the*

13th Annual ACM International Conference on Multimedia (MULTIMEDIA '05), pp. 535–538, Singapore, November 2005.

[20] S. Winkler and C. Faller, "Audiovisual quality evaluation of low-bitrate video," in *Human Vision and Electronic Imaging X*, vol. 5666 of *Proceedings of SPIE*, pp. 139–148, San Jose, Calif, USA, January 2005.

[21] H. Knoche, H. de Meer, and D. Kirsh, "Extremely economical: how key frames affect consonant perception under different audio-visual skews," in *Proceedings of the 16th World Congress on Ergonomics (IEA '06)*, Maastricht, The Netherlands, July 2006.

[22] M. M. Hannuksela, V. K. Malamal Vadakital, and S. Jumisko-Pyykkö, "Synchronized audio redundancy coding for improved error resilience in streaming over DVB-H," in *Proceedings of the 3rd International Mobile Multimedia Communications Conference (MobiMedia '07)*, Nafpaktos, Greece, August 2007.

[23] M. M. Hannuksela, V. K. Malamal Vadakital, and S. Jumisko-Pyykkö, "Comparison of error protection methods for audio-video broadcast over DVB-H," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, Article ID 71801, 12 pages, 2007.

[24] N. Ravaja, K. Kallinen, T. Saari, and L. Keltikangas-J"arvinen, "Suboptimal exposure to facial expressions when viewing video messages from a small screen: effects on emotion, attention, and memory," *Journal of Experimental Psychology: Applied*, vol. 10, no. 2, pp. 120–113, 2004.

[25] H. O. Knoche, J. D. McCarthy, and M. A. Sasse, "Reading the fine print: the effect of text legibility on perceived video quality in mobile tv," in *Proceedings of the 14th Annual ACM International Conference on Multimedia (MULTIMEDIA '06)*, pp. 727–730, Santa Barbara, Calif, USA, October 2006.

[26] A. Köpke, A. Willig, and H. Karl, "Chaotic maps as parsimonious bit error models of wireless channels," in *Proceedings of the 22nd Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '03)*, vol. 1, pp. 513–523, San Francisco, Calif, USA, March-April 2003.

[27] A. Willig, M. Kubisch, C. Hoene, and A. Wolisz, "Measurements of a wireless link in an industrial environment using an IEEE 802.11-compliant physical layer," *IEEE Transactions on Industrial Electronics*, vol. 49, no. 6, pp. 1265–1282, 2002.

[28] I. S. Reed and G. Solomon, "Polynomial codes over certain finite fields," *SIAM Journal of Applied Mathematics*, vol. 8, no. 2, pp. 300–304, 1960.

[29] K. Grill-Spector and R. Malach, "The human visual cortex," *Annual Review of Neuroscience*, vol. 27, pp. 649–677, 2004.

[30] M. S. Lewicki, "Efficient coding of natural sounds," *Nature Neuroscience*, vol. 5, no. 4, pp. 356–363, 2002.

[31] S. T. Fiske and S. E. Taylor, *Social Cognition*, McGrow-Hill, Singapore, 1991.

[32] U. Neisser, *Cognition and Reality: Principles and Implications of Cognitive Psychology*, W.H. Freeman, San Francisco, Calif, USA, 1976.

[33] K. Oatley and J. M. Jenkins, *Understanding Emotions*, Blackwell, Oxford, UK, 2003.

[34] S. Jumisko-Pyykkö, J. H"akkinen, and G. Nyman, "Experienced quality factors: qualitative evaluation approach to audiovisual quality," in *Multimedia on Mobile Devices 2007*, vol. 6507 of *Proceedings of SPIE, 65070M*, pp. 1–12, San Jose, Calif, USA, January 2007.

[35] M. C. Meilgaard, G. V. Civille, and B. T. Carr, *Sensory Evaluation Techniques*, CRC Press, New York, NY, USA, 1999.

[36] R. G. Picard, "Mobile telephony and broadcasting: are they compatible for consumers," *International Journal of Mobile Communications*, vol. 3, no. 1, pp. 19–28, 2005.

[37] R. G. Picard, "Interacting forces in the development of communication technologies," *European Media Management Review*, vol. 1, no. 1, pp. 18–24, 1998.

[38] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS Quarterly*, vol. 13, no. 3, pp. 319–340, 1989.

[39] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis, "User acceptance of information technology: toward a unified view," *MIS Quarterly*, vol. 27, no. 3, pp. 425–478, 2003.

[40] M. Amberg, M. Hirschmeier, and J. Wehrmann, "The compass acceptance model for the analysis and evaluation of mobile services," *International Journal of Mobile Communications*, vol. 2, no. 3, pp. 248–259, 2004.

[41] E. Kaasinen, *User acceptance of mobile services—value, ease of use, trust and ease of adoption*, Doctoral thesis, VTT Information Technology, Helsinki, Finland, 2005, VTT publications 566.

[42] M. Pagani, "Determinants of adoption of third generation mobile multimedia services," *Journal of Interactive Marketing*, vol. 18, no. 3, pp. 46–59, 2004.

[43] G. C. Bruner II and A. Kumar, "Explaining consumer acceptance of handheld Internet devices," *Journal of Business Research*, vol. 58, no. 5, pp. 553–558, 2005.

[44] S. Sarker and J. D. Wells, "Understanding mobile handheld device use and adoption," *Communications of the ACM*, vol. 46, no. 12, pp. 35–40, 2003.

[45] R. Aldridge, J. Davidoff, M. Ghanbari, D. Hands, and D. Pearson, "Regency effect in the subjective assessment of digitally-coded television pictures," in *Proceedings of the 5th International Conference on Image Processing and Its Applications (ICIP '95)*, pp. 336–339, Edinburgh, UK, July 1995.

[46] A. D. Baddley, *Working Memory*, Oxford University Press, New York, NY, USA, 1998.

[47] M. Ries, R. Puglia, T. Tebaldi, O. Nemethova, and M. Rupp, "Audiovisual quality estimation for mobile streaming services," in *Proceedings of the 2nd International Symposium on Wireless Communications Systems (ISWCS '05)*, pp. 173–177, Siena, Italy, September 2005.

[48] H. Knoche and J. D. McCarthy, "Good news for mobile TV," in *Proceedings of the 14th Wireless World Research Forum Meeting (WWRF14)*, San Diego, Calif, USA, July 2005.

[49] R. P. Aidridge, D. S. Hands, D. E. Pearson, and N. K. Lodge, "Continuous quality assessment of digitally-coded television pictures," *IEE Proceedings: Vision, Image and Signal Processing*, vol. 145, no. 2, pp. 116–123, 1998.

[50] H. de Ridder and R. Hamberg, "Continuous assessment of image quality," *SMPTE Journal*, vol. 106, no. 2, pp. 123–128, 1997.

[51] R. Hamberg and H. de Ridder, "Time-varying image quality: modeling the relation between instantaneous and overall quality," *SMPTE Journal*, vol. 108, no. 11, pp. 802–811, 1999.

[52] A. Bouch and M. A. Sasse, "Case for predictable media quality in networked multimedia applications," in *Multimedia Computing and Networking 2000*, K. Nahrstedt and W. Feng, Eds., vol. 3969 of *Proceedings of SPIE*, pp. 188–195, San Jose, Calif, USA, January 2000.

[53] D. S. Hands and S. E. Avons, "Recency and duration neglect in subjective assessment of television picture quality," *Applied Cognitive Psychology*, vol. 15, no. 6, pp. 639–657, 2001.

[54] R. T. Apteker, J. A. Fisher, V. S. Kisimov, and H. Neishlos, "Video acceptability and frame rate," *IEEE Multimedia*, vol. 2, no. 3, pp. 32–40, 1995.

[55] D. Wijesekera, J. Srivastava, A. Nerode, and M. Foresti, "Experimental evaluation of loss perception in continuous media," *Multimedia Systems*, vol. 7, no. 6, pp. 486–499, 1999.

[56] R. Steinmetz, "Human perception of jitter and media synchronization," *IEEE Journal on Selected Areas in Communications*, vol. 14, no. 1, pp. 61–72, 1996.

[57] N. Kitawaki, Y. Arayama, and T. Yamada, "Multimedia opinion model based on media interaction of audiovisual communications," in *Proceedings of the 4th International Conference on Measurement of Speech and Audio Quality in Networks (MESAQIN '05)*, pp. 5–10, Prague, Czech Republic, June 2005.

[58] A. Watson and M. A. Sasse, "The good, the bad, and the muffled: the impact of different degradations on Internet speech," in *Proceedings of the 8th ACM International Conference on Multimedia (MULTIMEDIA '00)*, pp. 269–276, Los Angeles, Calif, USA, October-November 2000.

[59] R. Pastrana, J. Gicquel, C. Colomes, and H. Cherifi, "Sporadic Signal Loss Impact on Auditory Quality Perception," 2004, http://wireless.feld.cvut.cz/mesaqin2004/contributions.html.

[60] R. R. Pastrana-Vidal, J. C. Gicquel, C. Colomes, and H. Cherifi, "Sporadic frame dropping impact on quality perception," in *Human Vision and Electronic Imaging IX*, vol. 5292 of *Proceedings of SPIE*, pp. 182–193, San Jose, Calif, USA, January 2004.

[61] ITU-R BT.500-11, "Methodology for the subjective assessment of the quality of television pictures," International Telecommunications Union - Radiocommunication sector, 2002.

[62] E. M. Rogers, *Diffusion of Innovations*, Free Press, New York, NY, USA, 5th edition, 2003.

[63] A. Watson and M. A. Sasse, "Measuring perceived quality of speech and video in multimedia conferencing applications," in *Proceedings of the 6th ACM International Conference on Multimedia (MULTIMEDIA '98)*, pp. 55–60, Bristol, UK, September 1998.

[64] A. Watson and M. A. Sasse, "Evaluating audio and video quality in low-cost multimedia conferencing systems," *Interacting with Computers*, vol. 8, no. 3, pp. 255–275, 1996.

[65] C. Carlsson and P. Walden, "Mobile TV—to live or die by content," in *Proceedings of the 40th Annual Hawaii International Conference on System Sciences (HICSS '07)*, p. 51, Waikoloa, Hawaii, USA, January 2007.

[66] C. Södergård, Ed., "Mobile television—technology and user experiences," VTT Publications 506, VTT Information Technology, Espoo, Finland, 2003.

[67] ETSI, "Digital Video Broadcasting (DVB); Specification for the use of video and audio coding in DVB services delivered directly over IP," ETSI standard, ETSI TS 102 005 V1.2.0, 2005.

[68] G. Faria, J. A. Henriksson, E. Stare, and P. Talmola, "DVB-H: digital broadcast services to handheld devices," *Proceedings of the IEEE*, vol. 94, no. 1, pp. 194–209, 2006.

[69] ETSI, "Digital Video Broadcasting (DVB): Transmission systems for handheld terminals," ETSI standard, EN 302 304 V1.1.1, 2004.

[70] ETSI, "Digital Video Broadcasting (DVB): DVB specification for data broadcasting," ETSI standard, EN 301 192 V1.4.1, 2004.

[71] ITU-T P.920, "Interactive test methods for audiovisual communications," International Telecommunications Union - Telecommunication sector, 2002.

[72] H. Coolican, *Research Methods and Statistics in Psychology*, J. W. Arrowsmith, London, UK, 4th edition, 2004.

[73] E. A. Styles, *The Psychology of Attention*, Psychology Press, Hove, UK, 1997.

[74] U. Reiter and S. Jumisko-Pyykkö, "Watch, press, and catch—impact of divided attention on requirements of audiovisual quality," in *Proceedings of the 12th International Conference on Human-Computer Interaction (HCI '07)*, pp. 943–952, Beijing, China, July 2007.

[75] V. K. Malamal Vadakital, M. M. Hannuksela, M. Rezaei, and M. Gabbouj, "Method for unequal error protection in DVB-H for mobile television," in *Proceedings of the 17th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '06)*, pp. 1–5, Helsinki, Finland, September 2006.

[76] S. Jumisko-Pyykkö, M. V. Vinod Kumar, and J. Korhonen, "Unacceptability of instantaneous errors in mobile television: from annoying audio to video," in *Proceedings of the 8th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '06)*, pp. 1–8, Helsinki, Finland, September 2006.

[77] R. M. Hogarth and H. J. Einhorn, "Order effects in belief updating: the belief-adjustment model," *Cognitive Psychology*, vol. 24, no. 1, pp. 1–55, 1992.

Hindawi Publishing Corporation
International Journal of Digital Multimedia Broadcasting
Volume 2008, Article ID 171385, 13 pages
doi:10.1155/2008/171385

Research Article

Adapting Content Delivery to Limited Resources and Inferred User Interest

Cezar Plesca,[1] Vincent Charvillat,[2] and Romulus Grigoras[2]

[1] *Computer Science Department, Military Technical Academy, 050141 Bucharest, Romania*
[2] *Computer Science Department, National Polytechnic Institute of Toulouse, 31071 Toulouse, France*

Correspondence should be addressed to Cezar Plesca, cezar.plesca@gmail.com

Received 3 March 2008; Accepted 1 July 2008

Recommended by Harald Kosch

This paper discusses adaptation policies for information systems that are subject to dynamic and stochastic contexts such as mobile access to multimedia web sites. In our approach, adaptation agents apply sequential decisional policies under uncertainty. We focus on the modeling of such decisional processes depending on whether the context is fully or partially observable. Our case study is a movie browsing service in a mobile environment that we model by using Markov decision processes (MDPs) and partially observable MDP (POMDP). We derive adaptation policies for this service, that take into account the limited resources such as the network bandwidth. We further refine these policies according to the partially observable users' interest level estimated from implicit feedback. Our theoretical models are validated through numerous simulations.

1. INTRODUCTION

Access alternatives to computer services continue to progress, facilitating our interaction with family, friends, or workplace. These new access alternatives encompass a wide range of mobile and distributed devices that our technological environment becomes truly pervasive. The execution contexts in which these devices operate are naturally heterogeneous. The resources offered by wireless networks vary with the number and the position of connected users. The available memory and the processing power also fluctuate dynamically. Last but not least, the needs and expectations of users can change at any instant. As a consequence, there are numerous research projects that aim to provide modern information systems with adaptation capabilities according to context variability.

In order to handle highly dynamic contexts, the approach that we propose in this paper is based on an adaptation agent. The agent perceives the successive states of the context, thanks to observations, and carries out adaptation actions. Often, the adaptations approaches proposed in literature suppose that the contextual data is easy to perceive or at least that there is no possible ambiguity to identify the state of the current context. One calls this an *observable context*. In this work, we relax this hypothesis and therefore deal with *partially observable contexts*.

Our case study is an information system for browsing multimedia descriptions of movies on mobile devices. The key idea is to show how a given adaptation strategy can be refined according to the estimation of user interest. User interest is clearly not directly observable by the system.

We build upon research on "implicit feedback" in order to allow the adaptation agent to estimate the user interest level while interacting with the context [1, 2]. The first section of this paper reviews important elements of the state of the art and details our adaptation approach. Next, we introduce the two formalisms used by our model: the Markov decision processes (MDPs) and the partially observable MDP (POMDP). The following section presents our case study and establishes the operational principles of this information system. Thanks to an MDP, we formalize an adaptation policy for our information system seen as an observable context. Then we show how to refine this policy according to user interest using a POMDP (refined itself from an MDP). Various experiments validate this approach and give a practical view of the behavior of an adaptation agent. We conclude this paper with some perspectives on this work.

2. RELATED WORK

This section introduces useful current literature in the field of adaptation to dynamic execution contexts which helps to position our adaptation approach. Adaptive systems commonly provide adaptation capabilities and therefore, these systems can be categorized according to available resources, user preferences, or more generally, to the context.

2.1. Resource-based adaptation

Given the heterogeneous nature of modern networks and mobile devices, there is an obvious need for adaptation to limited resources. Networks' QoS parameters vary in terms of available bandwidth, loss rate, or latency. The capabilities of the terminal are also very heterogeneous in terms of memory size, processing power, and display area.

To manage these limitations, one can adapt the content to be displayed or the access/distribution modalities. When considering content adaptation, several authors propose classifications [3] where the elementary components of the content (a media, e.g.) or the entire document's structure is to be transformed. A media can thus be transcoded [4], converted into another modality [5], or summarized [6]. The distribution or the access can also be adapted, for example, by optimizing the streaming [7] or by modifying the degree of interactivity of the service.

2.2. User-aware adaptation

In addition to adaptation capabilities to the available resources, one should also consider an application's adaptation according to human factors which are a matter of user preferences and satisfaction. Henceforth, we describe three main research directions as given by the literature.

The first research direction consists of switching the adaptation mechanisms for maximizing the quality of the service perceived by the user. A typical scenario is the choice of the transcoding strategy of a stream (e.g., a video stream) in order to maximize the perceptual quality given a limited bandwidth [8]. What is the best parameter to adapt: the size of the video, its chromatic resolution, or the frame-rate? Models had been proposed [9, 10] to assess quality variation both from technical and user perspectives. They are organized on three distinct levels: network, media, and content levels. For this line of research, the key factor for consideration is how variation in objective multimedia quality impacts on user perception.

A second active direction is related to user modeling. Here, the idea is to customize an application by modeling user profiles in order to recognize them later. For example, adaptive hypermedia contents or services [11] provide a user with navigation support for "easier/better learning using an on-line educational service" or support for "more efficient selling on an e-commerce site" according to the user profile. Very often, these systems use data mining techniques to analyze access patterns and discover interesting relations in usage data [12]. Such knowledge may be useful to recognize profiles and select the most appropriate modifications to improve content effectiveness.

The third research direction finds its motivation in the first two. In order to learn a user model or to evaluate the perceptual impact of a content adaptation solution, it is necessary to either explicitly ask users for evaluations or to obtain implicit feedback information. Research aiming to evaluate "implicit feedback" (IF) is experiencing a growing interest, since it avoids bringing together significant collections of explicit returns (which is intrusive and expensive) [1]. These IF methods are used in particular to decode user reactions in information search systems [2]. The idea is to measure the user interest for a list of query results, in order to adapt the search function. Among the studied implicit feedback signals one can consider: the total browsing time, the number of clicks, the scrolling interactions, and some characteristic sequences of interactions. In our work, we estimate user interest using IF by interpreting interaction sequences [2, 13]. Moreover, from a metadata perspective, IF can provide implicit descriptors like user interest descriptor as shown in [14].

2.3. Mixing resources and user-aware adaptation

More general adaptation mechanisms can be obtained by combining resource-based with user-based adaptation. The characteristics of users and resources are mixed to design an adaptation strategy for a given *context*. For example, streaming of a heavy media content can be adapted by prefetching while considering both users characteristics and resource constraints [15].

For mobile and pervasive systems, the link between resources and users starts by taking into account the geolocalization of the user, that can be traced in time and even predicted [16].

In the MPEG-21 digital item adaptation (DIA) standard, the context descriptors group the network's and the terminal's capabilities together with the user's preferences and the authors' recommendations to adapt multimedia productions. Given this complexity, the normative works only propose tools simply for describing the running context as a set of carefully chosen and extensible descriptors [17]. This is an approach by metadata that leaves free the conception of adaptation components while authorizing a high level of interoperability [18].

Naturally, the elements of the context vary in time. Therefore, one speaks of a dynamic context and, by extension, of a dynamic adaptation. It is important to note that static adaptation to static context elements is possible as well: one can negotiate once for all and always in the same manner the favorite language of a user at the moment of access to a multilingual service. On the contrary, the adaptation algorithm itself and/or its parameters can be dynamically changed according to the context state [19]. Our adaptation approach is in line with the latter case.

An important element of research in context adaptation is also the distinction between the adaptation decision and its effective implementation [18]. In a pervasive system, one can decide that a document must be transcoded into

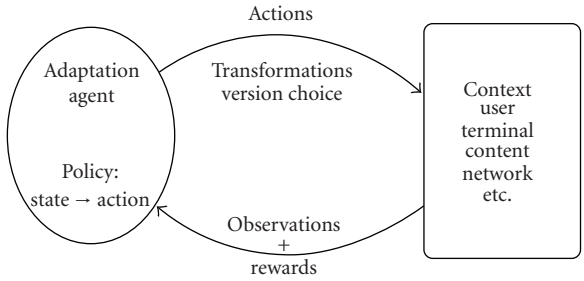


Figure 1: Context-based adaptation agent.

another format, but some questions still need to be answered. Is a transcoding component available? Where can it be found? Should one compose the transcoding service? In order to find solutions to these questions, many authors propose to use artificial learning techniques to select the right decision and/or the appropriate implementation of adaptation mechanisms (see [20] for a review). In this case, a description of the running context is given as input to a decision-making agent that predicts the best adaptation actions according to what it has previously learned. We extend this idea in line with a reinforcement learning principle.

We model the context dynamics by a Markov decision process whose states are completely or partially observable. This approach provides means to find the optimal decision (adaptation action) according to the current context. Next section introduces our MDP-based adaptation approach.

3. MARKOV DECISION PROCESSES-OUR FORMAL APPROACH

Figure 1 summarizes our adaptation approach that has been introduced in [21] and is further refined in this article. In this paper, an adaptation strategy for dynamic contexts is applied by an adaptation agent. This agent perceives sequentially, over a discrete temporal axis, the variations of the context through observations.

From its observations, the agent will compute the context state in order to apply an adaptation policy. Such a policy is simply a function that maps context states to adaptation decisions. Therefore, the agent acts on the context while deciding an adaptation action: it consumes bandwidth, influences the future user's interactions, increases, or reduces the user's interest. It is therefore useful to measure its effect by associating a reward (immediate or delayed) with the adaptation action decided in a given context state. The agent can thus learn from its interaction with the context and perform a "trial-and-error" learning called reinforcement learning [22]. It attempts to reinforce the actions resulting in a good accumulation of rewards and, conversely, avoids renewing fruitless decisions. This process represents a continuous improvement of its "decision policy."

This dynamic adaptation approach is common to frameworks of *sequential decisional policies under uncertainty*. In these frameworks, the uncertainty comes from two sources. On the one hand, the dynamic of the context can be random as a consequence of available resources' variability (e.g., the bandwidth); on the other hand, the effect of an agent's decision can be itself random. For example, if an adaptation action aims to anticipate user interactions, the prediction quality is obviously uncertain and subject to the user's behavior variations.

In this situation, by adopting a Markov definition of the context state, the agent's dynamics can be modeled as a Markov decision process (MDP). This section introduces this formalism.

We initially assume that context state variables are observable by the agent which makes it a sufficient condition to identify the decision state without any ambiguity. This paper takes a step forward by refining adaptation policies according to user interest. We estimate sequentially this hidden information through user behavior as suggested by research on the evaluation of "implicit feedback." Therefore, the new decision-making state contains at the same time observable variables as well as a hidden element associated with user interest.

We then move on from an MDP to a partially observable Markov decision process (POMDP). To the best of our knowledge, the application of the POMDP to the adaptation problem in partially observable contexts has not been studied before. To give concrete expression to this original idea, a case study will be presented in Section 4.

3.1. MDP definition

An MDP is a stochastic controlled process that assigns rewards to transitions between states [23]. It is defined as a quintuple $(S; A; T; p; r_t)$, where S is the state space, A is the action space, T is the discrete temporal axis of instants when actions are taken, $p()$ are the probability distributions of the transitions between states, and $r_t()$ is a function of reward on the transitions. We rediscover in a formal way the ingredients necessary to understand Figure 1: at each instant $t \in T$, the agent observes its state $\sigma \in S$, applies the action $a \in A$ that brings the system (randomly, according to $p(\sigma' \mid \sigma, a)$) to a new state σ', and receives a reward $r_t(\sigma, a)$.

As previously mentioned, we are looking for the best policy with respect to the accumulated rewards. A policy is a function π that associates an action $a \in A$ with each state $\sigma \in S$. Our aim is to find the best one: π^*.

The MDP theoretical framework assigns a *value function* V_π to each policy π. This value function associates each state $\sigma \in S$ with a global reward $V_\pi(\sigma)$, obtained by applying π beginning with σ. Such a value function allows to compare policies. A policy π outperforms another policy π' if

$$\forall \sigma \in S, \quad V_\pi(\sigma) \geq V_{\pi'}(\sigma). \tag{1}$$

The expected sum of rewards obtained by applying π starting from σ is weighted by a parameter γ in order to limit the influence of infinitely distant rewards,

$$\forall \sigma \in S, \quad V_\pi(\sigma) = E\left[\sum_{t=0}^{\infty} \gamma^t r_t \mid \sigma_0 = \sigma\right]. \tag{2}$$

In brief, for each state, this value function gives the expected sum of future rewards that can be obtained if the policy π

is applied from this state on. This value function allows to formalize the research of the optimal policy π^* which is the one associated with the best value function $V^* = V_{\pi^*}$.

Bellman's optimality equations characterize the optimal value function V^* and an optimal policy π^* that can be obtained from it. In the case of the γ-weighted criterion and stationary rewards, they can be written as follows:

$$V^*(\sigma) = \max_{a\in A}\left(r(\sigma,a) + \gamma \sum_{\sigma'\in S} p(\sigma' \mid \sigma, a)V^*(\sigma')\right),$$
$$\pi^*(\sigma) = \operatorname*{argmax}_{a\in A}\left(r(\sigma,a) + \gamma \sum_{\sigma'\in S} p(\sigma' \mid \sigma, a)V^*(\sigma')\right). \quad (3)$$
$$\forall \sigma \in S$$

3.2. Resolution and reinforcement learning

When considering to solve an MDP, we can distinguish between two cases, according to whether the model is known or unknown. When the model (probabilities $p()$) and the rewards are known, a dynamic programing solution can be found.

The operator L verifying $V_{n+1} = L\cdot V_n$ according to

$$V_{n+1}(\sigma) = \max_a\left(r(\sigma,a) + \gamma\sum_{\sigma'} p(\sigma' \mid \sigma, a)V_n(\sigma')\right) \quad (4)$$

is a contraction. The Bellman equation in $V^*(\sigma)$ can be solved by using a fixed point iterative method while choosing randomly V_0, then applying repeatedly the operator L that improves the current policy associated to V_n. If the rewards are bounded, the sequence converges to V^* and allows to compute π^*.

If the model is unknown, we can solve the MDP using a reinforcement learning algorithm [22]. The reinforcement learning approach aims to find an optimal policy through iterative estimations of the optimal value function. The *Q-learning* algorithm is a reinforcement learning method that is able to solve the Bellman equations for the γ-weighted criterion. It uses simulations to iteratively estimate the value function V^*, based on the observations of instantaneous transitions and their associated reward. For this purpose, Puterman [23] introduced a function Q, that carries a significance similar to that of V but makes it easier to extract the associated policy because it does not need transition probabilities any more. We can express the "Q-value" as a function of a given policy π and its value function,

$$\forall \sigma \in S,\ a \in A, \quad Q_\pi(\sigma,a) = r(\sigma,a) + \gamma\sum_{\sigma'} p(\sigma' \mid \sigma, a)V_\pi(\sigma'). \quad (5)$$

Therefore, it is easy to see that, in spite of the lack of transition probabilities, we can trace back to the optimal policy,

$$\forall \sigma \in S, \quad V^*(\sigma) = \max_a Q^*(\sigma,a)\ \ \pi^*(\sigma) = \operatorname*{argmax}_a Q^*(\sigma,a). \quad (6)$$

```
Initialize Q_0
for n = 0 to N_tot − 1 do
        σ_n = chooseState
        a_n = chooseAction
        (σ'_n, r_n) = simulate(σ_n, a_n)
        /* update Q_{n+1} */
        Q_{n+1} ← Q_n
        d_n = r_n + (γ max_b Q_n(σ'_n, b)) − Q_n(σ_n, a_n)
        Q_{n+1}(σ_n, a_n) ← Q_n(σ_n, a_n) + α_n(σ_n, a_n)d_n
end for
return Q_{N_tot}
```

Algorithm 1: The *Q-learning* algorithm.

The principle of the *Q-learning* Algorithm 1 says that after each observed transition $(\sigma_n, a_n, \sigma_{n+1}, r_n)$, the current value function Q_n for the couple (σ_n, a_n) is updated, where σ_n represents the current state, a_n the chosen and applied action, σ_{n+1} the resulted state, and r_n the immediate reward.

In this algorithm, N_{tot} is an initial parameter that represents the number of iterations. The *learning rate* $\alpha_n(\sigma,a)$ is particular to each pair state action, and decreases toward 0 at each iteration. The function "`stimulate`" returns a new state and its associated reward according to the dynamics of the system. The choice of the current state and of the action to execute is made by the functions "`chooseState`" and "`chooseAction`." The function "`intialize`" is used to initialize the values Q_0 to 0.

The convergence of this algorithm has been thoroughly studied and is now well established. We assume the following.

(i) S and A are finite, $\gamma \le 1$.

(ii) Each pair (σ,a) is visited an infinite number of times.

(iii) $\sum_n \alpha_n(\sigma,a) = \infty$, $\sum_n \alpha_n(\sigma,a)^2 < \infty$.

Under these hypotheses, the function Q_n converges almost surely to Q^*. Let us recall that the almost-sure convergence means that for all σ, a, the sequence $Q_n(\sigma,a)$ converges to $Q^*(\sigma,a)$ with a probability equal to 1. Practically, the sequence $\alpha_n(\sigma,a)$ is often defined as follows:

$$\alpha_n(\sigma,a) = \frac{1}{n_{\sigma,a}}, \quad (7)$$

where $n_{\sigma,a}$ represents the number of times the state σ was visited and the decision a was made.

3.3. Partial observation and POMDP definition

In many cases, the observations that a decision agent is able to capture (see Figure 1) are only partial and do not allow the identification of the context state without ambiguity. Therefore, a new class of problems needs to be solved: partially observable Markov decision processes. The states of the underlying MDP are hidden and only the observation process will help to rediscover the running state of the process.

A partially observable Markov decision process is defined by:

(i) $(S; A; T; p; r_t)$ the underlying MDP;

(ii) $\mathcal{O}$ a set of observations;

(iii) $O : S \to \Pi(\mathcal{O})$ an observation function that maps every state s to a probability distribution on the observations' space. The probability to observe o knowing the agent's state s will be referred to as follows: $O(s, o) = P(o_t = o \mid s_t = s)$.

Non-Markovian behavior

It is worth to note that, in this new model, we loose a widely used property for the resolution of the MDPs, namely that the observation process is Markovian. The probability of the next observation o_{t+1} may depend not only on the current observation and action taken, but also on previous observations and actions,

$$P(o_{t+1} \mid o_t, a_t) \neq P(o_{t+1} \mid o_t, a_t, o_{t-1}, a_{t-1}, \dots). \quad (8)$$

Stochastic policy

It has been proved that the results obtained for the V and Q convergence using MDP resolution algorithms are not applicable anymore. The POMDPs will need the use of stochastic policies and not deterministic ones, as for MDP [24].

3.4. Resolution

The POMDP classic methods attempt to bring back the resolution problem to the underlying MDP. Two situations are possible. If the MDP model is known, one cannot determine the exact state of the system but a distribution probability on the set of the possible states (a *belief state*). In the second situation, without knowing the model parameters, the agent attempts to construct the MDP model relying only on observations' history.

Our experimental test bed uses the resolution software package provided by Cassandra et al. [25] that works in the potentially infinite space of belief states using linear programing methods.

4. CASE STUDY: A MOVIE PRESENTATION SYSTEM FOR MOBILE TERMINALS

We introduce here a system for browsing movie descriptions on mobile devices. For this system, our strategy aims to adapt the presentation of a multimedia content (i.e., movie description) and not to transform the media itself. This case study is intended to be both simple and pedagogical, while integrating a degree of realistic interactivity.

4.1. Interactive access to a movie database

Figure 2 introduces an information system accessible from mobile terminals such as PDAs. A keyword search allows the user to obtain an ordered list of links to various movie descriptions. Within this list, the user can follow a link toward an interesting movie (the associated interaction will be referred to as *clickMovie*); then, he or she can consult details regarding the movie in question. This consultation will call on a full screen interactive presentation and a navigation scenario detailed below. Having browsed the details for one movie, the user is able to come back to the list of query results (interaction *back* in Figure 2). It is then possible to access the description of a second interesting film. The index of the accessed movie description will be referred to as n.

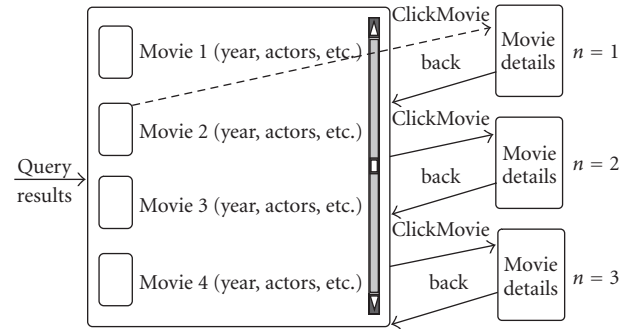


Figure 2: Information system of movie descriptions.

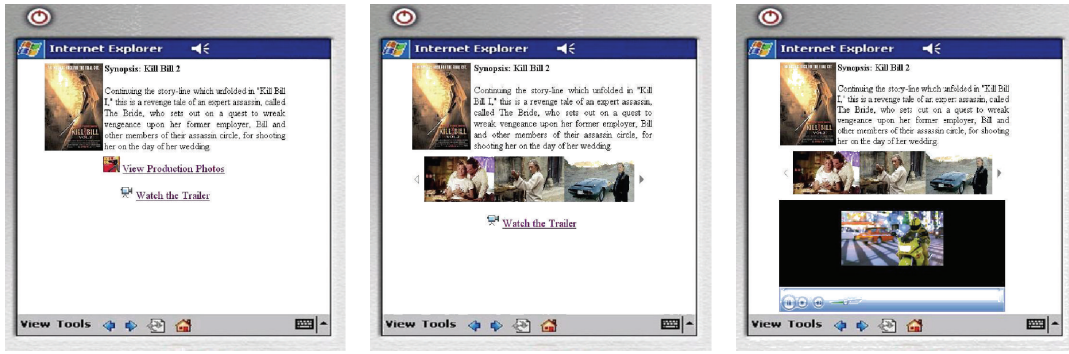
To simplify the context modeling, we choose to consider the browsing sequence indexed by n. Our problem becomes one that aims at adapting the content (movie descriptions) presented during this sequence. Our execution environment is dynamic because of the bandwidth's (bw) variability, a very frequent problem in mobile networks. For simplicity reasons, we do not take into account other important parameters of mobile terminals such as signal strength, user's mobility, and power constraints.

As we consider the browsing session at a high level, we do not need to provide special specifications for the final goal of the service that can be renting/buying a DVD, downloading a media, and so forth. Properly managing the download or the streaming of the whole media is a separate problem and is not considered here.

4.2. From the simplest to the richest descriptions

To present the details of a movie, three forms of descriptions are possible (see Figure 3). The poor "textual" version (referred to as T) groups together with a small poster image, a short text description, and links pointing to more production photos as well as a link to the video trailer. The intermediary version (I) provides a slideshow of still photos and a link to the trailer. The richest version (V) includes, in addition, the video trailer.

As the available bandwidth (bw) is variable, the usage of the three versions is not equivalent. The bandwidth required to download the content increases with the complexity of the versions ($T \to I \to V$). In other words, for a given bandwidth, the latencies perceived by the user during the download of



(a) (b) (c)

Figure 3: Basic (T), intermediary (I), and rich (V) versions of movie details.

the different versions grow proportionally with the size of the content.

More precisely, we now point out two problems generated by the inexistence of dynamic adaptation of the content when the available bandwidth varies. The adaptation strategy could systematically select only one of the three possible alternatives mentioned above. If it always selects the richest version (V), this impacts the behavior of the user who experiences bad network conditions (low bandwidth). Although strong latencies could be tolerated while browsing the first query results (small index n), it becomes quickly unacceptable if n grows. If the adaptation strategy selects systematically the simplest version (T), this would also have a harmful impact on the behavior of the user. Despite the links toward the other resources (I)mages and (V)ideo, the lack of these visual components, which normally stimulate interest, will not encourage further browsing. An important and a legitimate question to be raised is what can be called an "appropriate" adaptation policy.

4.3. Properties of appropriate adaptation policies

The afore-mentioned two examples of policies (one "too ambitious," the other "too modest") show how complex is the relationship among the versions, the number of browsed films, the time spent on the service, the quality of service, the available bandwidth, and the user interest. An in-depth analysis of these relationships can represent a research project in itself. We do not claim to deliver such an analysis in this paper, but we simply want to show how a policy and an adaptation agent can be generated automatically from a model where the context state is observable or partially observable.

Three properties of a good adaptation policy can be identified as follows.

(1) The version chosen for presenting the content must be simplified if the available bandwidth bw decreases (T is simpler than I, itself simpler than V).

(2) The version must be simplified if n increases: it is straightforward to choose rich versions for the first browsed movie descriptions that are normally the most pertinent ones (as we have already mentioned, we should avoid large latencies for big values of n and small bw).

(3) The version must be enriched if the user shows a high interest for the query results. The simple underlying idea is that a very interested user is more likely to be patient and to tolerate more easily large downloading latencies.

The first two properties are related to the variation of the context parameters, that we consider observable (n and bw), while the third one is related to a hidden element, namely, user interest. At this stage, given these three properties, an adaptation policy for our case study can be expressed: the selection of the version (T, I, or V) knowing n and bw and having a way to estimate the interest.

4.4. On navigation scenarios

This paragraph introduces by examples some possible navigation scenarios. Figure 4 illustrates different possible steps during navigation and introduces different events that are tracked. In this figure, the user chooses a film (event *clickMovie*), the presentation in version T is downloaded (event *pageLoad*) without the user interrupting this download. Interested in this film, the user requests the production photos, following the link toward the pictures (event *linkI*). In the one case, the downloading seems too long and the user interrupts it (event *stopDwl* means stopDownload) then returns to the movie list (event *back*). In the other case, the user waits for the downloading of the pictures to finish, then starts viewing the slideshow (event *startSlide*). Either this slideshow is shown completely and then an event *EI* (short for EndImages) is raised, or the visualization is incomplete, leading to the event *stopSlide* (not represented in the figure). Next, the link to the trailer can be followed (event *linkV*); here again an impatient user can interrupt the downloading (*stopDwl*) or start playing the video (*play*). Then the video can be watched completely (event *EV* for EndVideo) or stopped (*stopVideo*), before a return (event *back*).

Obviously, this example does not introduce all the possibilities, especially if the video is not downloaded but

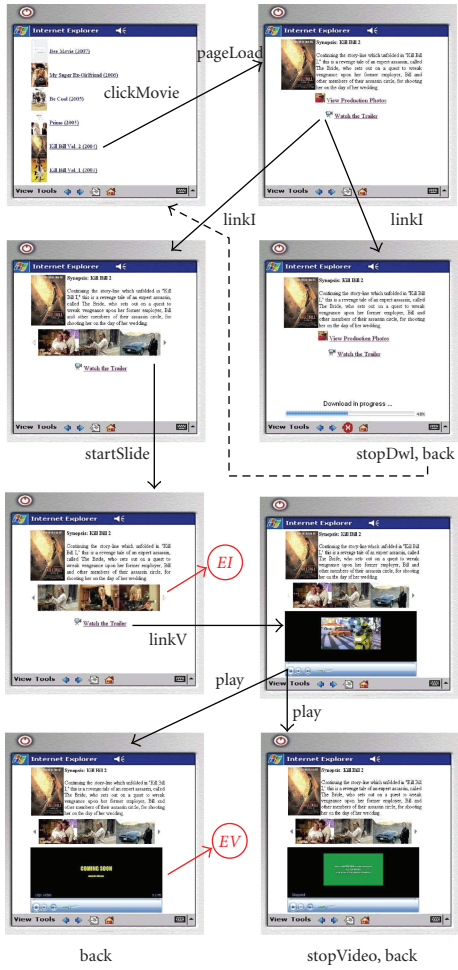


FIGURE 4: Example of navigations and interactions.

streamed. Streaming scenarios introduce different challenges and require a playout buffer that enriches the set of possible interactions (e.g., *stopBuffering*). Meanwhile, the user may choose not to interact with the proposed media: we introduce a sequence of events *pageLoad*, *noInt* (no interaction), *back*. Similarly, a *back* is possible just after a *pageLoad*, a *stopDwl* may occur immediately after the event *clickMovie*, watching the video before the pictures is also possible.

5. PROBLEM STATEMENT

5.1. Rewards for well-chosen adaptation policies

From the previous example and the definitions of associated interactions, it is possible to propose a simple mechanism aiming at rewarding a pertinent adaptation policy. A version (T, I, or V) is considered well chosen in a given context, if it is not questioned by the user. The reassessment of a version T as being too simple is suggested, for example, by the full consumption of the pictures. In the same way, the reassessment of a version V as being too rich is indicated by a partial consumption of the downloaded video. Four simple principles that guide our rewarding system are as follows.

(i) We reward the event *EI* for versions I and V.

(ii) We reward the event *EV* if the chosen version was V.

(iii) We penalize upon arrival of interruption events ("stops").

(iv) We favor the simpler versions for no or little interaction.

Thus, a version T is sufficient if the user does not request (or at least does not completely consume) the pictures. A version I is preferable if the user is interested enough and has access to enough resources to download and view the set of pictures (rewards *EI*). Similarly, a version I is adopted if the user views all the pictures (reward *EI*) and, trying to download the video, is forced to interrupt it because of limited bandwidth. Finally, a rich version V is adopted if the user is in good condition to consume the video completely (reward *EV*). The following decision-making models formalize these principles.

5.2. Toward an implicit measure of the interest

The previously introduced navigations and interactions make it possible to estimate the interest of the user. We proceed by evaluating "implicit feedback" and use the sequences of events to estimate the user's interest level. Our approach is inspired by [26] and is based on the two following ideas.

The first idea is to identify two types of interactions according to what they suggest: either an increasing interest (*linkI, linkV, startSlide, play, EI, EV*) or a decreasing interest (*stopSlide, stopVideo, stopDwl, noInt*). Therefore, the event distribution (seen as the probability of occurrence) depends on the user's interest in the browsed movie.

The second idea is to consider not only *a single running event* to update the estimation of user interest but also to regard *an entire sequence of events* as being more significant. In fact, it has been recently established that the user actions on a response page to a search (e.g., on Google) depend not only on the relevance of the current response but also on the global relevance of the set of the query results [2].

Following the work of [26], it is natural to model the sequences of events or observations produced by a hidden Markov model (HMM) for which we do not detail here the definition (e.g., see [27]). One can simply translate the two previous ideas by using an HMM with several (hidden) states of interest. The three states of interest shown in Figure 5 are referred as S, M, and B, respectively, for a small, medium, or big interest. The three distributions of observable events in every state are different as stressed in the first idea mentioned above. These differences explain the occurrences of different sequences of observations in terms of sequential interest evolutions (second idea). These evolutions are encoded thanks to transition probabilities (stippled) between hidden states of interest. Given a sequence of observations, an HMM can thus provide the most likely underlying sequence of hidden states or the most likely running hidden state. At this point, the characteristics of our information system are rich enough to define an adaptation agent applying decision

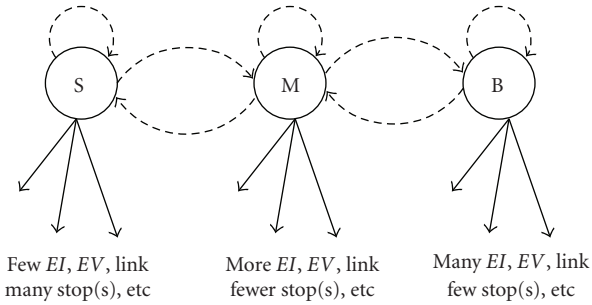


Figure 5: A hidden Markov model.

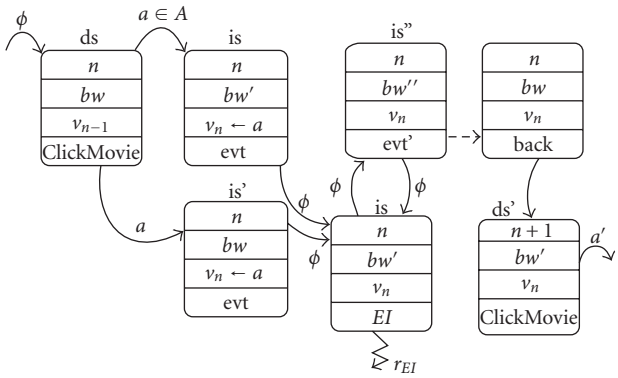


Figure 6: MDP dynamics illustration.

policies under uncertainty. These policies can be formalized in the framework presented in Section 3.1.

6. MODELING CONTENT DELIVERY POLICIES

In this section, we model the dynamic context of our browsing system (Section 4) in order to obtain the appropriate adaptation agents. Our goal is to characterize the adaptation policies in terms of Markov decision processes (MDPs or POMDP).

6.1. MDP modeling

Firstly, an observable context is considered. Let us introduce the proposed MDP that models it. The aim is to characterize adaptation policies which verify properties 1 and 2 described in Section 4.3: the presented movie description must be simplified if the bandwidth available bw decreases or if n increases.

A state (observable) of the context is a tuple $s = \langle n, bw, v, evt\rangle$ with n being the rank of the film consulted, bw the bandwidth available, v the version proposed by the adaptation agent, and evt the running event (see Figure 6). With $n \in [1, 2, \ldots, N_{\max}]$, $bw \in [bw_{\min}; bw_{\max}]$, $v \in \{T, I, V\}$, and $evt \in E$ (where $E = \{$*clickMovie, stopDwl, pageLoad, noInt, linkI, startSlide* $\cdots$ *stopSlide, EI, linkV, play, stopVideo, EV, back*$\}$).

To obtain a finite and reasonable number of such states (limiting thus the MDP size), we will quantize the variables according to our needs. Thus n (resp., bw) can be quantized according to three levels $n \in \{B, M, E\}$ meaning begin, middle, and end (resp., $bw \in \{L, A, H\}$ for low, average, and high) while segmenting in three regions the interval $[1, \ldots, N_{\max}]$ (resp., $[bw_{\min}; bw_{\max}]$).

The temporal axis of MDP is naturally represented by the sequence of events, every event implying a change of state.

The dynamics of our MDP is constrained by the dynamics of the context, especially by the user navigation. Thus, a transition from a movie index n to $n - 1$ is not possible. Similarly, every *back* is followed by an event *clickMovie*. The bandwidth's own dynamics will have also an impact (according to quantized levels) on the dynamics between the states of the MDP.

The choice of the movie description version (T, I, or V) proposed by the adaptation agent is done when the user follows the link to the film. This is encoded in the model by the event *clickMovie*. The states of the MDP can be classified in:

(i) decision states (ds) in which the agent executes a real action (it effectively chooses among T, I, or V);

(ii) nondecision or intermediary (is) states where the agent does not execute any action.

In an MDP framework, the agent decides an action in every single state. Therefore, the model needs to be enriched with an artificial action (ϕ) as well as an absorbent state of strong penalty income ($-\infty$). Thus, any valid action $a \in \mathcal{A} = \{T, I, V\}$ chosen in an intermediary state brings the agent in the absorbent state where it will be strongly penalized. Similarly, the agent will avoid deciding ϕ in a decision-making state where a valid action is desired. Thus, the valid actions mark out the visit of the decision states while the dynamics of the context (subject to user navigation and bandwidth variability) are captured by the transitions between intermediary states for which the action ϕ (the nonaction) is carried out. These properties are clearly illustrated in Figure 6.

In other words, there is no change of version during the transitions between intermediary states. The action a (representing the proposed version) chosen in a decision-making state is therefore, memorized ($v_n \leftarrow a$) in all the following intermediary states, until the next decision state. Thus, the MDP captures the variation of the context dynamics *according to the chosen version*. Therefore, it will be able to identify which are the good choices of versions (to reproduce later in similar conditions), if it is rewarded for them.

The rewards are associated with the decision states according to the chosen action. Intermediary states corresponding to the occurrences of the events *EI* and *EV* are rewarded as well, according to Section 5.1. The rewards (other formulations are possible as well including, e.g., negative rewards for interruption events) are defined as follows:

$$\begin{aligned} r(\langle n, bw, v, clickMovie\rangle, a) &= r_a, \quad a \in \{T, I, V\}, \\ r(\langle n, bw, v, EI\rangle, \phi) &= r_{EI} \quad \text{iff } v \in \{I, V\}, \\ r(\langle n, bw, v, EV\rangle, \phi) &= r_{EV} \quad \text{iff } v = V. \end{aligned} \tag{9}$$

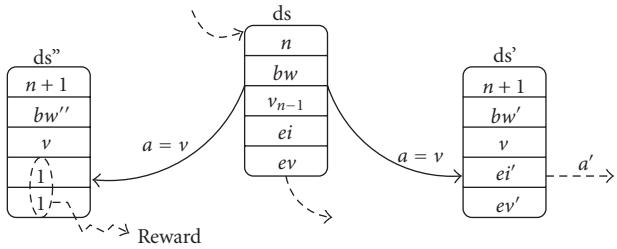


Figure 7: Simplified MDP.

To favor simpler versions for users who do not interact with the content and do not view any media (c.f. Section 5.1), let us choose $r_T > r_I > r_V$. To summarize, the model behaves in the following manner: the agent starts with a decision state *ds*, where it decides a valid action a for which it receives an "initial" reward r_a; the simpler the version, the bigger is the reward. According to the transitions probabilities based on context dynamics, the model goes through intermediary states where it can receive new rewards r_{EI} or r_{EV} at the time of the occurrences of EI (resp., EV), if the taken action a was I or V, (resp., V). As these occurrences are more frequent for small n and high bw, while the absence of interactions is more likely if n is big and bw low, then the MDP

(i) will favor the richest version for small n and high bw;

(ii) will favor the simplest version for big n and low bw;

(iii) will establish a tradeoff (optimum according to the rewards) for all the other cases.

The best policy given by the model is obviously related to the chosen values for r_T, r_I, r_V, r_{EI}, r_{EV}. In order to control this choice in the experimental section, a simplified version of the MDP will be defined.

A simplified MDP can be obtained by memorizing the occurrence of the events EI and EV during the navigation between two events *clickMovie*. Thus, we can delay the rewards r_{EI} or r_{EV}. This simplified model does not contain non decision-making states, if two booleans (ei and ev) are added to the state structure (Figure 7). The boolean ei (resp., ev) passes to 1 if the event EI (resp., EV) is observed between two states. The simplified MDP is defined by its states ($s = \langle n, bw, v, ei, ev\rangle$), the actions $a \in \{T, I, V\}$, the temporal axis given by the sequence of events *clickMovie*, and the rewards r redefined as follows:

$$\begin{aligned} r(\langle *, *, T, *, *\rangle, *) &= r_T, \\ r(\langle *, *, I, ei, *\rangle, *) &= r_I + ei \cdot r_{EI}, \\ r(\langle *, *, V, ei, ev\rangle, *) &= r_V + ei \cdot r_{EI} + ev \cdot r_{EV}. \end{aligned} \tag{10}$$

This ends the presentation of our observable model and we continue by integrating user interest in a richer POMDP model.

6.2. POMDP modeling

The new partially observable model adds a hidden variable (It) to the state. The value of It represents the user's interest quantized on three levels (Small, Average, Big). To be able to estimate user interest, we follow the principles described in Section 5.2 and Figure 5. The events (interactions) are taken out from the previous MDP state to become observations in the POMDP model. These observations are distributed according to It (the interest level). A sequence of observations provides an implicit measure of It, following the same principle described for the HMM in Figure 5. Therefore, it becomes possible for the adaptation agent to refine its decisions according to the probability of the running user's interest: s*mall, average, big*. In other words, this refinement is done according to a belief state. The principle of this POMDP is illustrated in Figure 8.

A hidden state of our POMDP becomes a tuple $s = \langle n, bw, v, ei, ev, It\rangle$. The notations are unchanged including the booleans ei and ev.

The temporal axis and the actions $\{T, I, V, \phi\}$ are unchanged.

The dynamics of the model. When an event *clickMovie* occurs, the adaptation agent is in a decision state *ds*. It chooses a valid action a and moves, according to the model's random transitions, to an intermediary state *is* where ei and ev are equal to 0. The version proposed by the agent is memorized in the intermediary states *is* during the browsing of the current film. The booleans ei and ev become 1, if the events EI or, respectively, EV are observed and preserve this value until the next decision state ds'. During the browsing of the running film, n and v remain constant while the other factors (bw, It, and the booleans) can change.

The observations o are the occurred events: $o \in E$. They are distributed according to the states. In Figure 8, the event *clickMovie* can be observed in *ds* and ds' (probability 1.0) and cannot be observed elsewhere (*is* and is').

In every intermediary state, the event distribution characterizes the value of the interest. Thus, just as the HMM of Figure 5, the POMDP will know how to evaluate, from the sequence of events, the current belief state. The most likely interest value will evolve therefore, along with the events occurred; increase if $linkI, EI, linkV, EV, \ldots$, decrease in case of *stopDwl*, *stopSlide*, *stopVideo*. To preserve the interest level throughout the decision states, the interest of the current *ds* receives the value corresponding to the last *is* (Figure 8).

The rewards associated with the actions taken in a decision-making state *ds* are collected in the following decision-making state ds' where we have all necessary information: v, ei, and ev;

$$\begin{aligned} r(\langle *, *, T, *, *, *\rangle, *) &= r_T, \\ r(\langle *, *, I, ei, *, *\rangle, *) &= r_I + ei \cdot r_{EI}, \\ r(\langle *, *, V, ei, ev, *\rangle, *) &= r_V + ei \cdot r_{EI} + ev \cdot r_{EV}. \end{aligned} \tag{11}$$

7. EXPERIMENTAL RESULTS

Simulations are used in order to experimentally validate the models. The developed software simulates navigations such as the one illustrated in Figure 4. Every transition probability between two successive states of navigation is a stochastic function of three parameters: bw, $It(n)$, and v. The

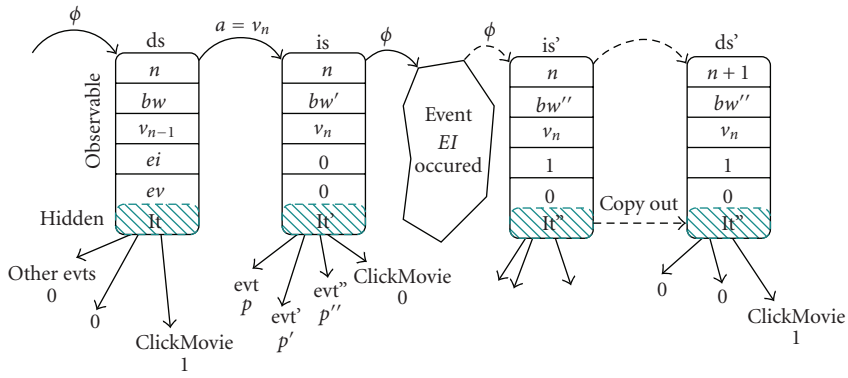


Figure 8: POMDP dynamics between hidden states.

bandwidth bw is simulated as a random variable uniformly distributed in an interval compatible with today mobile networks. $It(n)$ represents a family of random variables, whose expectation decreases with n. The parameter v is the movie version proposed to the user. Meanwhile, other experimental setups involving different distribution lows (e.g., normal distribution) for bandwidth dynamics or user's interest conduct to similar results.

7.1. MDP validation for observable contexts

To validate the MDP model of Section 6.1, let us choose a problem with $N_{\max} = 12$ and $[bw_{\min}; bw_{\max}] = [0; 128\,\text{Kbps}]$. Initially, the intervals of n and bw are quantized on 2 granularity levels: $[1,\ldots,N_{\max}] = N_S \cup N_L$ and $[bw_{\min}; bw_{\max}] = BW_S \cup BW_L$. Rather than proceeding to an arbitrary choice of values $r_T, r_I, r_V, r_{EI}, r_{EV}$ that define the rewards, we can look for the ones driving to the optimal policy shown in Table 1. In fact, this policy π_p^* respects the principles formulated in Section 4.3 and could be proposed beforehand by an expert (Table 1 gives π_p^* only for the pairs n, bw since $\pi_p^*(\langle n, bw, *, \ldots\rangle) = \pi_p^*(n, bw)$.)

The value functions Q corresponding to the simplified MDP, estimated over on a 1length horizon, (between two decision-making states ds and ds') can be written as follows:

$$Q_1(ds, a) = r(ds) + \gamma \sum_{ds'} p(ds' \mid ds, a) r(ds'), \tag{12}$$

because, for all $ds, r(ds, a)$ does not depend on action a.

$$\begin{aligned} Q_1(ds, T) &= r(ds) + \gamma \cdot r_T \left(\sum_{ds'} p(ds' \mid ds, T) \right) \\ &= r(ds) + \gamma \cdot r_T, \\ Q_1(ds, I) &= r(ds) + \gamma (r_I + p_{EI|I} \cdot r_{EI}), \\ Q_1(ds, V) &= r(ds) + \gamma (r_V + p_{EI|V} \cdot r_{EI} + p_{EV|V} \cdot r_{EV}), \end{aligned} \tag{13}$$

where $p_{EI|a}$ and $p_{EV|a}$ represent the probabilities to observe the events EI, respectively, EV, knowing the version a.

Table 1: Policy π_p^* stated for two-level granularity n and bw).

	$N_S(\{1,2,\ldots,6\})$	$N_L(\{7,8,\ldots,12\})$
BW_S[0–64] Kbps	I	T
BW_L[64–128] Kbps	V	I

Table 2: Policy π_p^* refinement for R_1 rewards.

	N_1	N_2	N_3	N_4
BW_1	I	$\underline{T}$	T	T
BW_2	I	I	I	T
BW_3	V	$\underline{I}$	I	$\underline{T}$
BW_4	V	V	$\underline{I}$	I

For every pair (n, bw) we have computed, based on simulations, the probabilities $p_{EI|I}, p_{EI|V}, p_{EV|V}$. The respect of the policy π_p^* is assured if and only if

$$\begin{gathered} \forall a \in \{T, I, V\}, \quad \forall ds = \langle n, bw, \ldots\rangle, \\ Q(ds, \pi_p^*(ds)) \geq Q(ds, a). \end{gathered} \tag{14}$$

Writing these inequalities for the 4 pairs (n, bw) from Table 1 and using the estimations $Q_1(ds, a)$ for Q, we obtain a 12-linear inequations system in the variables $r_T, r_I, r_V, r_{EI}, r_{EV}$. Two solutions of the system among an infinity are as follows:

$$\begin{aligned} R_1 &: r_T = 2,\ r_I = 1,\ r_V = 0,\ r_{EI} = 6,\ r_{EV} = 6, \\ R_2 &: r_T = 2,\ r_I = 1,\ r_V = 0,\ r_{EI} = 7,\ r_{EV} = 7. \end{aligned} \tag{15}$$

Starting from these values, we can experimentally check the correct behavior of our MDP model. Table 2 shows the policy obtained automatically by dynamic programing or Q-learning algorithm, with 4 granularity levels for n and bw and the rewards R_1. This table refines the previous coarse-grained policy; this is not a simple copy of π_p^* actions (e.g., see the pairs (N_2, BW_1): change from I to T, (N_2, BW_3): change from V to I, etc.). This new policy is optimal with respect to the rewards R_1, for this finer granularity level.

Resolving the MDP for the second set of rewards (R_2) gives a different refinement (Table 3) that shows richer versions (underlined) comparing to R_1. The explanation stays in the growth of the rewards associated to the events

Table 3: Policy π_p^* refinement for R_2 rewards.

	N_1	N_2	N_3	N_4
BW_1	I	$\underline{I}$	T	T
BW_2	I	I	I	T
BW_3	V	$\underline{V}$	I	$\underline{I}$
BW_4	V	V	$\underline{V}$	I

EI, EV that induce the choice of a more complex versions, for a long time (V lasts for 3 classes of n, when $bw = BW_4$).

7.2. POMDP validation: interest-refined policies

Once MDPs are calibrated and return appropriate adaptation policies, their rewards can be reused to solve the POMDP models. The goal is to refine the MDP policies for the observable case by estimating user interest.

Two experimental steps are necessary. The first step consists of learning the POMDP model and the second in solving the decision-making problem.

For the learning process, the simpler method consists of empirically estimating the transitions and observations probabilities from the simulator's traces. Starting from these traces, the probabilities are obtained from the frequencies' computation

$$p(o \mid s) = \frac{\#\text{ emissions of } o \text{ from } s}{\#\text{visits of } s},$$
$$p(s' \mid s, a) = \frac{\#\text{ transitions from } s \text{ to } s' \text{ taking action } a}{\#\text{visits of couple } (s, a)}. \tag{16}$$

Having a POMDP model, the resolution is the next step. Solving a POMDP is notoriously delicate and computationally intensive (e.g., see the tutorial proposed at www.pomdp.org). We used the software package *pomdp-solve 5.3* in combination with *CPLEX* (with the more recent strategy called finite grid).

The results returned by pomdp-solve is an automaton that implements a "near optimal" deterministic policy, represented by a decision-making graph (*policy graph*). The nodes of the graph contain the actions ($\{T, I, V, \phi\}$) while the transitions are done according to the observations. Only the transitions made possible by the navigation process are to be exploited.

To illustrate this form of result, let us show one of the automata that is small enough to be displayed on an A4 page (Figure 9). We choose a single granularity level for n and bw and three levels for It. Additionally, we consider that the consumption of the slideshow precedes the consumption of the video. The obtained adaptation policy therefore takes into account only the variation of the estimated user interest (n and bw do not play any role).

Figure 9 shows that the POMDP agent learns to react in a coherent way. For example, starting from a version T, and observing *pageLoad, linkI, startSlide, EI, noInt, back* the following version decided by the POMDP agent is I, which translates the sequence into an interest rise. This rise is even stronger if, after the event EI, the user follows the link *linkV*. This is enough to make the agent select the version V further.

Conversely, starting from version V, an important decrease in interest can be observed on the sequence *startSlide, stopSlide, play, stopVideo, back*, so the system decides T. A smaller decrease in interest can be associated with the sequence *startSlide, stopSlide, play, EV, back*, the next version selected being I. These examples show that there exists a natural correlation between the wealth of the selected versions and the implicit user interest. For this problem, where n and bw are not involved, the version given by the *policy graph* translates the estimation of the running interest (growing with $T \rightarrow I \rightarrow V$). For each movie, the choice of version is therefore based only on the events observed while browsing the previous movies.

Other sequences cause the decisions to be less intuitive or harder to interpret. For example, the sequence *pageLoad, linkI, startSlide, stopSlide, noInt, back* leaving T leads to the decision I. In this sequence, a compromise between interest rise (suggested by *linkI, startSlide*) and decrease (suggested by *stopSlide, noInt*) must be established. Thus, a decision T would not be illegitimate. The POMDP trades off this decision according to its dynamics and its rewards. To obtain a modified graph leading to a decision T for this sequence, it would be sufficient that the product $r_{EI}\xi$ decreases, where ξ represents the probability to observe EI in the version I, for a medium interest. In this case, *stopSlide*, instead of provoking a loopback on the node 5, would bring the agent to the node 1. Then the agent would decide T since the expectation of the gains associated to I would be smaller.

In general, the decision-making automaton depends on n and bw. When n, bw, and It vary, the automaton becomes too complex to be displayed. The results of the POMDP require a different presentation. Henceforth, working with 3 granularity levels on n, 2 on bw, 3 on It and the set of rewards R_1 leads to a p olicy graph of more than 100 nodes. We apply it during numerous sequences of simulated navigations. Table 4 gives the statistics on the decisions that have been taken. For every triplet (n, bw, It), the decisions—the agent not knowing It—are counted and translated into percentages.

We notice that the proposed content becomes statistically richer when the interest increases, proving again that the interest estimation from the previous observations is as expected. Let us take an example and consider the bottom-right part of Table 4 (corresponding to BW_L and N_3). The probability of the policy proposing version V increases with the interest: from 0% (small interest) to 2% (average interest) then 10% (big interest).

Moreover, when n and/or bw increase, the interest trend is correct. For example, for a given set of It and n (It = *average* and $n = N_2$), the proposed version becomes richer with the bandwidth's increase from (1%T, 99%I, 0%V) to (0%T, 51%I, 49%V).

The POMDP capacity to refine adaptation policies according to the user interest is thus validated. Once the POMDP model is solved (*offline* resolution), the obtained automaton is easily put into practice *online* by encoding it into an adaptation agent.

Table 4: Actions' distribution for the POMDP solution policy.

	Interest	$N_1\{1,2,3,4\}$			$N_2\{5,6,7,8\}$			$N_3\{9,10,11,12\}$		
		T	I	V	T	I	V	T	I	V
BW_S	Small		100%		5%	95%		96%	4%	
	Average		100%		1%	99%		86%	14%	
	Big		100%			100%		76%	24%	
BW_L	Small		4%	96%		73%	27%	67%	33%	
	Average		2%	98%		51%	49%	30%	68%	2%
	Big			100%		33%	67%	9%	81%	10%

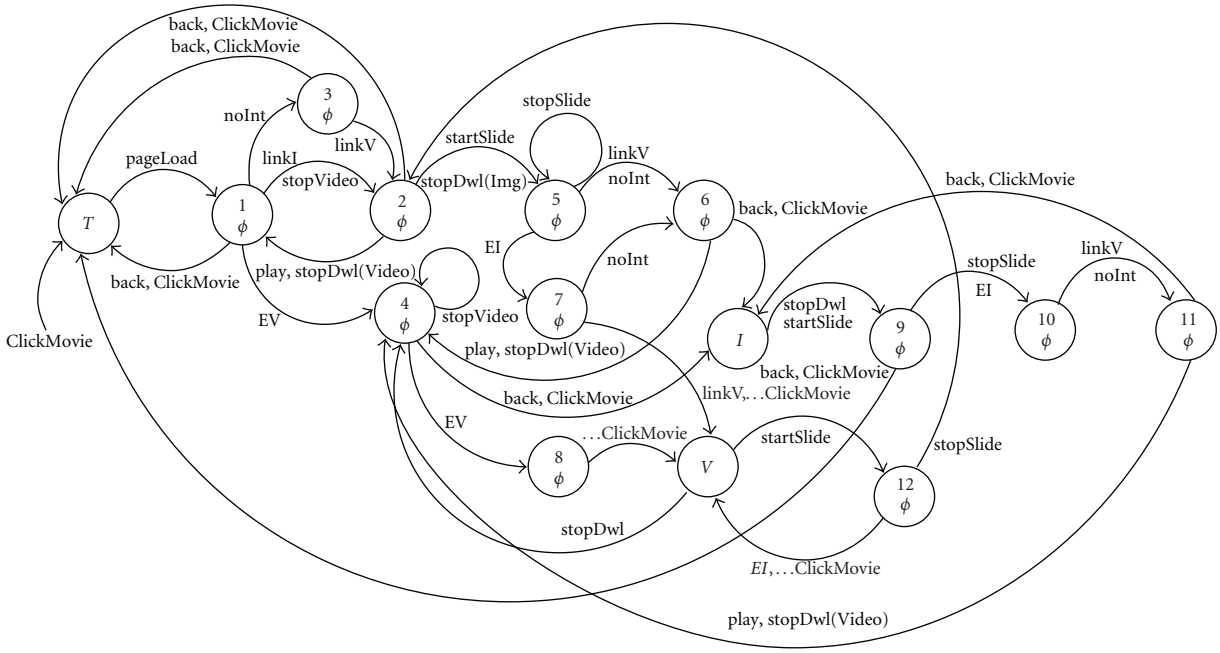


Figure 9: Decision-making automaton (policy graph), POMDP solution. Please note the different *stopDwl*, *stopDwl(Img)*, and *stopDwl(Video)*.

8. CONCLUSION

This paper has shown that sequential decision processes under uncertainty are well suited for defining adaptation mechanisms for dynamic contexts. According to the type of the context state (observable or partially observable), we have shown how to characterize adaptation policies by solving Markov decision processes (MDPs) or partially observable MDP (POMDP). These ideas have been applied to adapt a movie browsing service. In particular, we have proposed a method for refining a given adaptation policy according to user interest. The perspectives of this work are manifold. Our approach can be applied to cases where rewards are explicitly related to the service (e.g., to maximize the number of rented DVDs). It will also be interesting to extend our model by coupling it with functionalities from recommendation systems and/or from multimedia search systems. In the latter case, we would benefit a lot from a collection of real data, that is, navigation logs. These are the research directions that will guide our future work.

REFERENCES

[1] D. Kelly and J. Teevan, "Implicit feedback for inferring user preference: a bibliography," *ACM SIGIR Forum*, vol. 37, no. 2, pp. 18–28, 2003.

[2] T. Joachims, L. Granka, B. Pan, H. Hembrooke, and G. Gay, "Accurately interpreting clickthrough data as implicit feedback," in *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '05)*, pp. 154–161, Salvador, Brazil, August 2005.

[3] T. Lemlouma and N. Layaïda, "Media resources adaptation for limited devices," in *Proceedings of the 7th International Conference on Electronic Publishing (ICCC/IFIP '03)*, pp. 209–218, Minho, Portugal, June 2003.

[4] M. Margaritidis and G. C. Polyzos, "Adaptation techniques for ubiquitous internet multimedia," *Wireless Communications and Mobile Computing*, vol. 1, no. 2, pp. 141–163, 2001.

[5] T. C. Thang, Y. J. Jung, and Y. M. Ro, "Dynamic programming based adaptation of multimedia contents in UMA," in *Proceedings of the 5th Pacific Rim Conference on Advances in Multimedia Information Processing (PCM '04)*, vol. 3332 of

Lecture Notes in Computer Science, pp. 347–355, Springer, Tokyo, Japan, November-December 2004.

[6] A. Divakaran, K. A. Peker, R. Radhakrishnan, Z. Xiong, and R. Cabasson, *Video Summarization Using MPEG-7 Motion Activity and Audio Descriptors in Video Mining*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2003.

[7] B. Girod, M. Kalman, Y. J. Liang, and R. Zhang, "Advances in channel-adaptive video streaming," in *Proceedings of IEEE International Conference on Image Processing (ICIP '02)*, vol. 1, pp. 9–12, Rochester, NY, USA, September 2002.

[8] G. Ghinea and G. Magoulas, "Quality of service for perceptual considerations: an integrated perspective," in *Proceedings of IEEE International Conference on Multimedia and Expo (ICME '01)*, pp. 571–574, Tokyo, Japan, August 2001.

[9] S. R. Gulliver, T. Serif, and G. Ghinea, "Pervasive and standalone computing: the perceptual effects of variable multimedia quality," *International Journal of Human Computer Studies*, vol. 60, no. 5-6, pp. 640–665, 2004.

[10] S. R. Gulliver and G. Ghinea, "Defining user perception of distributed multimedia quality," *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 2, no. 4, pp. 241–257, 2006.

[11] P. Brusilovsky and E. Millán, "User models for adaptive hypermedia and adaptive educational systems," in *The Adaptive Web: Methods and Strategies of Web Personalization*, vol. 4321 of *Lecture Notes in Computer Science*, pp. 3–53, Springer, Berlin, Germany, 2007.

[12] C. Romero, S. Ventura, and P. De Bra, "Knowledge discovery with genetic programming for providing feedback to courseware authors," *User Modelling and User-Adapted Interaction*, vol. 14, no. 5, pp. 425–464, 2004.

[13] T. Syeda-Mahmood and D. Ponceleon, "Learning video browsing behavior and its application in the generation of video previews," in *Proceedings of the ACM International Multimedia Conference and Exhibition (Multimedia '01)*, vol. 9, pp. 119–128, Ottawa, Canada, September-October 2001.

[14] C. Pleşca, V. Charvillat, and R. Grigoras, "User-aware adaptation by subjective metadata and inferred implicit descriptors," in *Multimedia Semantics—The Role of Metadata*, vol. 101 of *Studies in Computational Intelligence*, pp. 127–147, Springer, Berlin, Germany, 2008.

[15] R. Grigoras, V. Charvillat, and M. Douze, "Optimizing hypervideo navigation using a Markov decision process approach," in *Proceedings of the 10th ACM International Conference on Multimedia*, pp. 39–48, Juan-les-Pins, France, December 2002.

[16] G. Yavaş, D. Katsaros, Ö. Ulusoy, and Y. Manolopoulos, "A data mining approach for location prediction in mobile environments," *Data and Knowledge Engineering*, vol. 54, no. 2, pp. 121–146, 2005.

[17] H. Kosch, L. Böszörményi, M. Döller, M. Libsie, P. Schojer, and A. Kofler, "The life cycle of multimedia metadata," *IEEE Multimedia*, vol. 12, no. 1, pp. 80–86, 2005.

[18] C. Timmerer and H. Hellwagner, "Interoperable adaptive multimedia communication," *IEEE Multimedia*, vol. 12, no. 1, pp. 74–79, 2005.

[19] O. Layaïda, S. B. Atallah, and D. Hagimont, "A framework for dynamically configurable and reconfigurable network-based multimedia adaptations," *Journal of Internet Technology*, vol. 5, no. 4, pp. 363–372, 2004.

[20] P. M. Ruiz, J. A. Botía, and A. Gómez-Skarmeta, "Providing QoS through machine-learning-driven adaptive multimedia applications," *IEEE Transactions on Systems, Man, and Cybernetics B*, vol. 34, no. 3, pp. 1398–1411, 2004.

[21] V. Charvillat and R. Grigoras, "Reinforcement learning for dynamic multimedia adaptation," *Journal of Network and Computer Applications*, vol. 30, no. 3, pp. 1034–1058, 2007.

[22] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, MIT Press, Cambridge, Mass, USA, 1998.

[23] M. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, Wiley-Interscience, New York, NY, USA, 1994.

[24] S. P. Singh, T. Jaakkola, and M. I. Jordan, "Learning without state-estimation in partially observable markovian decision processes," in *Proceedings of the 11th International Conference on Machine Learning (ICML '94)*, pp. 284–292, New Brunswick, NJ, USA, July 1994.

[25] A. R. Cassandra, L. P. Kaelbling, and M. L. Littman, "Acting optimally in partially observable stochastic domains," in *Proceedings of the 12th National Conference on Artificial Intelligence (AAAI '94)*, vol. 2, pp. 1023–1028, Seattle, Wash, USA, July-August 1994.

[26] T. Syeda-Mahmood, "Learning and tracking browsing behavior of users using hidden markov models," in *Proceedings of IBM Make It Easy Conference*, San Jose, Calif, USA, June 2001.

[27] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, Wiley-Interscience, New York, NY, USA, 2nd edition, 2000.

Hindawi Publishing Corporation
International Journal of Digital Multimedia Broadcasting
Volume 2008, Article ID 851628, 10 pages
doi:10.1155/2008/851628

Research Article

Efficient Execution of Service Composition for Content Adaptation in Pervasive Computing

Yaser Fawaz,[1] Girma Berhe,[2] Lionel Brunie,[1] Vasile-Marian Scuturici,[1] and David Coquil[3]

[1] *Laboratoire LIRIS-UMR-CNRS 5205, INSA de Lyon, 7 Avenue Jean Capelle, 69621 Villeurbanne Cedex, France*
[2] *Department of Computer Science and Communication (CSC), University of Luxembourg, Campus Kirchberg, 6 Rue Richard Coudenhove-Kalergi, 1359 Luxembourg, Luxembourg*
[3] *University of Passau, Innstraße 43, 94032 Passau, Germany*

Correspondence should be addressed to Yaser Fawaz, yaser.fawaz@insa-lyon.fr

Received 5 April 2008; Accepted 25 June 2008

Recommended by Harald Kosch

Multimedia content adaptation has been proved to be an effective mechanism to mitigate the problem of devices and networks heterogeneity and constraints in pervasive computing environments. Moreover, it enables to deliver data taking into consideration the user's preferences and the context of his/her environment. In this paper, we present an algorithm for service composition and protocols for executing service composition plan. Both the algorithm and the protocols are implemented in our distributed content adaptation framework (DCAF) which provides a service-based content adaptation architecture. Finally, a performance evaluation of the algorithm and the protocols is presented.

1. INTRODUCTION

There is a huge amount of multimedia information being captured and produced for different multimedia applications, and the speed of generation is constantly increasing. While providing the right information to a user is already difficult for structured information, it is much harder in the case of large volume of multimedia information. This is further complicated with the emergence of new computing environments. For example, a user may want to access content on the network through a handheld device connected by wireless link or from a high-end desktop machine connected by broadband network. The content that can be presented on one device might not be necessarily viewable on another device unless some content transformation operations are applied [1].

An efficient content delivery system must be able to adapt the delivered content for every client in every situation in order to address the wide range of clients, minimal bandwidth requirement, and fast real-time delivery [2]. As a consequence, content adaptation is one of the research topics that have attracted a number of multimedia research works. Here, we focus on issues related to content adaptation in pervasive computing systems.

Several adaptation approaches have been developed to perform content adaptation for pervasive computing. These approaches are generally classified into: server-based [3, 4], client-based [5, 6], and proxy-based [7–9]. As server-based adaptation approach degrades the performance of the server, and client-based adaptation approach is very difficult and sometimes impossible due to the limited processing power of pervasive devices (e.g., smartphones, PDAs), most of existing adaptation systems implement a proxy-based approach. Furthermore, to alleviate the overload problem of content adaptation processing, distributed approaches were proposed such as Ninja [10], MARCH [11], DANAE [12], and DCAF [13].

Jannach and Leopold [14] proposed a server-side multimedia content adaptation framework which performs the content adaptation by composing adaptation services resident on the server itself. This approach results in computational load and resource consumption on the server so it decreases the performance of content delivery. However, in our architecture (DCAF), content adaptation is performed externally using Internet accessible adaptation services which enhance the performance of the system in terms of content delivery time, resources consumption, and network overhead.

While the DCAF architecture [13] provides a distributed adaptation mechanism, its centralized control manner of adaptation services execution impacts the overall performance of the system in delivering the adapted data to the user. In order to enhance the performance of the system, a decentralized service execution protocol is incorporated to the DCAF architecture. In this paper, we present the centralized and decentralized service execution protocols, and the result of the experiments that have been done to compare their performances.

The rest of the paper is organized as follows. In Section 2, we discuss related works. Section 3 presents an overview of the distributed content adaptation framework (DCAF). The multimedia adaptation graph generator (MAGG) algorithm is presented in Section 4. In Section 5, we present the composite service execution protocols. The results of the experiments which have been done to evaluate the performance of the MAGG algorithm and the composite service execution protocols are presented in Section 6. In Section 7, we discuss fault tolerance issues in DCAF architecture. Finally, in Section 8, we conclude the paper and highlight some future works.

2. RELATED WORKS

As we have mentioned in Section 1, the existing adaptation frameworks are categorized into three groups: server-based approach, proxy-based approach, and client-side approach. In the following, we present each of these approaches.

In the case of server-based approach (e.g., [3, 4, 15]), the functionality of the traditional server is extended by adding content adaptation. In this approach, both static (offline) and dynamic (on-the-fly) content adaptations can be applied and better adaptation results could be achieved as it is close to the content; however clients experience performance degradation due to additional computational load and resource consumption on the server [16].

In proxy-based approach (e.g., [8, 16, 17]), a proxy, that is between the client and the server, acts as a transcoder for clients with similar network or device constraints. The proxy makes a request to the server on behalf of the client, intercepts the reply from the server, decides on and performs the adaptation, and then sends the transformed content back to the client. In this approach, there is no need of changing the existing clients and servers. The problem of proxy-based adaptation approaches is that most of them focus on a particular type of adaptation such as image transcoding, HTML to WML conversion, and so forth, and that they are application specific. In addition, if all adaptations are done at the proxy, it results in computational overload as some adaptations are computational intensive and this degrades the performance of information delivery like the server-based approach.

Client-based approach (e.g., [5, 6, 18]) can be done in two ways: transformation by the client device or selection of the best representation after receiving the response from the origin server. This approach provides a distributed solution for managing heterogeneity since all clients can locally decide and employ adaptations most appropriate to them. However, adaptations that can benefit a group of clients with similar request can be more efficiently implemented with server-based or proxy-based approaches. Furthermore, all of the clients may not be able to implement content adaptation techniques due to processor, memory resource constraint, and limited network bandwidth [19].

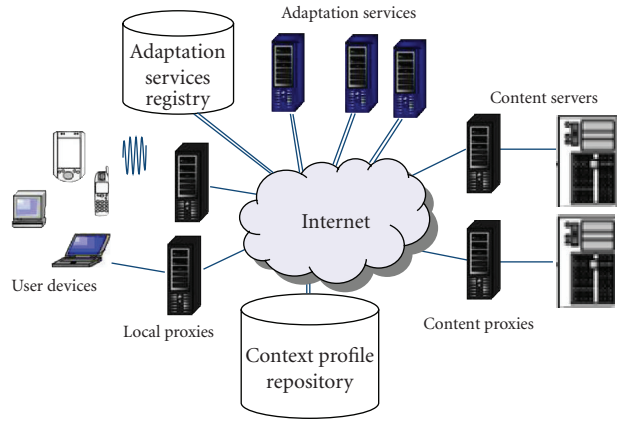


Figure 1: Distributed content adaptation framework (DCAF).

The above three approaches do not deal with the problem of content adaptation from a service perspective that can be commercialized and utilized by users, content providers, or other service providers (like Internet service providers). Introducing content adaptation as a service (service-based adaptation approach) distributes the activities and results in performance enhancement. It also opens new opportunities to service providers as additional revenue.

A service-based content adaptation approach is quite recent. There are very few works on distributed content adaptation mechanism. Nevertheless, we can cite: Ninja [10], MARCH [11], DANAE [12], and DCAF [13]. In our previous work [13], we proposed a service-based architecture (DCAF). In this architecture, the adaptation tools are developed externally by third party or service providers. While the architecture provides a distributed adaptation mechanism, the centralized control of the execution of these services impacts the overall performance of the system in delivering the adapted data to the user. In order to enhance the performance of the system, a decentralized service execution protocol is introduced. In this paper, we present this protocol and the result of the comparison done with the centralized service execution protocol.

3. OVERVIEW OF THE DCAF ARCHITECTURE

3.1. Components of the DCAF architecture

As displayed in Figure 1, the DCAF architecture is composed of six components. The description of these components is summarized as follows.

Content servers (CSs): they are standard data repositories such as web sites, databases, and media servers.

Content proxies (CPs): they provide access to content servers, formulate user request to source format, manage and provide content description (meta-data).

Context profile repository (CPR): it stores the users' preferences and the device characteristics. Users can update and modify their profiles at any time. Dynamic data such as the user location and the network conditions are determined at request execution.

Adaptation service registry (ASR): it is like Universal Description, Discovery and Integration (UDDI) registry; it stores multimedia adaptation services description including the semantic information (e.g., the type of data the service handles) and allows for service look up.

Adaptation services proxies (ASPs): they host adaptation tools. In our framework, ASPs are implemented as Web Services.

Local proxies (LPs): they are access points to information delivery systems. They are in charge of retrieving and processing context profile (user, device, and network), decide the type and number of adaptation processes, discover appropriate adaptation services, and plan execution of the services.

3.2. Context, content, and service descriptions

The decision of the adaptation process depends on the quality of information gathered from various sources. This information consists of context description, content description, and adaptation service descriptions.

(1) Context description (*context metadata*)

It includes the device profile, the user profile, the network profile, and other dynamic context data like location, sensor information, and so forth. We have used the CSCP [20] standard to represent the context information. Figure 2 shows an example of partial context profile for a sample device.

(2) Content description (*content metadata*)

The metadata contains both the information about the object like object's title, description, language, and so forth, and feature data about the media itself including the media type, the file format, the file size, the media dimensions, the color information, the location, and so forth. For content description, the XML form of MPEG-7 description is used. Figure 3 shows an example of the description for an image data.

(3) Adaptation service description

It contains information about the service such as name, identifier, function, processing rate, processing rate is the amount of data processed per second. It is expressed in Kbytes per second, cost, and so forth. In order to describe adaptation services, we developed multimedia adaptation service ontology [21]. This ontology facilitates describing adaptation services semantically so that they can be discovered, selected, composed, and invoked automatically. Figure 4 presents an example of a description of audio to text adaptation service using the described ontology.

```
<?xml version="1.0" encoding="UTF-8"?>
<ContextProfile rdf:ID=”Context”>
  <device>
  <dev:DeviceProfile>
    <dev:Hardware>
      <dev:ScreenWidth>320</dev:ScreenWidth>
      <dev:ScreenHeight>200</dev:ScreenHeight>
    </dev:Hardware>
    <dev:Software>
      <dev:OperatingSystem>
        <dev:Name>EPOC </dev:Name>
        <dev:Version>2.0</dev:Version>
        <dev:Vendor>Symbian</dev:Vendor>
      </dev:OperatingSystem>
      <dev:UserAgent>
        <dev:Type>Browser</dev:Type>
        <dev:Name>Mozilla</dev:Name>
        <dev:Version>5.0</dev:Version>
        <dev:Vendor>Symbian</dev:Vendor>
      </dev:UserAgent>
    </dev:Software>
  </dev:DeviceProfile>
  </device>
</ContextProfile>
```

Figure 2: Partial example of context profile.

```
<?xml version="1.0" encoding="UTF-8"?>
<ContentDescription>
  <MultimediaContent type=“Image”>
    <MediaInformation ID=“x_ray_image”>
    <MediaUri> http:/localhost/imagefiles/gorge .jpeg
    </MediaUri>
    <MediaTitle>Examen de la gorge</MediaTitle>
    <MediaDescription> Patient atteint d'un cancer de la
gorge…  </MediaDescription>
    <MediaLanguage>French</MediaLanguage>
    </MediaInformation>
    <MediaProfile>
      <MediaFormat>JPEG</MediaFormat>
      <MediaSize>61.6 </MediaSize>
      <MediaWidth>660</MediaWidth>
      <MediaHeight>445</MediaHeight>
      <MediaColor>24</MediaColor>
    </MediaProfile>
  </MultimediaContent>
</ContentDescription>
```

Figure 3: Sample multimedia content description using MPEG-7 standard.

4. MULTIMEDIA ADAPTATION GRAPH GENERATOR (MAGG)

In a multimedia content adaptation framework like DCAF, the challenge is that there is no single complete software solution that can satisfy all adaptation needs. In order to solve this problem, adaptation services are composed to realize the required adaptation. Service composition is defined as the process of putting together atomic/basic services to perform

complex tasks. For example, to transform English text into audio in French, we need the composition of a language translation service and a text to audio conversion service. Since an adaptation process can be carried out in a number of adaptation steps (adaptation tasks) and there could be several adaptation services that execute each adaptation task that leads to different service composition possibilities and makes services composition difficult. In order to solve this problem, we have developed an algorithm called a multimedia adaptation graph generator (MAGG) that can compose distributed multimedia adaptation services.

4.1. Service composition modelling: definitions and notations

Definition 1 (media object). A media object is a multimedia data item which can be a text, an image, an audio, or a video represented as $M(m_1, m_2, \ldots, m_n)$ where $m_1, m_2, \ldots, m_n$ are media features or metadata.

Definition 2 (state). The state S of a media object M, denoted as $S(M)$, is described by the metadata values. For example, for an image object the state holds the values for the format, the color, the height, the width, and so forth.

For example, $S(M)$ = (bmp, 24 bits, 245 pixels, 300 pixels).

Definition 3 (adaptation task). An adaptation task is an expression of the form $t(a_1 a_2 \cdots a_n)$ where t is a transformation and $a_1, a_2, \ldots, a_n$ are parameters.

For example, ImageFormatConversion (imageIn, imageOut, oldFormat, newFormat), where

(i) imageIn: image input file,

(ii) imageOut: image output file,

(iii) oldFormat: old file format,

(iv) newFormat: new file format.

Definition 4 (adaptation service). An adaptation service is a service described in terms of inputs, outputs, preconditions, and effects. An adaptation service is represented as $s = (R\ I\ O\ Pre\ Eff\ Q)$, *where*

(i) *R:* an atomic process that realizes an adaptation task,

(ii) *I:* input parameters of the process,

(iii) *O:* output parameters of the process,

(iv) *Pre:* preconditions of the process,

(v) *Eff:* effects of the process,

(vi) *Q:* quality criteria of the service.

Definition 5 (operator (plan operator)). A plan operator is an expression of the form $o = (h(a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_m), Pre\ Eff\ Q)$, where

(i) h: an adaptation task realized by an adaptation service with input parameters $a_1, a_2, \ldots, a_n$ and output parameters $b_1, b_2, \ldots, b_m$. h is called the head of identification of the operator,

```
<AtomicProcess rdf:ID="AudioToTextAdapationService">
  <hasInput rdf:ID="#InputAudio">
    <parameterType datatype="&xsd;#anyURI">
  </hasInput>
  <hasOutput rdf:ID="#OutputText">
    <parameterType datatype="&xsd;#anyURI">
  </hasOutput>
  <hasPrecondition>
    <hasExpression rdf:ID="InputAudioFormat">
      <hasSubject rdf:resource="#InputAudio"/>
      <hasProperty rdf:resource="#Audio"/>
      <hasFormat rdf:datatype="&xsd; #string">#WAV
      </hasFormat>
    </hasExpression>
  </hasPrecondition>
  <hasServicePrice>
    <parameterType datatype="&xsd;#float">25
  </hasServicePrice>
  <hasServiceTime>
    <parameterType datatype="&xsd;#float">200
  </hasServiceTime>
</AtomicProcess>
```

Figure 4: An expert of service description for Audio to Text adaptation service.

(ii) *Pre*: represents the operator's preconditions,

(iii) *Eff*: represents the effect of executing the operator,

(v) $Q = \{q_1, q_2, \ldots, q_n\}$: represents quality attributes (e.g., cost, processing rate, etc.).

Let S be a state, t be an adaptation task, and M be a media object. Suppose that there is an operator o with head h that realizes t such that *Pre* of o is satisfied in S. Then, we say that o is applicable to t, and the new state is given by

$$S(M_o) = \text{Executing } (o, M, S). \tag{1}$$

Example: for the above given adaptation task, we can have an adaptation operator instance as follows.

Operator:ImageFormatConversionOperator (http://media-adaptation/imagefiles/image1, http://media-adaptation/imagefiles/image2, mpeg, bmp).

(i) Input: http://media-adaptation/imagefiles/image1.

(ii) Output:http://media-adaptation/imagefiles/image2.

(iii) Precondition: hasFormat (http://media-adaptation/imagefiles/image1, mpeg).

(iv) Effect: hasFormat (http://media-adaptation/image files/image2, bmp).

(v) Quality: (cost = 30 units, processing rate = 1500 kbyte/s).

Definition 6 (adaptation graph). An adaptation graph $G(V, E)$ is a directed acyclic graph (DAG), where

(i) V is the set of nodes that represent the adaptation operators,

(ii) E is the set of edges that represent the possible connections between the adaptation operators.

The start node $A \in V$ is a pseudo operator with effect (initial state) but no precondition.

The end node $Z \in V$ is a pseudo operator with precondition (goal state) but no effect.

Remark 1. Let $o_i \in V$ and $o_j \in V$, a link or an edge e_{ij} exists from o_i to o_j if the following condition is satisfied:

$$o_j \cdot Pre \subseteq o_i \cdot Eff, \tag{2}$$

where

(i) $o_j \cdot Pre$ denotes preconditions of o_j,

(ii) $o_i \cdot Eff$ denotes effects of o_i.

Definition 7 (adaptation planning problem). An adaptation planning problem is a four-tuple (S_A, S_Z, T, D), where S_A is the initial state of the media object, S_Z is the goal state of the media object, T is an adaptation task list, and D is the adaptation operators. The result is a graph $G = (V, E)$.

Definition 8 (adaptation path). An adaptation path is a path in the adaptation graph G that connects the start node to the end node. It is represented as a list of the form $p = (A, o_1, o_2, \ldots, o_n, Z)$, where A and Z are the start and the end nodes and o_i is an adaptation operator instance.

The MAGG algorithm consists of different procedures and functions. In Algorithm 1, we present only the structure of the algorithm. See [21] for complete listing of the algorithm. This algorithm is used to construct a multimedia adaptation graph which gives all service composition possibilities that satisfy the required adaptation needs.

4.2. Optimal adaptation path search

Since an adaptation task can be achieved by more than one service, the services are represented by operators in the graph, and each service has different QoS, choosing an appropriate service is an obvious requirement. Once the adaptation graph that consists of all possible compositions is generated, then the choice of the optimal adaptation path (also called the service composition plan) in the graph is done based on user specified QoS criteria [13]. The QoS is a multidimensional property which may include service response, service charge, quality of received data, and so forth. Here, the QoS represents only the service charge (cost) and waiting time. For a service s executing an adaptation task t, the QoS is defined as follows:

$$Q(s) = (s_{\text{cost}}(s,t), s_{\text{time}}(s,t)), \tag{3}$$

where

(i) $s_{\text{cost}}(s,t)$: the adaptation service execution cost,

(ii) $s_{\text{time}}(s,t)$: the adaptation service execution time and the data transmission time.

Let $p = (A, s_1, s_2, \ldots, s_n, Z)$ be a path in an adaptation graph, where n is the number of services in the adaptation path. We define the QoS of the path p, denoted as $Q(p)$, as follows:

$$Q(p) = (Q_{\text{cost}}(p), Q_{\text{time}}(p)), \tag{4}$$

where

$$\begin{aligned} Q_{\text{cost}}(p) &= \sum_{i=1}^{n} s_{\text{cost}}(s_i, t_i), \\ Q_{\text{time}}(p) &= \sum_{i=1}^{n} s_{\text{time}}(s_i, t_i). \end{aligned} \tag{5}$$

To aggregate the quality values, we define scaled qualities $Qs_{\text{cost}}(s_i)$ and $Qs_{\text{time}}(s_i)$ as

$$\begin{aligned} Qs_{\text{cost}}(s_i) &= \begin{cases} \dfrac{Q_{\text{cost}}^{\max} - Q_{\text{cost}}(s_i)}{Q_{\text{cost}}^{\max} - Q_{\text{cost}}^{\min}}, & \text{if } Q_{\text{cost}}^{\max} - Q_{\text{cost}}^{\min} \neq 0, \\ 1, & \text{if } Q_{\text{cost}}^{\max} - Q_{\text{cost}}^{\max} = 0, \end{cases} \\ Qs_{\text{time}}(s_i) &= \begin{cases} \dfrac{Q_{\text{time}}^{\max} - Q_{\text{time}}(s_i)}{Q_{\text{time}}^{\max} - Q_{\text{time}}^{\min}}, & \text{if } Q_{\text{time}}^{\max} - Q_{\text{time}}^{\min} \neq 0, \\ 1, & \text{if } Q_{\text{time}}^{\max} - Q_{\text{time}}^{\max} = 0, \end{cases} \end{aligned} \tag{6}$$

where $Q_{\text{cost}}^{\max}$ and $Q_{\text{cost}}^{\min}$ are the values of the maximum and the minimum costs, respectively. $Q_{\text{time}}^{\max}$ and $Q_{\text{time}}^{\min}$ are the values of the maximum and the minimum times, respectively.

Users can give their preferences on QoS by specifying weight values for each criterion. The score of a path with weighted values is calculated as in

$$\text{Score } (p) = \sum_{i=1}^{n} (Qs_{\text{cost}}(s_i) * w_{\text{cost}} + Qs_{\text{time}}(s_i) * w_{\text{time}}), \tag{7}$$

where $w_{\text{cost}} \in [0,1]$ and $w_{\text{time}} \in [0,1]$ represent the weight values assigned to the cost and the time, respectively, and $w_{\text{cost}} + w_{\text{time}} = 1$.

Let $P_{\text{set}} = \{p_1, p_2, \ldots, p_K\}$ be the set of all possible paths in an adaptation graph, then the optimal path is the path with the maximum score value $\text{score}_{\max}$, where $\text{score}_{\max}$, is defined as follows: $\text{Score} = \max\{\text{Score}(p_i)\}; i \in \{1, k\}$.

Dijkstra's algorithm [22] was used to find the optimal path, that is, the path in the graph with the maximum score value $\text{score}_{\max}$. More information about the optimal path selection using Dijkstra's algorithm is found in [21].

5. COMPOSITE SERVICE EXECUTION PROTOCOLS

The execution of the composite service (also called the service composition plan) can be done in centralized or decentralized approaches. In the centralized approach (also called a star-based approach), the exchange of data between the services is done through the use of a broker as an intermediator [23, 24]. In the decentralized approach (also called a mesh-based approach), however, the exchange of data is done directly from one service to another one without the need to an intermediator [25]. In the following, we incorporate the two approaches with our architecture DCAF.

5.1. Centralized (star-based)

In the DCAF architecture, the local proxy acts as a broker. As presented in Figure 5, the local proxy sends the data to each

```
Algorithm: graph (S_A, S_Z, T, D)
Input: Initial state S_A, final state S_Z, adaptation task list T and adaptation operators D
Output: an adaptation graph G(V, E)
// Global constant
// Limit  maximum number of neutral operators allowed in a connection
// Global variables
// V          a set of nodes in a graph
// E          a set of edges in a graph
// ao         start node
// zo         end node
// NO         a set of neutral operators available in the system
// Local variables
// T          a set of adaptation tasks
// t          an adaptation task element of T
// O          a set of nodes for adaptation operators realizing an adaptation task
// PO         a set of parent nodes
// P_z        a set containing the end node
Var
V, E, T, t, O, PO, ao, zo, Pz, NO
Begin
   ao = ConstructStartNode(S_A) // constructs the start node from the initial state
   zo = ConstructEndNode(S_Z) // constructs the end node from the goal state
   NO = ConstructNeutralOperators() // returns the list of the neutral operators available in // the system
   V = {ao}      // initialization
   E = Ø         // initialization
   PO = {ao}     // initialization
   for each t ∈ T
      Begin
        Construct nodes O from D with operators realizing t
        // several operators can realize a task
        Connect (O, PO)
        // after the process PO holds the value of O
      End // T is processed
   Pz = {zo}
   Connect (Pz, PO) // connects the end node
    Return G(V, E)
End // graph
```

Algorithm 1: The MAGG algorithm structure.

adaptation service proxy (ASP) in the service composition plan and gets the result from the ASPs.

The data that the local proxy gets from the ASPs are partially adapted before it is sent to the last ASP in the service composition plan. The number of communications between the local proxy and the ASPs increases due to the number of the ASPs involved in performing the content adaptation process. Hence, the forward and backward communications between the local proxy and the ASPs incur additional overhead on the overall performance of data delivering to the user.

5.2. Decentralized (mesh-based)

As shown in Figure 6, in the decentralized service execution protocol, the ASPs communicate with each other and with the local proxy by exchanging a record message (RM) [25]. The RM contains

(1) the addresses of the services that are in the optimal path of the graph,

(2) the address of the local proxy,

(3) the data to be adapted.

The exchange of records messages between the ASPs is done by using service execution and data forwarding (SEDF) modules. The SEDF module is in charge of executing local services to perform content adaptation and forwarding the record message (denoted RM_i) containing the partially adapted data for subsequent adaptation. When the ASP receives RM_i, its SEDF module does the following.

If the first service S in RM_i is found locally, the SEDF executes S by assigning the data in RM_i as input of S. Then, it removes S from RM_i and puts the output of S in the data field of RM_i.

It repeats step 1 until the first service of RM_i is not found locally or the last service in RM_i is executed.

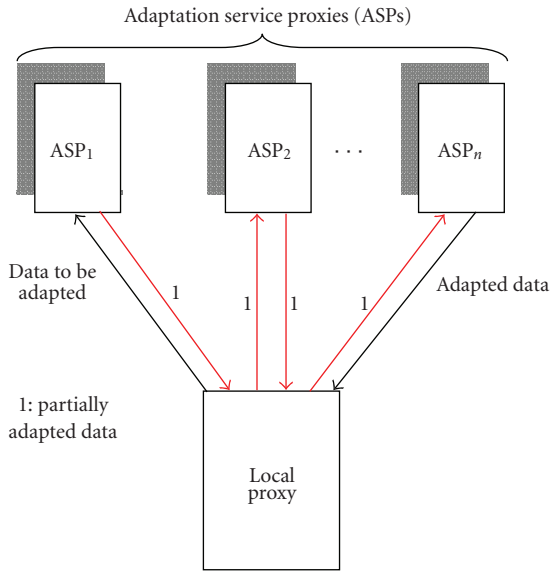


Figure 5: Centralized services execution pattern (star-based).

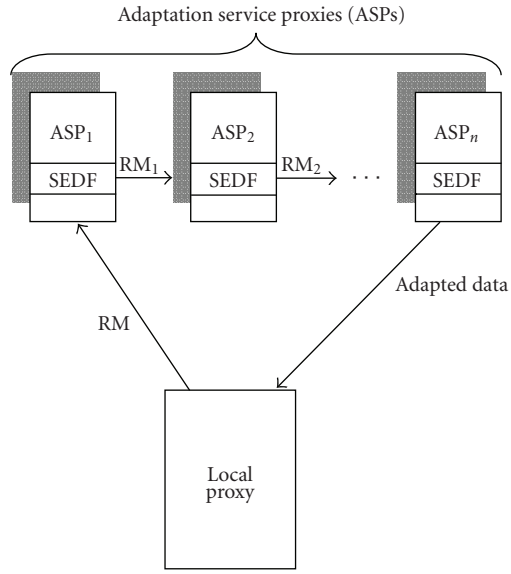


Figure 6: Decentralized services execution pattern (mesh-based).

The modified record message RM_i is forwarded to the ASP that owns the first service in RM_i. If RM_i does not contain any service which means that the adaptation process is finished, the adapted data are sent to the local proxy.

We have compared the centralized and decentralized service execution protocols using two metrics: data transmission time and size of exchanged data. The data transmission time as well as the size of exchanged data is less for the decentralized protocol since there are less number of communications. For a scenario with three ASPs, we need six and four communications for the centralized and decentralized protocols, respectively. Therefore, the decentralized protocol gives better performance than the centralized one especially when the bandwidth is small and the number of ASPs is big.

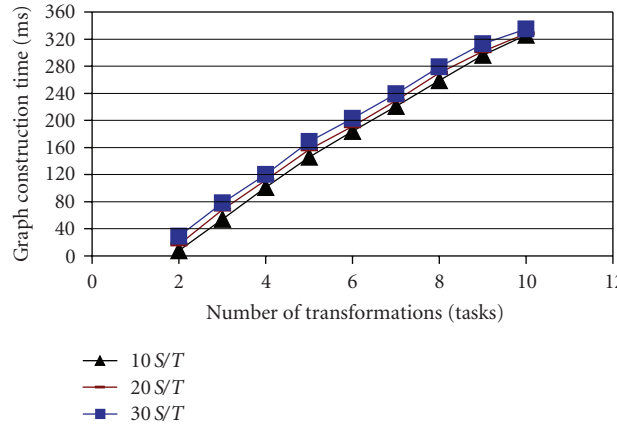


Figure 7: Graph construction time.

6. EXPERIMENTATIONS

Two major experiments were conducted. The first experiment is to study the behaviour of the graph generation algorithm with respect to the depth of the graph and the number of services per transformation. The second experiment is to measure the performance of the centralized and decentralized execution protocols. The experiments were performed on a 1.9 GHZ Pentium 4 with 256 MB RAM running Microsoft Windows 2000. In the following, we present the results of these experiments.

6.1. Graph construction

As depicted in Figure 7, the relationship between graph construction time and the number of the adaptation tasks (the adaptation transformations) is linear. Moreover, the construction time progresses slowly with the number of services per transformation (S/T). The graph construction time does not include the services execution time. It was also observed that the progress both for the depth (number of transformations) and the width (number of services per transformation) was almost constant with average increase of 40 milliseconds for each depth and 10 milliseconds for each 10 additional services. This implies that having several services per transformation does not affect much on the total construction time, while it provides the possibility to select the best service among the candidates.

The performance of the graph construction algorithm is reasonable even for the maximum depth of the graph (graph depth equals 10). For example, if we consider a graph with depth and width equal to 10 and 30, respectively, the construction time is only 335 milliseconds which is really acceptable. Nevertheless, most adaptation scenarios have 5 or 6 depth graph which is enough to realize any possible type of adaptation.

6.2. Service composition plan execution

A simulation has been made to compare the performance of the composite service execution protocols in terms of data delivery time and exchanged data size. The data delivery

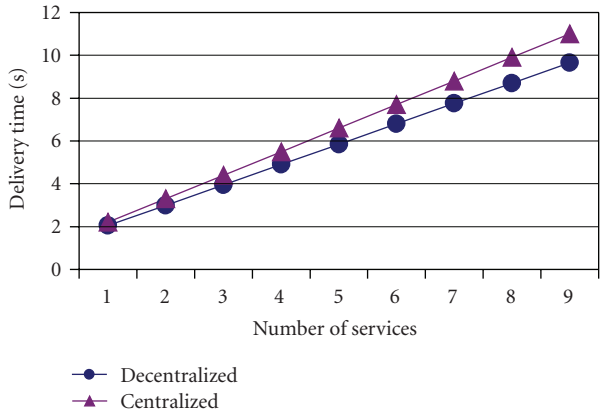


Figure 8: Data delivery time versus number of services. (Bandwidth = 54 Mbps, file size = 1 Mbyte.)

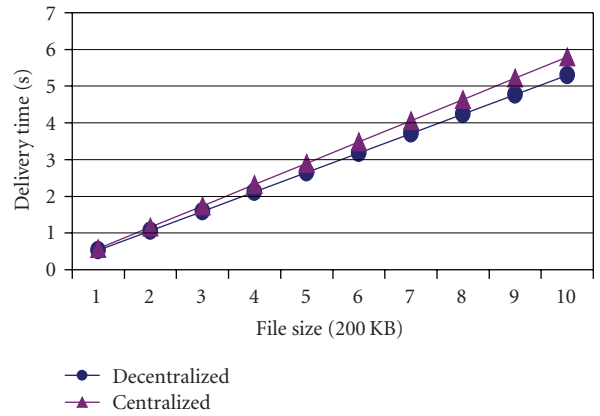


Figure 9: Data delivery time versus file size. (Bandwidth = 54 Mbps, three services are considered.)

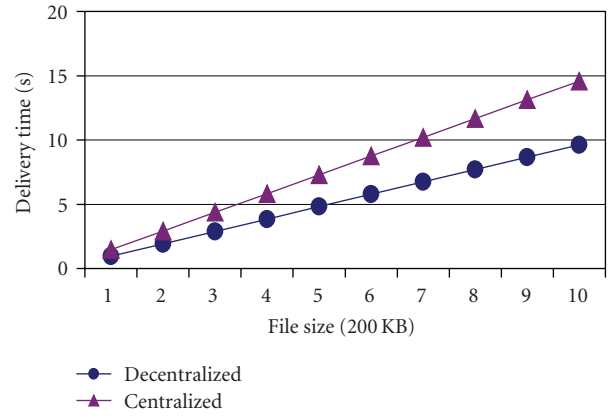


Figure 10: Data delivery time versus file size. (Bandwidth = 5.5 Mbps, three services are considered.)

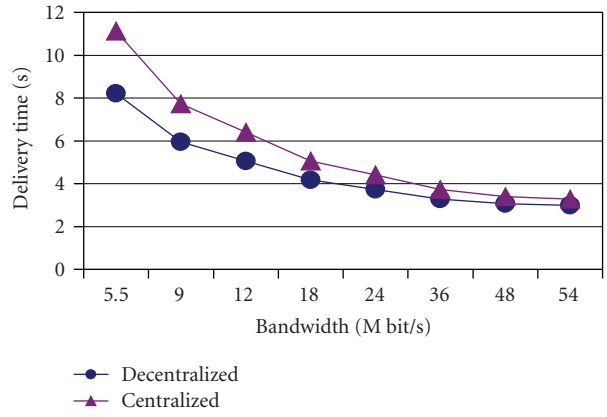


Figure 11: Delivery time versus bandwidth. (File size = 1 Mbyte, three services are considered.)

time is calculated as the sum of services execution time and the data transmission time. The data delivery time and the exchanged data size are measured with respect to the file size, the bandwidth which is considered as uniform, and the number of services involved in the content adaptation process. Figure 8 presents a relationship between the data delivery time and the number of services for the two protocols. The relationship is perceived to be linear for both protocols. As the number of services increases the performance gap between the protocols gets wide. This means, the decentralized protocols perform better with increasing the number of services in the service composition plan.

The data delivery time versus file size presented in Figure 9 and Figure 10 behaves in the same way as in Figure 8. The advantage of the decentralized service execution protocol becomes more visible when the data delivery time is analyzed with respect to the bandwidth. As illustrated in Figure 11, the performance gap between the two protocols is very significant when the bandwidth changes from 5.5 Mbits per second to 54 Mbits per second.

In addition, the analysis of the exchanged data size versus the number of services and the file size (see Figures 12 and 13) reflects similar behaviour to the data delivery time for the two protocols. Hence, the decentralized protocol performs better than the centralized one with respect to the data delivery time and the size of exchanged data.

To summarize, the decentralized execution protocol can enhance the performance of the DCAF architecture by decreasing the network load and the data delivery time and make it more scalable.

7. FAULT TOLERANCE IN DCAF ARCHITECTURE

Content adaptation process is accomplished successfully if there is no fault during services execution. However, the execution of the composite service can fail due to different causes such as network disruption and service discovery failure.

To tackle this problem, the local proxy can replace the failed service with an equivalent one. Identifying the failed service is straightforward for the centralized protocol as compared to the decentralized one since the local proxy controls the execution of each service in the centralized protocol. However, in the decentralized protocol, as discussed in

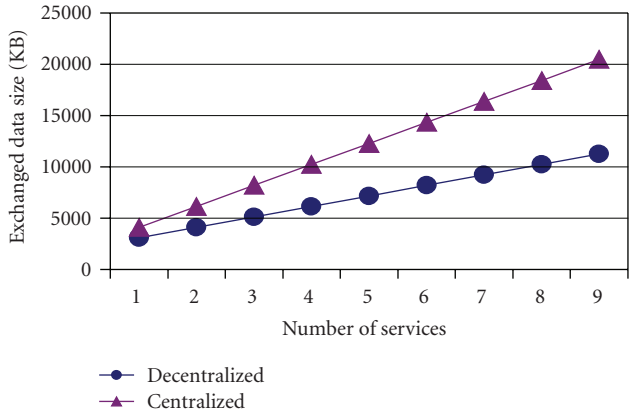


Figure 12: Exchanged data size versus number of services. (Bandwidth = 54 Mbps, file size = 1 Mbyte.)

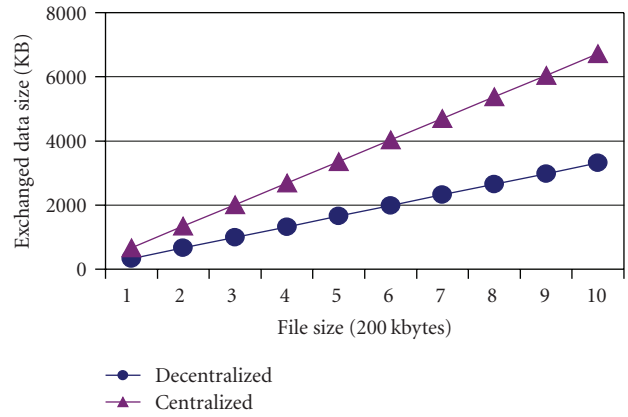


Figure 13: Exchanged data size versus file size. (Bandwidth = 54 Mbps, three services are considered.)

[26], the failed service can be identified through the use of acknowledgment messages (ACKs). When an ASP executes an adaptation service S and forwards the partially adapted data to the next ASP, it sends an ACK message to the local proxy to inform that the service S has been executed. If the local proxy does not receive an ACK message from such ASP, it concludes that the composite service execution is failed. The implementation of the fault detection and recovery mechanism within DCAF architecture is in progress.

8. CONCLUSION AND FUTURE WORK

In this paper, we have presented a multimedia adaptation graph generator (MAGG) algorithm and service composition plan execution protocols, that is, centralized and decentralized protocols. The algorithm constructs an adaptation graph which gives all service composition possibilities that satisfy the required adaptation needs. The selection of the optimal path (service composition plan) from the graph is done based on the QoS model which consists of the adaptation service execution cost, the adaptation service execution time, and the data transmission time.

The MAGG algorithm and the service execution protocols have been experimented in the DCAF architecture. The experiments on the graph construction algorithm (service composition algorithm) show that the graph construction time increases linearly as the number of adaptation tasks increases. The number of service per task, however, has not an important impact on the graph construction time. In addition, the experiments on the decentralized execution protocol show better performance than the centralized one, especially when the bandwidth is small, which is actually not surprising.

Successful content delivery does not only depend on effective execution of adaptation services but also on detection and recovery of failed adaptation services during services execution. For this purpose, we are planning to implement a fault detection and recovery mechanism within the DACH architecture in order to develop a fault tolerant content adaptation system.

ACKNOWLEDGMENT

The authors would like to thank the reviewers for their valuable comments that helped them improving the paper.

REFERENCES

[1] A. Held, S. Buchholz, and A. Schill, "Modeling of context information for pervasive computing applications," in *Proceedings of the 6th World Multiconference on Systemics, Cybernetics and Informatics (SCI '02)*, Orlando, Fla, USA, July 2002.

[2] C.-H. Chi, Y. Cao, and T. Luo, "Scalable multimedia content delivery on Internet," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '02)*, vol. 1, pp. 1–4, Lusanne, Switzerland, August 2002.

[3] M. Margaritidis and G. C. Polyzos, "Adaptation techniques for ubiquitous Internet multimedia," *Wireless Communications and Mobile Computing*, vol. 1, no. 2, pp. 141–163, 2001.

[4] R. Mohan, J. R. Smith, and C.-S. Li, "Adapting multimedia Internet content for universal access," *IEEE Transactions on Multimedia*, vol. 1, no. 1, pp. 104–114, 1999.

[5] K. Marriott, B. Meyer, and L. Tardif, "Fast and efficient client-side adaptivity for SVG," in *Proceedings of the 11th International Conference on World Wide Web (WWW '02)*, pp. 496–507, ACM Press, Honolulu, Hawaii, USA, May 2002.

[6] Z. Lei and N. D. Georganas, "Context-based media adaptation in pervasive computing," in *Proceedings of the IEEE Canadian Conference on Electrical and Computer Engineering (CCECE '01)*, vol. 2, pp. 913–918, Toronto, Canada, May 2001.

[7] A. Singh, A. Trivedi, K. Ramamritham, and P. Shenoy, "PTC: proxies that transcode and cache in heterogeneous web client environments," *World Wide Web*, vol. 7, no. 1, pp. 7–28, 2004.

[8] J.-G. Kim, Y. Wang, and S.-F. Chang, "Content-adaptive utility-based video adaptation," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '03)*, vol. 3, pp. 281–284, Baltimore, Md, USA, July 2003.

[9] S. Wee and J. Apostolopoulos, "Secure scalable streaming and secure transcoding with JPEG-2000," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '03)*, vol. 1, pp. 205–208, Barcelona, Spain, September 2003.

[10] S. D. Gribble, M. Welsh, R. von Behren, et al., "The Ninja architecture for robust Internet-scale systems and services," *Computer Networks*, vol. 35, no. 4, pp. 473–497, 2001.

[11] S. Ardon, P. Gunningberg, B. Landfeldt, Y. Ismailov, M. Portmann, and A. Seneviratne, "MARCH: a distributed content adaptation architecture," *International Journal of Communication Systems*, vol. 16, no. 1, pp. 97–115, 2003.

[12] A. Hutter, P. Amon, G. Panis, E. Delfosse, M. Ransburg, and H. Hellwagner, "Automatic adaptation of streaming multimedia content in a dynamic and distributed environment," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '05)*, vol. 3, pp. 716–719, Genova, Italy, September 2005.

[13] G. Berhe, L. Brunie, and J.-M. Pierson, "Content adaptation in distributed multimedia systems," *Journal of Digital Information Management*, vol. 3, no. 2, pp. 95–100, 2005.

[14] D. Jannach and K. Leopold, "Knowledge-based multimedia adaptation for ubiquitous multimedia consumption," *Journal of Network and Computer Applications*, vol. 30, no. 3, pp. 958–982, 2007.

[15] B. D. Noble, M. Price, and M. Satyanarayanan, "A programming interface for application-aware adaptation in mobile computing," in *Proceedings of the 2nd USENIX Symposium on Mobile and Location-Independent Computing (MLICS '95)*, pp. 57–66, Ann Arbor, Mich, USA, April 1995.

[16] R. Han, P. Bhagwat, R. LaMaire, T. Mummert, V. Perret, and J. Rubas, "Dynamic adaptation in an image transcoding proxy for mobile web browsing," *IEEE Personal Communications*, vol. 5, no. 6, pp. 8–17, 1998.

[17] W. Y. Lum and F. C. M. Lau, "A context-aware decision engine for content adaptation," *IEEE Pervasive Computing*, vol. 1, no. 3, pp. 41–49, 2002.

[18] C. Yoshikawa, B. Chun, P. Eastham, A. Vahdat, T. Anderson, and D. Culler, "Using smart clients to build scalable services," in *Proceedings of the USENIX Annual Technical Conference*, pp. 105–117, Anaheim, Calif, USA, January 1997.

[19] V. Cardellini, P. S. Yu, and Y.-W. Huang, "Collaborative proxy system for distributed Web content transcoding," in *Proceedings of the 9th International ACM Conference on Information and Knowledge Management (CIKM '00)*, pp. 520–527, McLean, Va, USA, November 2000.

[20] S. Buchholz, T. Hamann, and G. Hübsch, "Comprehensive structured context profiles (CSCP): design and experiences," in *Proceedings of the 2nd IEEE Annual Conference on Pervasive Computing and Communications Workshops (PerCom '04)*, pp. 43–47, Orlando, Fla, USA, March 2004.

[21] G. Berhe, *Access and adaptation of multimedia content for pervasive systems*, Ph.D. thesis, INSA de Lyon, Lyon, France, September 2006, http://docinsa.insa-lyon.fr/these/2006/girma/these.pdf.

[22] E. W. Dijkstra, "A note on two problems in connection with graphs," *Numerische Mathematik*, vol. 1, no. 1, pp. 269–271, 1959.

[23] D. Chakraborty, A. Joshi, T. Finin, and Y. Yesha, "Service composition for mobile environments," *Mobile Networks & Applications*, vol. 10, no. 4, pp. 435–451, 2005.

[24] G. Berhe, L. Brunie, and J.-M. Pierson, "Service-based architectural framework of multimedia content adaptation for pervasive computing environment," in *Proceedings of the Communication Networks and Distributed Systems Modeling and Simulation Conference (CNDS '04)*, San Diego, Calif, USA, January 2004.

[25] Y. Fawaz, A. Negash, L. Brunie, and V.-M. Scuturici, "Service composition-based content adaptation for pervasive computing environment," in *Proceedings of the International Conference Wireless Applications and Computing (IADIS '07)*, Lisbon, Portugal, July 2007.

[26] Y. Fawaz, C. Bognanni, V.-M. Scuturici, and L. Brunie, "Fault tolerant content adaptation for a dynamic pervasive computing environment," in *Proceedings of the 3rd International Conference on Information and Communication Technologies: from Theory to Applications (ICTTA '08)*, Damascus, Syria, April 2008.

Hindawi Publishing Corporation
International Journal of Digital Multimedia Broadcasting
Volume 2008, Article ID 863613, 21 pages
doi:10.1155/2008/863613

Research Article

Two-Level Automatic Adaptation of a Distributed User Profile for Personalized News Content Delivery

Maria Papadogiorgaki,[1] Vasileios Papastathis,[1] Evangelia Nidelkou,[1] Simon Waddington,[2] Ben Bratu,[3] Myriam Ribiere,[3] and Ioannis Kompatsiaris[1]

[1] *Centre for Research and Technology Hellas (CERTH), Informatics and Telematics Institute (ITI), 1st Km Thermi-Panorama Road, Thessaloniki 57001, Greece*
[2] *Motorola Labs, Motorola Ltd., Jays Close, Viables Industrial Estate, Basingstoke, Hampshire, RG22 4PD, UK*
[3] *Motorola Labs, Parc Les Algorithmes, Saint Aubin, 91193 Gif sur Yvette Cedex, France*

Correspondence should be addressed to Ioannis Kompatsiaris, ikom@iti.gr

Received 28 February 2008; Accepted 24 June 2008

Recommended by Harald Kosch

This paper presents a distributed client-server architecture for the personalized delivery of textual news content to mobile users. The user profile consists of two separate models, that is, the long-term interests are stored in a skeleton profile on the server and the short-term interests in a detailed profile in the handset. The user profile enables a high-level filtering of available news content on the server, followed by matching of detailed user preferences in the handset. The highest rated items are recommended to the user, by employing an efficient ranking process. The paper focuses on a two-level learning process, which is employed on the client side in order to automatically update both user profile models. It involves the use of machine learning algorithms applied to the implicit and explicit user feedback. The system's learning performance has been systematically evaluated based on data collected from regular system users.

1. INTRODUCTION

The increasing popularity of mobile devices, such as laptops, mobile phones and personal digital assistants, and the advances in wireless networking technologies allow information to be accessed almost anywhere, at any time. As part of this trend, several personalized news services are emerging, such as systems that enable the distribution and delivery of news content to the individual users, from heterogeneous networks of devices. These environments raise challenging problems for the development of personalization applications. These problems concern the requirements of matching the user preferences while preserving privacy issues, as well as being aware of the limitations, for example, in network traffic.

The focus of this paper is to cover the personalization requirements of mobile users in the news domain, taking into account the user's personal preferences and interests but also attempting to preserve the privacy of the user preferences. To this aim, our system architecture performs a management of a distributed user profile across client and server. The high-level user preferences reflecting the long-term user interests are stored in a skeleton (high-level) profile, which is managed by the server, while the low-level preferences representing the short-term user interests are stored in a detailed (low-level) profile in the handset. This distribution enables a two-level matching process between the user profile and the news content, which uses semantic metadata extracted from the textual content and aims at the same time at a minimal computational and communication cost. Thus the available content is initially filtered on the server to derive a list of recommended items in all preferred categories, while the matching of detailed user preferences in the handset results in the displaying of the items in a ranked order.

Apart from the distributed architecture, the novelty in the proposed approach lies in the fact that the distributed user profile on both sides is automatically updated by means of machine learning processes, which are performed in the handset and by exploiting both explicit and implicit user feedback. In addition, the paper emphasizes the exploitation

of named entities in the learning process. The motivation behind the automatic adaptation of the user profile is that the latter should be consistent with the user interaction, that is, it should follow the long/short-term changes of the user interests. For instance, assume that a user has denoted several categories as topics of high interest such as "Markets" and "Politics," in order to receive interesting news items. If the user demonstrates stronger preference for the topic "Markets" and more specifically for the subtopic "Equity Markets" through her interaction with the system during a short time period, the news items from "Equity Markets" will be displayed higher than the other news items in the incoming list. On the other hand, if the user constantly selects to read news from another topic, which had been initially denoted with a low degree of preference, for example, "Society," during a long time period, then the degree of preference of this topic will be automatically increased in order to receive more "Society" news items.

This paper is organized as follows. In Section 2, the server- and client-side components of our system architecture are briefly described. Following this, in Section 3, the distribution of the user modeling is presented along with the user profile initialization process. In Section 4, the semantic annotation of the incoming news items on the server is presented. The two-level matching processes, that is, the initial content filtering performed on the server and the low-level matching in the handset, are described in detail in Section 5. The short-term and long-term learning algorithms for the automatic adaptation of the user profile are presented in Section 6 and evaluated in Section 7. In Section 8, related work addressing issues raised in this paper is reported. Finally, in Section 9, conclusions regarding the proposed system architecture and the learning processes are drawn.

2. SYSTEM ARCHITECTURE

In this section, we present the distributed system architecture across the server and the client. A general diagram of the distributed system architecture is depicted in Figure 1, while the main server and client side components of the system are illustrated in detail in Figure 2. These components and the related processes are described below.

2.1. Server-side components

The incoming to the server news items, which in our application are articles that typically include a headline and a short abstract, is first stored in the content repository. Next, they are analyzed by the metadata generation module in order to be semantically annotated with the appropriate metadata. High-level semantic information about the content typically consists of relevant topic categories. Thus, the key step in the metadata extraction is the server-side classification of each news item according to a hierarchical news taxonomy. Low-level information comprises specific terms such as nouns and Named Entities and associated weights. Hence, the semantic annotation of each news item concerns its classification to a topic, as well as the extraction of the topic-related low-level

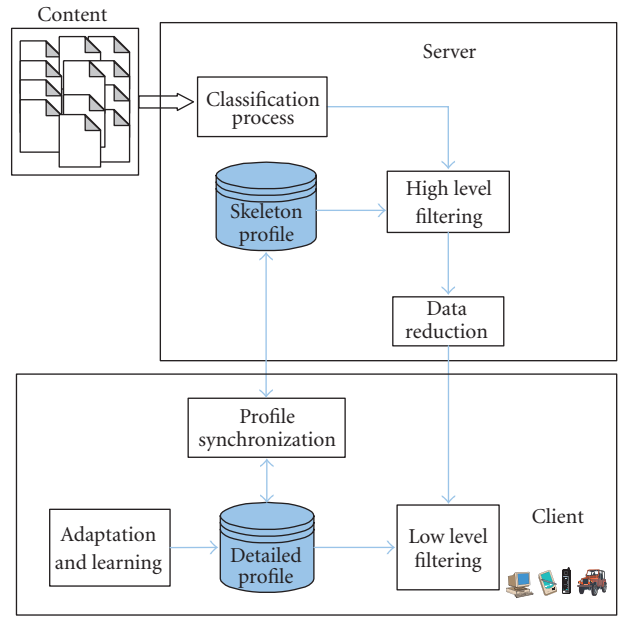


Figure 1: Distributed system architecture.

terms, that is, nouns and Named Entities contained in this item.

Following the classification process, a metadata reduction process is applied for each article, aiming to identify the most significant nouns according to the classification topic. Additionally, Named Entities are identified in each article and reduced by using a constantly evolving knowledge base. The reduction process is applied for both nouns and Named Entities of each news item, in order to significantly reduce the unnecessary metadata before their transmission to the client side, since contextual constraints exist, such as network overloading.

The news items are initially filtered on the server, based on the high-level general interests of the users stored on both the server and the handset. The high-level filtering algorithm matches entries in the skeleton user profile (long-term user interests) to the high-level metadata (e.g., topic categories, preferred sources of content). Thus, an initial set of recommended items for delivery to the user is computed, which will be subject to a further filtering step on the client. That set of news items is then transmitted to the client along with the corresponding final reduced metadata.

2.2. Client-side components

The main objective of client-side filtering apart from preserving the privacy of the user preferences (detailed profile) is to reduce the loading on the server infrastructure. The reduced metadata of each article are matched against the detailed user profile (short-term user interests) in the handset, in order to display the news items on the user screen of the mobile device with the appropriate ranking order (i.e., taking into account that the most interesting items should be displayed at the top of the screen).

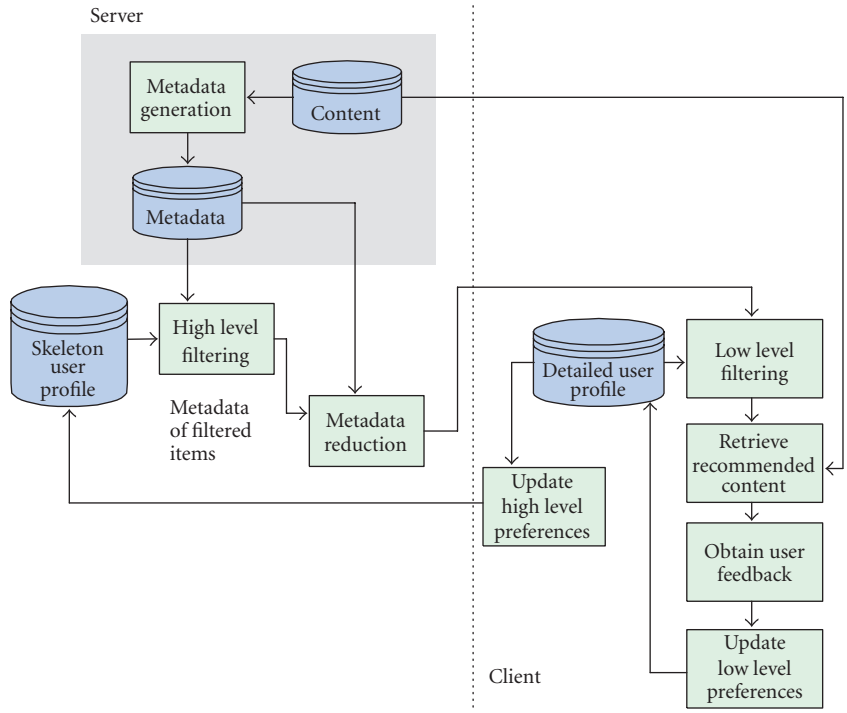


Figure 2: Diagram of client-server interactions.

The automatic adaptation of the distributed user profile concerns a two-level learning process, which is performed in the handset. More specifically, as far as the detailed user profile is concerned, usage tracking monitors the user interactions with content items and employs a short-term learning process to add significant terms (i.e., nouns and Named Entities) to the user profile. The usage tracking includes information on which items were read, what is the proportion of the length of a read item and the time spent on it. The updates of the high-level user profile by the long-term learning process are also employed in the handset, according to the classification topic of each read and nonread news item contained in a long time period set. The high-level profile can also be explicitly adapted by the user through an appropriate user interface at any time of the automatic learning process. Finally the high-level user preferences are transmitted to the server to, respectively, update the skeleton profile, whenever an adaptation takes place.

3. USER PROFILE MODELING

The main objective of the user profile modeling is to allow for the distributed semantic matching process that takes place both on the server and in the handset. This is enabled by the distribution of the user profile across client and server [1]. More specifically, the long-term user preferences are stored in a high-level (skeleton) profile on the server, and the short-term preferences in a low-level (detailed) profile in the handset.

The high-level user preferences refer to the broad news topics or categories that the users are interested in (e.g., sports, politics, business, etc.). They express the long-term interests of the user that are likely to remain the same through a long time period and they are not subject to abrupt changes [2, 3]. In contrast, the low-level user preferences refer to the more detailed aspects of the news and represent the short-term user interests, which are subject to abrupt changes, depending on the daily news.

3.1. *Initialization of the high-level profile*

Several news web sites organize their content according to taxonomies, for example, Yahoo News http://news.yahoo.com/rss. In our system, the high-level (skeleton) user profile is represented as a three-level hierarchy of topics, which corresponds to our sample categorization of the news domain and is illustrated in Figure 3. The hierarchy was defined to be unambiguous and intuitive for users, through which they can reliably identify which category a news item falls under. However, it is not an exhaustive news topics hierarchy, but it contains only a subset of news domain topics, for purposes of fast prototyping the approach presented in this paper. The hierarchy consists only of three levels, in order to reduce the user overload when she tries to explicitly fill in her personal profile. Most of the topics correspond to the categories defined in the Reuters Corpus Volume 1 (RCV1) comprising news texts produced in 1996-1997, while the other topics were added for the completion of the hierarchy. RCV1 was chosen, since it provides a wide and adequate news categorization and additionally it was freely available in XML format, being appropriate for developing and evaluation purposes. The hierarchy consists of four general topics, namely, the "Business and Finance," the "Lifestyle," the "Government/Social" and the "World Crises." Each of the

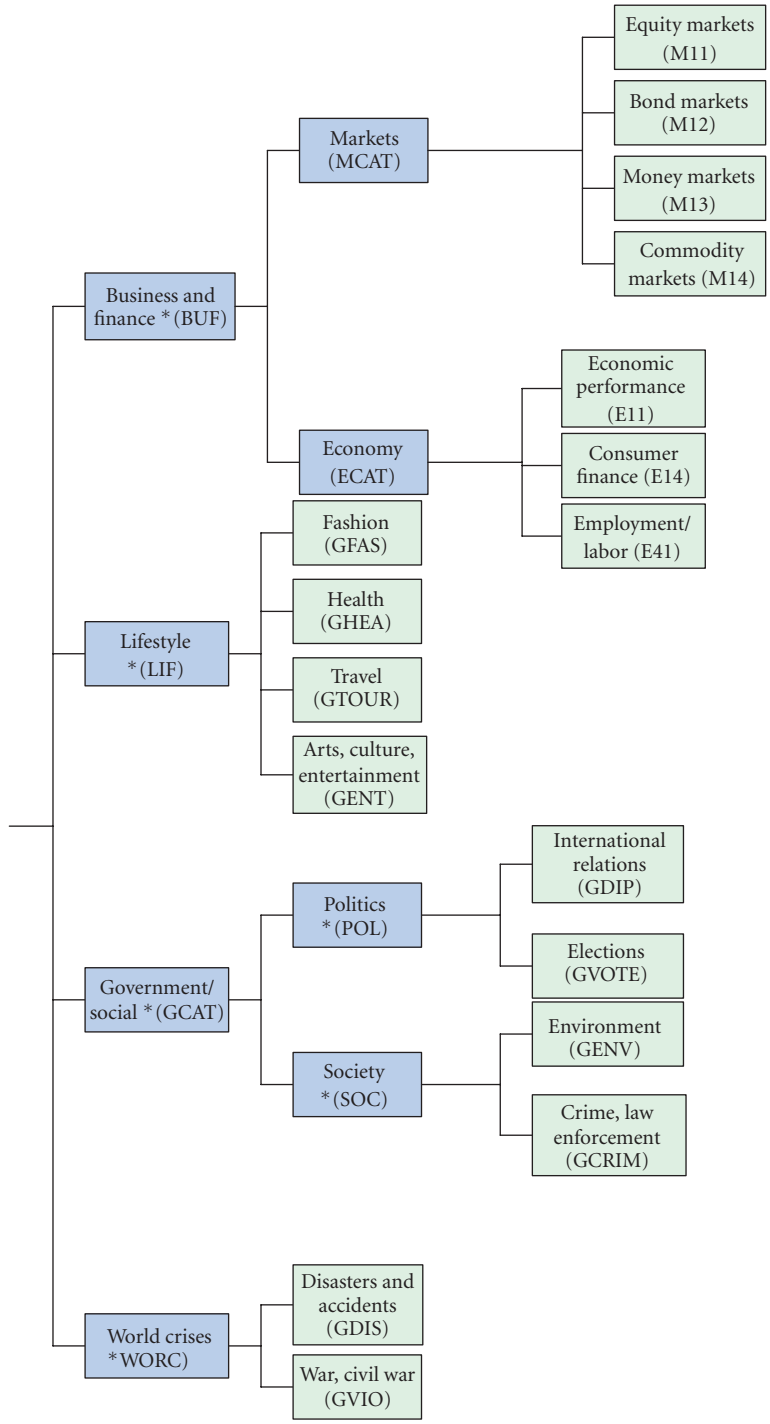
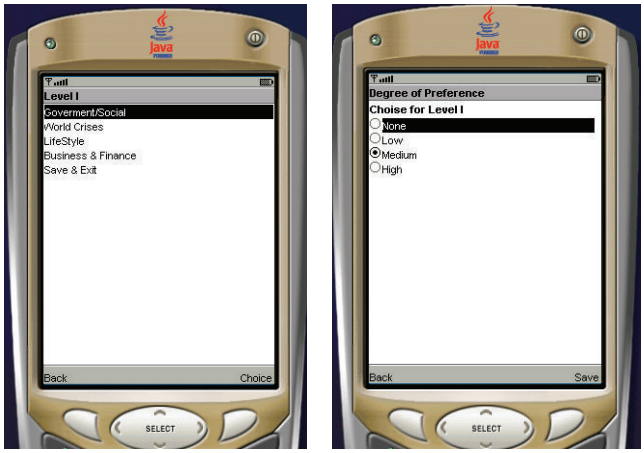


Figure 3: Hierarchy of topics. The topics, which do not belong to the Reuters corpus are marked with "∗."

above first-level topics consists of subtopics, down to the last level. Hence the hierarchy consists of four trees each of which starts from a general (first-level) topic and ends to the last level (second-or third-level) topics, namely, the leaf topics.

The server transmits the three-level hierarchy to the client in order to be presented in the handset. The initialization of the high-level user profile is performed on the mobile device where the user should explicitly express her degree of interest (i.e., high, medium, low, or none) for each particular topic based on the described hierarchy of topics. An appropriate user interface, which allows the user to browse the topics and denote her preference has been developed and is presented in Figure 4.

The initial (default) degree of interest for all the topics is "Medium" and the user may denote her preference either for an individual leaf topic, or for a higher-level topic. In the second case, all the subtopics down to the last level of



(a) (b)

Figure 4: Snapshots from handset's screens: (a) the user is allowed to choose a preferred topic at any level of hierarchy: using the scroll up/down arrows she can move to the several topics of the list, while with the "Select" button the user is able to navigate the respective next level of the hierarchy, that is, the list of the subtopics, which correspond to the selected topic. (b) This screen can be viewed through the "Choice" button of screen (a) for a marked topic. The user is allowed to denote a degree of preference for that topic using the "Select" button.

the hierarchy inherit the same degree of preference. If the user do not explicitly denote a specific degree of preference for one or more topics the default ("Medium") degree is kept for the corresponding topics. The symbolic degrees of preferences are converted to numerical values in the 0-1 scale according to almost uniform intervals (Table 1) and are stored locally in the handset (in the corresponding vector) for all the leaf topics. Regarding the definition of the numerical values of Table 1, the "Medium" degree of preference was initially considered, to be equal to the middle weight value of the interval [0, 1], that is, 0.5, and the "None" degree equal to 0. Then the "Low" and "High" degrees were approximately set, while their correspondence with the numerical values and the intervals of this table was validated through long-term learning experimental evaluation, described in Section 7.2.

The high-level user profile consists of the following vectors related to the leaf topics.

(i) A Leaf Topics Vector $\overrightarrow{\text{LTopic}}$, which contains all the leaf topics.

(ii) A Topic Weights Vector $\overrightarrow{\text{WLTopic}}$ containing the corresponding weights of the leaf topics.

The initialized high-level user profile (i.e., the Leaf Topics Vector $\overrightarrow{\text{LTopic}}$ along with a vector containing the corresponding symbolic degrees of preference) is transmitted to the server and stored in a users' database in order to formulate the skeleton user profile. It will then allow for the semantic matching process described in detail in Section 5.1. It is noted that only the Leaf Topics Vector is transmitted, since as will be described in Section 4.2.1 the news items are finally classified only to leaf topics.

Table 1: Degrees of preference and the corresponding numerical values.

Degree of preference	Numerical value	Numerical interval
None	0	0
Low	0.3	(0, 0.3]
Medium	0.5	(0.3, 0.7)
High	0.7	[0.7, 1]

Additionally when the user browses the nonleaf hierarchical topics a bottom-up propagation of the degrees of preference is performed. More specifically, the adaptation of the degree of preference for all the nonleaf topics is determined from the average preference (numerical) value of their corresponding subtopics (as described in Section 6.2.3). This average value is quantized according to the existing symbolic degrees of preference as presented in Table 1.

Finally, the user is allowed to explicitly adapt her high-level profile whenever she wills, through the above-described user interface of the mobile device and following the same steps presented for the initialization of the profile.

3.2. Initialization of the detailed (low-level) profile

The detailed user profile, which is implicitly initialized and adapted by the personalization system, consists of textual low-level features, that is, nouns and Named Entities, that play a key role in the news personalization domain, as they capture a major part of the semantics in a news item. More specifically, the detailed user profile consists of the following vectors related to the nouns and Named Entities (Figure 5):

(i) A Terms Vector $\vec{T}$, which contains the nouns and the Named Entities.

(ii) A Weights Vector $\vec{W}$ containing the corresponding weights.

(iii) A Usage History Vector $\overrightarrow{\text{UH}}$ containing the corresponding usage history of each term, which is a counter of how many times the term has been selected (integer number).

(iv) A Terms Type Vector $\overrightarrow{\text{TT}}$, which contains for each term a character which expresses its corresponding type, that is, "N" for "Noun" is assigned to each noun and "P" for "Person," "O" for "Organization" and "L" for "Location" are assigned to Named Entities (see example in Figure 5).

The initialization of the detailed user profile refers to a transitive period starting the first time the user interacts with the personalization system until the vector of low-level terms is sufficiently large according to a defined maximum number of terms (100 terms were used in our experiments). During this period, all extracted low-level features contained in the news items selected by the user will be inserted in the Terms

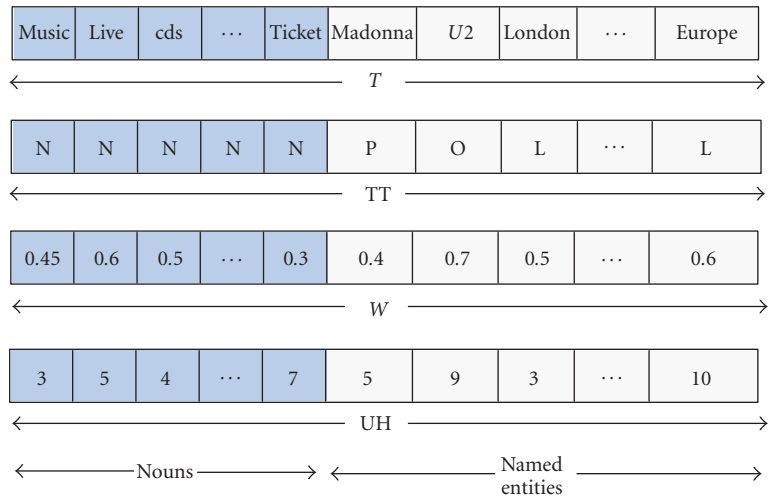


Figure 5: Example of detailed user profile vectors.

Vector $\vec{T}$ of the detailed user profile, having an initial weight equal to 0.5 (corresponding to a "Medium" preference), in the Weights Vector $\vec{W}$. Additionally, the values contained in the Usage History Vector $\overrightarrow{\text{UH}}$ for both nouns and Named Entities are updated throughout the initialization process according to the number of news items that the user selects.

In addition, during the initialization process, an important characteristic of the personalization system is initialized, which is the user behavior determined by the reading rate, that is, the number of news items selected per day. This is used in the proposed mathematical formulas (3) and (4), of Section 6.1.1, aiming to adjust the weights of the terms.

4. SEMANTIC ANNOTATION OF NEWS CONTENT

In order to apply the personalization system based on the distributed user profile modeling described in the previous section, the news content should be semantically annotated with the appropriate metadata. More specifically, the analysis of the textual news content results in the extraction of the leaf topic to which each news item is classified and also of the significant (according to the classification topic) low-level terms, that is, nouns and Named Entities contained in this item. The semantic annotation of the news items takes place on the server side and the extracted metadata are transmitted to the handset following a reduction process that takes into account the communication and computational costs.

4.1. Construction of training sets

Both the classification and the metadata reduction processes, described in Sections 4.2 and 4.3, respectively, require a training stage, which involves the use of the Reuters Corpus Volume 1 (RCV1) to provide the articles that can be used for training.

The training sets used for the semantic representation of each leaf topic in the proposed approach are generated according to the "Same Subtree" filtering. More specifically, the news item is included in the training set of a leaf topic, only if it is categorized from Reuters, apart from the particular leaf topic, to sibling leaf topics (only if they belong to the third-level), or to its ancestors up to the first level. For example, a news item which according to Reuters is categorized to the topics MCAT (Markets), M11 (Equity Markets), and M13 (Money Markets), can be included in the training set of the leaf topic M11, while on the contrary, it is excluded from that training set if it is categorized to the topics MCAT (Markets), M11 (Equity Markets), E14 (Consumer Finance). The training set of each non-leaf hierarchical topic is constructed using the training sets of their subtopics. This criterion is applied for the semantic representation of the topics for the classification purpose.

The training sets, which are used for the data reduction process are constructed according to the "Only One Leaf Topic" filtering, that is, the news item is included in the training set of a leaf topic, if it is categorized from Reuters, only to this particular leaf topic and to any other higher-level topic. Since the data reduction process involves only the leaf topics' representation, there is no training set construction for the non-leaf hierarchical Topics.

4.2. Metadata generation

4.2.1. News item classification according to topic category

The current classification method is based on machine learning along with vector representation techniques and uses the fixed hierarchical taxonomy of content categories, presented in Section 3.1. The automatic framework for the text classification is composed of an offline training process and an online classification process. An overview of the overall topic classification process is shown in Figure 6.

The textual analysis process includes the typical NLP preprocessing steps (i.e., tokenization, sentence splitting, parts of speech tagging, stemming), which are performed with the use of the appropriate GATE components [4, 5] and the text search engine Lucene http://lucene.apache.org/.

A statistic text analysis follows according to the vector space model, where each news article is represented as a vector of feature-value pairs. The features used are the extracted nouns in the text and the values are the corresponding weights based on their frequency of appearance in the text (term frequency—TF). Accordingly, each topic in the hierarchy is also represented as a feature-value vector that best expresses the semantics of this topic. The features are the semantically significant nouns while the values correspond to the term frequency-inverse document frequency (TF-IDF) weights [6]. This is referred to as the Topic Prototype Vector, statistically constructed from the appropriate training set.

The construction of the Topic Prototype Vectors is based on the relevance feedback algorithm originally proposed by Rocchio [7] for the vector space model. A generalization of the Rocchio algorithm that can be used for text categorization with more than two categories has been proposed by [8]. The Topic Prototype Vectors in our approach have been constructed using the Rocchio formula along with the topics hierarchy and the training sets generated according to the "Same Subtree" criterion described in Section 4.1. Thus, for each topic in the formula, as relevant articles (positive terms) are taken into account the training articles belonging to that topic, whereas as nonrelevant articles (negative terms) are considered the training articles belonging to all subtrees apart from the subtree of the topic.

Following the offline construction of the Topic Prototype Vectors, the online classification process depicted in Figure 6 results in the categorization of the news item in only one leaf topic. It involves the assignment of each incoming news article to the leaf topic with the shortest distance between the Topic Prototype Vector and the article's vector of feature-value pairs. Thus the news item is classified to the leaf topic with which its noun terms vector has the highest cosine similarity, given by the following formula:

$$\mathrm{Sim}\,(\mathrm{PV}, I) = \frac{\overrightarrow{\mathrm{PV}} * \vec{I}}{|\overrightarrow{\mathrm{PV}}| \cdot |\vec{I}|}, \tag{1}$$

where $\overrightarrow{\mathrm{PV}}$ is the vector of the TF-IDF weights of the Topic Prototype Vector, while $\vec{I}$ is the vector of TF weights of the noun terms extracted from the news item.

4.2.2. Extraction of low-level metadata

The extracted low-level features include only the common nouns between those which were initially identified in the news item with the aid of GATE (the type of noun "N" was assigned to each of them) and the Prototype Vector of the topic category where it was classified. Along with these nouns their TF weights in the news item are extracted.

The analysis goes further by identifying the Named Entities contained in the news item, as well as their corresponding type (such as person, organization, location) both using GATE software. According to this, an additional term-frequency vector of Named Entities is generated for each content item in the online process. Furthermore, an enhanced semantic identification process for Named Entities is performed, as described in detail in Section 4.3.2.

4.3. Metadata reduction

Following the extraction of metadata, the next stage concerns their transmission to the client side in order to be used for the low-level filtering and learning processes. However, it is sensible to significantly reduce those metadata before the transmission, due to contextual constraints, such as the network overloading as well as, rarely in nowadays, limited memory space and processing capability in the client device. Thus, the unnecessary terms are eliminated for reducing the communication cost along with the computational cost in the handset as much as possible, aiming to allow the personalization process to be the most efficiently performed.

4.3.1. Reduction of nouns using adapted TF-IDF method

The reduction of the noun terms is made based on the presumption that after different incoming documents are classified in a given topic, the differentiation between them can be made using only a subset of the extracted metadata, which were described in Section 4.2.2.

The reduction of nouns involves an offline training process aiming at a representation of the leaf topics, which is different from the ones used in the classification process in Section 4.2.1. During this stage, only the corpus of articles pre-classified in the same leaf topic is used in the training sets for each leaf topic, that is, the training set used has been constructed according to the "Only One-Leaf-Topic" criterion, described in Section 4.1. Hence, the above-mentioned corpus of each leaf topic is employed for the extraction of the new set of nouns along with the corresponding TF-IDF weights. In order to maintain the terms with the highest relevance a threshold has been defined, which corresponds to the percentage of the highest weighted nouns that will construct the reduced representation of the leaf topics. To this end, the 10% of the extracted nouns along with their TF-IDF weights will be contained in the resultant feature-value vector, namely the Adapted TF-IDF Prototype Vector.

The online step of the reduction process concerns the identification of the common nouns between the incoming document and the Adapted TF-IDF Prototype Vector of the leaf topic where it has been classified. Thus, the document's metadata representation will be reduced only to those noun terms that are also present in the corresponding Adapted TF-IDF Prototype Vector.

The reduction aims for the new metadata to be able to identify the particular sub-area of the given leaf topic in which a user is interested. In this case, if a noun term is not relevant for the topic category in which the document was classified, it would be unlikely to be relevant for a particular sub-area from that topic. Additionally, the document's metadata representation will eliminate all the noun terms that are not relevant for an intratopic classification. This reduction is made according to the fact that the Adapted TF-IDF Prototype Vector contains noun terms that are relevant in making a differentiation with other topics but may have a low differentiation value for the intratopic classification. For example, if an article was classified into the topic *Tennis* in order to determine the sub-area of the Wimbledon event,

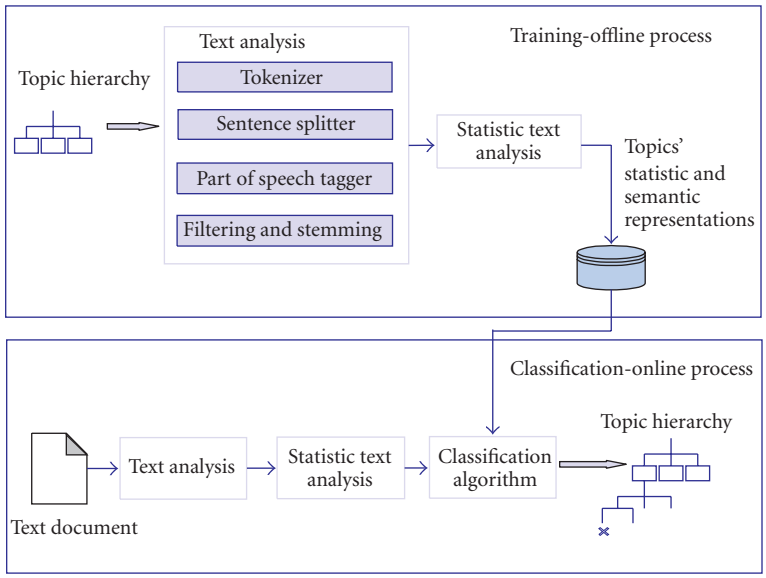


Figure 6: Diagram of news items classification process.

terms such as *tennis*, *set*, *game* will have less relevance than *grass*, *July*, *slam* and of course *Wimbledon*.

As the result of the data reduction process, the nouns which are contained in both the news item and the Adapted TF-IDF Prototype Vector of the corresponding leaf topic along with their TF weights in the news item will be sent to the user's device as the most significant topic-related nouns. These nouns will be called Adapted TF-IDF nouns.

4.3.2. *Reduction of named entities with a construction of a named entities knowledge base*

During the metadata generation process there is no process aiming at the semantic identification of the Named entities. Thus, the output of the process is limited to Named Entities recognition and classification into a particular type (i.e., person, organization, location). To overcome this limitation, a methodology for constructing a Named Entities knowledge base has been defined, aiming both at semantically identifying Named Entities and also reducing the amount of data transmitted to the client.

More specifically, the semantic identification of a Named Entity concerns its association along with it's corresponding type, with the particular topic where the news item, which contains the Named Entity has been classified. Additionally, the intended use of the knowledge base of Named Entities is to deal with cases of Named Entities having more than one representations. To this end, an ontology based learning approach has been followed to handle multiple interpretations such as follows.

(i) Identify that two or more different representations refer to the same Entity. For example, identify that "Greenspan" and "The Federal Reserve Chairman" refer to the same person.

(ii) Associate a Named Entity with its abbreviation. For example, "U.S." and "United States" refer to the same country.

In order to reach the aforementioned goals, the knowledge base of Named Entities is constructed following two complementary processes.

(i) The process dealing with abbreviated Named Entities connecting to external abbreviations databases. Two such databases have been investigated, one concerning Locations and the other concerning organizations. The locations database is actually a list of countries acronyms, whereas the organizations database consist of a number of organizations and companies.

(ii) An ontology-based discovery of multiple representations. In order to handle different representations, they are initially regarded as distinct Entities and gradually identify their associations with other existing Named Entities, according to the learning criteria of the co-occurrence in the same context, and of the belonging to the same type.

The reduction process of Named Entities includes the assigning of all possible representations of a Named Entity to a particular code, corresponding to a unique character sequence. Additionally, the system recomputes the frequency of appearance of each Named Entity in the news item, with respect to the new reduced Named Entities vector (where each one is represented by a unique code).

4.4. *Metadata storage and transmission to the handset*

Following the metadata extraction and reduction processes, the metadata are stored in a news items repository (Figure 7)

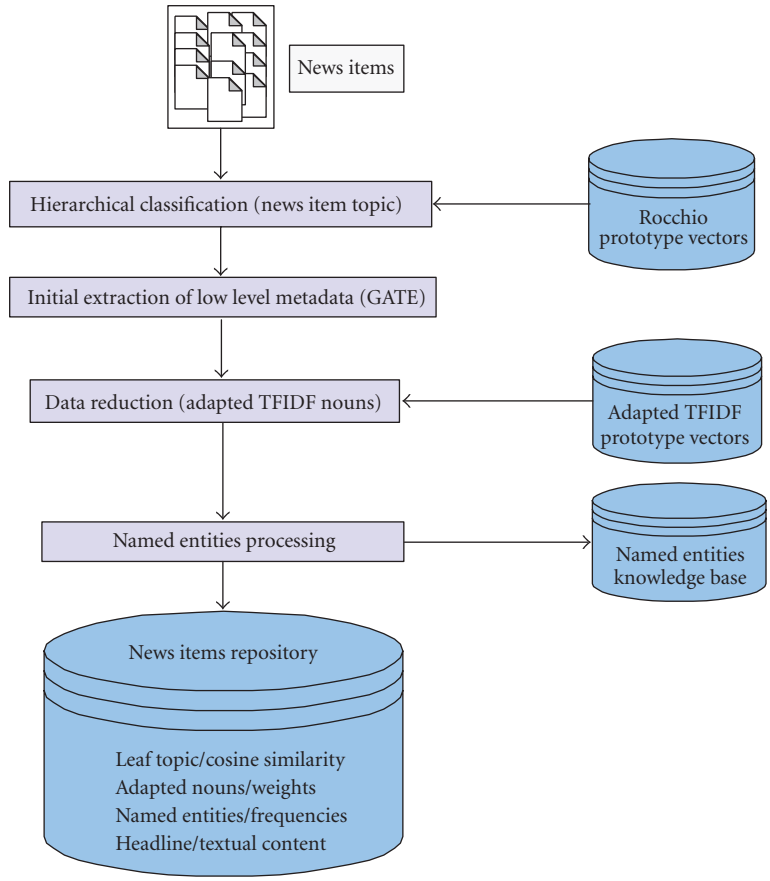


Figure 7: Metadata extraction and storage in the news items repository.

and are finally transmitted to the user device to be used in the low-level filtering and learning processes in the handset. The stored metadata arethe following.

(i) The Adapted TF-IDF nouns.

(ii) The weights of the Adapted TF-IDF nouns corresponding to their frequency of appearance (TF) in the news item.

(iii) The codes of Named Entities.

(iv) The frequencies of the Named Entities' appearance in the news item.

(v) The corresponding type of each term (i.e., the vector which assigns the type "N" to the nouns and "P," "O," or "L" to Named Entities).

(vi) The leaf topic where the news item was classified.

(vii) The cosine similarity value that has been computed for the news item and the classification topic; this is not transmitted to the handset but it is stored in the repository in order to be used in the server-side initial content filtering process, presented in Section 5.1.

(viii) The headline and the textual content of the news item.

Apart from the above-mentioned metadata, a factor referring to the effect of each Named Entity type to a specific topic is also transmitted. This is independent of each individual incoming news item and is calculated from the Named Entities knowledge base for each leaf topic. More specifically, it is measured by the percentage of the Named Entities belonging to a particular type (i.e., person, organization, or location) in a particular leaf topic, with regard to the total number of Named Entities belonging to this topic. This metric is used in the short-term learning process in the handset described in Section 6.1.

5. DISTRIBUTED SEMANTIC MATCHING

The main idea of this section is to present the high-level filtering of available content on the server, followed by matching of detailed user preferences in the handset. The output of the first filtering step is a list of recommended items for each user in all preferred leaf topic categories. Then, following the second filtering step, the content is displayed to the user in a ranked order. The distributed semantic matching process is described in detailed in this section.

5.1. *Server-side initial content filtering*

There are two inputs to the high-level filtering algorithm: the output of the topic classification process, that is, the only one leaf topic for the incoming to the server news items

(news items repository) and the explicit high-level user preferences stored in the skeleton profile (users' database). The main idea is that the content items are sent for further processing only to the users whose high-level profiles are related to the leaf topic of the content item. Hence, each user is assigned a set of news items that are classified to topics denoted as topics of interest. More specifically, the high-level filtering is based on a matching, implemented by simple queries, between the classification topic of each semantically annotated news item stored in the news items repository, and the preferences in the skeleton profile of each user in the users' database. Then according to simple rules which take into account this matching, the user will receive

(i) the 100% (all) of the incoming to the server news items, which were classified to a leaf topic denoted with "High" degree of preference;

(ii) the 50% of the incoming to the server news items, which were classified to a leaf topic denoted with "Medium" degree of preference;

(iii) the 30% of the incoming to the server news items, which were classified to a leaf topic denoted with "Low" degree of preference; and

(iv) none of the incoming to the server news items, which were classified to a leaf topic denoted with "None" degree of preference.

The 50% and 30% of the incoming news items in the cases of "Medium" and "Low" preference, respectively, are selected according to their ranking using the cosine similarity value estimated during the server-side classification. Namely, the 50% and 30% top ranked articles of the leaf topic are transmitted to the client. The motivation behind the rules regarding these 50% and 30% cases, lies in the following assumptions:

(i) A user who has denoted high interest in a topic would read all the news articles related to that topic.

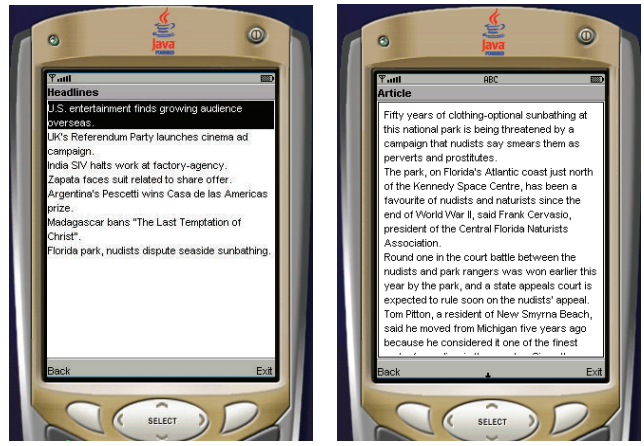
(ii) On the other hand a user who has denoted medium or low interest to a topic would be satisfied to receive some news items concerning that topic, a percentage of them in the case of medium interest (i.e., the 50%) and a smaller one in the case of low interest (i.e., the 30%), in order to keep herself informed about the topic.

(iii) Finally if the user is not interested at all in a particular topic, then she would be annoyed to receive related news items.

The output of the high-level filtering algorithm will be a list of content items for each user that will be submitted to the process of retrieval of the appropriate metadata from the news items repository.

5.2. Client-side low-level filtering-ranking

The semantic matching in the handset involves the semantic similarity between the detailed user profile stored in the handset and the significant low-level terms extracted from the article, that is, the Adapted TF-IDF nouns and the Named Entities, measured according to the cosine similarity metric.



(a) (b)

Figure 8: Snapshots from handset's screens (a) the ranked list of the headlines, and (b) the textual content of a selected article.

The cosine similarity measure between the detailed user profile and the vector containing the terms of the document is calculated:

$$\mathrm{Sim}\left(W, I'\right) = \frac{\vec{W} * \vec{I'}}{|\vec{W}| \cdot |\vec{I'}|}, \tag{2}$$

where $\vec{W}$ is the Weights Vector of the detailed user profile, while $\vec{I'}$ is the vector of TF weights of the terms (i.e., the Adapted TF-IDF nouns and reduced Named Entities) extracted from the news item.

After calculation of the cosine similarity measures for all the incoming news items, the headlines of the news items are displayed on the user's screen based on the descending order of their corresponding cosine similarity measures, that is, the headlines of the articles with the highest cosine similarities results are displayed higher on the list, as illustrated in Figure 8. This ranked order corresponds to the short term user preferences, which are recorded in the detailed user profile. The user is able to view the textual content of each news item by selecting each headline in the list.

6. USER PROFILE LEARNING AND ADAPTATION

The user profile learning process is necessary for the personalization system, since user information needs are constantly changing, particularly in the context of the news domain. In this framework, the implicit user profile learning in the handset aims at identifying two different types of interest changes:

(i) Abrupt interest changes: Abrupt interest changes may occur when new information needs arise due to user curiosity/immediate thoughts (internal) or motivation by the question of another person (external). Those changes refer to the need for adaptation of the short-term model.

(ii) Gradual interest changes: User interests are widely recognized as changing slowly and gradually over time, for example, as conditions, goals and knowledge change. Gradual changes happen as consequences of continuous progress, for example, the user gaining experience or growing older. Those changes motivate the need for adaptation of the long-term model. However, as was already mentioned in Section 3.1, abrupt changes in the high-level profile can also be explicitly inserted by the user through the user interface employed in the initialization phase.

Our system performs a two-level learning process in order to automatically update the detailed and high-level user profile in the handset.

6.1. Short-term learning

The short term learning process exploits the implicit user feedback (i.e., monitoring of the user interactions with content items) for the adaptation of the detailed user profile, where the two main types of semantic metadata, that is, the nouns and Named Entities are involved. More specifically, the short-term user profile learning supports two main functionalities:

(i) The adaptation of the values contained in the Weights Vector $\overrightarrow{W}$

(ii) The insertion and elimination of terms into and from the Terms Vector $\overrightarrow{T}$ of the detailed user profile, respectively.

Both functionalities take place after the initialization process described in Section 3.2 has been completed, and thus the nouns and Named Entities corresponding to the news items selected by the user has been initially inserted in the Terms Vector $\overrightarrow{T}$ of the detailed user profile.

6.1.1. Weights adaptation

In several systems, which perform learning processes in order to update the user profiles, mathematical formulas are used for the adaptation of the different weight values [3, 9]. In our approach, the values contained in the Weights Vector $\overrightarrow{W}$ of the detailed user profile, that is, the weight of each term in the Terms Vector $\overrightarrow{T}$, is updated according to a formula depending on whether the user selects or ignores news items that contain the term. This formula incorporates factors related to the particular term, such as the previous weight and the usage history of the term. Additionally, factors related to the selected content items where the term is contained are participating in the formula, such as the similarity measure of the item with the detailed user profile, the explicitly denoted weight of the leaf topic where it belongs, and the proportion of the amount of time spent to the item to its length [10]. Apart from the aforementioned factors the overall user behavior towards the personalization system is taken into account in order to adapt the weights of the terms in the detailed profile, that is, the average number of read news item per day.

More specifically, the weights of the noun terms are adapted according to the following formula:

$$W_{\text{new}} = W_{\text{old}} \pm W_{\text{LT}} \cdot \text{Sim}(W, I') \cdot e^{-\beta * U_b * U_h} \cdot \log \frac{\text{time}}{\log \text{length}}. \tag{3}$$

Correspondingly the weights of the Named Entities, are updated according to a similar formula:

$$W_{\text{new}} = W_{\text{old}} \pm W_{\text{LT}} \cdot W_{\text{NEType}} \cdot e^{-\beta * U_b * U_h} \cdot \log \frac{\text{time}}{\log \text{length}} \tag{4}$$

where:

(i) W_{old}: represents the current term weight to be updated contained in the Weights Vector $\overrightarrow{W}$.

(ii) $\pm$: is used to increase or decrease the current weight in case of positive or negative feedback, respectively. The articles that the user clicks to read are considered to be positive feedback for a term which exists in them, while the rest of the documents that contain the term but are not selected by the user, are considered to be negative feedback for the term.

(iii) W_{LT}: is the explicitly denoted high-level weight of the leaf topic to which the news item has been classified (contained in the topic weights vector $\overrightarrow{\text{WLTopic}}$ of the high-level profile).

(iv) $\text{Sim}(W, I')$: is the cosine similarity measure between the Weights Vector $\overrightarrow{W}$ of the detailed user profile and the vector of TF weights of the terms (i.e., the Adapted TF-IDF nouns and reduced Named Entities) extracted from a news item.

(v) The W_{NEType} is a factor referring to the effect of each Named Entity type (person, organization, or location) to a specific topic. It is calculated based on the Named Entities knowledge base and represents the semantic information, which is gathered from there.

(vi) $\log(\text{time}/\log \text{length})$: incorporates the amount of time spent reading a news item in seconds and the length of the article in bytes, which operates as the normalizing factor. In the case of negative feedback, the time-length factor is set to 1, that is, it has no effect in the weight adaptation since the user does not spend time on the corresponding article.

(vii) $e^{-\beta * U_b * U_h}$: is used to follow the personalized nonlinear change of the term weight according the usage history of the term. The changing rate of the weight is inversely proportional to the value of the parameter U_h that stands for the integer number of the selected articles where the term exists (contained in the Usage History Vector $\overrightarrow{\text{UH}}$) and U_b, which represents the

indicative mean number of articles that the user selects to read per day. The more articles a user reads per day, for example, the more slowly the weights increase in the low level profile.

(viii) β: is a constant that is used to differentiate between the changing rate of the weight if the update is performed in relation with an interesting article or a non-interesting one. Thus it takes different values in the two opposite scenarios of positive/negative user feedback. More specifically, in the case of non-read articles (i.e., negative feedback from the user), the changing rate (i.e., the decreasing rate) should be much slower, since an unread news item does not constitute an explicit indication for non-interest. This is because an unread news item apart from considering it as not interesting, it can be interpreted as already read from another source, or it is possible that the user had no time to spend on it. On the contrary, in the case of read articles (i.e., positive feedback from the user) the changing rate (i.e., the increasing rate) should be faster, since a read news item demonstrates a strong indication for interest. Based on the numerical values produced by applying the formula, the proposed values for the beta constant were experimentally set to the following:

(a) $\beta = 0.01$ for read news items (positive feedback);

(b) $\beta = 0.02$ for non-read news items (negative feedback).

6.1.2. Insertion/elimination of terms into/from the detailed user profile

Apart from weights adaptation, a mechanism has been developed to update the terms (both nouns and Named Entities) contained in the Terms Vector $\vec{T}$ of the detailed user profile. This ensures that the detailed user profile does not remain static after the initialization process but is constantly updated based on specific criteria.

When a user reads an article, which contains new terms (i.e., terms not existing in the current detailed profile), each of these terms is placed in a subordinate waiting stack as depicted in Figure 9. Then, each time the user selects a news item that contains any of those terms, the corresponding usage history value of each term in the waiting stack changes (i.e., it increases by one for each selection). The metric that determines the insertion of a new term into the Terms Vector $\vec{T}$ in the detailed user profile is whether the term usage history exceeds a certain threshold. This threshold is determined by the user attitude towards the personalization system, namely it is proportional to the average number of news items that the user reads per day (e.g., for a user who reads approximately 20 news items per day, the usage history threshold of the terms in order to be inserted into the detailed user profile corresponds to 5). When a term is inserted into the Terms Vector $\vec{T}$, its initial weight in the Weights Vector $\vec{W}$ has the default value of 0.5. Thus, the default values for the initial entry into the system are similar to those used during the initialization process described in Section 3.2.

While the user interacts with the system, there may also be a need to remove terms from the detailed user profile, which imply low or non-interest from the user. In order to remove a term, both of the following two criteria should be satisfied:

(i) whether the value of the term in the Usage History Vector $\overrightarrow{\mathrm{UH}}$ is lower than a certain threshold, which similarly to the insertion depends on the average number of read news items per day and it is lower than the insertion threshold, and

(ii) whether the value of the term in the Weight Vector $\vec{W}$ is lower than another certain threshold, which corresponds to a weight value around the medium preference (i.e., 0.5). It is additionally noted that if the weight of a term has turned to zero after several negative feedbacks, it is removed from the detailed profile anyway, that is, independently of its usage history.

6.2. Long-term learning

During the initialization process of the high-level user profile described in Section 3.1, the user explicitly denotes her high-level preferences, which are then transmitted to the server to allow for the initial content filtering. However, even the long-term user interests are subject to slow and gradual changes.

Thus, a long-term learning process has been developed to allow the system to follow any changes in the user preferences by automatically update the high-level user profile. This process involves a long-term user model, which consists of the following vectors, illustrated in Figures 10 and 11:

(i) A Long-Term Noun Vector $\overrightarrow{\mathrm{LTN}}$, which contains the nouns of a long-term set of articles.

(ii) A Long-Term Correlation Values Vector $\overrightarrow{\mathrm{LTCV}}$, which contains the corresponding calculated correlation values of the nouns in the long-term set of articles.

(iii) The Prototype Nouns Vector $\overrightarrow{\mathrm{PN}}$ containing the Adaptive TF-IDF Prototype nouns for each leaf topic.

(iv) The Prototype Weights Vector $\overrightarrow{\mathrm{PW}}$ containing the corresponding Prototype weights.

(v) A Prototype Correlation Values Vector $\overrightarrow{\mathrm{PCV}}$, which contains the calculated correlation values of the Prototype nouns for each leaf topic.

(vi) A Long-Term Weights Vector $\overrightarrow{\mathrm{LTW}}$, which contains long-term weights of the prototype nouns, which are constantly updated during the long-term learning process.

The long-term learning process involves three stages:

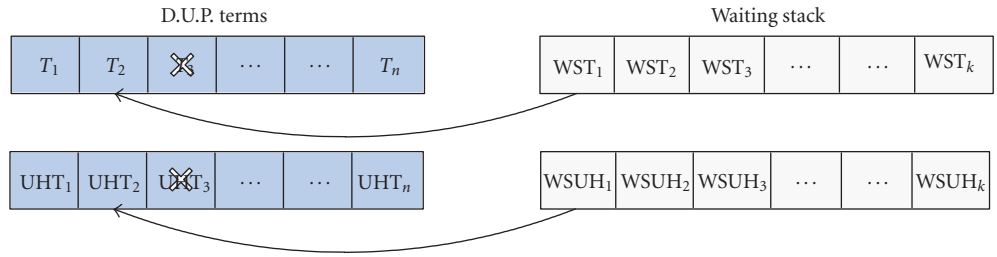


Figure 9: Insertion/elimination of terms into/from the detailed user profile.

(i) The collection of nouns contained in a long-term set of articles.

(ii) The association of the collected noun terms with the long-term user model and the adjustment of their weights according to a long-term learning formula.

(iii) The updating of the skeleton profile on the server based on a client-server synchronization process.

6.2.1. *Collection of nouns contained in a long-term set of articles*

While the user is interacting with the system, all the articles displayed to the user (either selected or not), constitute a long-term set, on which the long-term learning process is based. The number of articles belonging to this set is predefined in the system to ensure that it covers a long-term period. All the noun terms that are contained in the long-term set of articles participate in the learning process.

In order to identify the effect of all those nouns on the long-term model, their correlation values have been investigated. More specifically, the relation between the change in the weights of the noun terms belonging both to the detailed user profile and a long-term set and their corresponding correlation values has been examined. After experimentation it has been found that:

(i) When the correlation value is positive, there is an increase of the weight. Additionally, the greater the correlation value is, the larger the increase in the weight.

(ii) When the correlation value is negative, there is a decrease in the weight.

(iii) When the correlation value is zero, there is no change in the weight.

The motivation behind this investigation concerns the exploitation of the correlation value in order to update the long-term weights of the Prototype nouns corresponding to each leaf topic (Long-Term Weights Vector $\overrightarrow{\text{LTW}}$), aiming at the automatic adaptation of the weight of each leaf Topic contained in the high-level profile, that is, in the topic Weights Vector $\overrightarrow{\text{WLTopic}}$. Therefore, by estimating the correlation values of the nouns in a long-term set of articles the weights of the leaf topics in the high-level profile can adapted.

The correlation values are calculated for all the terms contained in the long-term set of articles, inserted in the Long-Term Nouns Vector $\overrightarrow{\text{LTN}}$, and they are stored in the Long-Term Correlation Values Vector $\overrightarrow{\text{LTCV}}$ (Figure 10). If the training set contains N incoming news items, the two binary discrete random variables taking 0 or 1 for values are defined:

(i) X = Event that a randomly selected article contains the term w.

(ii) Y = Event that the user selects to read an article from the incoming set.

If P_X and P_Y are the probability density functions of X, Y, respectively, the joint probability density function is P_{XY}. The function to be computed is the correlation $\rho(X, Y)$ between X, Y, which produces a value between -1 and 1. The positive value for the term w indicates a dependency of the user selecting the article on the occurrence of w, while the negative value would tend to indicate that the user would not read articles containing w. A value of 0 would indicate that the two events of the user selecting an article and the occurrence of the term w in a news item are independent. The joint probability density functions for $(x, y) = (0,0), (1,0), (0,1), (1,1)$ are considered as:

$P_{XY}(1,1)$ = {articles containing w that the user selects}/N

$P_{XY}(1,0)$ = {articles containing w that the user does not select}/N

$P_{XY}(0,1)$ = {articles not containing w that the user selects}/N

$P_{XY}(0,0)$ = {articles not containing w that the user does not select}/N as well as the marginal probabilities $P_X(x)$ and $P_Y(y)$:

$$\begin{aligned} P_X(1) &= P_{XY}(1,0) + P_{XY}(1,1) \\ P_X(0) &= P_{XY}(0,0) + P_{XY}(0,1) \\ P_Y(1) &= P_{XY}(0,1) + P_{XY}(1,1) \\ P_Y(0) &= P_{XY}(0,0) + P_{XY}(1,0). \end{aligned} \tag{5}$$

The correlation coefficient between X, Y is [11]:

$$\rho(X,Y) = \frac{E(XY) - E(X)\cdot E(Y)}{\sigma(X)\cdot\sigma(Y)}. \tag{6}$$

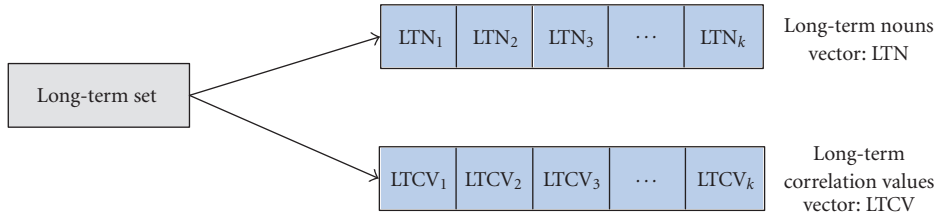


Figure 10: Vectors of nouns contained in the long-term set of articles.

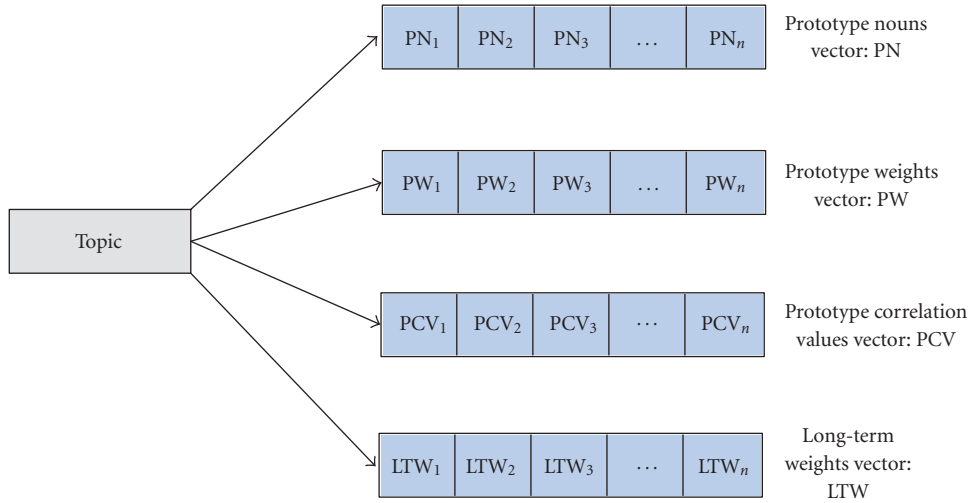


Figure 11: Vectors of nouns and weights corresponding to a particular leaf topic in the long-term learning model.

Analyzing the parts of this equation,

$$E(XY) - E(X)\cdot E(Y) = \sum_{xy} xy\cdot P_{XY}(x,y) - \left(\sum_{x} x\cdot P_X(x)\right)\cdot\left(\sum_{y} y\cdot P_Y(y)\right) = P_{XY}(1,1) - P_X(1)\cdot P_Y(1) \tag{7}$$

while

$$\sigma^2(X) = E(X^2) - E(X)^2 = P_X(1) - P_X(1)^2 = P_X(1)\cdot P_X(0). \tag{8}$$

Finally, the correlation between X, Y is calculated as follows:

$$\rho(X,Y) = \frac{P_{XY}(1,1) - P_X(1)\cdot P_Y(1)}{\sqrt{P_X(0)\cdot P_X(1)\cdot P_Y(0)\cdot P_Y(1)}}. \tag{9}$$

6.2.2. Association of low-level terms with the long-term learning model

Following the collection of nouns contained in the long-term set and the calculation of their correlation value, the next step is the association of those nouns with the leaf topics in the hierarchy of the high-level user profile. To this aim, the handset receives from the server the Adapted TF-IDF Prototype Vectors containing the nouns for each leaf topic and their corresponding weights, which are stored in the handset's memory. These are the Prototype Nouns Vector $\overrightarrow{\text{PN}}$ and the Prototype Weights Vector $\overrightarrow{\text{PW}}$ depicted in Figure 11.

The long-term learning process for a particular leaf topic aims at adjusting the weight of the topic, which is initially specified from the user during the initialization of the high-level profile described in Section 3.1. This explicitly defined weight is propagated to all the nouns contained in the Prototype Nouns Vector $\overrightarrow{\text{PN}}$ in order to initialize the Long-Term Weights Vector $\overrightarrow{\text{LTW}}$ also displayed in Figure 11. Thus, the long-term learning process involves the adaptation of all the weights in the Long-Term Weights Vector $\overrightarrow{\text{LTW}}$. This is performed, according to the following steps:

(1) All the Adapted TF-IDF Prototype nouns of this topic (in the Prototype Nouns Vector $\overrightarrow{\text{PN}}$), which also belong to the Long-Term Nouns Vector $\overrightarrow{\text{LTN}}$ are identified.

(2) For the Adapted TF-IDF Prototype nouns that are NOT present in the Long-Term Nouns Vector $\overrightarrow{\text{LTN}}$, there is no change in the long-term weight, since the correlation value is assumed to be zero (due to the lack of information).

(3) For the Adapted TF-IDF Prototype nouns that are also present in the long-term set, the new long-term weight is computed according to the following mathematical formula, after their corresponding correlation values are identified in the Long-Term Correlation Values Vector $\overrightarrow{\text{LTCV}}$ and stored in the Prototype Correlation Values Vector $\overrightarrow{\text{PCV}}$:

$$W_{\text{new}} = W_{\text{old}} + \text{CV} \cdot W_{\text{Prototype}} \cdot U_b, \tag{10}$$

where:

(i) W_{new} is the updated long-term weight to be stored in the Long-Term Weights Vector $\overrightarrow{\text{LTW}}$.

(ii) W_{old} is the current value of the long-term weight contained in the Long-Term Weights Vector $\overrightarrow{\text{LTW}}$.

(iii) CV is the computed correlation value of the noun contained in the Prototype Correlation Values Vector $\overrightarrow{\text{PCV}}$.

(iv) $W_{\text{Prototype}}$ is the topic related Adapted TF-IDF Prototype weight contained in the Prototype Weights Vector $\overrightarrow{\text{PW}}$.

(v) U_b is the indicative mean number of articles that the user reads per day. This parameter represents the user's behavior towards the personalization system.

6.2.3. *Updating the skeleton profile*

After the computation of the adapted weights for all the (Adapted TF-IDF) Prototype nouns, the weight of each leaf topic in the topic weights vector $\overrightarrow{\text{WLTopic}}$ can now be updated using the following formula:

$$W_{\text{LT\_new}} = \frac{\sum_{i=1}^{N} W_{\text{ad}}}{N}, \tag{11}$$

where:

(i) N is the number of the prototype nouns corresponding to the topic.

(ii) $W_{\text{ad}} = W_{\text{new\_}i} \cdot (1 \pm (1 - C_p))$, where:

(a) $W_{\text{new\_}i}$ is the long-term weight of each Prototype noun, which has been updated according to Formula (10) and stored in the Long-Term Weights Vector $\overrightarrow{\text{LTW}}$, that is, W_{new}.

(b) C_p is a coefficient used to increase the influence of the nouns, which are common to both the Prototype Nouns Vector $\overrightarrow{\text{PN}}$ and the Long-Term Nouns Vector $\overrightarrow{\text{LTN}}$. This is equal to the percentage of those nouns in the Prototype Nouns Vector $\overrightarrow{\text{PN}}$.

(c) The + or − sign is applied when the correlation value of the Prototype Noun is positive, or negative, respectively.

(iii) For the Prototype nouns that are not contained in the current Long-Term Nouns Vector $\overrightarrow{\text{LTN}}$, as well as for the Prototype nouns having zero correlation values, the following equation holds: $W_{\text{ad}} = W_{\text{new\_}i}$, where $W_{\text{new\_}i}$ has not been updated.

Finally, the adapted weights of the leaf topics are stored in the handset's memory and then transmitted to the server to allow for the high level news content filtering.

When the user browses the non-leaf topics in the hierarchy, the changes in the weights of the leaf topics are propagated to their corresponding supertopics. Thus, the weights of the non-leaf topics are determined using the adapted weights of the leaf topics according to the following formula:

$$W_{T\_\text{new}} = \frac{\sum_{i=1}^{M} W_{T\_\text{new\_}i}}{M}, \tag{12}$$

where:

(i) $W_{T\_\text{new\_}i}$ is the updated weight of each subtopic of the non-leaf topic ($W_{T\_\text{new\_}i}$ corresponds to $W_{\text{LT\_new}}$ if the subtopic is a leaf topic).

(ii) M is the number of subtopics corresponding to the non-leaf topic.

7. EVALUATION OF THE PERSONALIZATION ENGINE

In this section, the experimental results are presented following the evaluation of the personalization engine, which includes the automatic adaptation for both the detailed and the high-level user profile. The evaluation tests concern each of the distinct learning processes performed in the handset, that is, the short-term learning and the long-term learning process. The evaluation experiments were conducted using news content from the Reuters corpus, and collecting data from regular system users. It should be noted that the user is aware of the system's personalization capabilities:

(i) of automatically updating the high-level profile according to her long-term interests (thus she can explicitly alter the adapted symbolic degree of preference according to her choice when she does not approve the system's changes).

(ii) of ranking the headlines of the incoming news items based on the user implicit feedback so she expects that the higher a headline is displayed in the list the more the corresponding news item falls under her interest.

7.1. *User evaluation of short-term learning component*

The evaluation of the short-term learning process is performed in order to demonstrate the overall performance of the short-term learning component. Moreover, the effectiveness of the low-level filtering that results in the ranking of the news items, is shown through this evaluation. Two versions of the short-term learning component have been

compared, namely the complete approach, which uses nouns and Named Entities, and a variant of this approach that uses only nouns. This comparative evaluation aims in demonstrating the contribution of Named Entities in the learning performance. In the variant of the system, Named Entities do not participate neither in the low-level matching process nor in the learning of the detailed user profile, since they do not exist at all in the profile.

For the evaluation of the two learning versions, 500 articles were semantically annotated and their metadata were stored in a news items repository. The user evaluation group consisted of 25 individuals. Each user was asked to manually rank according to his/her preferences a test set of 20 articles (5 articles from each topic) that belong to 4 different leaf topics:

(i) 2 leaf topics chosen by the users belonging to 2 different trees in the hierarchy. A tree is defined as a group of topics sharing the same first-level topic.

(ii) 2 other leaf topics chosen randomly from the 2 remaining trees.

For each user, a set of 100 articles was collected (4 topics of 25 items) that were used in a short-term learning process involving the interaction of the user. During this process, the user receives 4 different sets of 25 articles per day and the system constructs a detailed user profile for the current user, exploiting the user implicit feedback. The created profile is used to automatically rank the initial manually ranked test set of the 20 articles for which the user has provided explicit feedback. The above-described process has been repeated by each user twice, namely once for each variation of the short-term learning system.

In order to evaluate the learning system's performance, the ranking output of the system is compared to the manual ranking of the user. For this purpose one or more performance measures are needed. The standard IR performance measures precision and recall, rely for their calculation on the identification of each retrieved result as either a positive or a negative one. However, in our case only the ranking of the 20 articles for the different users is known; an item that has been ranked, for example, 8th by our algorithm is only known to have been ranked, for example, 10th by a user or a pool of users. Hence it can be clearly identified as relevant (positive) nor irrelevant (negative) to a given subject. Consequently, precision and recall are not the most suitable measures for quantifying the agreement between these two ranked lists. Instead a standard IR metric is used, which measures the correlation between two ranked lists, in our case the manually ranked by the user list and the automatically ranked by the system one. This metric is Spearman's rank correlation coefficient [12]:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}, \tag{13}$$

where d_i represents the difference of each article's ranking between the two lists, and n the number of articles in each list. In our case $n = 20$. Indicative correlation results concerning the two ranked lists for both our short-term learning approaches, that is, the complete short-term learning system and the variant of the system with absence of Named Entities, are depicted in Table 2.

Table 2: Correlation between the two (user's and system's) ranked lists.

User	Correlation of ranked lists using nouns and NEs	Correlation of ranked lists without NEs
1	0.85	−0.09
2	0.67	0.35
3	0.83	0.38
4	0.61	0.22
5	0.79	0.67
...	...	...
25	0.83	0.15
Average	0.70	0.48

Additionally, the percentage error for the position of each of the ranked articles by the system, according to the manual ranking per user, is defined::

$$\text{error} = \left(\frac{\left| N_{\text{Manual Order}} - N_{\text{System Order}} \right|}{\max\left[N_{\text{Manual Order}} - 1, N_{\text{Total Articles}} - N_{\text{Manual Order}}\right]}\right) \cdot 100\%, \tag{14}$$

where $N_{\text{Total Articles}} = 20$. Indicative results along with the average percentage error per user are shown in Table 3 for both our short-term learning approaches, that is, the complete short-term learning system and the variant of the system with absence of Named Entities.

Precision can be applied for measuring the learning system's performance for the N top recommendations of the system, that is, the percentage of the N top ranked articles according to the system ranking, which were manually ranked also within the N top ones. In this case recall is equal with precision. Hence indicative results for the precision of the complete short-term learning system and the variant of the system without Named Entities, for the 10 top recommendations are depicted in Table 4.

The results of the evaluation process through all of the three different metrics seem promising. Additionally, they demonstrate the strong contribution of Named Entities in the high short-term learning performance, since the results of the complete learning component are much better than the ones of the variation without Named Entities. However, they could be further improved if certain limitations are handled, which do not concern the personalization system, but are mostly related to user perception. More specifically, the limitations arising from the user feedback are the following:

(i) The users belong to a specific "social" group and most of them are not familiar with certain topics (particularly the economic related ones).

Table 3: %Error for the exact ranking position of each article in comparison to the manual ranking per user (25 users).

	User	1	2	3	4	5	⋯	25
	1	0%	63.16%	15.79%	15.79%	21.05%	⋯	5.26%
Articles	2	5.56%	22.22%	5.56%	27,28%	0%	⋯	11.11%
ranking	3	0%	41.18%	23.53%	82.35%	0%	⋯	17.65%
position	4	0%	37.50%	12.50%	6.25%	18.75%	⋯	12.50%
	5	6.67%	6.67%	13.33%	6.67%	6.67%	⋯	13.33%
Short-term	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯
Learning	16	6.67%	13.33%	20%	33.33%	13.33%	⋯	26.67%
using	17	37.50%	18.75%	6.25%	18.75%	12.50%	⋯	12.50%
nouns & NEs	18	11.76%	17.65%	11.76%	11.76%	11.76%	⋯	11.76%
(Formula (3), (4))	19	5.56%	0%	44.44%	0%	11.11%	⋯	16.67%
	20	21.05%	21.05%	15.79%	31.58%	47.37%	⋯	10.53%
Average		16.69%	24.79%	20.01%	28.35%	23.10%	⋯	17.23%
	1	10.53%	44.44%	62.50%	26.67%	33.33%	⋯	83.33%
Articles	2	33.33%	27.28%	6.67%	12.50%	34.50%	⋯	67.40%
ranking	3	41.18%	55.70%	39.20%	37.50%	33.33%	⋯	34.50%
position	4	62.50%	22.30%	58.67%	34.50%	22.22%	⋯	17.65%
	5	13.33%	10.53%	12.50%	33.33%	6.67%	⋯	42.80%
Short-term	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯
Learning	16	6.67%	10.53%	39.20%	59.70%	27.28%	⋯	37.50%
using only	17	93.75%	44.44%	27.28%	55.70%	6.67%	⋯	22.22%
nouns	18	100%	12.50%	28.35%	44.44%	26.67%	⋯	17.23%
(Formula (3))	19	83.33%	23.10%	31.58%	13.33%	58.67%	⋯	12.50%
	20	36.84%	83.33%	21.05%	28.35%	10.53%	⋯	39.20%
Average		46.60%	33.33%	29.20%	41.18%	25.75%	⋯	44.44%

Table 4: Precision of the short-term learning system for the 10 top ranked articles.

User	Correctly ranked articles using nouns and NEs	Correctly ranked articles without NEs
1	100%	60%
2	90%	70%
3	90%	60%
4	80%	50%
5	80%	70%
⋯	⋯	⋯
25	90%	50%
Average	83.20%	61.50%

(ii) Some users raised the issue that they do not find the articles consistent with some topics (i.e., different "interpretation" of the topics from the users). Hence some topics are harder to predict than others.

(iii) The users have chosen certain topics, but the corresponding existing articles are not interesting for them, so this has affected their manual ranking to the test articles. Additionally their interaction with the system and consequently the final ranking of the system has been affected.

7.2. Experimental evaluation of long-term learning component

In order to evaluate the long-term learning process, 500 articles were offline annotated and their metadata were stored in a news items repository. The articles constituted 5 long-term sets of 100 articles each, used for evaluation purposes. These sets were constructed using articles that are classified to all the different leaf topics of the hierarchy. The 20 articles in each set belong to a specific leaf topic, which was selected for the evaluation purposes. More specifically, the "Disasters & Accidents" ("GDIS") category was selected in order to observe its weight adaptation, which is induced by the interaction with the personalization system on the news items contained in the long-term sets.

Initially the topic "GDIS" was explicitly denoted with a "Medium" degree of preference. Thus, its initial weight corresponds to the 0.5 value. The following experiments were conducted for evaluating the long-term adaptation (i.e., increase or decrease) of the initial weight during the 5 sets.

(i) All the "GDIS" along with several news items from other topics were selected in each set. An increase of the "GDIS" weight is expected.

(ii) Approximately half of the "GDIS" along with several news items from other topics were selected in each set. A small increase of the "GDIS" weight is expected.

Table 5: Weight adaptation of "GDIS" topic after the completion of 5 long-term evaluation sets concerning the user selections of all, half, and none of the "GDIS" news items contained in these sets.

Long-term evaluation set	All (100%)	Half (50%)	None (0%)
1	0.5623	0.5197	0.4705
2	0.6203	0.5405	0.4325
3	0.6308	0.5654	0.4047
4	0.6762	0.5598	0.3520
5	0.7015	0.5740	0.2982

(iii) None of the "GDIS" but only news items from other topics were selected in each set. A decrease of the "GDIS" weight is expected.

The experiments showed that when the 100% of the "GDIS" articles were selected, there was a constant increase of the weight of the topic after the completion of each set. When the last set was completed, the final weight just exceeded the 0.7 value, that is, the degree of preference changed from "Medium" to "High." In the second case, when half of the "GDIS" articles were selected, there was also a constant increase of the weight, but in a reduced rate, so that the degree of preference did not finally change to "High." Finally, when none of the "GDIS" articles was selected, there was a constant decrease of the weight of the topic until the last set where the weight became lower than the 0.3 value, that is, the degree of preference changed from "Medium" to "Low." In Table 5, the weight adaptation of the topic "GDIS" during the 5 long-term sets is depicted.

As a conclusion, the changing rate of the weight of a particular topic is sufficient in order to change its symbolic degree of preference (e.g., from Medium to High, or from Medium to Low). This happens when the user demonstrates strong interest for this topic, or she keeps ignoring it through her interaction with the personalization system during a large number of news items.

8. RELATED WORK

In this section, related work is described addressing issues raised in this paper, such as distributed personalization architectures, methods for acquiring/adapting user profiles from implicit/explicit feedback and user modeling in the news personalization domain.

In recent years, machine learning techniques have been developed for application to a distributed architecture consisting of a server and a client machine (i.e., a cell phone, or a pocket personal computer). Reference [1] presents a distributed architecture for personalized news access, consisting of a central server, which handles a variety of functions and two clients, a web-based adaptive news service that learns from users' explicit feedback, and another, which is geared towards wireless information devices (i.e., wireless organizers, PDAs, cell phones) and learns by observing the user. The learning process still resides, contrary to our approach, for both clients, on the server. A distributed learning approach in a PDA, which uses a Bayesian classifier for the selection of articles of interest according to the user profile, is presented by [13]. The articles are extracted from web pages and displayed in a zoomable interface-based browser on a PDA. For keeping the profile up to date, the user provides implicit feedback to the system, which monitors her reading behaviors. The [14] approach uses a two step filtering, with a first filter on the server, and a second filter on the device. However, the server filter in that case is often reduced to a simple filtering linked to content sources.

Several systems that have recently been developed for personalized news access, use the explicit or implicit feedback that the user provides for the construction and the updating of the user profile [15]. In the implicit user input, the user has no direct access to the information in the user profile or its construction. The acquisition of the profile as terms, categories or sets of relevant documents must be made implicitly by interpreting user actions on the system such as the number of key clicks in a document, the amount of scrolling through the document, or the amount of time spent reading the document. The types of implicit feedback that can be reliably extracted from observed user behavior in web search are investigated by [16]. Furthermore, [17] explores different approaches for ranking web search results by exploiting user interactions with the search engine. On the other hand, in explicit profile construction, the user has the responsibility to give the required information to the personalization engine for the construction of the user profile representation, normally through a graphical user interface. The acquisition of the profile can be made by asking the user to enter "terms" or "categories" corresponding to her preferences [18, 19], or by applying a supervised learning algorithm on a training set of "documents," which the user regards as relevant. Reference [20] proposes an adaptive personalized web browser, monitors the user's access behavior such as history, bookmarks, content of pages and access logs to model her interests. A user model dealing with an explicit definition provided by the user through a profile editor, and an implicit part maintained by intelligent services is presented by [21]. Explicit feedback provides more accurate estimates of user interest [22], since there are many reasons why a user would spend time on a particular document other than being interested in it, for example, the user decides that she is not interested in a document after the careful analysis of it. In our work, both explicit and implicit user feedback have been exploited. The explicit feedback is being used to ensure that the user profile is being properly initialized, while the implicit feedback provides the means to reduce the user overload by exploiting a combination of different types of metadata, that is, hierarchical topic categories, and low-level terms.

Recently, the advances in the Semantic Web technologies have enabled the representation of user profiles in a variety of ways. Semantic annotation of content with domain concepts, combined with semantic user preferences, enable inferring user preferences for content. Content semantics are typically based on hierarchies (taxonomies) of categories. The majority of the wireless content providers adopt this

type of hierarchical structures. References [23, 24] also use concept hierarchies for user profiles. References [25–27] create a list of concepts of interest, while [23, 28, 29] create a hierarchically-arranged collection of concepts, or ontology. References [10, 30] build user profiles consisting of specific concepts of a hierarchy, which is represented by an ontology. The system automatically monitors the user's browsing habits in requested web pages from search engines. The initial profile is constructed by assigning the visited web pages to specific concepts of a predefined reference hierarchy-ontology. Semantic user preferences often form the basis of user profiles and they may be divided in two categories, namely records of thematic categories indicating user preference for specific categories or classification schemes of content, and records of simple concepts or weighted sets, indicating the level of the user interest for each concept [31–33].

The above-mentioned approaches, which use Semantic Web tools in order to represent user profiles, are closely related to our work with respect to the use of hierarchical long-term user models, and the classification of content to topic categories. However, there are two main limitations compared to our work. First, they focus on the detection only of long-term user interests and second, they do not propose how these methods could be applied on constrained environments such as mobile devices.

Regarding user interests, [34] distinguishes between short-term interests, which are determined by a particular user query and long-term interests, which are determined by the user preferences over a long time period. He argues that longer-term user properties should also be taken into account when a system filters the content to be delivered. The two separate user models, that is, a long-term and a short-term user model are applied in several systems. In [2], a user interest hierarchy is learnt from a set of web pages visited by the user. The higher-level interests (more general), correspond to long-term interests, while the lower-level ones (more specific), correspond to short-term interests. Reference [3] describes a scheme for dynamic learning of user interests from user feedback in an automated information filtering Internet system using a 3-descriptor scheme for the representation of each category of interests in a profile, which also allows learning of long/short-term interests. Reference [9] captures user interests in order to build and update user profiles exploiting low-level features such as keywords, extracted from text using language processing techniques. Generally, the user profiles are adapted using various learning techniques including the exploitation of vector space model [28, 35], genetic algorithms [36], the probabilistic model [25], or clustering [37].

Our research combines aspects from several systems regarding the separation of the user model into short-term and long-term and the user profile learning. The novelty in our approach, apart from the distributed nature of the architecture, is that the learning process for both models is employed exclusively on the client side following the user explicit and implicit feedback. Additionally, the short-term model is not limited to the use of terms as keywords, but it also exploits the semantic information arising from the association of the noun terms with the topic classification of the news articles.

9. CONCLUSIONS

In this paper, a distributed architecture for personalized news content delivery has been presented. It consists of a two-stage semantic matching process, enabling a high-level filtering of available content on the server, followed by matching of detailed user preferences in the handset. This is enhanced with a learning and adaptation process based on explicit and implicit user feedback. The learning process for both the short-term and long-term models takes place in the handset and the adaptation in the long-term model is also transmitted to the server through a client-server synchronization process.

Both user models exploit the semantic annotation of the news content with different types of metadata such as the topic category of the news item, the identified Named Entities and the most significant noun terms according to the classification topic.

The evaluation results of both the short-term and long-term learning processes are very promising for the implementation of the system in a commercial environment, not only because they are consistent with the user expectations, but also because they are achieved with a minimal user overload and taking into account the communication and computational cost.

In the future, another challenge would be to automatically learn topic hierarchies from the textual content rather than use the manually constructed ones, as in the current case. Furthermore, the learning process in the handset could be extended to take into account the contextual information of the user, such as time and location, which are key inputs in the current mobile environments.

ACKNOWLEDGMENT

The work presented in this paper was fully supported by the project "Distributed Knowledge Management for Personalized Content Delivery" funded by Motorola UK Ltd.

REFERENCES

[1] D. Billsus and M. J. Pazzani, "User modeling for adaptive news access," *User Modelling and User-Adapted Interaction*, vol. 10, no. 2-3, pp. 147–180, 2000.

[2] H. R. Kim and P. K. Chan, "Learning implicit user interest hierarchy for context in personalization," in *Proceedings of the 8th International Conference on Intelligent User Interfaces(IUI '03)*, pp. 101–108, ACM Press, Miami, Fla, USA, January 2003.

[3] D. Widyantoro, T. Ioerger, and J. Yen, "An adaptive algorithm for learning changes in user interests," in *Proceedings of the 8th International Conference on Information and Knowledge Management (CIKM '99)*, pp. 405–412, Kansas City, Mo, USA, November 1999.

[4] H. Cunningham, D. Maynard, K. Bontcheva, and V. Tablan, "GATE: a framework and graphical development environment for robust NLP tools and applications," in *Proceedings of the*

40th Anniversary Meeting of the Association for Computational Linguistics (ACL '02), Philadelphia, Pa, USA, July 2002.

[5] H. Cunningham, D. Maynard, K. Bontcheva, et al., "Developing language processing components with GATE (a user guide)," University of Sheffield, Sheffield UK, 2005.

[6] G. Salton, A. Wong, and C. S. Yang, "A vector space model for automatic indexing," *Communications of the ACM*, vol. 18, no. 11, pp. 613–620, 1975.

[7] J. Rocchio, "Relevance feedback in information retrieval," in *The Smart Retrieval System—Experiments in Automatic Document Processing*, G. Salton, Ed., chapter 14, pp. 313–323, Prentice-Hall, Englewood Cliffs, NJ, USA, 1971.

[8] T. Joachims, "A probabilistic analysis of the Rocchio algorithm with TFIDF for text categorization," in *Proceedings of 14th International Conference on Machine Learning (ICML '97)*, pp. 143–151, Nashville, Tenn, USA, July 1997.

[9] E. Bloedorn and I. Mani, "Using NLP for machine learning of user profiles," *Intelligent Data Analysis*, vol. 3, no. 2, pp. 3–18, 1998.

[10] S. Gauch, J. Chaffee, and A. Pretschner, "Ontology-based personalized search and browsing," *Web Intelligence and Agent Systems*, vol. 1, no. 3-4, pp. 219–234, 2003.

[11] J. Cohen, P. Cohen, S. G. West, and L. S. Aiken, *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, Lawrence Erlbaum Associates, Hillsdale, NJ, USA, 3rd edition, 2003.

[12] E. L. Lehmann and H. J. M. D'Abrera, *Nonparametrics: Statistical Methods Based on Ranks*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1998.

[13] R. Carreira, J. M. Crato, D. Gonçalves, and J. A. Jorge, "Evaluating adaptive user profiles for news classification," in *Proceedings of the 9th International Conference on Intelligent User Interfaces (IUI '04)*, pp. 206–212, Madeira, Portugal, January 2004.

[14] W. Matz, "System and method for filtering content," Bellsouth Intellect PTY Corp, Application no. US2003726727A, Filed 20031202 Published 20050602, 2005.

[15] A. Kobsa, "Generic user modeling systems," *ACM User Modeling and User Adapted Interaction*, vol. 2, no. 1-2, pp. 49–63, 2001.

[16] T. Joachims, L. Granka, B. Pang, H. Hembrooke, and G. Gay, "Accurately interpreting clickthrough data as implicit feedback," in *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '05)*, pp. 154–161, Salvador, Brazil, August 2005.

[17] E. Agichtein, E. Brill, and S. Dumais, "Improving web search ranking by incorporating user behavior information," in *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '06)*, pp. 19–26, Seattle, Wash, USA, August 2006.

[18] R. Guha, R. Kumar, P. Raghavan, and A. Tomkins, "Propagation of trust and distrust," in *Proceedings of the 13th International Conference on World Wide Web (WWW '04)*, pp. 403–412, New York, NY, USA, May 2004.

[19] L. Kerschberg, W. Kim, and A. Scime, "WebSifter II: a personalizable meta-search agent based on weighted semantic taxonomy tree," in *Proceedings of the International Conference on Internet Computing (IC '01)*, pp. 14–20, Las Vegas, Nev, USA, June 2001.

[20] P. K. Chan, "Constructing web user profiles: a non-invasive learning approach," in *Proceedings of the International Workshop on Web Usage Analysis and User Profiling (WEBKDD '99)*, pp. 39–55, San Diego, Calif, USA, August 1999.

[21] L. Razmerita, A. Angehrn, and A. Maedche, "Ontology-based user modeling for knowledge management systems," in *Proceedings of 9th International Conference on User Modeling (UM '03)*, pp. 213–217, Springer, Johnstown, Pa, USA, June 2003.

[22] L. M. Quiroga and J. Mostafa, "Empirical evaluation of explicit versus implicit acquisition of user profiles in information filtering systems," in *Proceedings of 4th ACM Conference on Digital Libraries*, vol. 37, pp. 238–239, Berkeley, Calif, USA, August 1999.

[23] T. Kurki, S. Jokela, R. Sulonen, and M. Turpeinen, "Agents in delivering personalized content based on semantic metadata," in *Proceedings of the AAAI Spring Symposium on Intelligent Agents in Cyberspace*, pp. 84–93, Stanford, Calif, USA, March 1999.

[24] Y. Labrou and T. Finin, "Yahoo! as an ontology—using Yahoo! categories to describe documents," in *Proceedings of the 8th International Conference on Information Knowledge Management (CIKM '99)*, pp. 180–187, Kansas City, Mo, USA, November 1999.

[25] D. Mladenic, "Text-learning and related intelligent agents: a survey," *IEEE Intelligent Systems and Their Applications*, vol. 14, no. 4, pp. 44–54, 1999.

[26] M. Pazzani, J. Muramatsu, and D. Billsus, "Syskill & Webert: identifying interesting web sites," in *Proceedings of the 13th National Conference on Artificial Intelligence and Eighth Innovative Applications of Artificial Intelligence Conference (AAAI '96)*, vol. 1, pp. 54–61, Portland, Ore, USA, August 1996.

[27] S. E. Middleton, D. C. De Roure, and N. R. Shadbolt, "Capturing knowledge of user preferences: ontologies in recommender systems," in *Proceedings of the 1st International Conference on Knowledge Capture (K-Cap '01)*, pp. 100–107, Victoria, Canda, October 2001.

[28] A. Pretschner, *Ontology based personalized search*, M.S. thesis, The University of Kansas, Lawrence, Kan, USA, 1999.

[29] F. Tanudjaja and L. Mui, "Persona: a contextualized and personalized web search," in *Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS '02)*, pp. 1232–1240, Big Island, Hawaii, USA, January 2002.

[30] S. Gauch and J. Trajkova, "Improving ontology-based user profiles," in *Proceedings of Computer Assisted Information Retrieval (RIAO '04)*, pp. 380–389, Vaucluse, France, April 2004.

[31] K. Bradley, R. Rafter, and B. Smyth, "Case-based user profiling for content personalisation," in *Proceddings of International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH '00)*, pp. 62–72, Trento, Italy, August 2000.

[32] R. Prestes, G. Carvalho, R. Paes, C. J. P. Lucena, and M. Endler, "Applying ontologies in open mobile systems," in *Proceedings of the Workshop on Building Software for Pervasive Computing (OOPSLA '04)*, Vancouver, Canada, October 2004.

[33] A. Calì, D. Calvanese, S. Colucci, T. Di Noia, and F. M. Donini, "A description logic based approach for matching user profiles," in *Proceedings of the 17th International Workshop on Description Logics (DL '04)*, vol. 104, pp. 196–202, Whistler, Canada, June 2004.

[34] A. Jameson, "Modeling both the context and the user," *Personal and Ubiquitous Computing*, vol. 5, no. 1, pp. 29–33, 2001.

[35] L. Chen and K. Sycara, "Web mate: a personal agent for browsing and searching," in *Proceedings of the 2nd International Conference on Autonomous Agents and Multi Agent Systems*

(AGENTS '98), pp. 132–139, ACM Press, Minneapolis, Minn, USA, May 1998.

[36] M. Mitchell, *An Introduction to Genetic Algorithms*, A Bradford Book, The MIT Press, Cambridge, Mass, USA, 1996.

[37] B. Mobasher, H. Dai, T. Luo, M. Nakagawa, Y. Sun, and J. Wiltshire, "Discovery of aggregate usage profiles for web personalization," in *Proceedings of Web Minning for E-Commerce Workshop, in Conjunction with the ACM-SIGKDD Conference on Knowledge Discovery in Databases (KDD '00)*, Boston, Mass, USA, August 2000.

Hindawi Publishing Corporation
International Journal of Digital Multimedia Broadcasting
Volume 2008, Article ID 835438, 12 pages
doi:10.1155/2008/835438

Research Article

Context-Aware UPnP-AV Services for Adaptive Home Multimedia Systems

Roland Tusch, Michael Jakab, Julius Köpke, Armin Krätschmer, Michael Kropfberger, Sigrid Kuchler, Michael Ofner, Hermann Hellwagner, and Laszlo Böszörmenyi

M3-Systems Research Laboratory, Institute of Information Technology, University of Klagenfurt, 9020 Klagenfurt, Austria

Correspondence should be addressed to Roland Tusch, roland.tusch@m3-systems.com

Received 25 June 2008; Accepted 15 July 2008

Recommended by Harald Kosch

One possibility to provide mobile multimedia in domestic multimedia systems is the use of Universal Plug and Play Audio Visual (UPnP-AV) devices. In a standard UPnP-AV scenario, multimedia content provided by a Media Server device is streamed to Media Renderer devices by the initiation of a Control Point. However, there is no provisioning of context-aware multimedia content customization. This paper presents an enhancement of standard UPnP-AV services for home multimedia environments regarding context awareness. It comes up with context profile definitions, shows how this context information can be queried from the Media Renderers, and illustrates how a Control Point can use this information to tailor a media stream from the Media Server to one or more Media Renderers. Moreover, since a standard Control Point implementation only queries one Media Server at a time, there is no global view on the content of all Media Servers in the UPnP-AV network. This paper also presents an approach of multimedia content integration on the Media Server side that provides fast search for content on the network. Finally, a number of performance measurements show the overhead costs of our enhancements to UPnP-AV in order to achieve the benefits.

1. INTRODUCTION

During the last 25 years, a number of digital audio and video broadcasting standards and systems for large-scale broadcast scenarios have been developed. Some of these standards are *Digital Audio Broadcasting* (DAB) and its successor DAB+ for digital audio transmissions [1], South Korea's DAB-based *Digital Multimedia Broadcasting* (DMB) for broadcasting multimedia data to mobile devices [2], and the complete suite of *Digital Video Broadcasting* (DVB) standards [3, 4]. Similar to DMB, DVB also specified its own transmission system for handheld terminals entitled *DVB-H* [5]. Moreover, the DVB suite also includes a set of Java-based open middle-ware specifications for interactive digital television, called the DVB *Multimedia Home Platform* (DVB-MHP) [6]. DVB-MHP is designed to work across all DVB transmission technologies and requires an additional return channel for each interactive TV application.

However, such broadcast systems are usually not applicable to small-scale environments like single-site home entertainment systems for the following two reasons. First, multimedia broadcasting to mobile devices in a domestic multimedia environment is not economical, since the information coverage area usually is simply too small. Second, in a home multimedia environment, maybe some but not all users are usually interested in the same content at the same time. These reasons result more in the need for multimedia unicasting and multicasting than for multimedia broadcasting in domestic multimedia systems. Therefore, for home multimedia environments, the widely accepted *Universal Plug and Play Audio Visual* (UPnP-AV) [7, 8] standard may be of interest, which is an extension of the original *Universal Plug and Play* (UPnP) [9] standard.

While *UPnP* enables automatic discovery of common devices and services in a local area network, *UPnP-AV* deals with multimedia devices and especially multimedia content. UPnP-AV specifies device and service descriptions for *Media Servers* and *Media Renderers*, which represent multimedia sources and multimedia sinks, respectively. In between these two device classes, a *control point* acts as a dispatcher of multimedia content. Metadata about the available multimedia content is provided to the Control

Point via the Media Server's *Content Directory Service* (CDS). The Control Point queries the *CDS* for the desired content and initiates the playback of the appropriate streams on a Media Renderer, which in turn is responsible for the correct decoding and rendering of the streams.

However, today's *UPnP-AV* implementations have two major drawbacks which make their use difficult in a heterogenous home multimedia environment with several Media Servers and many different Media Renderers. First, standard Control Point implementations only query one Media Server at a time. If there is a larger number of Media Servers in the local area network, there is no global view on the content of all Media Servers in the network. This makes a search for specific content very difficult. Typically, the search is performed by browsing the content directories of all Media Servers. Second, the multimedia content must be consumed by the Media Renderers as provided by the Media Servers. There is no provisioning for customization of the media content to the capabilities of a Media Renderer device. A typical workaround to this problem in most *UPnP-AV* implementations is that if a Media Renderer is not able to render a specific format, the rendering of the stream can not be initiated at the Control Point.

This paper addresses these two drawbacks of today's *UPnP-AV* implementations. In Section 2, the notion of *Usage Context* for customizing multimedia content to different profiles like user and device profiles is introduced. Section 3 presents our *Integrating Media Server* which integrates multimedia metadata from all available Media Servers in the local network. Our extensible *Context-aware Media Renderer* is presented in Section 4. In Section 5, the internal behavior of our *Context-aware Control Point* is described by an example of a control and data flow. Section 6 comes up with performance evaluations of the *Integrating Media Server* and *Context-aware Media Renderer* implementations. Finally, Section 7 concludes the contribution of this paper to context-aware provisioning of mobile multimedia in domestic multimedia systems.

2. THE NOTION OF USAGE CONTEXT

Customization of multimedia content in multicasting or broadcasting systems is not an easy task, since multicasting implies that the delivered data is to be consumed by a number of consumers simultaneously, whereas personalization is rather a powerful content adaptation method for unicast content delivery scenarios. The usual goal of personalization is to deliver a customized version of multimedia content for exactly one consumer. However, there are already approaches of multimedia personalization in large-scale environments like 3DTV and terrestrial DMB (T-DMB), which are mainly based on multiview video and multichannel audio broadcasting techniques [10, 11]. Currently, these approaches are limited to adapting the content according to specific user profiles (i.e., the language, age, or interests of the consumers).

However, in mobile multimedia systems, customization is not limited to the user profile only. besides, the user profile, there are several other profiles which may also require an adaptation of the multimedia content. Especially in environments with a number of different mobile devices, additional constraints to the content delivery are imposed, for example, by the *terminal* and *location profiles* [12]. The terminal capabilities and the current location of the consumer device play an important role in distributed mobile multimedia systems.

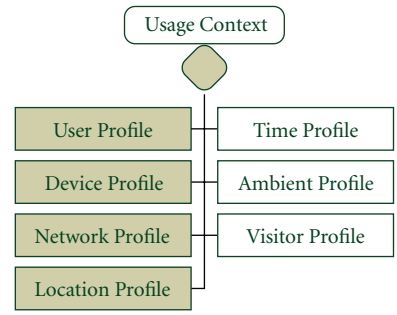


Figure 1: General usage context model for our context-aware multimedia services.

To cope with all relevant profiles for multimedia content tailoring, we introduced the notion of *Usage Context*. The following two subsections provide an overview of our usage context profiles, and of the three possibilities to add context awareness to *UPnP-AV*.

2.1. Usage context profiles

Basically, a universal set of context profiles that is valid for all application areas does not exist, since context is always an issue of the interaction between a user and an application [13]. For example, in a small-scale *UPnP-AV*-based home multimedia environment, it may not be relevant which content the user consumed at what time. However, in a large-scale multimedia tour guide environment as described in [14] this information is definitively of interest to the system provider, since the content consumption history affects possible content demand in the future.

In general, there are three aspects of context which are valid for all application areas: (i) where you are, (ii) who you are with, and (iii) what resources are nearby [15]. For multimedia applications, these aspects have already been addressed in the *MPEG-21 Digital Item Adaptation* standard by the means of *Usage Environment Descriptions* [16–18]. For ubiquitous mobile devices like mobile phones, the *User Agent Profile* (UAProf) [19] has been developed as a de facto standard for describing the resource aspect (i.e., the Device Profile). Moreover, UAProf also describes the preferences aspect, since it is based on the *Composite Capability/Preference Profiles* (CC/PP) [20] vocabulary extension of the *Resource Description Framework* (RDF) [21].

However, neither the MPEG-21 DIA's Usage Environment nor the CC/PP-based UAProf profiles contain sufficient information regarding the context profiles needed by our context-aware application domains, including the context-aware UPnP-AV services and the context-aware large-scale tour guide application [14]. Thus, we derived an own

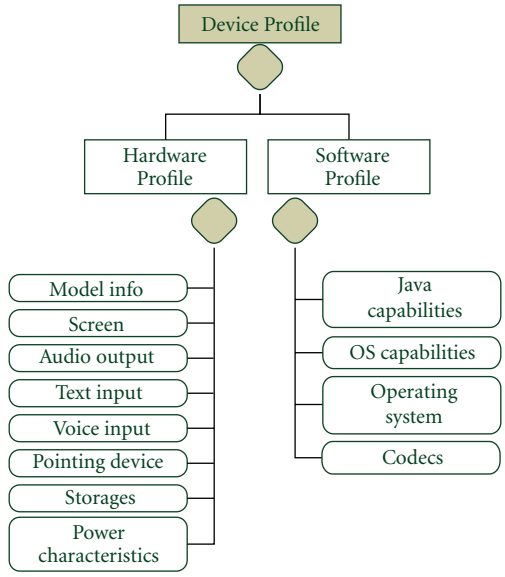


Figure 2: The Device Profile as a composition of Hardware and Software profiles.

context model for our context-aware multimedia services. An overview of this *XML Schema*-based model is given in Figure 1.

Basically, this context model includes the four profiles *User*, *Device*, *Network*, and *Location Profile*. These profiles are required for all of our application domains and represent the four profiles needed by the context-aware UPnP-AV services. The *User Profile* collects data about the characteristics of the user like preferred language, age, interests, and presentation preferences (like audio only and video only for handicapped persons, or mixed). The *Device Profile* delivers information about the hardware and software properties of the consumer device (see Figure 2). For context-aware UPnP-AV services, the *Device Profile* is one of the most important profiles for content adaptation.

The *Network Profile* gathers data about available networks including average bit rate, reliability, latency, and transmission costs. And the *Location Profile* provides information about the current location of the consumer device, according to the used localization technique [22]. For context-aware UPnP-AV services in domestic multimedia environments, localization techniques using Bluetooth and/or Radio Frequency ID (RFID) technology are suitable.

In addition to these four basic profiles, three further profiles (*Time*, *Ambient*, and *Visitor Profiles*) are required for our large-scale multimedia tour guide application. They are not used by the context-aware UPnP-AV services, but for the sake of application-domain comparison they are mentioned here. The *Time Profile* tracks the current system time of the content consumption. Current environmental conditions including weather, temperature, and air conditions are collected by the *Ambient Profile*. And finally, the *Visitor Profile* tracks the content consumption history of a tourist, as well as the data about her/his vacation, such as location, date of arrival, date of departure, and number of persons.

2.2. Adding context awareness to UPnP-AV

As mentioned in Section 1, one major drawback of the UPnP-AV specification is that the multimedia content must be consumed by the Media Renderers as provided by the Media Servers. The only adaptation step existing Control Point implementations typically perform is to avoid the rendering of a Media Server's media stream on a Media Renderer if the renderer is not able to deal with the coding format of the stream or if the renderer does not support any provided transport protocol of the Media Server.

To overcome this drawback, two improvements can be added to UPnP-AV services. First, the Media Renderer can be extended with means for querying the current usage context (see previous Section 2.1) on the renderer device. Second, the Control Point may incorporate a content transcoding application, which adapts a Media Server's media stream to a given usage context, before the content is delivered to the Media Renderer.

Basically, there are three possibilities to enhance a standard UPnP-AV Media Renderer with context provisioning [23]. First, a selected service of the Media Renderer can be extended by an additional action which returns the current context information. An *enhanced* Control Point may then call this action to acquire the desired information. Second, a selected service of the Media Renderer can be extended with a set of state variables describing the context. A Control Point may query these variables to acquire the context via UPnP-AV's eventing mechanism. And third, the Media Renderer can be extended with a new service which encapsulates the aforementioned variables, provides actions to query them and offers eventing for the notification of value changes to these variables.

All three possibilities have their advantages and disadvantages. Our decision to extend the Media Renderer's *Connection Manager* service by an additional *GetContextInfo* action is described in Section 4. The Control Point's usage of the context data for customizing a media stream is presented in Section 5.

3. THE INTEGRATING MEDIA SERVER

To overcome the first drawback of standard UPnP-AV mentioned in Section 1, an *Integrating UPnP-AV Media Server* was implemented [24]. Figure 3 illustrates the steps undertaken to integrate multimedia metadata from *n* UPnP-AV Media Servers in the network.

When a Media Server starts up, it first fills its own *Content Directory* with metadata obtained from locally available multimedia sources (i.e., files from the file system or live sources). The gathered metadata is then provided to a Control Point via the Media Server's *Content Directory Service* (CDS). In the next step, the *Integrating Media Server* browses the CDS of each detected Media Server, and integrates its metadata into its own CDS. In order to be able to connect to other Media Servers, the Integrating Media Server implements its own Control Point, which listens to arrivals and departures of Media Servers in the UPnP-AV network, and performs metadata integration or segregation steps, respectively.

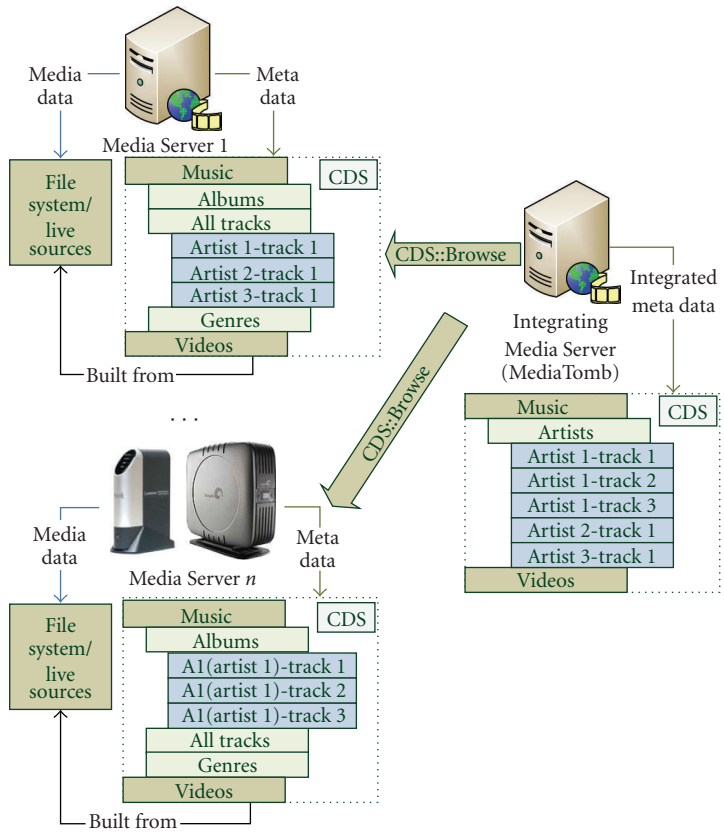


Figure 3: Metadata integration from UPnP-AV Media Servers.

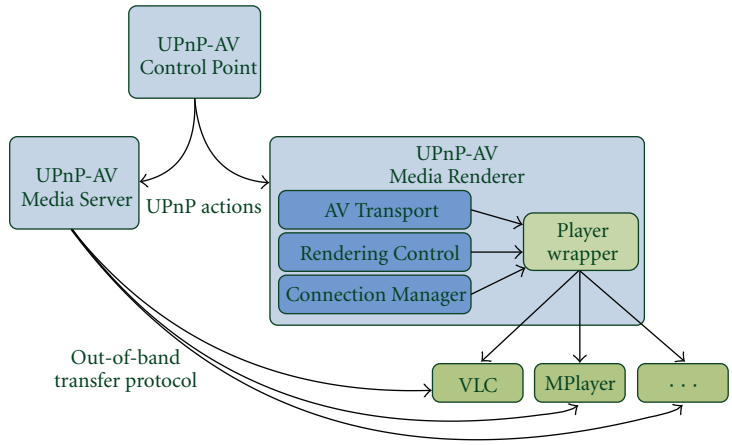


Figure 4: Conceptual view on the generic UPnP-AV Media Renderer architecture.

Integrating metadata views from a remote Media Server does not simply mean an exact import of them to the local CDS. While the latter approach is also known as *metadata mirroring* [25], *metadata integration* enhances metadata mirroring by reorganizing the mirrored metadata into one unified view [24]. For example, in Figure 3 the metadata view *All Tracks* on Media Server 1 and the metadata view *Albums* on Media Server n are integrated to a metadata view *Artists* on the Integrating Media Server.

The multimedia data itself remains on the origin Media Servers and is referenced as a resource in *DIDL-Lite*-based media item descriptions. DIDL-Lite [26] is a subset of MPEG-21's *Digital Item Declaration Language* (DIDL) [27] used in UPnP-AV. It is also based on XML as DIDL, but its schema restricts the possible metadata fields to the UPnP and *Dublin Core* [28] namespaces. Algorithm 1 shows an example DIDL-Lite response of Media Server 1 to a *Browse* action call of the Integrating Media Server's Control Point.

```
<DIDL-Lite>
  <container id="100" parentID="10"
      childCount="3" restricted="1">
    <dc:title>All Tracks</dc:title>
    <upnp:class>
      object.container.musicContainer
    </upnp:class>
  </container>
  <item id="101" parentID="100" restricted="1">
    <dc:title>Dancing Queen</dc:title>
    <upnp:artist>Abba</upnp:artist>
    <upnp:album>Arrival</upnp:album>
    <upnp:genre>Pop</upnp:genre>
    <res size="3482846" duration="0:03:52.110"
      protocolInfo="http-get:*:audio/mpeg:*" >
      http://192.168.1.5:9001/disk/101.mp3</res>
    <upnp:class>
      object.item.audioItem.musicTrack
    </upnp:class>
  </item>
</DIDL-Lite>
```

Algorithm 1: An example media item description.

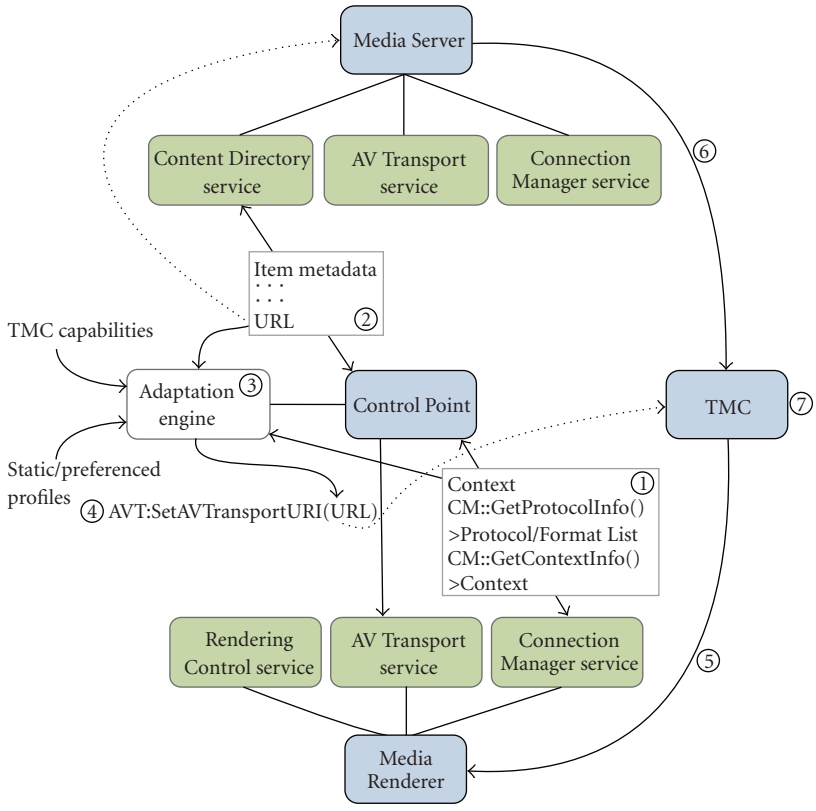


Figure 5: Action call sequence of the Context-aware UPnP-AV Control Point.

The *res* element provides information about the media item's encoding properties, as well as the URL to use for requesting the media item.

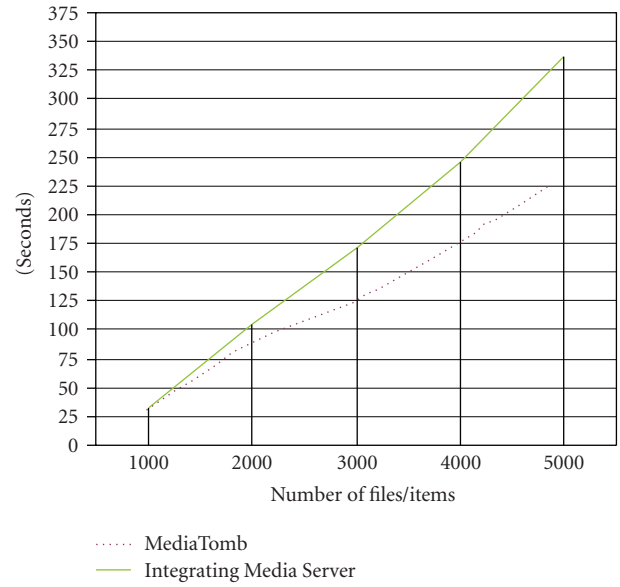
For the implementation of the Integrating Media Server, the open source UPnP-AV Media Server *MediaTomb* [29] and the open source UPnP-AV stack *libupnp* [30] were used.

The presented approach of metadata integration brings two major advancements for Control Points compared to standard UPnP-AV. First, the Integrating Media Server can easily implement the optional *Search* action to enable a Control Point to query the integrated view of metadata for certain multimedia content, even though some Media Servers do not implement this action themselves. And second, the integration step allows to reorganize the metadata in customizable views to customize the multimedia content provisioning according to the profiles of the Usage Context.

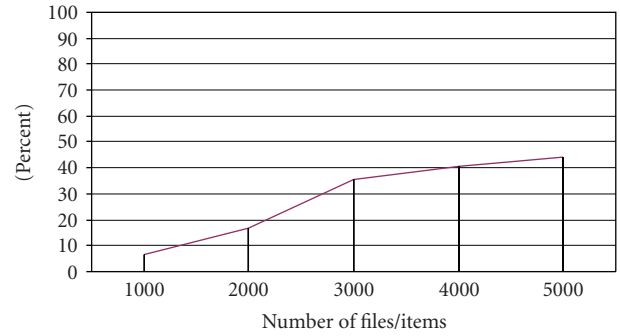
4. THE CONTEXT-AWARE MEDIA RENDERER

The architecture of our context-aware UPnP-AV Media Renderer is kept generic in order to be able to use any existing UPnP or non-UPnP enabled media player as media renderer [31]. Figure 4 provides a conceptual view on this architecture. Existing media players like the *MPlayer* [32] or the *VLC* [33] can be used by our Media Renderer to playback media streams from a UPnP-AV Media Server. While the VLC itself is already UPnP-AV enabled, the standard MPlayer currently does not have support for UPnP-AV. Neither of them is context-aware by default. Our generic Media Renderer architecture allows to integrate the binary version of any supported media player and hence turns it into UPnP-AV enabled. This is achieved by a *Player Wrapper* inside the Media Renderer, which delegates selected action requests to the Media Renderer's *AV Transport*, *Rendering Control*, and *Connection Manager* services to a concrete player instance. In our prototype implementation of the Media Renderer, the MPlayer has been chosen as media player, and an *MPlayer Wrapper* delegates all requests to a running MPlayer instance. The wrapper is also responsible for returning all results of the MPlayer instance to the Media Renderer's calling services.

The ability to publish its context is added to the Media Renderer by extending its *Connection Manager* service with an additional *GetContextInfo* action. This approach has the advantages that (i) it is the responsibility of the Control Point to obtain the context from the Media Renderer, and (ii) the Control Point is also able to control the additional traffic overhead generated by context data. If context awareness was instead realized by eventing, the additional generated network load would have depended on the change frequency of the evented context properties. Considering very dynamic context properties such as *Storages* of the *Device Profile*, UPnP-AV's eventing mechanism may generate a large number of events during a Media Renderer session, since the available memory in the local storage subsystem may often change during a session. Although UPnP-AV provides the concept of *deferred eventing* by the use of thresholds, the eventing behavior is not desirable in many cases, since the transmission of each event generates considerable additional load on the network. On the other hand, the *pull-based* approach to receiving context information from the Media Renderer has the drawback that the Control Point has to periodically query the *GetContextInfo* action. But comparing the advantages and the disadvantages of both approaches, using an additional action is the right decision. Section 6.2.2 provides some performance figures about the costs of calling this action on two different renderer devices.



(a) Build times from file system versus integration times



(b) Integration overhead

Figure 6: Overhead of integrating media items.

5. THE CONTEXT-AWARE CONTROL POINT

Besides the Control Point implementation for metadata integration in the Integrating Media Server presented in Section 3, we also developed a full implementation of a *Context-aware Control Point* [23]. This Control Point operates as dispatcher of media streams from a Media Server to available Context-aware Media Renderers. Figure 5 illustrates an action call sequence for initiating the rendering of a media stream on a Context-aware Media Renderer.

First, the Control Point queries the Usage Context from the Connection Manager service as soon as it connects to a new Media Renderer. In step two, it queries the metadata

Table 1: Test stream variations.

Variation	Resolution	Video bit rate	Audio bit rate	Total bit rate
V_1	320 × 240	200 kbps	64 kbps	264 kbps
V_2	352 × 208	464 kbps	64 kbps	528 kbps
V_3	352 × 208	500 kbps	64 kbps	564 kbps
V_4	800 × 480	200 kbps	64 kbps	264 kbps

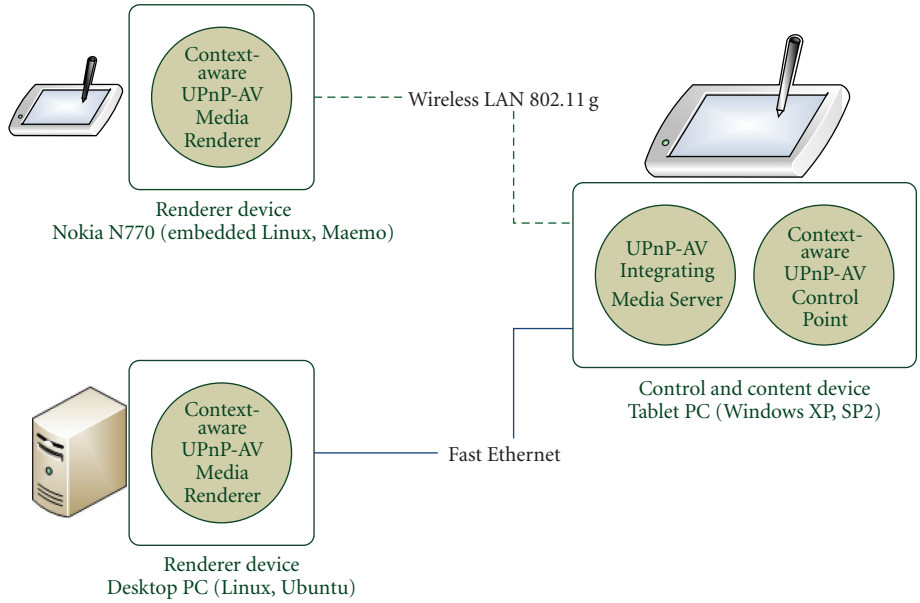


Figure 7: UPnP-AV Media Renderer performance test setup.

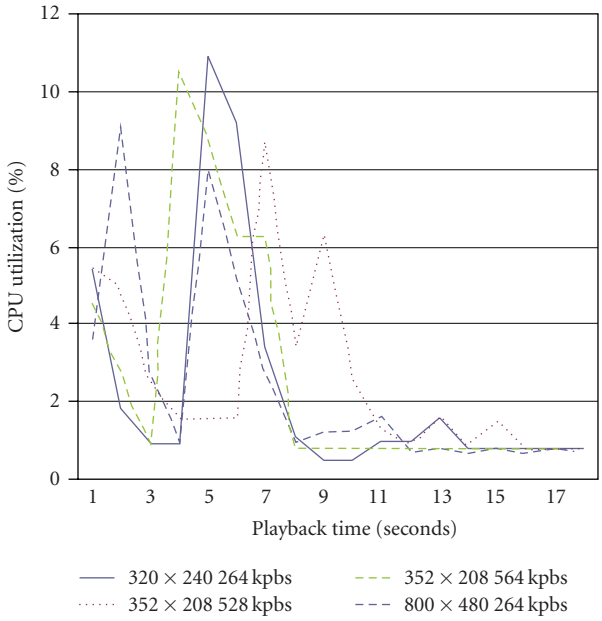
item for the selected media item from the Integrating Media Server. At this point, it has to be mentioned that our Context-aware Control Point provides a Web-based user interface for browsing and searching content. In the example of Figure 5, the Media Renderer selects a media item via this interface. Thus, this example illustrates a pull-based unicast scenario. However, a push-based multicast scenario can also be realized. In the third step, an internal *Adaptation Engine* of the Control Point is used to initiate an adaptation of the media item according to the given Usage Context. This is achieved by incorporating our transcoding service *Transcoding Media Cache* (TMC), which is able to transcode and compose multimedia streams from various input to various output formats [34]. The transcoding of the media item does not take place at this step, it is only initiated by the Adaptation Engine by rewriting the URL in the item's metadata to point to the TMC, and subsequently calling the action *SetAVTransportURI* of the Media Renderers *AV Transport* service with this rewritten URL as parameter (step four). In step five, the Context-aware Media Renderer requests the media item from the TMC, which in turn fetches the original media item from the corresponding Media Server (step six), and transcodes the media item according to the transcoding parameters set by the Adaptation Engine (step seven). Finally, the adapted stream is sent to the Media Renderer by the use of an out-of-band (i.e., non-UPnP) transport protocol. Steps six and seven can be omitted if the TMC already has a cached version of the requested stream for the given Usage Context.

6. PERFORMANCE EVALUATION

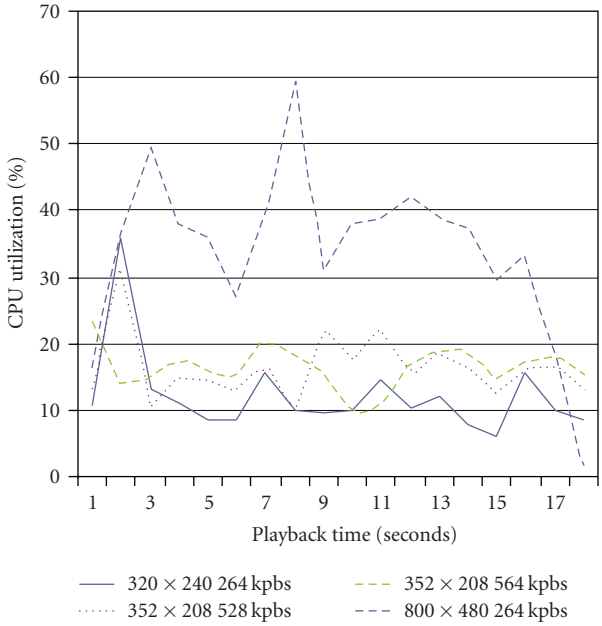
The following two subsections evaluate the performance overhead costs of our approach to extend the UPnP-AV Media Server *MediaTomb* with media item integration capabilities and to implement a UPnP-AV Media Renderer by the use of the non-UPnP-AV-enabled third party media player *MPlayer*, respectively.

6.1. Integrating Media Server

Figures 6(a) and 6(b) illustrate the overhead of the integration of media items in the UPnP-AV Integrating Media Server, compared to the construction of the content directory from a file system as performed by the *MediaTomb* Media Server. The comparison is based on five data sets containing 1000, 2000, 3000, 4000, and 5000 media items from a remote MediaTomb Media Server and the same number of files from a local file system, respectively [24]. As test items, common media items and files from music albums were used. The integration times were measured in an insulated Fast Ethernet LAN.

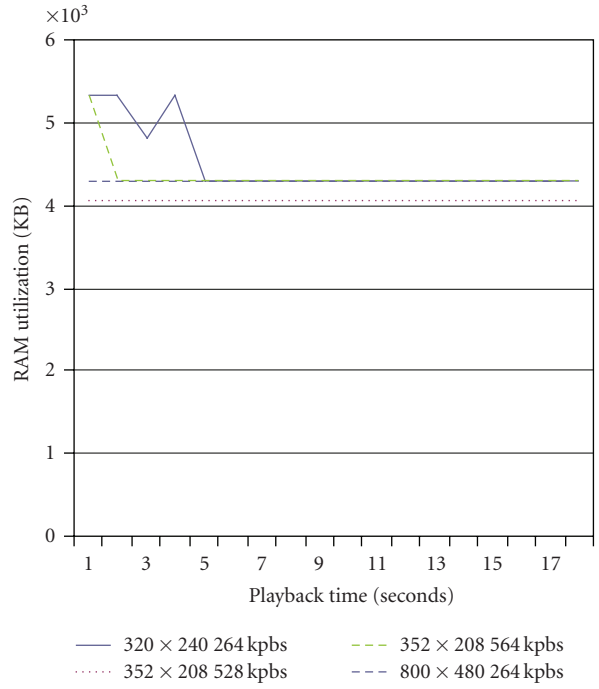


(a) Renderer

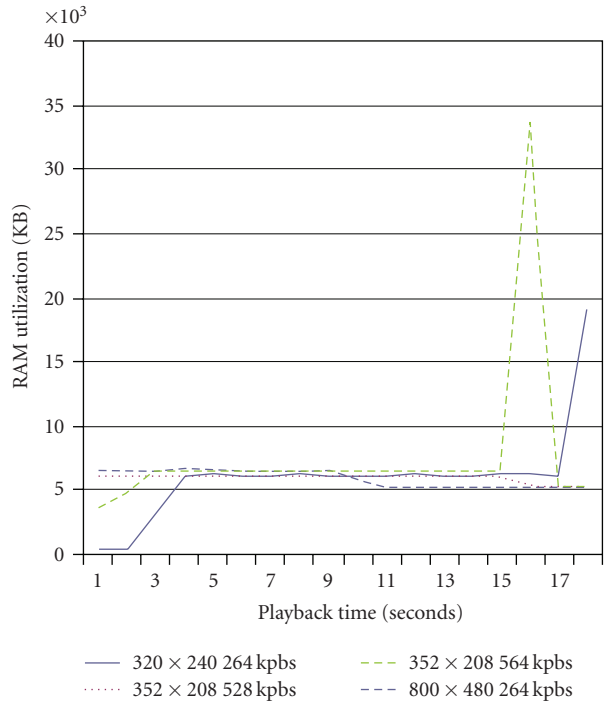


(b) MPlayer

Figure 8: CPU utilization on the Nokia N770.



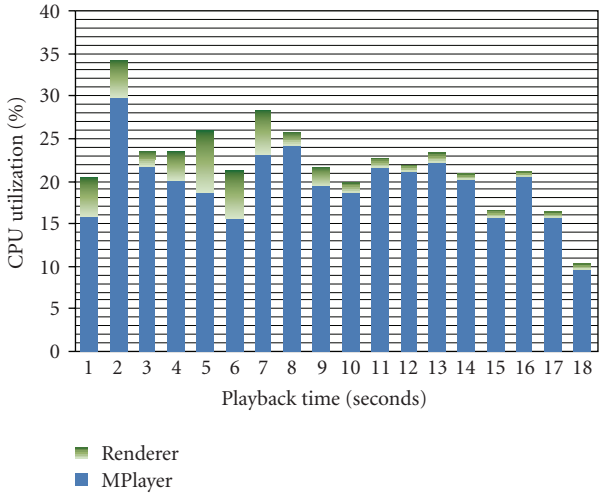
(a) Renderer



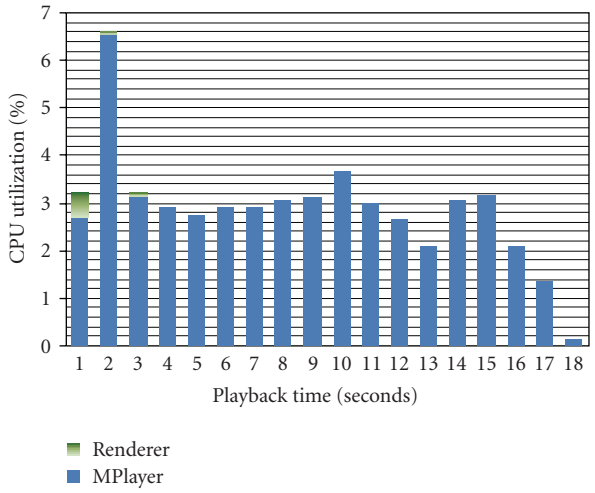
(b) MPlayer

Figure 9: RAM utilization on the Nokia N770.

The integration overhead steadily increases with the number of media items to retrieve from the remote MediaTomb Media Server. While the integration overhead for 1000 items is rather low with about 7%, it increases to about 44% for 5000 items. Although the increase is not strictly linear, it shows a linear gradient of 9.2% for 1000 additional media items on average. This overhead increase is mainly due to the protocol overhead imposed by UPnP-AV, accompanied by the continuous UPnP-AV *Browse* actions, which have to be called on the remote MediaTomb Media Server's *CDS* to query for all available remote media items. This integration overhead is acceptable for an adaptive domestic multimedia system, where the number of media items is seldom higher than 10000 media items (causing an integration overhead of at least 92%) and the frequency of integration activities is rather low.

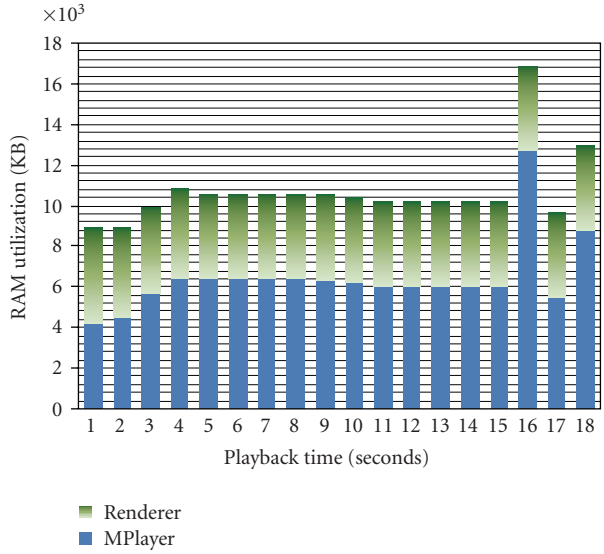


(a) Nokia 770

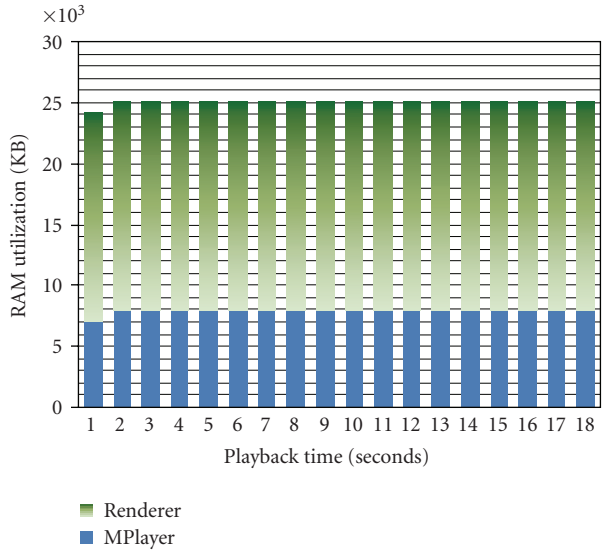


(b) Desktop PC

Figure 10: CPU utilization overhead.



(a) Nokia 770



(b) Desktop PC

Figure 11: RAM utilization overhead.

6.2. Context-aware Media Renderer

6.2.1. CPU and RAM utilization overhead

The CPU and RAM utilization overhead of the context-aware UPnP AV Media Renderer wrapping the MPlayer as player instance was evaluated using the test setup depicted in Figure 7.

On the right side, a Tablet PC running Windows XP (SP2) is used as control and content device. This device runs one instance of the Integrating UPnP-AV Media Server and one instance of the Context-aware UPnP-AV Control Point. In this test case, the media server does not integrate any content from other media servers. Instead, its *CDS* simply offers one system stream (i.e., a composed stream of one video and one audio elementary stream) in four variations regarding bit rate and resolution. Table 1 illustrates the properties of these four stream variations. Note that even though V_4 has the highest resolution, it has the lowest bit rate (actually equal to V_1). Video elementary streams are encoded as MPEG-4 SimpleProfile@Level3, audio streams are commonly encoded as MPEG-1@Layer3. All stream variations have a duration of 18 seconds in playback time. The context awareness of the Control Point is not used in this test case either, since the video variations are already prepared and available to the *CDS* of the Media Server.

On the left side of Figure 7, two renderer devices are used for evaluating the performance of the Context-aware UPnP-AV Media Renderer. The first renderer device is a Nokia N770 Internet Tablet running the Internet Tablet OS 2006 (Maemo 2.2 [35]). It ships with a 252 MHz *Texas Instruments* CPU, 64 MB RAM, 128 MB Flash memory, and a widescreen display with a maximum video resolution of 800 × 480 pixels. The second renderer device is a common Desktop PC running Ubuntu Linux 6.1, with an

Intel Pentium 4 2.53 GHz processor and 128 MB RDRAM installed, providing a maximum video resolution of 1920 × 1200 pixels.

Before discussing the CPU and RAM utilization overhead of the Context-aware Media Renderer, the CPU and RAM utilization of both the Media Renderer and the MPlayer on the Nokia N770 device are illustrated in Figures 8(a)-8(b) and 9(a)-9(b), respectively. Figures 8(a) and 9(a) confirm our expectation that the different stream variations do not have considerable impacts on the CPU and RAM utilization of the Media Renderer, since the Media Renderer is only the *UPnP-AV wrapper* of the MPlayer and hence does not directly operate on the media streams. In contrary, Figures 8(b) and 9(b) clearly show the impacts of the different stream variations on the CPU and RAM utilization of the MPlayer, respectively. While the stream variation with the highest video resolution (V_4) generates the highest load on the CPU, the stream with the highest bit rate (V_3) shows the highest RAM usage peak. Whereas the latter result is not surprising, the former is a bit unexpected since the CPU load of stream variation V_4 is about two times higher than those of the other variations, although it has the smallest bit rate. This is due to the higher computational requirements for larger video resolutions especially during the double buffering and bit-blitting operations.

The CPU and RAM utilization overhead of the Context-aware Media Renderer on both renderer devices is illustrated in Figures 10(a)-10(b) and 11(a)-11(b), respectively. The overhead is calculated on averaged values of CPU and RAM utilizations of all stream variations. Figure 10(a) shows a mean CPU utilization overhead of the Media Renderer of 12.5% on the Nokia N770 device, compared to the average CPU load generated by the MPlayer. On the Desktop PC, this overhead accounts for only 1.3%, as shown in Figure 10(b). Interestingly, the RAM utilization overhead shows a diametrical result. While the RAM utilization overhead of the Media Renderer on the Nokia N770 results in a mean value of 66.8% (see Figure 11(a)), the overhead on the Desktop PC accounts for 220%, as depicted in Figure 11(b). The latter result is due to the used *libupnp* library, which uses more dynamically linked libraries on the Maemo platform.

6.2.2. Response times to context queries

Figure 12 illustrates the response times (in millisecond) of calls to the *GetContextInfo* action in a run of 20 subsequent measurements on both, the Nokia N770 and the Desktop PC renderer devices. It is obvious that the execution of this action is much more expensive than other UPnP-AV actions like *SetAVTransportURI* or *Play*, which take about 10 milliseconds on average. This is due to the dynamic collection of context information each time this action is called. On the Desktop PC the initial call to this action results in a response time of about 1.8 seconds. This is the time the renderer needs on this device when all context information is queried by either directly invoking system calls or by executing shell scripts. However, since some context information is static and does not change during the whole life cycle of the Renderer (like the manufacturer and the device model of the Device Profile), this static context information is only queried once and cached for later action calls. The caching of static context information results in a significant response time reduction of about 80% to about 350 milliseconds.

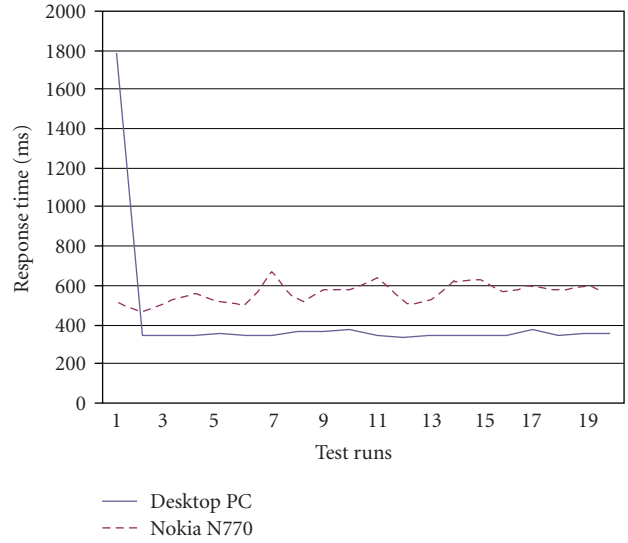


Figure 12: Response times to *GetContextInfo* action calls.

In comparison to the Desktop PC, the Renderer on the Nokia N770 device does not show this kind of *slow start*. Although equipped with a much less powerful CPU and I/O subsystem, the Renderer on the Nokia N770 starts faster than that on the Desktop PC. This is due to the statical assignment of some context properties like screen resolution, text and voice input capabilities in the program code on the Nokia N770. This, of course, results in a much faster collection of the required context data.

7. CONCLUSIONS

This paper presented a novel approach of realizing context-aware multimedia services for domestic multimedia systems by using the *Universal Plug and Play Audio Visual* (UPnP-AV) technology. Since UPnP-AV is designed for local area networks, it is also suitable for multimedia multicasting and broadcasting scenarios. However, standard UPnP-AV does not provide means for tailoring multimedia content to different context properties like the user, device, or network profile. To overcome this drawback, an extension to the Media Renderer has been realized which enables the Control Point to periodically query context information from the Media Renderers. This *Usage Context* information in turn is used by the Control Point to customize media streams from the Media Server by the use of a Transcoding Media Cache (TMC). This approach allows to optimize the multimedia content to the needs of the user with respect to the constraints of her/his usage environment.

The second enhancement of this contribution to standard UPnP-AV is the development of an Integrating Media Server which integrates media items from all other Media

Servers available on the local area network. This integration step provides a global view on all available multimedia content in the UPnP-AV network and allows the Control Point to perform fast queries on the available content. Finally, performance evaluations regarding the overhead of the integration step in the Integrating Media Server, as well as CPU and RAM utilization overheads of the Media Renderer implementation have shown that the overhead costs for achieving the benefits are rather low.

ACKNOWLEDGMENT

This work was supported by the Austrian Science Fund (FWF) under project L92-N13 (CAMUS: Context-Aware Multimedia Services).

REFERENCES

[1] European Telecommunications Standards Institute, "Digital Audio Broadcasting (DAB); Guide to DAB standards; Guidelines and Bibliography," ETSI TR 101 495, January 2005.

[2] European Telecommunications Standards Institute, "Radio Broadcasting Systems; Digital Audio Broadcasting (DAB) to Mobile, Portable and Fixed Receivers," ETSI EN 300 401, April 2000.

[3] European Telecommunications Standards Institute, "Digital Video Broadcasting (DVB); A Guideline for the Use of DVB Specifications and Standards," ETSI TR 101 200, September 1997.

[4] European Telecommunications Standards Institute, "Digital Video Broadcasting (DVB); Framing Structure, Channel Coding and Modulation for Digital Terrestrial Television," ETSI EN 300 744, June 2004.

[5] European Telecommunications Standards Institute, "Digital Video Broadcasting (DVB); Transmission System for Handheld Terminals (DVB-H)," ETSI EN 302 304, June 2004.

[6] European Telecommunications Standards Institute, "Digital Video Broadcasting (DVB); Multimedia Home Platform (MHP) Specification 1.1.1," ETSI TS 102 812 V1.2.2, August 2006.

[7] UPnP Implementers Corporation, UPnP AV Architecture:0.83. White Paper, June 2002, http://www.upnp.org/standardizeddcps/documents/UPnPAvArchtiecture0.83.pdf.

[8] Intel R&D, "Overview of UPnP AV Architecture: A Digital Media Distribution Technology for the Home," White Paper, July 2003, http://cache-www.intel.com/cd/00/00/21/87/218764_218764.pdf.

[9] UPnP Implementers Corporation, UPnP Device Architecture 1.0. White Paper, 2006, http://www.upnp-ic.org/resources/UPnP_device_architecture_docs/UPnP-DeviceArchitecture-v1_0-20060720.pdf.

[10] K.-J. Oh, M. Kim, J. S. Yoon, et al., "Multi-view video and multi-channel audio broadcasting system," in *Proceedings of the 3DTV Conference (3DTV-CON '07)*, pp. 1–4, Kos Island, Greece, May 2007.

[11] T. Lee, Y. J. Lee, J. H. Yoo, and D. Jang, "Personalized audio broadcasting system through the terrestrial-DMB system," in *Proceedings of the International Conference on Consumer Electronics (ICCE '07)*, pp. 1–2, Las Vegas, Nev, USA, January 2007.

[12] S. Panagiotakis and A. Alonistioti, "Context-aware composition of mobile services," *IT Professional*, vol. 8, no. 4, pp. 38–43, 2006.

[13] A. K. Dey, "Understanding and using context," *Personal Ubiquitous Computing*, vol. 5, no. 1, pp. 4–7, 2001.

[14] J. Köpke, R. Tusch, H. Hellwagner, and L. Böszörmenyi, "Context-aware hoarding of multimedia content in a large-scale tour guide scenario: a case study on scaling issues of a multimedia tour guide," in *Proceedings of the International Conference on Signal Processing and Multimedia Applications (SIGMAP '08)*, Porto, Portugal, July 2008.

[15] B. Schilit, N. Adams, and R. Want, "Context-aware computing applications," in *Proceedings of the Workshop on Mobile Computing Systems and Applications*, pp. 85–90, Santa Cruz, Calif, USA, December 1994.

[16] International Organization for Standardisation, "Information technology—multimedia framework (MPEG-21)—part 7: digital item adaptation," Tech. Rep. ISO/IEC 21000-7, ISO, 2004.

[17] A. Vetro and C. Timmerer, "Digital item adaptation: overview of standardization and research activities," *IEEE Transactions on Multimedia*, vol. 7, no. 3, pp. 418–426, 2005.

[18] A. Vetro, C. Timmerer, and S. Devillers, "Digital item adapation—tools for universal multimedia access," in *The MPEG-21 Book*, chapter 7, pp. 243–280, John Wiley & Sons, New York, NY, USA, 2006.

[19] OpenMobile Alliance, User Agent Profile - Approved Version 2.0, OMA-TS-UAProf-V2_0-20060206-A, February 2006.

[20] World Wide Web Consortium, "Composite Capability/Preference Profiles (CC/PP): Structure and Vocabularies 1.0. W3C Recommendation," 2004, http://www.w3.org/TR/CCPP-struct-vocab/.

[21] World Wide Web Consortium, "Resource Description Framework (RDF): Concepts and Abstract Syntax," W3C Recommendation, 2004, http://www.w3.org/TR/rdf-concepts/.

[22] M. Santner, R. Tusch, M. Kropfberger, L. Böszörmenyi, and H. Hellwagner, "Ein Ortserkennungssystem für mobile Touristenführer," in *Proceedings of the DACH Mobility*, pp. 84–98, Ottobrunn, Germany, October 2006.

[23] M. Ofner, *Design and implementation of a context-aware UPnP-AV control point*, M.S. thesis, Institute of Information Technology, University of Klagenfurt, Klagenfurt, Austria, 2008.

[24] M. Jakab, M. Kropfberger, M. Ofner, R. Tusch, H. Hellwagner, and L. Böszörmenyi, "Metadata integration and media transcoding in universal-plug-and-play (UPnP) enabled networks," in *Proceedings of the 15th Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP '07)*, pp. 363–372, Naples, Italy, February 2007.

[25] Intel R&D, "Designing a UPnP-AV MediaServer," White Paper, July 2003, http://cache-www.intel.com/cd/00/00/21/87/218762_218762.pdf.

[26] UPnP Implementers Corporation, "ContentDirectory:1 Service Template Version 1.01," White Paper, June 2002, http://www.upnp.org/standardizeddcps/documents/ContentDirectory1.0.pdf.

[27] International Organization for Standardisation, "MPEG-21 part 2: digital item declaration language (DIDL)," Technology Report ISO/IEC 21000-2, ISO, 2003.

[28] Dublin Core Metadata Initiative, "Dublin Core Metadata Element Set, Version 1.1," DCMI Recommendation, June 2008, http://dublincore.org/schemas/xmls/qdc/2008/02/11/dc.xsd.

[29] The MediaTomb Project, MediaTomb. http://mediatomb.cc/.

[30] "PUPnP SourceForge Community. Portable SDK for UPnP Devices (libupnp 1.6.6," SourceForge.net Project, June 2008,

http://pupnp.sourceforge.net/.

[31] S. Kuchler, *An extendable UPnP-AV media renderer*, M.S. thesis, Institute of Information Technology, University of Klagenfurt, Klagenfurt, Austria, August 2007.

[32] The MPlayer Project, MPlayer. http://www.mplayerhq.hu/.

[33] The VideoLAN Project, VLC media player. http://www.videolan.org/vlc/.

[34] M. Kropfberger, R. Tusch, M. Jakab, et al., "A multimedia-based guidance system for various consumer devices," in *Proceedings of the 3rd International Conference on Web Information Systems and Technologies (WEBIST '07)*, pp. 83–90, Barcelona, Spain, March 2007.

[35] Maemo Community, Maemo. White Paper, June 2008, http://maemo.org/intro/white_paper/.

Hindawi Publishing Corporation
International Journal of Digital Multimedia Broadcasting
Volume 2008, Article ID 492942, 13 pages
doi:10.1155/2008/492942

Research Article

Region-Based Watermarking of Biometric Images: Case Study in Fingerprint Images

K. Zebbiche and F. Khelifi

School of Electronics, Electrical Engineering and Computer Science, Queen's University of Belfast, Belfast BT7 1NN, UK

Correspondence should be addressed to K. Zebbiche, kzebbiche01@qub.ac.uk

Received 1 March 2008; Accepted 27 June 2008

Recommended by Harald Kosch

In this paper, a novel scheme to watermark biometric images is proposed. It exploits the fact that biometric images, normally, have one *region of interest*, which represents the relevant part of information processable by most of the biometric-based identification/authentication systems. This proposed scheme consists of embedding the watermark into the region of interest only; thus, preserving the hidden data from the segmentation process that removes the useless background and keeps the region of interest unaltered; a process which can be used by an attacker as a cropping attack. Also, it provides more robustness and better imperceptibility of the embedded watermark. The proposed scheme is introduced into the optimum watermark detection in order to improve its performance. It is applied to fingerprint images, one of the most widely used and studied biometric data. The watermarking is assessed in two well-known transform domains: the discrete wavelet transform (DWT) and the discrete Fourier transform (DFT). The results obtained are very attractive and clearly show significant improvements when compared to the standard technique, which operates on the whole image. The results also reveal that the segmentation (cropping) attack does not affect the performance of the proposed technique, which also shows more robustness against other common attacks.

1. INTRODUCTION

Biometric-based systems that use physiological characteristics and/or behavioral traits offer a good alternative to traditional systems such as token-based or knowledge-based systems. These systems are more reliable and more user friendly. However, there are many issues that need more attention, especially the security aspect of both biometric system and biometric data. Several researchers show the existence of many threats and attacks that may affect the security and the integrity of biometric-based systems [1–4]. The problems that may arise from the attacks on such systems are raising concerns as more and more biometric systems are deployed [5]. Some techniques such as cryptography and watermarking have been introduced to thwart some of these attacks. Watermarking techniques are gaining more interest by providing promising results [6–8]. For example, watermarking of fingerprint images can be used to secure central databases from which fingerprint images are transmitted on request to intelligence agencies in order to use them for identification purposes (see Figure 1).

In the literature, watermarking has been introduced and shown to be satisfying the need for the protection of digital data. It can be used for many security purposes such as copyright protection, fingerprinting, copy protection, data authentication, and so forth [9]. Depending on the application, the watermarking schemes can be cast in two classes. In the first class, often known as multibit watermarking, a specific data, such as ID or track number, is embedded into the host data. In this case, the embedded watermark communicates a multibit message which must be extracted accurately at the decoding side [10, 11]. In the second class, it is not known whether a candidate watermark is embedded in the input data. The task here is therefore to verify the presence of the watermark, usually referred to as watermark detection [12, 13].

In these applications, the basic requirement is that the watermark should remain in the host data, even if its quality is degraded, intentionally or unintentionally. Examples of unintentional degradations are applications involving storage or data transmission where lossy compression is used; also filtering, resampling, digital-analog (D/A), and

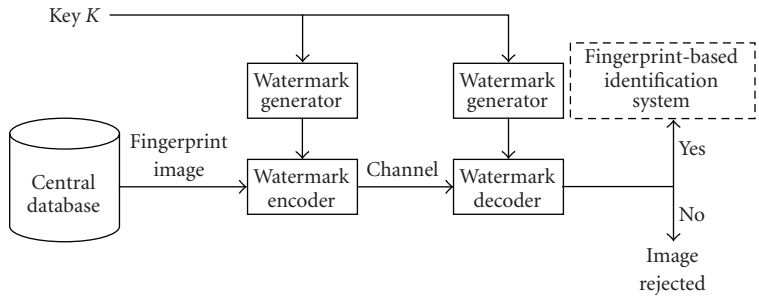


Figure 1: Block diagram of a watermarking application for fingerprint images.

analog-digital (A/D) conversion may affect the quality of the image. The host data can also be intentionally attacked in order to remove the watermark by using malicious data processing techniques such as noise addition, cropping, rotation, and translation.

The cropping technique, which consists of removing a portion of the image, remains one of the toughest attacks to deal with. Indeed, the attacker might apply it to take out parts of the image which are useless; hence, a portion of the watermark embedded within these regions is easily removable. Unfortunately, most of the watermarking algorithms are not robust enough to such an attack. Also, the watermark algorithms that make use of the human visual systems (HVSs) characteristics intend to maximize the inserted watermark, especially, in the texture areas but these algorithms do not make the difference between the useful textures and the useless noise. In order to overcome this problem, the watermark should be inserted into the most relevant part(s) of the image, that is, *region of interest* (ROI). However, this is difficult to apply to natural images since the ROI of such images is user-dependent or just undefined.

Several biometric-based systems, such as fingerprint, face, iris, or hand, use images as input data. A common characteristic of these images is that they have only one ROI, constituting the part processable by the identification/authentication algorithms. The segmentation technique is usually used to extract the ROI. However, this technique, which is basically used as a preprocessing step, can be used by an attacker as a special case of cropping since it removes the background area (i.e., removes the part of the watermark embedded in this area) while keeping ROI unchanged. The motivation is that the idea of inserting the watermark into the ROI is applicable to biometric images whose ROI can be extracted.

In this work, we propose a new scheme to embed the watermark into the ROI of biometric images. This is motivated by the following: (i) securing the embedded watermark against the segmentation process and increasing the robustness of the watermark against other attacks such as filtering, noise because even the attacker knows that the watermark is embedded in this region, concentrating his attacks on that area degrades significantly its quality, hence, making it useless; (ii) providing more transparency to the embedded watermark since the human eye is less sensitive to changes in textured areas.

Region-based method proposed in this work can be viewed as a special case of personalization because the proposed algorithm is adaptive and only a portion of the data (i.e., ROI) is watermarked. The proposed scheme is applied on fingerprint images. Note that fingerprint-based systems are regarded as the most powerful and widely deployed biometric systems. To extract the ROI of such images, referred here to as *ridges area*, the segmentation technique proposed by Wu et al. [14] is modified in order to use adaptive thresholding. For sake of completeness, the watermark is embedded into the discrete wavelet transform (DWT) and discrete Fourier transform (DFT), where the DWT coefficients are statistically modeled by the generalized Gaussian distribution (GGD) and the DFT coefficients are modeled by the Weibull distribution. Experiments were carried out on test images from real-fingerprint database and the results obtained clearly show the performance introduced by the proposed scheme. Also, the robustness of inserting the watermark into the ROI is assessed in the presence of attacks such as wavelet scalar quantization (WSQ) compression, mean filtering and additive white Gaussian noise (AWGN).

The paper is organized as follows: the proposed watermarking scheme for biometric images is explained in Section 2. Application of the proposed scheme to fingerprint images is described in Section 3. Experiments were carried out in Section 4 to assess the impact of the proposed technique on the overall performance of the optimum detector. Finally, conclusions are drawn in Section 5.

2. PROPOSED WATERMARKING SCHEME FOR BIOMETRIC IMAGES

The proposed watermarking scheme is depicted by Figure 2. At the encoder side, we aim to insert the watermark into the ROI only and exclude the background area; therefore, the ROI is first extracted. The extraction techniques can be either block-wise or pixel-wise and usually provide a binary image, called *region mask*, where 1 indicates that the block (or pixel) belongs to the ROI and 0 indicates that the block (or pixel) belongs to the background area. Then, the region mask is divided into nonoverlapping blocks to obtain a *watermarking mask*; where each block is classified based on the number of 1 in it. If the number of 1 exceeds a given threshold, then the block is classified as ROI block, otherwise, it is a background

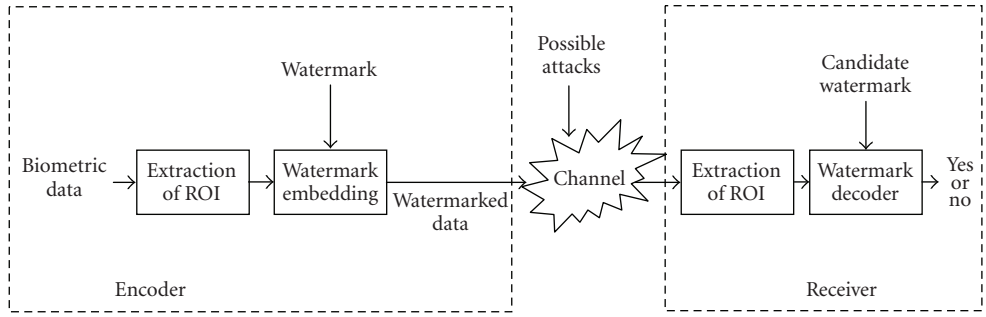


Figure 2: Proposed watermarking scheme for biometric data.

block. This watermarking mask is used to select the blocks that will hold the watermark.

It is worth noting that there are two issues to be taken into account when choosing the ROI extraction technique for watermarking purposes, which are as follows: (i) the robustness of the technique against possible attacks that may affect a watermarked image, that is, the ROI extraction technique must extract approximately the same ROI from the original and the watermarked images even after applying attacks; (ii) computational complexity. Indeed, the block-wise extraction scheme is less complex than the pixel-wise one. However, this comes at the cost of accuracy. From the view point of watermarking, pixel-wise extraction techniques are more powerful since they provide more accuracy of ROI at the detector side. This is obviously required in blind watermarking. The proposed watermarking scheme is equipped with an optimum watermark detector. In such a case, the false-alarm probability (P_{fa}) and the detection probability (P_{det}) are the natural performance measures.

3. APPLICATION TO FINGERPRINT IMAGES

The region-based method proposed in this paper can be viewed as a special case of personalization because the algorithm adaptively operates on a portion of the input data (i.e., ROI) as illustrated by Figure 3. As can be seen, the encoding system uses the ROI to insert the watermark and keeps the background image unchanged. The bigger the ROI, the larger the number of coefficients that can be used for watermarking. Once the watermark is embedded, the background area is used to reconstruct the watermarked image. At detection, the detector follows the same steps to extract the ROI and check the presence of the watermark. It is worth mentioning that the selected extraction method is first assessed on the original images by varying the attacks strength. This method should be robust enough to attacks that might alter the watermarked image. Although the watermarked image may undergo attacks that aim to remove the watermark, the visual quality should be kept useful so that the attacker can use it. We have carried out experiments on the original images to verify the efficiency of the extraction method against different attacks with various strengths controlled by a number of parameters such as compression ratio, noise variance, filtering window size.

3.1. Region of interest extraction

A fingerprint is a pattern of alternating convex skin called *ridges* and concave skin called *valleys* with a spiral-curve-like line shape. In fingerprint images, the ridges area is considered as the ROI and the noisy area around it and at the borders is the background area. In the literature, several methods have been proposed to extract the ROI from fingerprint images. These methods can be divided into two categories: block-wise and pixel-wise features classification. The algorithms that fall in the first category decompose the image into blocks. Then, some characterizing features, such as the local histogram of ridge orientation, gray-level variance, magnitude of the gradient, are calculated and based on these features, a classifier can be used to decide whether a block belongs to the ROI or to the background area. In the second category, pixel features are first extracted. This includes for example coherence, average gray level, variance and Gabor response, and then a simple classifier is chosen for classification. Such pixel-wise methods provide accurate results, but their computational complexity is higher than the commonly used block-wise methods.

In this work, Harris corner point features method [14] is adopted to extract the ridges area of fingerprint images. The Harris corner detector is based on the local autocorrelation function of a signal; where the local autocorrelation function measures the local changes of the signal with patches shifted by a small amount in different directions [15]. Wu et al. found in [14] that the strength of the Harris point in the ridges area is much higher than that of the background area. However, the authors used different thresholds, which are determined experimentally for each image. Also, they reported the existence of some noisy regions in the background area corresponding to high strength values, which cannot be eliminated even by using high threshold values and proposed to use a heuristic algorithm based on the corresponding Gabor response in order to discard these noisy regions.

To make this technique more flexible and practical, it has been modified by using the Otsu thresholding method [16] to adaptively determine adequate thresholds. Otsu method is based on maximizing the between-class variance to find the optimum threshold. This modification provides an excellent threshold for fingerprint images with different visual

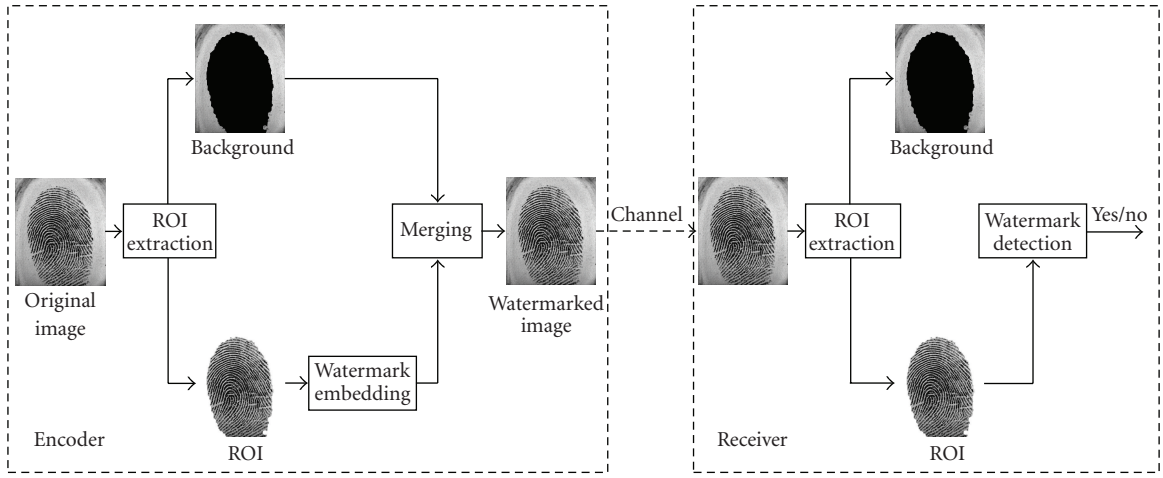


Figure 3: Personalized watermarking system applied to fingerprint images.

(a) (b) (c)

Figure 4: Example of fingerprint image: (a) original image, (b) region mask, (c) watermarking mask. The block size = 32.

qualities. To eliminate the noisy regions, some morphological methods are then applied, leading to excellent segmented images.

The Harris point is a pixel-wise method, the segmentation mask (Figure 4(b)) has the same size as the original image and it is partitioned to obtain the watermarking mask (Figure 4(c)). Note that in this paper, only blocks whose all pixels belong to the ridges area are taken into account, that is, 100% of the pixels belong to the ridges area for every selected block.

3.2. Watermark embedding

The watermark is embedded into the transform domain. In this paper, we consider two widely used transforms: the DWT and the DFT. These transforms can be applied to the entire image or in a block-wise manner. Also, the multiplicative rule is used to embed the watermark due to its advantages over the additive one, especially in exploiting the HVS characteristics. The watermark, denoted by $w = \{w_1, w_2, \ldots, w_N\}$, is a pseudorandom sequence uniformly distributed in $[-1, +1]$ and generated by using a secret key K. The embedding process is comprised of the steps described below.

(i) Extract the ROI for the input image I and obtain the region mask RM.

(ii) Determine the watermarking mask WM from the ROI by decomposing RM into nonoverlapping blocks of size $m \times m$.

(iii) Decompose the image I into nonoverlapping blocks B_{ij} of size $m \times m$ pixels and only the blocks that belong to the ROI are selected to carry the watermark, that is, if $\mathrm{WM}_{ij} = 1$ the corresponding block B_{ij} is selected; otherwise, it remains unchanged.

(iv) Transform the selected blocks using a transform, such as DWT and DFT, to obtain the original coefficients $x = \{x_1, x_2, \ldots, x_N\}$. The watermark is embedded into the original image using the multiplicative rule as follows:

$$y_i = (1 + \lambda w_i) x_i, \tag{1}$$

where $y = \{y_1, y_2, \ldots, y_N\}$ represents the watermarked coefficients and λ is the strength of the watermark.

3.3. Watermark detection

The goal of the optimum watermark detector is to verify whether or not there is a candidate watermark embedded in the received image, based on its statistical properties. This problem is usually formulated as a binary hypothesis test, in which, two hypotheses are used to represent the presence/absence of a given watermark within the host data. The two hypotheses can be established as follows:

H_0: the coefficients y are not watermarked by the candidate watermark w^*;

H_1: the coefficients y are watermarked by the candidate watermark w^*.

The decision rule for the binary test formulated above, denoted by $\Lambda(y)$, relies on maximum-likelihood method

based on Bayes' decision theory. The likelihood ratio can be written as

$$\Lambda(y) = \frac{f_y(y|H_1)}{f_y(y|H_0)}, \tag{2}$$

where $f_y(y|H_1)$ and $f_y(y|H_0)$ represent the probability distribution function (pdf) of vector y conditioned to the hypotheses H_1 and H_0, respectively. Following the same steps as described by Barni et al. in [12], the decision rule is defined as

$$l(y) = \sum_{i=1}^{N}\left[\ln\left(f_{x_i}\left(\frac{y_i}{1+\lambda w_i^*}\right)\right) - \ln(f_{x_i}(y_i))\right] \gtrless_{H_0}^{H_1} \eta', \tag{3}$$

where $l(y) = \ln(\Lambda(y))$. The decision rule reveals that H_1 is accepted (i.e., the coefficients y are marked by the sequence w^*) only if $l(y)$ exceeds the threshold η'. By employing the Neyman-Pearson criterion [17], the threshold is obtained in such a way that the detection probability P_{det} is maximized, subject to a fixed false-alarm probability P_{fa} [12]:

$$\eta' = \text{erfc}^{-1}(2P_{\text{fa}})\sqrt{2\sigma_0^2} + \mu_0, \tag{4}$$

where erfc$(\cdot)$ is the complementary error function, $\mu_0 = E[l(y)|H_0]$ and $\sigma_0^2 = V[l(y)|H_0]$ are the mean and the variance of $l(y)$ under hypothesis H_0, respectively.

3.3.1. Optimum watermark detector structure based on the GGD

To describe the probability characteristics of DWT coefficients, the GGD is widely used in the literature and some studies show that this distribution provides the closest approximation [18]. The GGD pdf of zero-mean is given by

$$f_X(x;\alpha,\beta) = \frac{\beta}{2\alpha\Gamma(1/\beta)}\exp\left(-\left(\frac{|x|}{\alpha}\right)^{\beta}\right), \tag{5}$$

where $\Gamma(\cdot)$ is the Gamma function, $\Gamma(z) = \int_0^{\infty} e^{-t}t^{z-1}dt$, $z > 0$. The parameter α is referred to as the scale parameter and it models the width of the pdf peak (standard deviation) and β is called the shape parameter and it is inversely proportional to the decreasing rate of the peak.

By substituting (5) in (3), the log-likelihood for the GGD is given by [19]

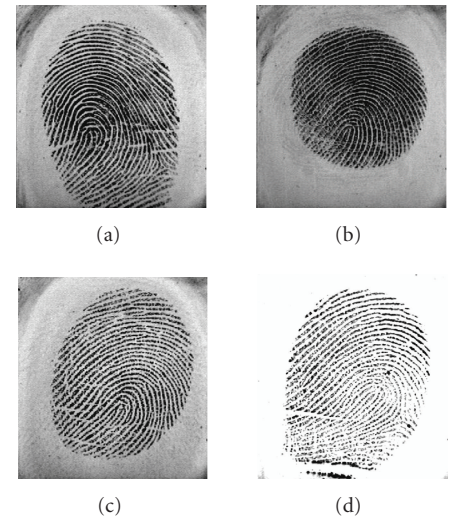
$$l(y) = \sum_{i=1}^{N}\left(\frac{|y_i|}{\alpha}\right)^{\beta}[1 - |1+\lambda w_i^*|^{-\beta}], \tag{6}$$

where α and β are the parameters of the GGD for the coefficients y.

The threshold η' can be obtained by using (4), where μ_0 and σ_0^2 are given by

$$\begin{aligned}\mu_0 &= \sum_{i=1}^{N}\frac{1}{\beta}[1 - |1+\lambda w_i^*|^{-\beta}],\\ \sigma_0^2 &= \sum_{i=1}^{N}\frac{1}{\beta}[1 - |1+\lambda w_i^*|^{-\beta}]^2.\end{aligned} \tag{7}$$

The parameters α and β can be estimated as described in [20].



(a) (b) (c) (d)

Figure 5: Test images with different visual quality from DB3: (a) Image 22_6, (b) Image 88_1, (c) Image 46_2, and (d) Image 24_3.

3.3.2. Optimum detector structure based on the Weibull model

The DFT coefficients are widely modeled by the Weibull distribution in the literature [12, 21]. Its pdf is defined as

$$f_X(x;\alpha,\beta) = \frac{\beta}{\alpha}\left(\frac{x}{\alpha}\right)^{\beta-1}\exp\left[-\left(\frac{x}{\alpha}\right)^{\beta}\right], \quad x \geq 0, \tag{8}$$

where $\beta > 0$ represents the shape parameter and $\alpha > 0$ is the scale parameter of the distribution. The detector structure for the Weibull distribution is defined by Barni et al. [12] and given by

$$l(y) = \sum_{i=1}^{N} y_i^{\beta}\left(\frac{(1+\lambda w_i^*)^{\beta} - 1}{\alpha^{\beta}(1+\lambda w_i^*)^{\beta}}\right), \tag{9}$$

where α_i and β_i are the parameters of the Weibull model for the coefficients y.

Equation (4) is used to derive the threshold η' where the mean μ_0 and the variance σ_0^2 are defined as

$$\begin{aligned}\mu_0 &= \sum_{i=1}^{N}\left(\frac{(1+\lambda w_i^*)^{\beta} - 1}{(1+\lambda w_i^*)^{\beta}}\right),\\ \sigma_0^2 &= \sum_{i=1}^{N}\left(\frac{(1+\lambda w_i^*)^{\beta} - 1}{(1+\lambda w_i^*)^{\beta}}\right)^2.\end{aligned} \tag{10}$$

4. EXPERIMENTAL RESULTS

In order to efficiently measure the actual performance of proposed technique, experiments were carried out on real fingerprint images of size 448×478 taken from Fingerprint Verification Competition "FVC 2000, DB3" database [22]. These images have been chosen with respect to their different visual quality (Figure 5). The performance of the proposed technique, which embeds the watermark in the ridges area

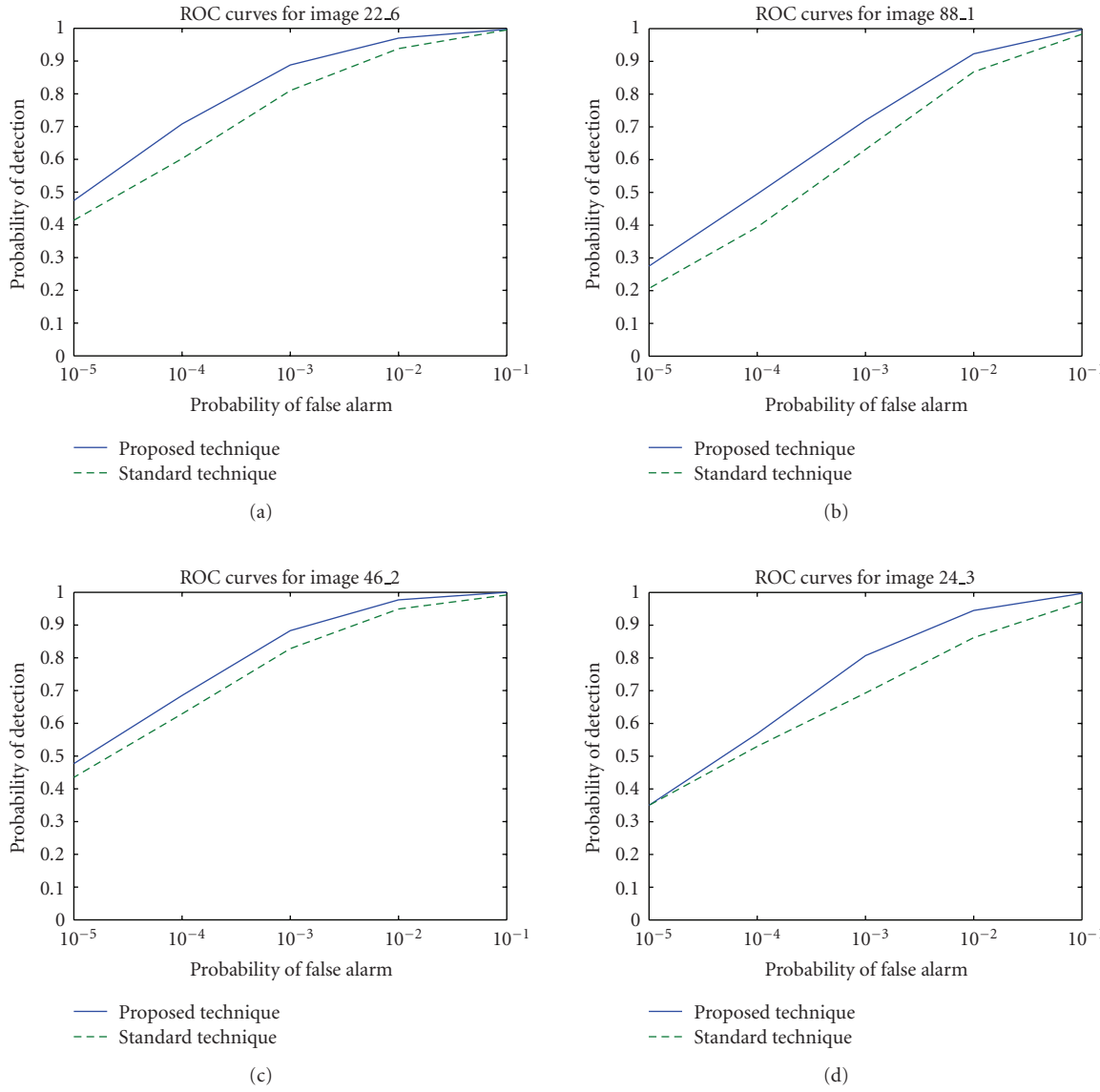


Figure 6: ROC curves of test images. Watermarking applied in the DWT domain. Strength $\lambda = 0.10$.

only, is compared against the standard technique which inserts the watermark into the whole image. In the DWT domain, Daubechies 9/7 wavelet is used. Note that such a wavelet has been adopted by the FBI as part of the wavelet scalar quantization (WSQ) compression standard for fingerprint images. The watermark is embedded in all coefficients in the third level subbands, except the approximation subband. An approach similar to that proposed in [12] is used to cast the watermark in the DFT domain, where the watermark is inserted into the magnitude of a set of full-frame coefficients. Blind detection is adopted for all experiments, that is, the statistical model parameters are directly estimated from the watermarked data. The receiver operating characteristics (ROCs) curves are used to assess the performance of both the proposed and the standard techniques. The ROC curves represent the variation of the detection probability (P_{det}) against the false-alarm probability (P_{fa}). Note that for our proposed technique, the number of coefficients to be watermarked (the length of the watermark sequence) is image dependent. The larger the ROI (i.e., ridges area), the higher the number of coefficients to be watermarked (the length of the watermark) and vice versa. For the size of the blocks m used to determine the watermarking mask, it has been set to 32 after extensive experiments held on many fingerprint images. This value allows the extraction of the ridges area even after applying severe attacks.

At the first stage, we investigate the performance of the proposed technique against the standard one without the presence of any attack. The probability of false alarm is varied in the range 10^{-5} to 10^{-1} and the value of the strength λ is fixed to value 0.10. The experimental ROC curves

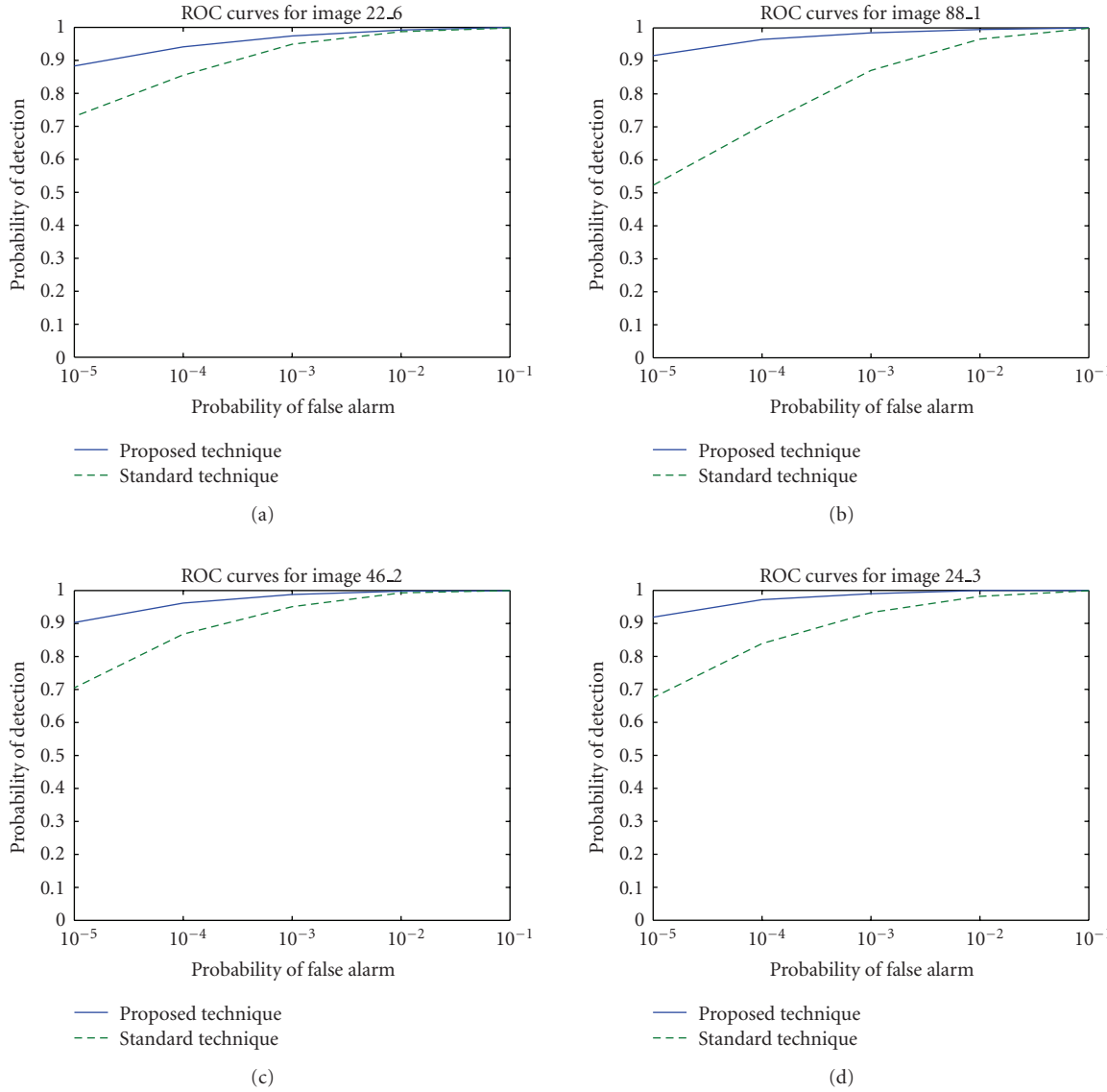


Figure 7: ROC curves of test images. Watermarking applied in the DFT domain. Strength $\lambda = 0.10$.

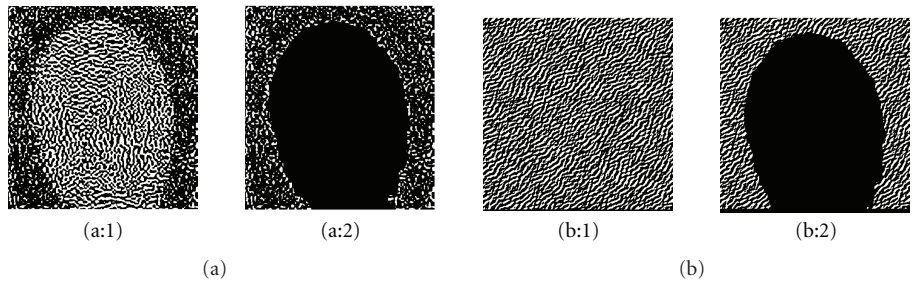


Figure 8: Standard watermarking of test image. Image 22_6: (a:1): difference image between original and watermarked images in the DWT domain; (a:2): difference image when removing ROI in the DWT domain; (b:1): difference image between original and watermarked images in the DFT domain; (b:2): difference image when removing ROI in the DFT domain.

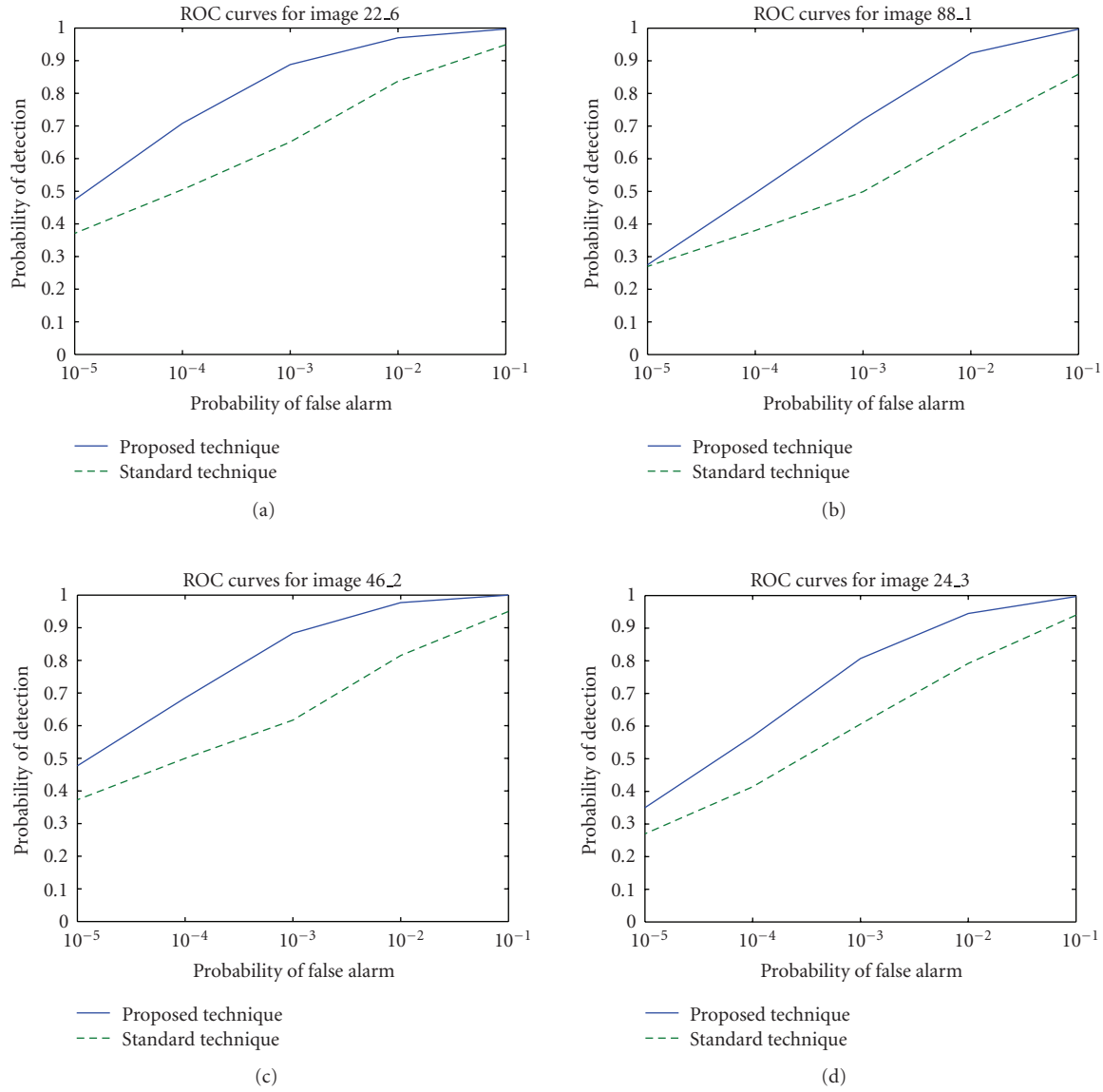


Figure 9: ROC curves of segmented, watermarked images. Watermarking applied in the DWT domain. Strength $\lambda = 0.10$.

are computed by measuring the performance of the actual watermark detection system by calculating the probability of detection from real-watermarked images. Experiments are then conducted by comparing the likelihood ratio with the corresponding threshold for each value of the false-alarm probability and for 1000 randomly generated watermark sequences. The results obtained for the DWT domain are plotted in Figure 6 and those obtained for the DFT domain are plotted in Figure 7.

As can be seen from Figures 6 and 7, the proposed technique outperforms the standard one even without applying any attack. This is justified by the fact that the transform coefficients are better suited to watermarking for the proposed technique since they correspond to a highly textured area (i.e., ridges area) only. These coefficients allow the embedding of strong watermarks.

As mentioned earlier, an attacker may use segmentation techniques on biometric images to remove a part of the watermark embedded within the background area without altering the ROI. To illustrate this, the spatial repartition of the watermark is plotted in Figure 8(a:1) for the DWT domain and in Figure 8(b:1) for the DFT domain in the case of a standard watermarking; it represents the difference between the watermarked image and the original one. The part of the watermark removed by the segmentation technique is plotted in Figure 8(a:2) for the DWT domain and in Figure 8(b:2) for the DFT domain. It represents the difference image without the ridges area. For the sake of illustration, only the results for one image is shown since the results for other images are very similar. As can be seen, an important part of the watermark is embedded into the background area, which can be removed easily by applying

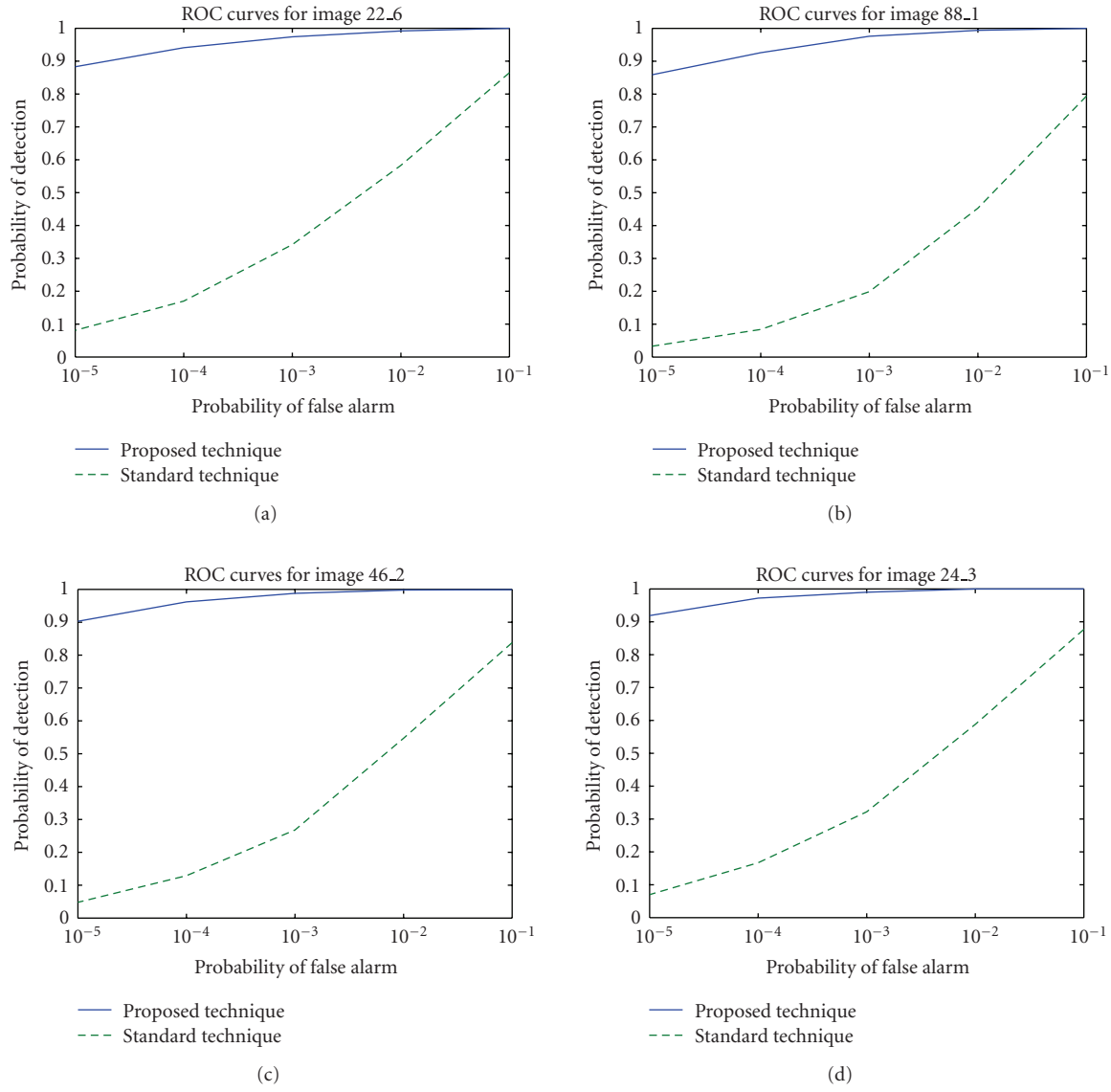


Figure 10: ROC curves of segmented, watermarked images. Watermarking applied in the DFT domain. Strength $\lambda = 0.10$.

segmentation. Comparing Figures 8(a:1) and 8(b:1), the watermark energy in the DWT domain is concentrated into the ridges area (i.e., textured area). However, in the DFT domain, the watermark energy is uniformly spread all over the image. Thus, a severe degradation of the standard detector performance in the DFT domain is expected when applying the segmentation attack, compared to the DWT domain.

After applying the segmentation process on watermarked images, the previous experiment has been carried out and the results obtained are plotted in Figure 9 for the DWT domain and Figure 10 for the DFT domain. For the proposed technique, the ROC curves are exactly the same as for the first experiment, thus, the segmentation process has no influence on the performance of the optimum detector. For the standard technique, the probability of detection decreases significantly and the segmentation process causes a deterioration of detection performance in both DWT and DFT domains. As expected for the DFT domain, the degradation in performance is more significant than that obtained in the DWT domain.

The performance of the proposed technique against common attacks, namely, mean filtering, WSQ compression, and additive white Gaussian noise (AWGN), is also evaluated. Each attack has been applied several times with different strength values. For each attack, the response of the detector to the embedded watermark is plotted along with the threshold. In this way, the influence of each attack strength on the detector response and the corresponding threshold is assessed. The theoretical P_{FA}, which is used to determine the decision threshold, has been fixed at 10^{-7} and the strength λ is set in such a way to obtain a peak signal-to-noise ratio

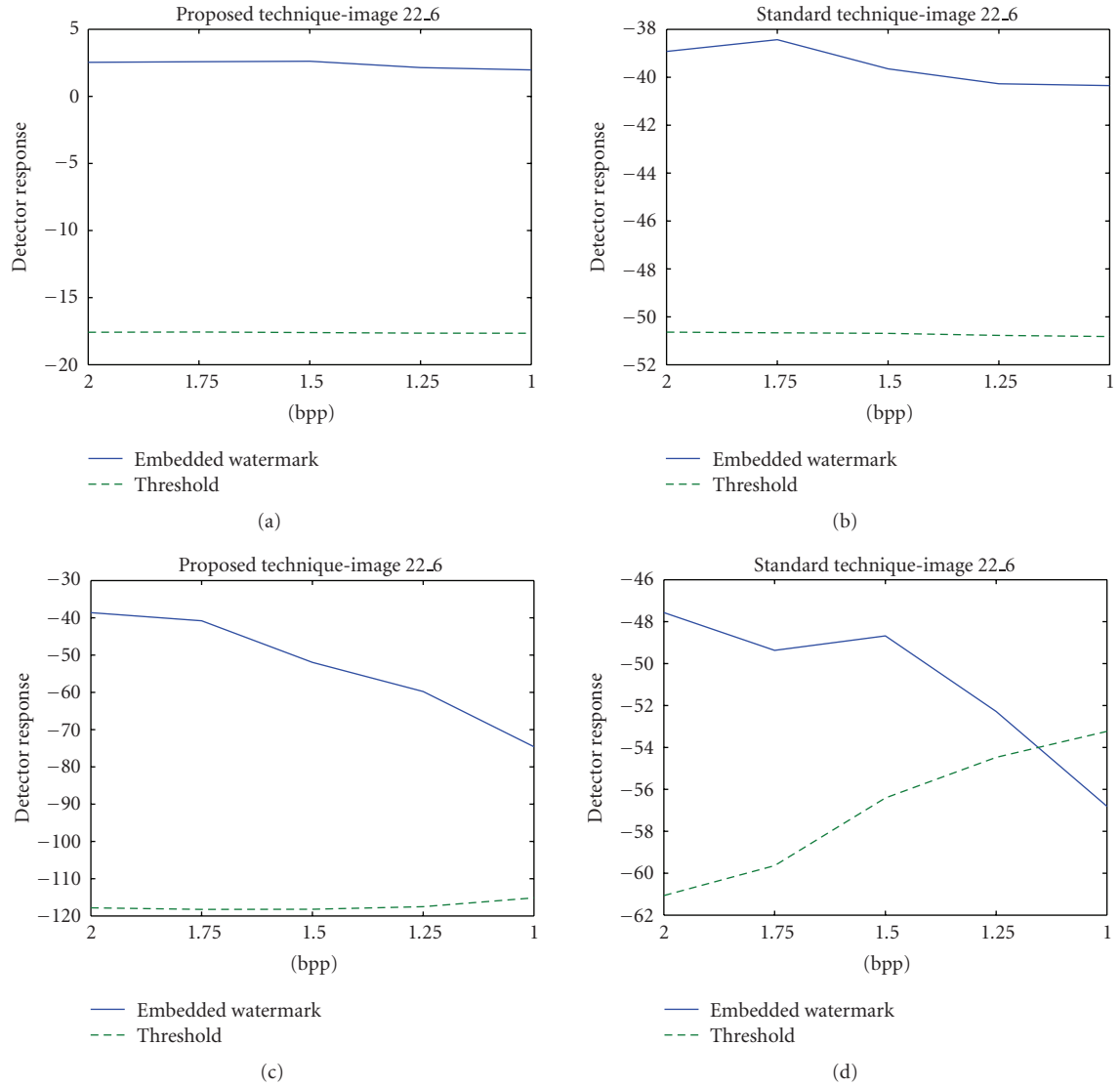


Figure 11: Robustness against WSQ compression. Top: DWT domain. Bottom: DFT domain. Left side graphs: Proposed technique. Right side graphs: Standard technique.

(PSNR) value ≈ 40 for all test images and in both DWT and DFT domains. Only results for one image are plotted since results obtained from other images are very similar.

Robustness against WSQ compression is assessed by iteratively applying the WSQ compression on the watermarked images using the WSQ viewer [23] and varying the bit-rate value measured by bits per pixel (bpp). The results obtained are reported in Figure 11. Obviously, the watermarking in the DWT domain is more robust for both the proposed and the standard techniques since the compression technique is operating in the same domain. On the contrary, the watermarks embedded in the DFT domain do not resist the WSQ compression. Again, the proposed technique outperforms the standard one.

The results of degradations due to AWGN are shown in Figure 12. The watermarked images were corrupted by AWGN with different value of signal-to-noise ratio (SNR). For all images and in both the DWT and the DFT domains, the watermarks are very robust for both the proposed and the standard techniques.

Figure 13 shows the results of watermarked fingerprint images corrupted by mean filtering. The watermarked images were blurred with different filter window size. Although the proposed technique is slightly better than the standard one, the mean filtering affects significantly the detector performance. Note that the detector for the standard technique in the DFT domain is unable to detect the embedded watermarks for all images and all filter window sizes. This is justified by the fact that this type of filtering smooths the image and attenuates the shape of edges and textures.

5. CONCLUSIONS

In this paper, a novel scheme has been proposed to watermark biometric images. This scheme exploits the fact

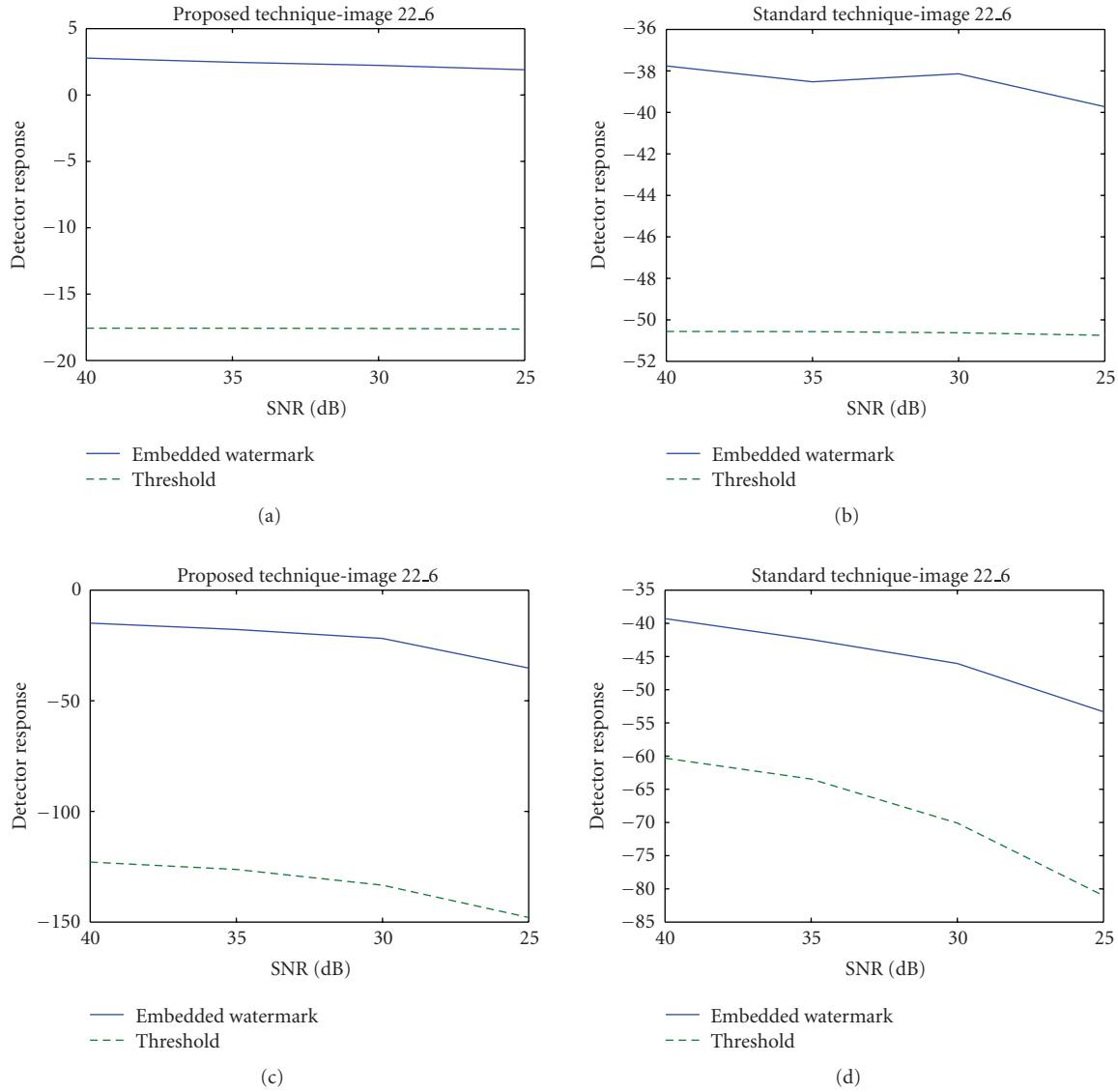


Figure 12: Robustness against additive white Gaussian noise. Top: DWT domain. Bottom: DFT domain. Left side graphs: Proposed technique. Right side graphs: Standard technique.

that biometric images have only one *region of interest*, which constitutes the useful and unique processed region by most of the biometric-based identification/authentication systems. This fact can also be exploited by watermarking techniques where the watermark should be embedded into the region of interest only, instead of spreading it into the whole image. This proposed scheme is motivated by the following: (i) increasing the robustness of the watermark against segmentation and other attacks such as filtering, noise because even the attacker knows that the watermark is embedded in this region, concentrating his attacks on that area degrades significantly its quality, hence, making it useless; (ii) providing more transparency to the embedded watermark since the region of interest is a highly textured area and the human eye is less sensitive to changes in that area. The embedding process for the proposed scheme starts by extracting the region of interest and then embeds the watermark in this area only. This scheme is applied to fingerprint images that are used by one of the most employed and widely deployed biometric systems. To extract the ROI of such images, known as *ridges area*, we modified the segmentation technique proposed by Wu et al. [14].

The proposed scheme is used with the classical optimum, multiplicative watermark detection. For sake of generality, the watermark is applied to the DWT and the DFT domains. The DWT coefficients modeled by the generalized Gaussian distribution, whereas, the DFT coefficients are modeled by the Weibull model. The influence introduced by the proposed scheme on the optimum detectors were assessed through experiments, carried out on real fingerprint images with different characteristics. The results obtained clearly show that the detector performance has been improved compared to the standard technique, which operates on the whole image, and this even in the absence of attacks.

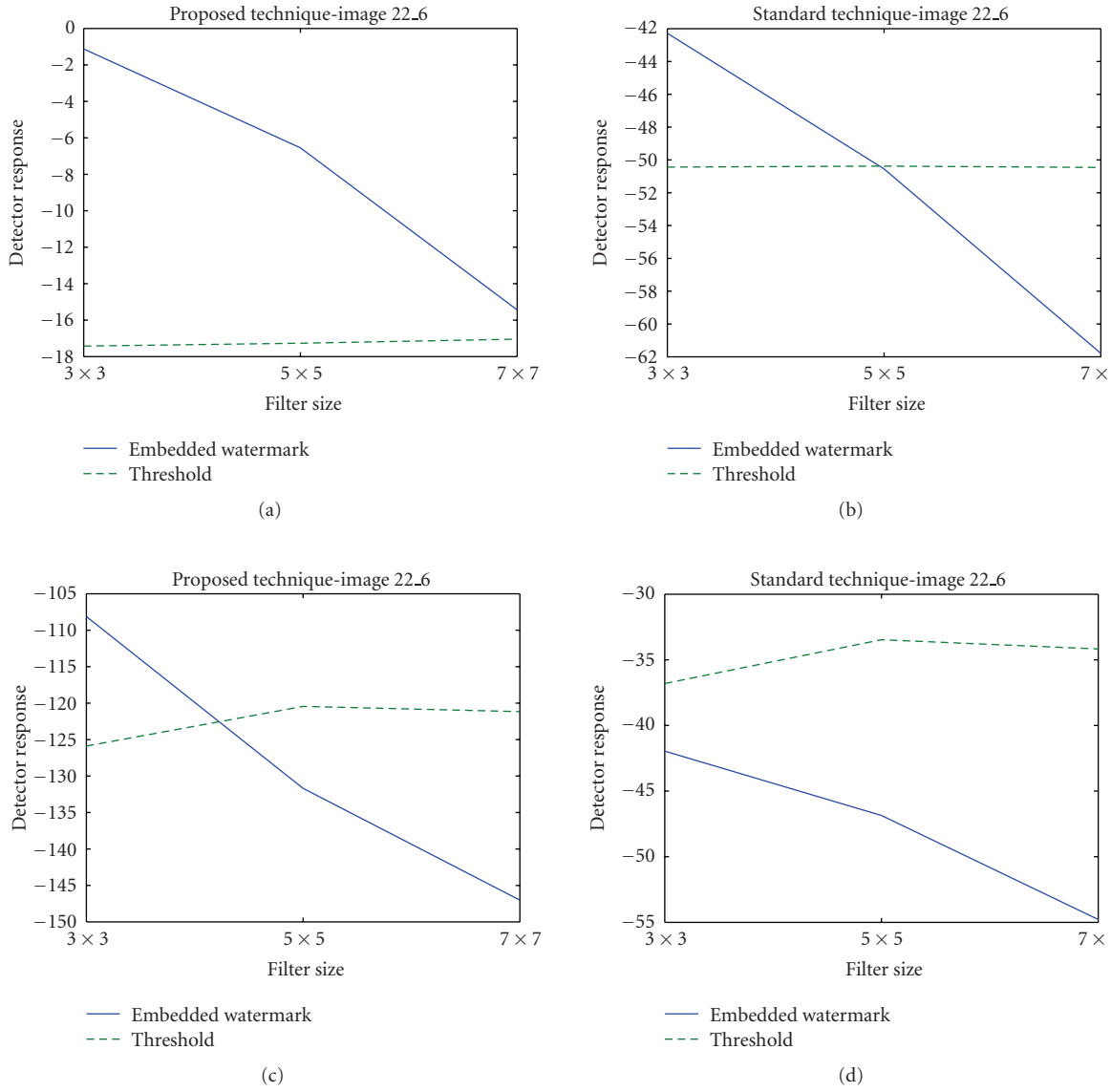


Figure 13: Robustness against mean filtering. Top: DWT domain. Bottom: DFT domain. Left side graphs: Proposed technique. Right side graphs: Standard technique.

In addition, the segmentation technique, which has been applied as a special case of cropping attack, affects the performance of the standard technique since it removes the part of the watermark embedded within the background area. However, this attack has no effect on the proposed technique. Furthermore, the watermarks embedded using the proposed scheme show to be more robust against some other common attacks such as WSQ compression, mean filtering, and white noise addition.

REFERENCES

[1] B. Schneier, "Inside risks: the uses and abuses of biometrics," *Communications of the ACM*, vol. 42, no. 8, pp. 136–139, 1999.

[2] N. K. Ratha, J. H. Connell, and R. M. Bolle, "An analysis of minutiae matching strength," in *Proceedings of the 3rd International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA '01)*, vol. 2091 of *Lecture Notes in Computer Science*, pp. 223–228, Halmstad, Sweden, June 2001.

[3] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar, *Handbook of Fingerprint Recognition*, Springer, New York, NY, USA, 2003.

[4] U. Uludag and A. K. Jain, "Attacks on biometric systems: a case study in fingerprints," in *Security, Steganography, and Watermaking of Multimedia Contents VI*, vol. 5306 of *Proceedings of SPIE*, pp. 622–633, San Jose, Calif, USA, January 2004.

[5] Congress of the United States of America, "Enhanced border security and visa entry reform act of 2002," http://www.unitedstatesvisas.gov/pdfs/Enhanced_Border_SecurityandVisa_Entry.pdf.

[6] N. K. Ratha, J. H. Connell, and R. M. Bolle, "Secure data hiding in wavelet compressed fingerprint images," in *Proceedings of the ACM Multimedia Workshops (MULTIMEDIA '00)*, pp. 127–130, Los Angeles, Calif, USA, October-November 2000.

[7] A. K. Jain and U. Uludag, "Hiding biometric data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 11, pp. 1494–1498, 2003.

[8] K. Zebbiche, L. Ghouti, F. Khelifi, and A. Bouridane, "Protecting fingerprint data using watermarking," in *Proceedings of the 1st NASA/ESA Conference on Adaptive Hardware and Systems (AHS '06)*, pp. 451–456, Istanbul, Turkey, June 2006.

[9] G. C. Langelaar, I. Setyawan, and R. L. Lagendijk, "Watermarking digital image and video data: a state-of-the-art overview," *IEEE Signal Processing Magazine*, vol. 17, no. 5, pp. 20–46, 2000.

[10] F. Pérez-González, J. R. Hernández, and F. Balado, "Approaching the capacity limit in image watermarking: a perspective on coding techniques for data hiding applications," *Signal Processing*, vol. 81, no. 6, pp. 1215–1238, 2001.

[11] M. Barni, F. Bartolini, A. De Rosa, and A. Piva, "Optimum decoding and detection of multiplicative watermarks," *IEEE Transactions on Signal Processing*, vol. 51, no. 4, pp. 1118–1123, 2003.

[12] M. Barni, F. Bartolini, A. De Rosa, and A. Piva, "A new decoder for the optimum recovery of nonadditive watermarks," *IEEE Transactions on Image Processing*, vol. 10, no. 5, pp. 755–766, 2001.

[13] Q. Cheng and T. S. Huang, "An additive approach to transform-domain information hiding and optimum detection structure," *IEEE Transactions on Multimedia*, vol. 3, no. 3, pp. 273–284, 2001.

[14] C. Wu, S. Tulyakov, and V. Govindaraju, "Robust point-based feature fingerprint segmentation algorithm," in *Proceedings of the International Conference on Advances in Biometrics (ICB '07)*, vol. 4642, pp. 1095–1103, Seoul, Korea, August 2007.

[15] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proceedings of the 4th Alvey Vision Conference*, vol. 15, pp. 147–151, Manchester, UK, August-September 1988.

[16] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[17] J. V. DiFranco and W. L. Rubin, *Radar Detection*, SciTech Publishing, Raleigh, NC, USA, 2004.

[18] K. Zebbiche, F. Khelifi, and A. Bouridane, "Maximum-likelihood watermarking detection on fingerprint images," in *Proceedings of the ECSIS Symposium on Bio-Inspired, Learning, and Intelligent Systems for Security (BLISS '07)*, vol. 9, pp. 15–18, Edinburgh, UK, August 2007.

[19] T. M. Ng and H. K. Garg, "Wavelet domain watermarking using maximum-likelihood detection," in *Security, Steganography, and Watermaking of Multimedia Contents VI*, vol. 5306 of *Proceedings of SPIE*, pp. 816–826, San Jose, Calif, USA, January 2004.

[20] M. N. Do and M. Vetterli, "Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance," *IEEE Transactions on Image Processing*, vol. 11, no. 2, pp. 146–158, 2002.

[21] Q. Cheng and T. S. Huang, "Optimum detection and decoding of multiplicative watermarks in DFT domain," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '02)*, vol. 4, pp. 3477–3480, Orlando, Fla, USA, May 2002.

[22] Fingerprint verification competition, http://biometrics.cse.msu.edu/fvc04db/index.html.

[23] The Wsq viewer (version 2.7), http://www.cognaxon.com/index.php?page=wsqview.

Hindawi Publishing Corporation
International Journal of Digital Multimedia Broadcasting
Volume 2008, Article ID 289837, 12 pages
doi:10.1155/2008/289837

Research Article

Extracting Moods from Songs and BBC Programs Based on Emotional Context

Michael Kai Petersen and Andrius Butkus

Department of Informatics and Mathematical Modeling, Technical University of Denmark, Richard Petersens Plads, Building 321, 2800 Kongens Lyngby, Denmark

Correspondence should be addressed to Michael Kai Petersen, mkp@imm.dtu.dk

Received 2 March 2008; Revised 2 July 2008; Accepted 4 August 2008

Recommended by Harald Kosch

The increasing amounts of media becoming available in converged digital broadcast and mobile broadband networks will require intelligent interfaces capable of personalizing the selection of content. Aiming to capture the mood in the content, we construct a semantic space based on tags, frequently used to describe emotions associated with music in the *last.fm* social network. Implementing latent semantic analysis (LSA), we model the affective context of songs based on their lyrics, and apply a similar approach to extract moods from BBC synopsis descriptions of TV episodes using TV-Anytime atmosphere terms. Based on our early results, we propose that LSA could be implemented as machinelearning method to extract emotional context and model affective user preferences.

1. INTRODUCTION

When both digital broadcast streams and the content itself are adapted to the small screen size of handheld devices, it will literally translate into hundreds of channels featuring rapidly changing mobisodes and location-aware media, where it might no longer be feasible to select programs by scrolling through an electronic program guide. In order to automatically filter media according to personalized preferences, this will require metadata which not only defines traditional genre categories but also incorporates parameters capturing the changing mobile usage contexts. Since 2005, the broadcaster BBC has made their program listings available as XML formatted TVA TV-Anytime [1] metadata, which allows for describing media using complementary aspects, such as content genre, format, intended audience, intention, or atmosphere. We have previously in a related paper [2] analyzed how especially atmosphere metadata describing emotions may facilitate identifying programs that might be perceived as similar even though they belong to different genre categories. Also in music it appears that despite the often idiosyncratic character of tags, defined by hundred thousands of users in social networks like *last.fm*, people tend to agree on the affective terms they attach to describe music [3, 4]. A mounting question might therefore be: could we possibly apply machine learning techniques to extract emotional aspects associated with media in order to model our perception, and thus facilitate an affective categorization which goes beyond traditional divides of genres?

2. RELATED WORKS

In usage scenarios involving DVB-H mobile TV, where shifting between a few channels might be even more time-consuming than watching the actual mobisode, new text mining approaches to content-based filtering have been suggested as a solution. Reflecting preferences for categories like "fun," "action," "thrill," or "erotic," topics and emotions are extracted from texts describing the programs and incorporated into the EPG electronic program guide data as a basis for generating user preferences [5]. In broadcast context, a similar approach has been implemented to extract both textual and visual concepts for automatic categorization of TV ad videos based on probabilistic latent semantic analysis (pLSA) [6]. As a machine learning method similar

to latent semantic analysis (LSA) [7], it captures statistical dependencies among distributions of visual objects or brand names, and thus enables unsupervised categorization of semantic concepts within the content. Recent neuroimaging experiments, focused on visualizing human brain activity reflecting the meaning of nouns, have demonstrated a direct relationship between the observed patterns in brain scans of regions being activated, and the statistics of word cooccurrence in large collections of documents. The distinct patterns of functional magnetic resonance images (fMRIs) triggered by specific terms seem not only to cause similar brain activities across different individuals [8], but also makes it possible to predict which voxels in the brain will be activated according to semantic categories based on word cooccurrence in a large text corpus [9]. Or in other words, the way LSA simulates text comprehension by modelling the meaning of words as the sum of contexts in which they occur appears to have neural correlates.

Over the past decade, advances in neuroimaging technologies enabling studies of brain activity have established that musical structure to a larger extent than previously thought is being processed in "language" areas of the brain [10]. Neural resources between music and language appear to be shared both in syntactic sequencing and also semantic processing of patterns reflecting tension and resolution [11–13], adding support for findings of linguistic and melodic components of songs being processed in interaction [14]. Similarly, there appears to be an overlap between language regions in the brain and mirror neurons, which transfer sensory information of what we perceive by reenacting them on a motor level. The mirror neuron populations mediate the inputs across audiovisual modalities and the resulting sensory-motor integrations are represented in a similar form, whether they originate from actions we observe in others, only imagine or actually enact ourselves [15, 16]. This has led to the suggestion that our empathetic comprehension of underlying intentions behind actions, or the emotional states reflected in sentences and melodic phrases are based on an imitative reenactment of the perceived motion [17].

Aspects of musical affect have been the focus of a wide field of research, ranging from how emotions arise based on the underlying harmonic and rhythmical hierarchical structures forming our expectations [18–20], to how we consciously experience these patterns empathetically as contours of tensions and release [21], in turn triggering physiological changes in heart rate or blood pressure as has been documented in numerous cognitive studies of the links between music and emotions [22]. But when listening to songs our emotions are not only evoked by low-level cognitive representations but also exposed to higher level features reflecting the words which make up the lyrics. Studies on retrieving songs from memory indicate that lyrics and melody appear to be recalled from two separate versions: one storing the melody and another containing only the text [23], while further priming experiments indicate that song memory is not organized in strict temporal order, but rather that text and tune intertwine based on reciprocal connections of higher-order structures [24].

Taking the above findings into consideration, could we possibly extract affective components from textual representations of media like song lyrics, and model them as patterns reflecting how we emotionally perceive media? Applying LSA as a machine learning method to extract moods in both song lyrics and synopsis descriptions of BBC programs, we describe in the following sections, the methodology used for extracting high level representations of media using emotional tags, the early results retrieved when mapping emotional components of song lyrics and synopsis descriptions, and conclude with a discussion of the potential for automatically generating affective user preferences as a basis for mood-based recommendation.

3. EMOTIONAL TAG SPACE

When investigating how unstructured metadata can be used to describe media, the social music network *last.fm* provides an interesting case. The affective terms which are frequently chosen as tags by *last.fm* users to describe the emotional context of songs seem to form clusters around primary moods like mellow, sad, or more agitated feelings like angry and happy. This correlation between social network tags and the specific music tracks they are associated with has been used in the music information retrieval community to define a simplified mood ground-truth, reflecting not just the words people frequently use when describing the perceived emotional context, but also which tracks they agree on attaching these tags to [3, 4]. We have selected twelve of these frequently used tags for creating an emotional semantic space. Drawing on standard psychological parameters for emotional assessment, we map these affective terms along the two primary dimensions of *valence* and *arousal* [25], and use these two axes to outline an emotional plane for dividing them within an affective semantic space containing four groups of frequently used *last.fm* tags:

(i) *happy, funny, sexy;*
(ii) *romantic, soft, mellow, cool;*
(iii) *angry, aggressive;*
(iv) *dark, melancholy, sad.*

Within this emotional plane, the dimension of *valence* describes how pleasant something is along an axis going from positive to negative associated with words like happy or sad, whereas *arousal* captures the amount of involvement ranging from passive states like mellow and sad to active aspects of excitation as reflected in tags like angry or happy. Applying the selected *last.fm* tags as emotional buoys to define a semantic plane of psychological valence and arousal dimensions, we apply latent semantic analysis (LSA) to assess the correlation between the lyrics and each of the selected affective terms. Applying these affective terms as markers also enables us to compare the LSA-retrieved values against the actual tags users have applied in the *last.fm* tag clouds associated with the songs in our analysis. Additionally, when analyzing the synopsis descriptions of BBC programs we have complemented the *last.fm* tags with a large number of TV-Anytime atmosphere terms similarly used as emotional

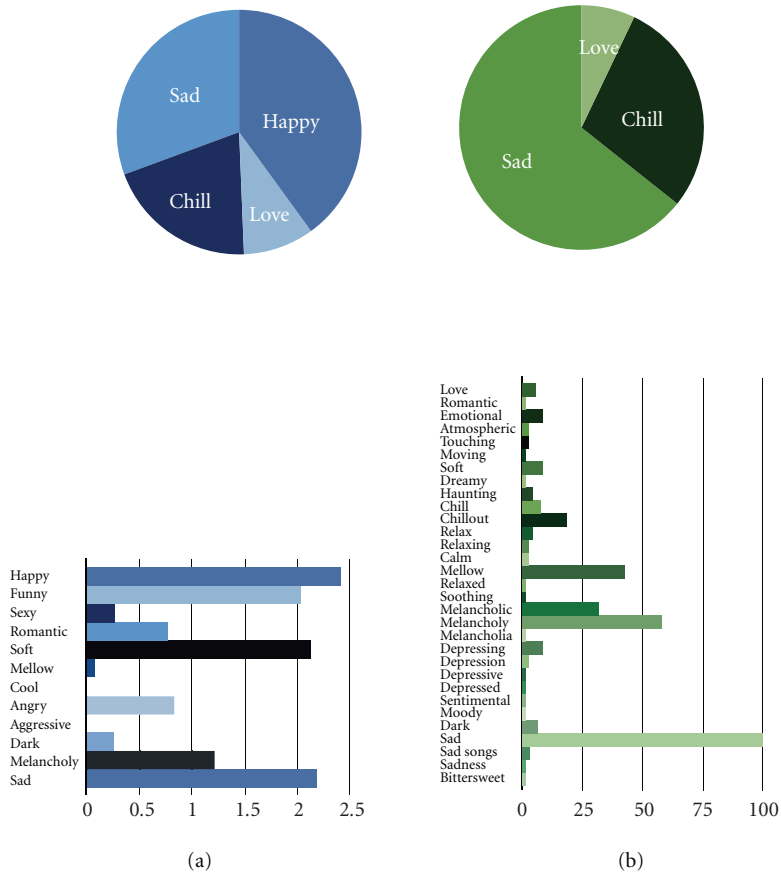


Figure 1: Accumulated LSA correlation between (a) the lyrics of the song "Nothing else matters" and 12 affective terms, compared to (b) the actual user-defined emotional tags at last.fm.

buoys. Though the two sets of markers are clearly affected differently by the synopsis, a comparison shows that despite the higher degree of detail in the TV-Anytime vocabulary, the overall emotional context is reflected similarly by the *last.fm* tags and the atmosphere terms. Or in other words, the *last.fm* and TV-Anytime markers provide different granularities for capturing emotions but the larger tendencies in the resulting patterns remain the same.

As a machine learning technique, LSA extracts meaning from paragraphs by modelling the usage patterns of words in multiple documents and represent the terms and their contexts as vectors in a high-dimensional space. The basis for assessing the correlations between lyrics and emotional words vectors in LSA is an underlying text corpus consisting of a large collection of documents which provides the statistical basis for determining the cooccurrence of words in multiple contexts. For this experiment, we chose the frequently implemented standard *TASA* text corpus, consisting of the 92409 words found in 37651 texts, novels, news articles, and other general knowledge reading material that American students are exposed to up to the level of their 1st year in college. The frequency at which terms appear and the phrases wherein they occur are defined in a matrix with rows made up of words and columns of documents. Many of the cells made up by rows and columns contain only zeroes, so in order to retain only the most essential features, the dimensionality of the original sparse matrix is reduced to around 300 dimensions. This makes it possible to model the semantic relatedness of song lyrics and affective terms as vectors, with values toward 1 signifying degrees of similarity between the items and low or minus values typically around 0.02 signifying a random lack of correlation. In this semantic space lines of lyrics or emotional words which express the same meaning will be represented as vectors that are closely aligned, even if they do not literally share any terms. Instead, these terms may cooccur in other documents describing the same topic, and when reducing the dimensionality of the original matrix, the relative strength of these associations can be represented as the cosine of the angle between the vectors.

4. RESULTS: SONG LYRICS

Whereas the user-defined tags at *last.fm* describe a song as a whole, we aim to model the shifting contours of tension and release which evoke emotions, and therefore project each of the individual lines of the lyrics into the semantic space. Analyzing individual lines on a timescale of seconds also reflects the cognitive temporal constraints applied by our brains in general when we bind successive events into perceptual units [26]. We perceive words as

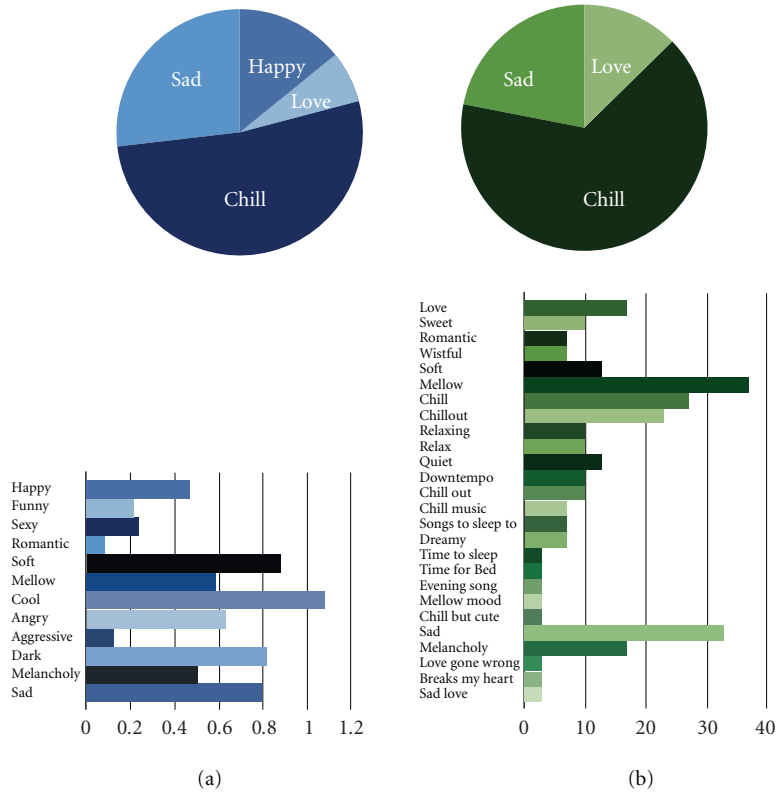


Figure 2: Accumulated LSA correlation between (a) the lyrics of the song "Now at last" and 12 affective terms, compared to (b) the actual user-defined emotional tags at last.fm.

successive phonemes and vowels on a scale of roughly 30 milliseconds, which are in turn integrated into larger segments with a length of approximately 3 seconds. We thus assume that lines of lyrics consisting of a few words each correspond to one of these high-level perceptual units. Viewed from a neural network perspective, projecting the lyrics into a semantic LSA space line by line, could also in a cognitive sense be interpreted as similar to how mental concepts are constrained by the amount of activation among the neural nodes representing events and associations in our working memory [27]. In that respect, the cooccurrence matrix formed by the word frequencies of *last.fm* tags and song lyrics might be understood as corresponding to the strengths of links connecting nodes in a mental model of semantic and episodic memory.

4.1. Accumulated emotional components

Projecting the lyrics of thirty songs selected from the weekly top track charts at *last.fm*, we compute the correlation between lyrics and tags against each of the twelve affective terms used as markers in the LSA space, while discarding cosine values below a threshold of 0.09. And in order to compare the retrieved LSA correlation values of lyrics and affective terms against the user-defined tags attached to the song at *last.fm*, we sum up the accumulated LSA values retrieved from each line of the lyrics.

Taking the song "Nothing else matters" as an example, the user defined tags attached to the song as at *last.fm*, include less frequently used tags like *love, love songs, chill, chillout, relaxing, relax, memories,* and *melancholic* which are not among the markers we used for our LSA analysis. We therefore subsequently combine these tags into larger segments of tags in order to facilitate a direct comparison with the LSA-retrieved values (Figure 1). Comparing the accumulated LSA values of emotional components against the user-defined tags at *last.fm*, the terms *melancholy,* and *melancholic*, which describe the most dominant emotions in the tag cloud, could be understood as captured by the affective term *sad* in the LSA analysis. Similarly, if interpreting *love* from the *last.fm* tag cloud as associated with the term *happy* (based on a cosine correlation of 0.56 between the words love and happy), the LSA analysis could be understood to retrieve also aspects of this emotion. Likewise, if *chill* in the *last.fm* tag cloud is understood as associated with *soft* and *mellow* (based on cosine correlations of 0.36 and 0.35, resp.), the LSA analysis also here appears to capture that mood.

Applying a similar approach to a set of thirty songs, we grouped semantically close *last.fm* tags into larger segments consisting of *sad, happy, love,* and *chill* aspects to facilitate a comparison with the LSA-derived correlations between song lyrics and the selected affective terms. Though there is an overlap between the retrieved LSA values and user-defined

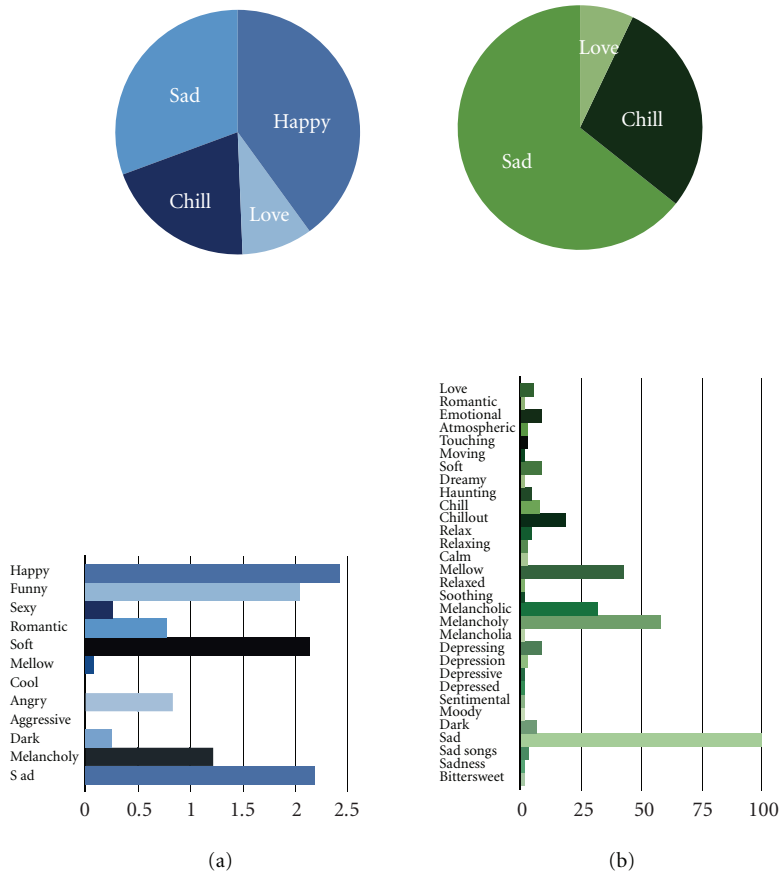


Figure 3: Accumulated LSA correlation between (a) the lyrics of the song "Mad world" and 12 affective terms, compared to (b) the actual user-defined emotional tags at last.fm.

last.fm tags in most of the songs, there is no overall significant correlation between LSA-retrieved values and the exact distribution of tags in the user-defined *last.fm* tag clouds. Essentially, the individual tags in a cloud are "one size fits all" and apply to the song as a whole, whereas the LSA correlation between lyrics and semantic markers reflects the changing degrees of affinity between the song lines and affective components over time. But for a third of the set of songs, as exemplified by "Now at last" (Figure 2), the distribution of *last.fm* tags resembled the LSA values if grouped into larger segments. While in the remaining two thirds of the set of songs, as exemplified by the song "Mad World" (Figure 3), the overall distribution in *last.fm* tags while clearly overlapping remain overly biased toward *sad* type of components.

4.2. Distribution of emotional components

Instead of grouping the emotional components into larger segments, we subsequently maintained the LSA values retrieved from each of the individual lines in the lyrics, and proceeded by plotting the values over time to provide a view of the distribution of emotional components. The plots can be interpreted as mirroring the structure of patterns of changing emotions in the songs along the horizontal axis. Vertically, the color groupings indicate which of the aspects of valence and arousal are triggered by the lyrics as well as their general distribution in relation to each other. Any color will signify an activation beyond the cosine similarity threshold level of 0.09, and the amount of saturation from light to dark signifies the degree of correlation between the song lyrics and each of the affective terms. The contribution of each emotional component apparent in the overall LSA values of the lyrics can be made out when considering their distribution as single pixels over time triggered by the individual lines in each of the songs. When analyzing which emotional components appear predominant and overall contribute the most, the LSA plots can roughly be grouped into three categories which can be characterized as *unbalanced distributions*, *centered distributions,* and *uniform distributions*.

Going back to the song "Nothing else matters," Figure 4, the plot exemplifies the first *unbalanced* category by in this case having a bottom-heavy distribution of emotional components biased toward *melancholy*. The below curve of accumulated LSA values indicates the contribution of each component over the entire song, where the significant aspects of *melancholy* are clearly separated from the other components.

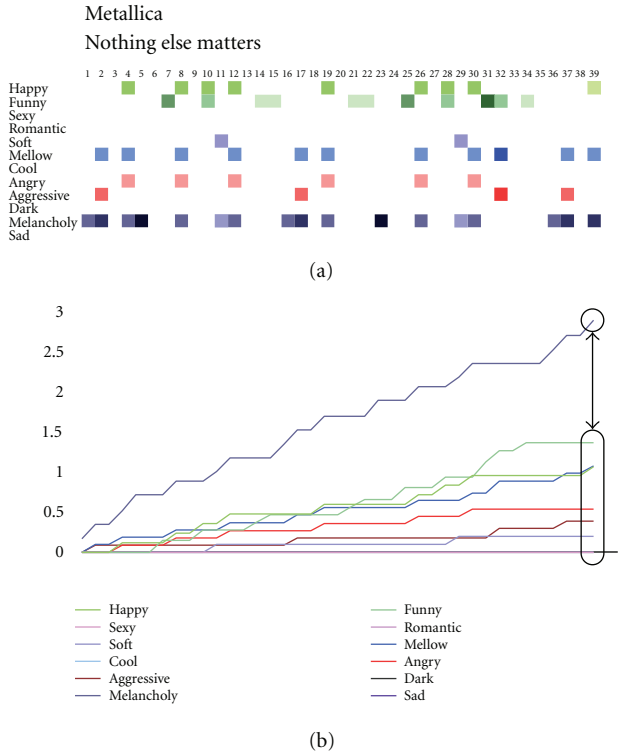


FIGURE 4: LSA correlation between (a) the lyrics of the song "Nothing else matters" and 12 affective terms, with (b) accumulated values plotted over the entire length of the song.

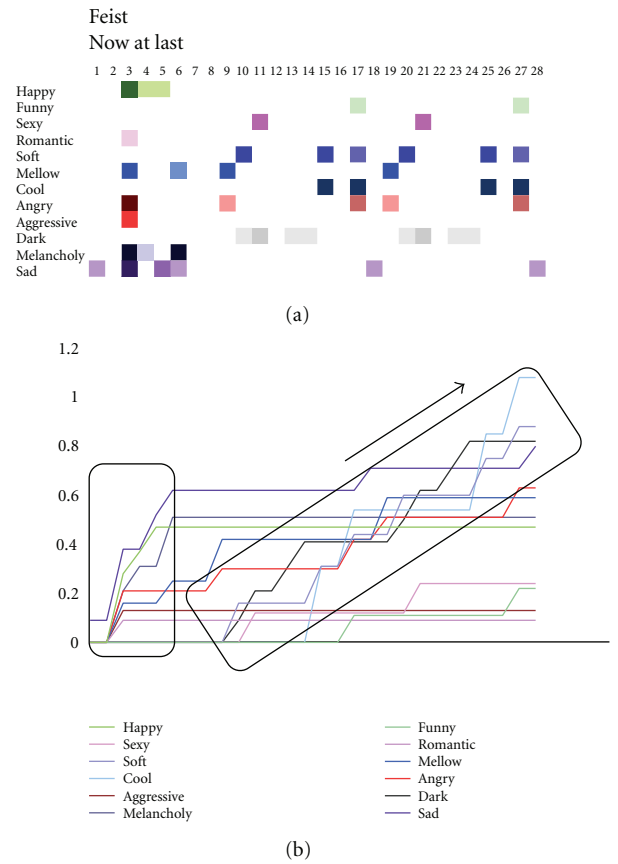


FIGURE 5: Summed up values of LSA correlation between (a) the lyrics of the song "Now at last" and 12 affective terms, with (b) accumulated values plotted over the entire length of the song.

The *centered distribution* distribution as found in "Now at last" (Figure 5) shows a lack of the more explicit emotions like "happy" or "sad" apart from the very beginning, while instead the main contribution throughout the song comes from more passive "mellow" and "soft" aspects. In contrast to the former example, the below curves of accumulated emotional contributions reflect a pattern combining the activation of "happy" or "sad" elements which remain at the initial level, whereas the more passive aspects "mellow" and "soft" are continuously accumulating throughout the song.

A *uniform distribution* of a wide range of simultaneous emotional components is exemplified by "mad world," Figure 6, simultaneously juxtaposing emotional areas around "happy" against "sad" components. This pattern can also be made out in the below curves, where additionally the sudden steep increase in accumulated values starting roughly a third into the song also illustrates how the emotional components reflect the overall structure in the song.

The overall saturation defining the amount of correlation between lyrics and emotional markers, as well as the distributional patterns of emotional components throughout the songs seem consistent. Lyrics that appear more or less saturated in relation to the emotional markers used for the LSA analysis remain so over the entire song. The distributional patterns of emotional elements seem throughout the songs to form consistent schemas of contrasting elements, which appear to form sustained lines or clusters that are preserved as pattern once initiated. We suggest that these elements form bags of features, which could be used to categorize and infer patterns as a basis for building emotional playlists. From these features, general patterns emerge, as in the distributions of emotional components in the songs "Wonderwall" and "My Immortal," Figure 7, which appear similar due to a sparsity of central aspects like "soft," while instead emphasizing the outer edges by juxtaposing elements around "happy" against "sad." The opposite character can be seen in the distributions of central elements stressed in the songs "Falling slowly" and "Stairway to heaven," Figure 8, which underline the aspects of "soft" and "mellow" at the expense of "happy" and "sad." Whereas these elements in the songs "Everybody hurts" and "Smells like teen spirit," Figure 9, appear as structural components grouped into clusters, either providing a strong continuous activation of complementary feelings or juxtaposing these emotional components against each other.

5. RESULTS: BBC SYNOPSIS

Repeating the approach, but this time to extract emotions from texts describing TV programs, we take a selection of short BBC synopses as input, and compute the cosine

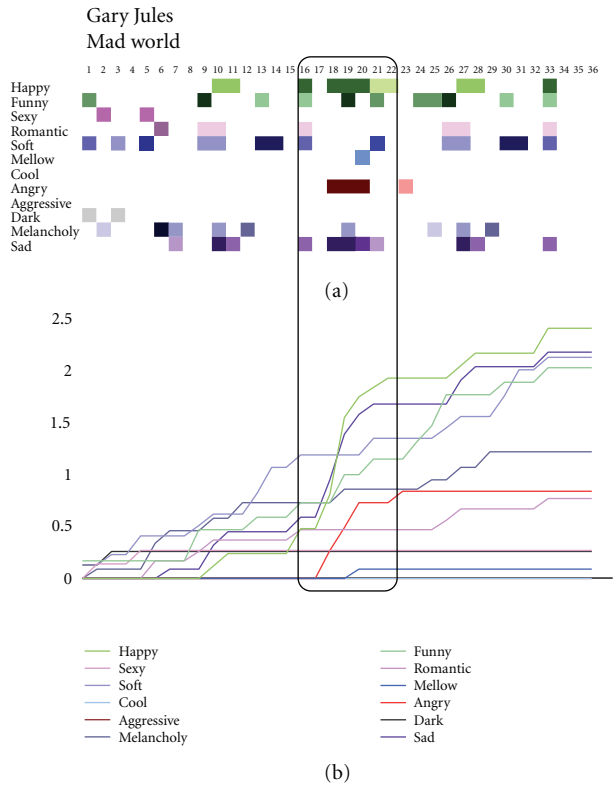


Figure 6: Summed up values of LSA correlation between (a) the lyrics of the song "Mad world" and 12 affective terms, with (b) accumulated values plotted over the entire length of the song.

similarities between a synopsis text vector and each of the selected *last.fm* emotional words. While the previously analyzed lyrics could be seen as integral parts of the original media, a synopsis description is clearly not. It only provides a brief summary of the program, but it nevertheless offers an actual description complementary to the associated *TV-Anytime* metadata genres. We initially analyzed a number of standalone synopsis descriptions to see if would be possible to capture emotional aspects of the BBC programs.

An analysis of the program "News night," based on the short description: *News in depth investigation and analysis of the stories behind the day('s) headline*, triggers the tags "funny" and "sexy" which might not immediately seem a fitting description, probably caused by these emotional terms being directly correlated with the occurrence of the words stories and news within the synopsis. The atmosphere of the lifestyle program "Ready Steady Cook!" might be somewhat better reflected in the synopsis: *Peter Davidson and Bill Ward challenge celebrity chefs to create mouth watering meals in minutes*, which triggers the tag "romantic" as associated with meals. Another singular emotion can be retrieved from the documentary "I am a boy anorexic," which based on the synopsis: *Documentary following three youngsters struggling to overcome their obsessive relationship with food as they recover inside a London clinic and then return to the outside world*, triggers the affective term "dark." We find a broader emotional spectrum reflected in the lifestyle program "The flying gardener" described by the text: *The flying gardener Chris travels around by helicopter on a mission to find Britain('s) most inspirational gardens. He helps a Devon couple create a beautiful spring woodland garden. Chris visits impressive local gardens for ideas and reveals breathtaking views of Cornwall from the air*. The synopsis triggers a concentration of passive pleasant *valence* elements related to the words "soft, mellow" combined with "happy." In this context also the tag "cool" comes out as it has a strong association to the word air contained in the synopsis, while the activation of the tag "aggressive" appears less explainable. This cluster of pleasant elements is lacking in the LSA analysis of the program "Super Vets" which instead evokes a strong emotional contrast based on the text: *At the Royal Vet College Louis the dog needs emergency surgery after a life threatening bleed in his chest and the vets need to find out what is causing the cat fits*, where both pleasant and unpleasant active terms like "happy" and "sad" stand out in combination with strong emotions reflected by the tag "romantic." And as can be seen from programs like "The flying gardener" and "Super Vets" (Figure 10), the correlation between the synopsis and the chosen tags might often trigger both complementary elements as well as contrasting emotional components.

We proceeded to explore whether we could sum up a distinct pattern reflecting an emotional profile pertaining to a TV series, by accumulating the LSA values of correlation between synopsis texts and emotional tags over several episodes. Similar to our previous approach when analyzing lyrics, where we held the LSA results against the user defined *last.fm* tag clouds, we here compare the LSA values of the synopsis against the *TV-Anytime* atmosphere genres used in the BBC metadata. This classification scheme offers 53 different terms which might be included in the genre metadata to express the atmosphere or perceived emotional response when watching a program. Projecting the synopsis descriptions against 53 *TV-Anytime* terms, used as emotional markers in the LSA analysis, allows for defining more differentiated patterns. At the same time also projecting the BBC synopsis against the previously used *last.fm* tags in the LSA analysis, makes it possible to compare to what extent the choice of using either *TV-Anytime* atmosphere terms or *last.fm* tags as emotional markers in the semantic space is influencing the results.

For analyzing the emotional context in a sequence of synopsis descriptions of the same program, we chose the soap "East Enders," the comedy "Two pints of lager," and sci-fi series "Doctor Who." Initially, plotting the LSA analysis of the soap "East Enders" and comedy "Two pints of lager" against 12 *last.fm* tags (Figures 1 and 2, increased color saturation corresponds to degree of correlation), the distributions of emotional components appear unbalanced in both cases. But whereas the soap has a bottom-heavy bias toward "sad" and "angry" outweighing "happy," the balance is reversed in the comedy which shifts towards predominantly "happy" and "funny" complemented by "soft" and "mellow" aspects. Overall, the distribution in "East Enders" is much more dense and emotionally saturated as

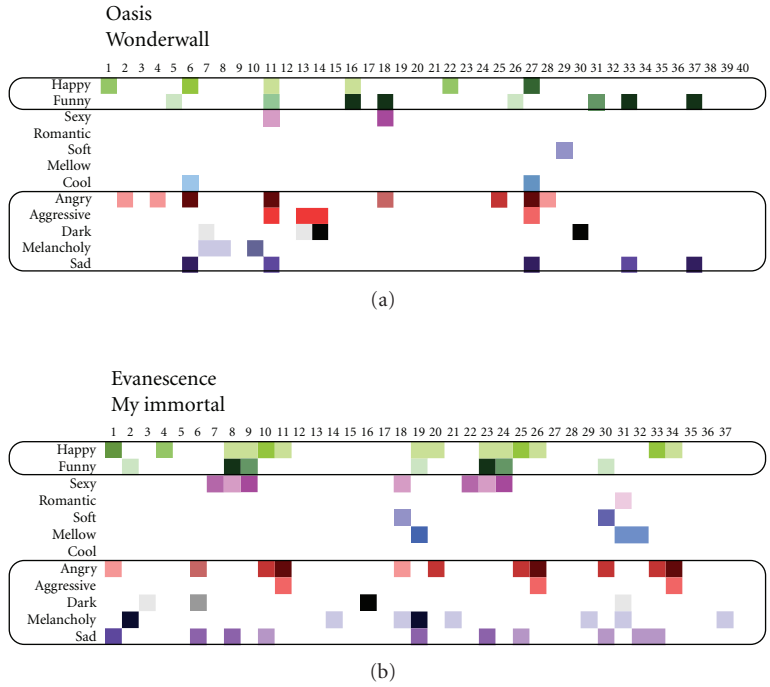


Figure 7: Pairwise comparison of patterns reflecting LSA correlation values in the lyrics of the songs (a) "Wonderwall", and (b) "My immortal" against 12 affective terms.

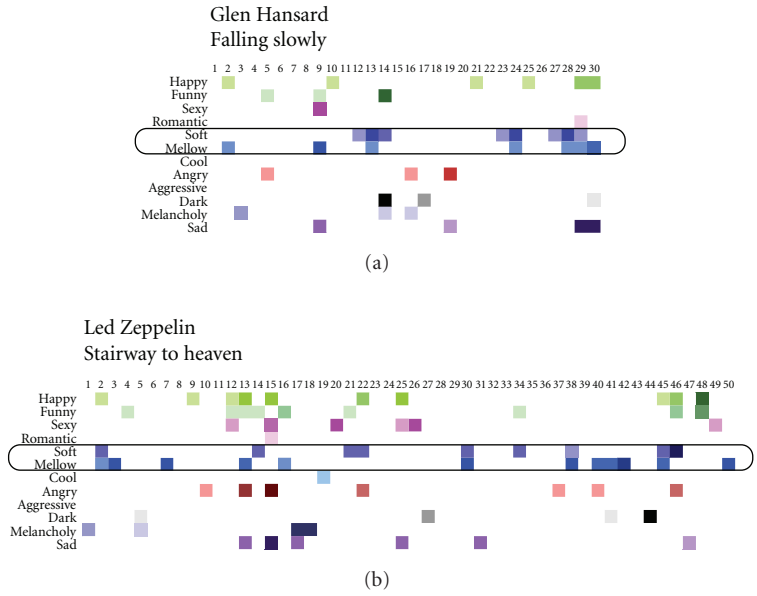


Figure 8: Pairwise comparison of patterns reflecting LSA correlation values in the lyrics of the songs (a) "Falling slowly", and (b) "Stairway to heaven" against 12 affective terms.

exemplified in elements like "angry" reflecting high arousal. In contrast, the lighter character of "Two pints of lager" comes out in the clustering of positive valence elements such as "happy" and "funny," coupled with a general sparsity of excitation within the matrix.

As a second step, projecting the synopsis descriptions against the 53 *TV-Anytime* atmosphere terms of course results in more differentiated patterns. Users at *last.fm* frequently describe tracks as "angry" but as music is rarely described as scary, feelings of fear are lacking. Otherwise,

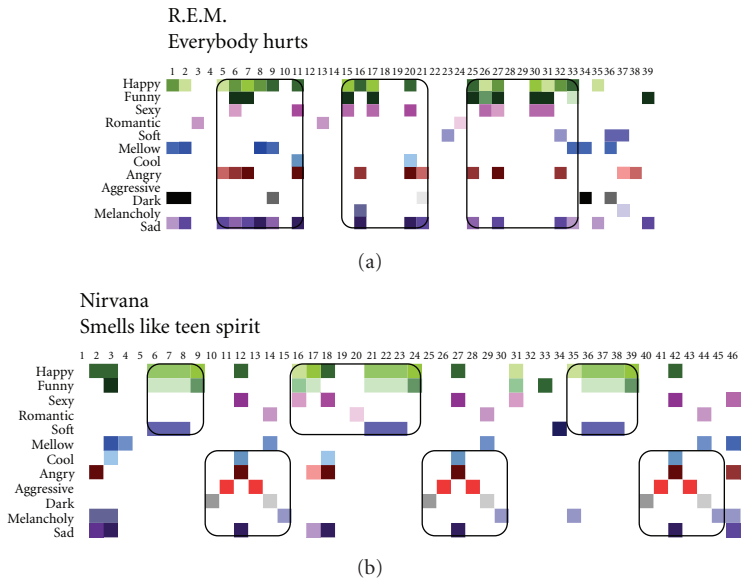


FIGURE 9: Pairwise comparison of patterns reflecting LSA correlation values in the lyrics of the songs (a) "Everybody hurts", and (b) "Smells like teen spirit" against 12 affective terms.

so with the *TV-Anytime* metadata which also captures these aspects in a synopsis with atmosphere terms like "terrifying." Some of these elements are essential for describing the content as is evident in the sci-fi series "Doctor Who," Figure 13. Lacking words for these feelings, the *last.fm* tags "Melancholy" and "dark" are triggered, whereas it takes the increased resolution of the *TV-Anytime* atmosphere terms to capture the equally "spooky" and "silly" aspects.

Altogether *TV-Anytime* adds a large number of terms, which rather than describing emotions capture attitudes or perceived responses like "stylish" or "compelling," and as such trigger vast amounts of elements contributing to the atmosphere. In "East Enders" adding elements like "frantic" and "exciting" to the pattern. Similarly, the larger number of comical elements exemplified by words like "crazy, silly," or "wacky" provides a much higher emotional granularity in the description of "Two pints of lager". However, the overall bias toward positive or negative valence and arousal within the distributions seem largely preserved, independent of whether *last.fm* or *TV-Anytime* terms are used as emotional markers in the LSA analysis.

Comparing the emotional components retrieved from the LSA analysis of the synopsis texts against the actual *TV-Anytime* atmosphere terms in the BBC metadata, they seem to be largely in agreement. The comedy has been indexed as "humorous, silly, irreverent, fun, wacky, crazy," while based on the synopsis texts alone, most of these components also come out in the LSA analysis. In the case of the soap "East Enders," the episodes are annotated as "gripping, gritty, gutsy." Although these terms are also triggered from the synopsis texts, these aspects might be even more reflected in the stark accumulated contrasts of "happy" and "sad" components retrieved by the LSA analysis. Similarly, in "Doctor Who" the actual *TV-Anytime* atmosphere terms applied in the BBC metadata *spooky, exciting* are also captured, while the grey patterns of perceived responses seem to add a lot more nuances to this description.

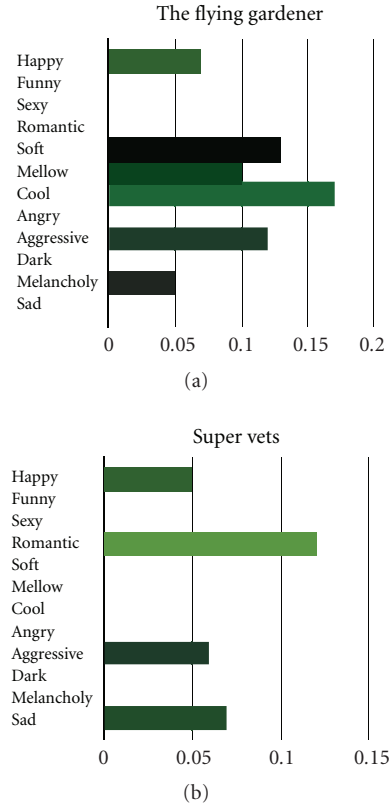


FIGURE 10: LSA cosine similarity between the synopsis descriptions of "The flying gardener" and "Super Vets" against 12 frequently used *last.fm* affective terms.

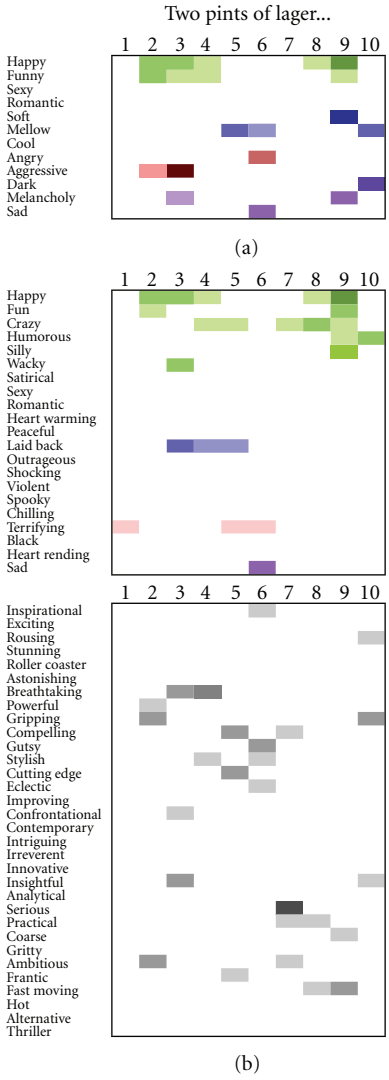


Figure 11: LSA correlation values of 10 episodes of (a) "Two Pints of lager" against 12 last.fm tags, and (b) 53 tva atmosphere terms.

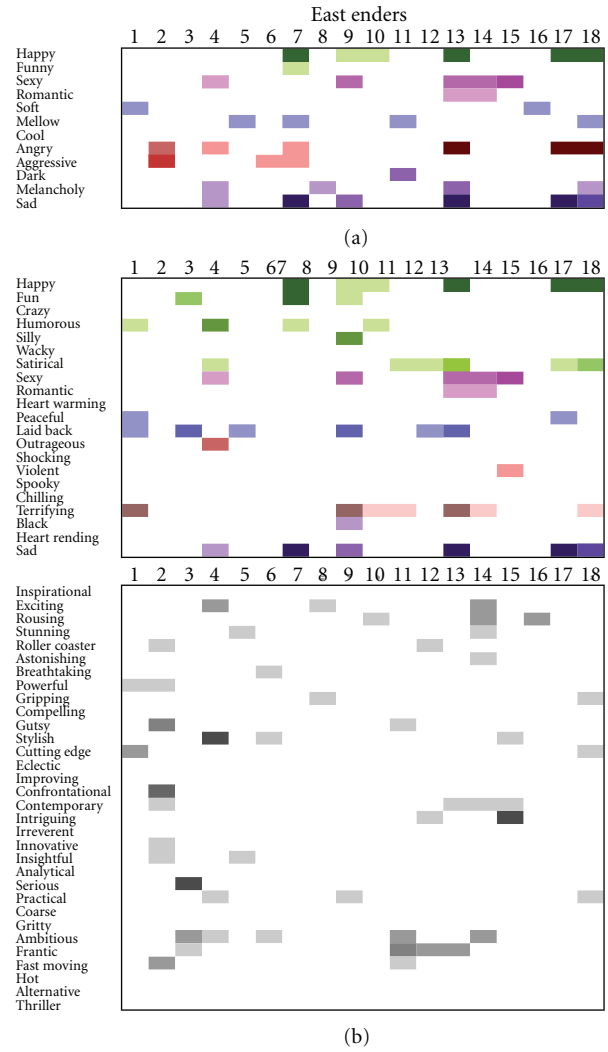


Figure 12: LSA correlation values of 18 episodes of (a) "East Enders" against 12 last.fm tags, and (b) 53 tva atmosphere terms.

6. CONCLUSIONS

Projecting BBC synopsis descriptions into an LSA space, using both *last.fm* tags and *TV-Anytime* atmosphere terms as emotional buoys Figures 11–13, we have demonstrated an ability to extract patterns reflecting combinations of emotional components. While each synopsis triggers an individual emotional response related to a specific episode, general patterns still emerge when accumulating the LSA correlation between synopsis and emotional tags over consecutive episodes, which enables us to differentiate between a comedy and a soap based on textual descriptions alone. Applying more semantic markers in the analysis allows for capturing additional elements of atmosphere in terms of perceived attitudes or responses to the media being consumed. However, the overall balance of affective components reflecting the media content seems largely preserved, independent of whether *last.fm* or *TV-Anytime* terms are used as emotional markers in the LSA analysis.

Moving beyond the static LSA analysis of consecutive synopsis descriptions, plotting the components over time might provide a basis for modelling the patterns of emotions evolving when we perceive media. We hypothesize that these emotional components reflect compositional structures perceived as patterns of tension and release, which form the dramatic undercurrents of an unfolding story line. As exemplified in the plots of song lyrics each matrix column corresponds to a time window of a few seconds, which is also the approximate length of the high-level units from which we mentally construct our perception of continuity within time [26]. Interpreted in that context, we suggest that the LSA analysis of textual components within a similar size of time window is able to capture a high level representation of the shifting emotions triggered by the media. Or from a cognitive perspective, the dimensionality reduction enforced by LSA might be interpreted as a simplified model of how mental concepts are constrained by the strengths of links connecting nodes in our working memory [27].

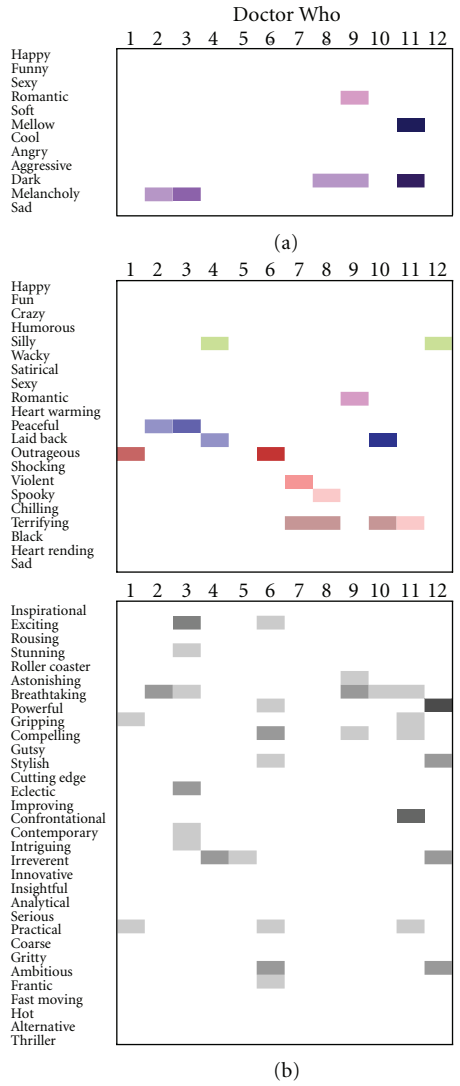


FIGURE 13: LSA correlation values of 12 episodes of (a) "Doctor Who" against last.fm tags, and (b) 53 tva atmosphere terms.

Finding that the emotional context of media can be retrieved by using affective terms as markers, we propose that LSA might be applied as a basis for automatically generating mood-based recommendations. It seems that even if we turn off both the sound and the visuals, emotional context as well as overall formal structural elements can still be extracted from media based on latent semantics.

REFERENCES

[1] ETSI, TV-Anytime. Part 3. Metadata 1. Sub-part 1. Part 1—Metadata schemas, TS 102822-3-1, 2006.

[2] A. Butkus and M. K. Petersen, "Semantic modelling using TV-anytime genre metadata," in *Proceedings of the 5th European Conference on Interactive TV: A Shared Experience (EuroITV '07)*, P. Cesar, K. Chorianopoulos, and J. F. Jensen, Eds., vol. 4471 of *Lecture Notes in Computer Science*, pp. 226–234, Springer, Amsterdam, The Netherlands, May 2007.

[3] M. Levy and M. Sandler, "A semantic space for music derived from social tags," in *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR '07)*, pp. 411–416, Vienna, Austria, September 2007.

[4] X. Hu, M. Bay, and S. J. Downie, "Creating a simplified music mood classification ground-truth set," in *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR '07)*, pp. 309–310, Vienna, Austria, September 2007.

[5] A. Bär, A. Berger, S. Egger, and R. Schatz, "A lightweight mobile TV recommender," in *Proceedings of the 6th European Conference on Interactive TV: A Shared Experience (EuroITV '08)*, M. Tscheligi, M. Obrist, and A. Lugmayr, Eds., vol. 5066 of *Lecture Notes in Computer Science*, pp. 143–147, Springer, Salzburg, Austria, July 2008.

[6] J. Wang, L. Duan, L. Xu, H. Lu, and J. S. Jin, "TV ad video categorization with probabilistic latent concept learning," in *Proceedings of the 9th ACM SIG Multimedia International Workshop on Multimedia Information Retrieval (MIR '07)*, pp. 217–226, Bavaria, Germany, September 2007.

[7] T. K. Landauer and S. T. Dumais, "A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge," *Psychological Review*, vol. 104, no. 2, pp. 211–240, 1997.

[8] K. Skreiner, "In the news: machine learning takes on the brain," *IEEE Intelligent Systems*, vol. 23, no. 3, pp. 7–8, 2008.

[9] T. M. Mitchell, S. V. Shinkareva, A. Carlson, et al., "Predicting human brain activity associated with the meanings of nouns," *Science*, vol. 320, no. 5880, pp. 1191–1195, 2008.

[10] D. J. Levitin and V. Menon, "Musical structure is processed in "language" areas of the brain: a possible role for Brodmann Area 47 in temporal coherence," *NeuroImage*, vol. 20, no. 4, pp. 2142–2152, 2003.

[11] S. Koelsch and W. A. Siebel, "Towards a neural basis of music perception," *Trends in Cognitive Sciences*, vol. 9, no. 12, pp. 578–584, 2005.

[12] N. Steinbeis and S. Koelsch, "Shared neural resources between music and language indicate semantic processing of musical tension-resolution patterns," *Cerebral Cortex*, vol. 18, no. 5, pp. 1169–1178, 2008.

[13] L. R. Slevc, J. C. Rosenberg, and A. D. Patel, "Language, music and modularity, evidence for shared processing of linguistic and musical syntax," in *Proceedings of the 10th International Conference on Music Perception & Cognition (ICMPC '08)*, Sapporo, Japan, August 2008.

[14] D. Schön, R. L. Gordon, and M. Besson, "Musical and linguistic processing in song perception," *Annals of the New York Academy of Sciences*, vol. 1060, pp. 71–81, 2005.

[15] V. Gallese, "Embodied simulation: from neurons to phenomenal experience," *Phenomenology and the Cognitive Sciences*, vol. 4, no. 1, pp. 23–48, 2005.

[16] V. Gallese and G. Lakoff, "The brain's concepts: the role of the sensory-motor system in conceptual knowledge," *Cognitive Neuropsychology*, vol. 22, no. 3-4, pp. 455–479, 2005.

[17] I. Molnar-Szakacs and K. Overie, "Music and mirror neurons: from motion to 'e' motion," *Social Cognitive and Affective Neuroscience*, vol. 1, no. 33, pp. 235–241, 2006.

[18] L. B. Meyer, "Meaning in music and information theory," *Journal of Aesthetics and Art Criticism*, vol. 15, no. 7, pp. 412–424, 1957.

[19] D. Temperley, *Music and Probability*, MIT Press, Cambridge, Mass, USA, 2007.

[20] D. Huron, *Sweet Anticipation: Music and the Psychology of Expectation*, MIT Press, Cambridge, Mass, USA, 2006.

[21] R. Jackendoff and F. Lerdahl, "The capacity for music: what is it, and what's special about it?" *Cognition*, vol. 100, no. 1, pp. 33–72, 2006.

[22] C. L. Krumhansl, "Music: a link between cognition and emotion," *Current Directions in Psychological Science*, vol. 11, no. 2, pp. 45–50, 2002.

[23] I. Peretz, R. Gagnon, and S. Hebert, "Singing in the brain: insights from cognitive neuropsychology," *Music Perception*, vol. 21, no. 3, pp. 71–81, 2004.

[24] I. Peretz, M. Radeau, and M. Arguin, "Two-way interactions between music and language: evidence from priming recognition of tune and lyrics in familiar songs," *Memory and Cognition*, vol. 32, no. 1, pp. 142–152, 2004.

[25] M. M. Bradley and P. J. Lang, "Afective norms for English words (ANEW): stimuli, instruction manual and affective ratings," Tech. Rep. C-1, The Center for Research in Psychophysiology, University of Florida, Gainesville, Fla, USA, 1999.

[26] E. Pöppel, "A hierarchical model of temporal perception," *Trends in Cognitive Sciences*, vol. 1, no. 2, pp. 56–61, 1997.

[27] W. Kintsch, *Comprehension—A Paradigm for Cognition*, Cambridge University Press, Cambridge, UK, 1998.