Universal Multimedia Experiences for Tomorrow

Fernando Pereira and Ian Burnett

©DIGITAL VISION

The previous decade has seen a variety of trends and developments in the area of communications and thus multimedia access. While individual, isolated developments produced small advances on the status quo, their combination and cross-fertilization resulted in today's complex but exciting landscape. In particular, we are beginning to see delivery of all types of data for all types of users in all types of conditions. This article discusses the current status of universal multimedia access (UMA) technologies and investigates future directions in this area.

Recent Key Developments

Key developments and trends from the last few years have set the scene for ubiquitous multimedia consumption. In summary, these are

▲ wireless communications and mobility
▲ standardized multimedia content
▲ interactive versus passive consumption
▲ the Internet and the World Wide Web (WWW).

Wireless Communications and Mobility

An important step in the progression of ubiquitous multimedia has been the accomplishment of (almost) full mobility. With the explosion of mobile networks and

1053-5888/03/$17.00©2003IEEE

terminals, users can now realize the dream of being constantly online. Originally, only voice communication was feasible, but limited Internet access and video telephony are now possible. The impact that 24/7 communication has had on everyday life can be gauged by the amazing levels of mobile phone penetration. Already, phone usage is reaching saturation in several countries. Moreover, mobile terminals (especially with the emerging third-generation mobile technology) are becoming increasingly sophisticated, offering additional services, notably multimedia-centered services.

Standardized Multimedia Content

A major factor in emerging communication networks and devices has been the explosion of multimedia data (image, video, graphics, and music) as opposed to simple text and speech. Until recently, and with the exception of broadcast television and radio, voice was still the sole communication mechanism. The diffusion of digital processing algorithms and hardware has brought images, music, and video into everyday life. The availability of open standards [such as JPEG, MPEG-1/-2 Audio (includes MP) and Video, H.261, and H.263] has had a major impact on this progression. Such standards have made the creation and communication of (digital) data aimed at our most important senses—sight and hearing—simple, inexpensive, and commonplace. Video and music are no longer the domain of special-purpose devices (analogue TVs and radios, records, and cassette tapes) but have invaded communication networks, computing, and consumer electronics devices. In this sense, digital multimedia has empowered consumers. They can create CDs and even DVDs in their homes with a quality and professionalism that was, only a few years ago, in the dreams of studio engineers.

Interactive Versus Passive Consumption

Another vital feature of human interface with the environment has also found its way into communications applications and devices: interactivity. While the consumption of audio and video has been passive for many decades, the Internet adventure and the diffusion of games have shown the importance of a deeper, interactive relationship between the user and multimedia content. This generated an expectation of multimedia beyond passive television viewing. Interactivity also invaded that media with hundreds of cable TV channels today providing interactive capabilities for hyperlinking, voting, chatting, and the like. In general, interactivity provides the capability of tuning the content according to the user's needs and wishes, in short, personalizing the content to offer more individual and private experiences.

The Internet and the WWW

Finally, there is the Internet. The way users think today about multimedia content is strongly determined by the Internet phenomena. Broad content availability, simple yet powerful interaction with the content, and easy access have transformed the Internet into a part of everyday life. Perhaps the next development stage is even more challenging: integration into other familiar devices and technology. Already the Internet refrigerator and air conditioner have appeared; but in terms of multimedia, the integration of the Internet with other forms of multimedia delivery is only just beginning. Currently, this is limited to linking to WWW sites from prepackaged media (such as CDs or DVDs) but the major challenge lies in broadcasting. The strength of the Internet is that it provides versatile bidirectional interactivity and allows peer-to-peer communication. This personalized experience marks a jump from broadcasting to narrowcasting, where the same content is distributed to everybody but is tuned to a consumer's personal context. As yet this is limited to WWW pages, but soon it will extend to full multimedia content.

The UMA Problem

While there is a wealth of audio and visual data on the Internet today, it is increasingly difficult to find the information we need (in spite of the significant advances in terms of search engines). The gap between our expectation of information and the delivered information increases daily, at least for those less familiar with search engines. Description, structure, and management of information is becoming critical. One solution is offered by portals, the Web virtual libraries and shopping centers, which provide users gateways to organized and specific multimedia domains. This growing body of structured information (even if that structure is still limited) increasingly needs to be accessed from a diverse set of networks and terminals. The latter range (with increasing diversity) from gigabit Ethernet-connected workstations and Internet-enabled TV sets to mobile watchlike video-enabled terminals (see Figure 1). The variety of delivery mechanisms to those terminals is also growing; currently, these include satellite, radio broadcasting, cable, mobile, and copper using xDSL. At the end of the distribution path are the users, with different devices, preferences, locations, environments, needs, and possibly disabilities.

Figure 1 highlights that, in a heterogeneous world, the delivery path for multimedia content to a multimedia terminal is not straightforward. The notion of UMA addressed in this issue of *IEEE Signal Processing Magazine* calls for the provision of different presentations of the same content/information, with more or less complexity, suiting the different usage environments (i.e., the context) in which the content will be consumed. (See "Universal Multimedia Definitions.") "Universal" applies here to the user location (anywhere) and time (anytime) but also to the content to be accessed (anything), even if that requires some adaptation to occur. UMA requires a general understanding of personalization involving not only the user's needs and preferences but also the capabilities of the user's environment (e.g., the network characteristics; the terminal where the content will be presented; and the natural environment where a user is located, such as the location, temperature, and altitude).

Technologies that will allow a UMA system to be constructed are just starting to appear. Among the most relevant are adaptation tools that process content to fit the characteristics of the consumption environment. These adaptation tools have to consider individual data types (e.g., video or music) as well as structured content, such as portals and MPEG-21 digital items [1]. Thus, adaptation extends from individual multimedia objects to the presentation of multiple, structured elements. Content and usage environment or context descriptions (both metadata) are central to content adaptation since they provide information that can control a suitable adaptation process. For interoperable adaptation, some tools will need to be or are being standardized; examples are content [2] and usage environment description [3], delivery protocols, and rights expression mechanisms [1]. Today, UMA service deployment is limited not only by network and terminals bottlenecks but also by the lack of standard technologies that allow some services to hit mass markets at acceptable prices, e.g., mobile video streaming.

From Access to Experiences

It's Better Than a Fish in Your Ear

Multimedia delivery is evolving from simple user content access to the delivery of a "best experience" to a user in a given context. This concept involves much more than just terminals, networks, and moves, for example, into the psychology of a user's experience of multimedia information. While today's UMA technologies are offering adaptation to a terminal, tomorrow's technologies will provide users with adapted and informative universal multimedia experiences (UMEs). This is the critical difference between UMA and UMEs: the latter clearly acknowledge that the end point of universal multimedia consumption is the user and not the terminal. With this, mass media becomes mass customization.

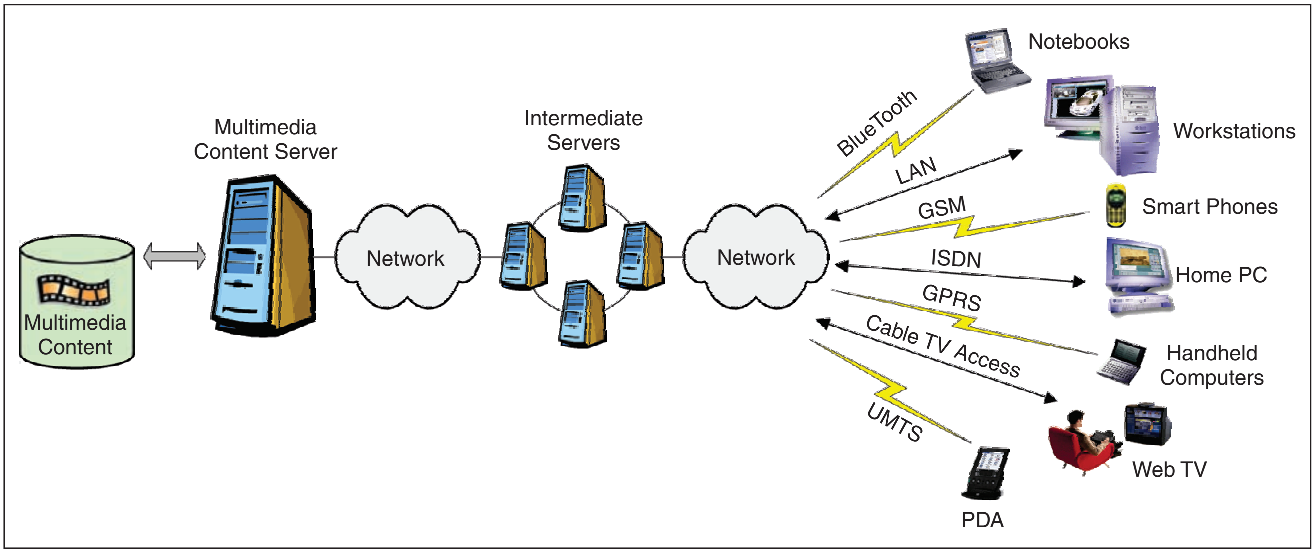
Universal Multimedia Definitions

Universal Multimedia Access (UMA): The notion (and associated technologies enabling) that any content should be available anytime, anywhere, even if after adaptation. This may require that content be transcoded from, for example, one bit rate or format to another or transcoded across modalities; e.g., text to speech. UMA concentrates on altering the content to meet the limitations of a user's terminal or network.

Universal Multimedia Experience (UME): The notion that a user should have an equivalent, informative experience anytime, anywhere. Typically, such an experience will consist of multiple forms of multimedia content. Each will be adapted as in UMA but rather than to the limits of equipment, to limits that ensure the user has a worthwhile, informative experience. Thus, the user is central and the terminal and network are purely vehicles of the constituent content.

In March 1978, BBC Radio broadcast the first part of the humorous science fiction radio series *The Hitchhiker's Guide to the Galaxy* (later a TV and book series) by the late Douglas Adams. The following is a short excerpt [4], in which the human Arthur finds himself in a spaceship hold with a new friend, Ford:

> Suddenly a violent noise leapt at them from no source that he could identify. He gasped in terror at what sounded like a man trying to gargle while fighting off a pack of wolves.
>
> "Shush!" said Ford. "Listen, it might be important."
>
> "Im.....important?"



▲ *1. Different terminals access rich multimedia content through different networks.*

> "It's the Vogon captain making an announcement on the Tannoy."
>
> "You mean that's how Vogons talk?"
>
> "Listen!"
>
> "But I can't speak Vogon."
>
> "You don't need to. Just put this fish in your ear."
>
> …
>
> [Arthur] was still somehow listening to the howling gargles, he knew that, only now it had somehow taken on the semblance of perfectly straightforward English.
>
> …
>
> "What's this fish doing in my ear?"
>
> "It's translating for you. It's a Babel fish. Look it up in the book if you like."
>
> "The Babel fish," said *The Hitchhiker's Guide to the Galaxy* quietly, "is small, yellow and leech-like, and probably the oddest thing in the Universe. It feeds on brainwave energy received not from its own carrier but from those around it. It absorbs all unconscious mental frequencies from this brainwave energy to nourish itself. It then excretes into the mind of its carrier a telepathic matrix formed by combining the conscious thought frequencies with nerve signals picked from the speech centers of the brain which has supplied them. The practical upshot of all this is that if you stick a Babel fish in your ear you can instantly understand anything said to you in any form of language …"

While Douglas Adams' *Hitchhiker* series is worthwhile reading for its humor content, the interesting point for us is that the Babel fish is a perfect example of a wearable computer and transparent multimedia adaptation system; it automatically translates speech to meet a user's requirements. But beyond simple adaptation, the Babel fish is so good at its job that it generates an adaptive multimedia experience based on both the content and usage. Today such systems are beginning to appear, though few have the ability to work in real time and certainly cannot cope with all languages. Thankfully, they also avoid the requirement for placing a fish in the user's ear!

While some of the 1970s science fiction in this excerpt remains just that—fiction—the message is clear. The "Holy Grail" of universal multimedia communications is to deliver any informative experience transparently adapted to a user's context. So, how does a user's experience relate to the multimedia content streams that we can deliver?

Experiences and Knowledge

We can easily overload people with vast quantities of multimedia data, but deriving useful information from this data is also then increasingly difficult (particularly if the data is delivered in a form unsuitable for the consumption environment). Instead, the ultimate objective of any multimedia communication system should be to provide the end user with the best meaningful "experience" of the data. Here we are loading the term "experience" with observation, participation, and sensory consumption dimensions. Thus, the primary difference between accessing the content (UMA) and ensuring a consistent user experience (UME) is a shift in focus from data delivery to the terminal to experience delivery to the users themselves. The key aspect is to ensure that our systems can deliver a meaningful experience that will deliver information to the user and, in turn, allow the user to build knowledge. Human knowledge evolves as a result of millions of individual experiences at different times and locations and in various contexts; it is this past that a user carries and adds to while consuming multimedia experiences. A profitable multimedia experience is thus one that takes the end user beyond the simple content presentation to information and knowledge; this may involve personal communication, entertainment, or education. The more powerful the experience, the higher its value in terms of resulting knowledge for the task.

Senses and Sensors

The success of telecommunications is largely related to its ability to convey and allow users to share experiences. For example, people want to tour places they have never visited. This has been the theme of many science fiction novels, and telecommunications now has the tools to deliver powerful experiences of remote environments. To capture the events and create the content, different types of sensors may be used. The closer the sensors match human senses, the more powerful and immediate the experiences that can be generated. In part, this explains why audiovisual content has to date been dominant. Sensors for touch, smell [5], [6], and taste are still comparatively early in their development. Content that includes these senses is likely to be increasingly used. Communication of information using senses other than sight and hearing will be especially relevant to delivering high-quality experiences to those with sight- and hearing-related disabilities.

While cross-modal adaptation has wide application, it will be especially useful to those with disabilities. Examples could range from the simple inclusion of a (automatically generated) sign language window for deaf people to the provision of visual experiences through the optical nerve and brain implants for blind people [7], [8]. In considering such possibilities, it is apparent that the notion of a multimedia experience mirrors the inter-relationship between senses and information; thus, the human-information interface plays a fundamental role in a UME system. While we may still be at the infancy of these interfaces for sight and hearing, other senses have hardly been considered.

As acquisition sensors and authoring tools become more sophisticated, content may become "smarter." Smart content includes programmable behavior that may comprise adaptation algorithms to be applied remotely as well as mechanisms to set the course of the experience based on events or user interaction. For example, the next advertisement may depend on the user location; if the user is close to a beach, he/she may receive ice cream advertisements.

Such adaptation algorithms may further exploit usage environment sensors to increase the level of user awareness of the experience. By detecting natural conditions such as location, time, and temperature, as well as user states such as mood and action, usage environment sensors and the information they provide will determine the efficacy of an adapted experience. For example, the sensors may provide the information that the user is at the cinema or taking a nap and, thus, the terminal will not "ring" (unless the matter is urgent). This solution where (a simple) part of the adaptation may be performed at the terminal may be an interesting way to overcome ethical and privacy issues since it avoids the need to send to the server rather private information about the user.

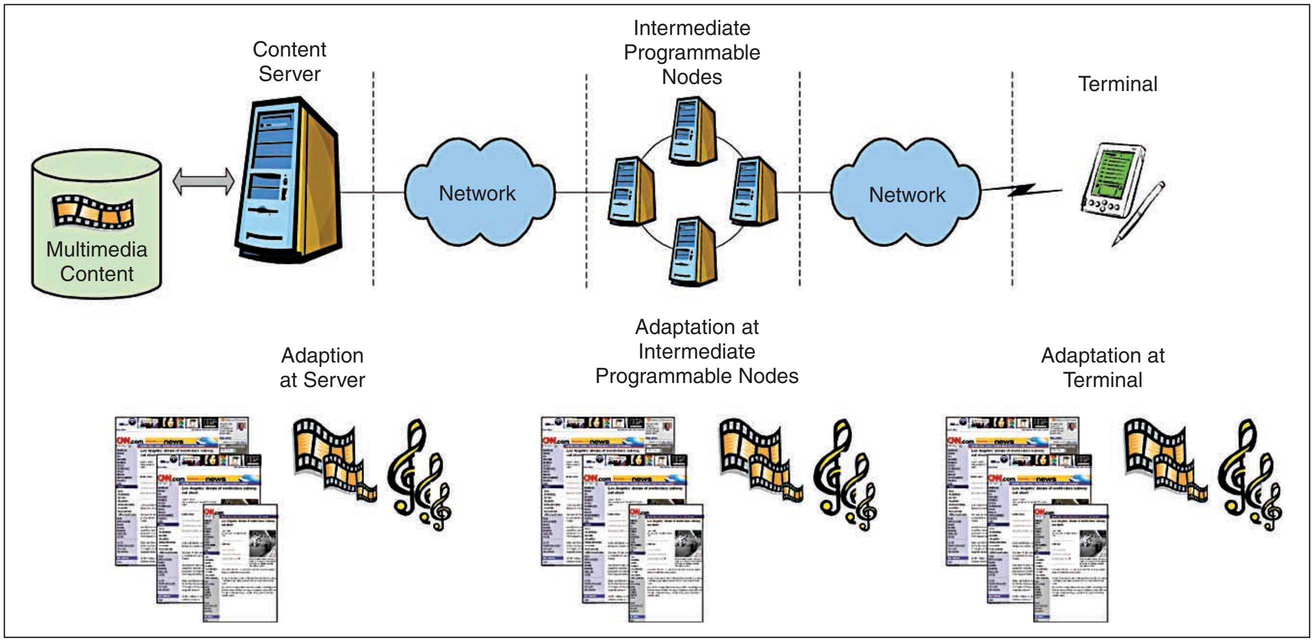
These examples quickly bring us to the notion of sensor networks. These are generally understood as a set of typically low-cost, low-power, tiny sensors that are densely deployed inside or close to the relevant environment, e.g., within a room or the human body. The sensors may have sensing, processing, and communication capabilities and should act in a cooperative way. This may include data fusion to minimize the required data sent to a central node, e.g., the global network enabled terminal. Some specific characteristics of these sensor networks (e.g., rapid deployment, self-organization, and fault tolerance) make them particularly relevant to the collection, processing, and fusion of distributed data. Although mobile ad hoc network protocols and algorithms provide the first solutions for sensor networks, many problems remain open [9], [10].

For the ultimate UME system, whatever the user's usage environment, there will always be an experience to be delivered (even if that requires many adaptation mechanisms). In some circumstances, the experience may be made more rewarding by augmenting it with virtual/synthetic content [11]. Since the key to the future is in the quality of the experience, advanced interfaces and devices will play a major role.

The processing of the content to provide the best user experience may be performed at one location or distributed over various locations. The candidate locations are the content server(s), any processing server(s) in the network, and the consumption terminal(s). The choice of the processing location(s) may be determined by several factors: transmission bandwidth, storage and computational capacity, acceptable latency, acceptable costs, and privacy and rights issues (see Figure 2).

Experience Limitations

Providing multimedia events and content anytime, anywhere, and by any means leads to the notions of ubiquitous multimedia and omnipresence of users at events. It is clear that mobile terminals have an essential role in this as they provide (almost) full mobility and clearly establish the "universal" attribute of UMEs [12]. These small, mobile terminals may also serve to highlight the limitations and boundaries of multimedia experiences in certain contexts. While it may be heartwarming to see one's children at home on a small mobile terminal, few would be enamored by watching *The Lord of the Rings* on such a tiny system. There are also rights issues involved with such adaptations, since the rights' owners of the content or events may wish to prevent adaptation to a format below some defined experiential threshold. This may have a substantial impact in terms of adaptation processing. For example, a higher-quality experience may have to be provided to the user even if they would prefer reduced cost and lower quality or, alternately, a low-quality experience may be provided to the user for free to convince them to pay for a higher-grade experience.



▲ *2. Adaptation may be performed at different places.*

Emerging and Future Trends and Technologies

Since there are many UME-relevant technologies and it is not possible to address all of them in this article, particular emphasis will be given to signal-processing-related technologies. While some of the technologies considered are already well established, providing a significant technological basis for the development and deployment of UME applications (such as content scalability and transcoding) there are still vital technologies missing for a complete UME system vision. Many of these technologies are directly related to particular usage environments. While multimedia adaptation for improved experiences is typically thought of in the context of more constrained environments (e.g., mobile terminals and networks), it is also possible that the content has to be adapted to more sophisticated environments, e.g., with three-dimensional (3-D) capabilities. Whether the adaptation processing is to be performed at the server, at the terminal, or partially at both, is something that may have to be determined case by case, depending on such criteria as computational power, bandwidth, interfacing conditions, and privacy issues.

Existing and Emerging Technologies

Scalable coding represents the original data such that certain subsets of the total coded bit stream still provide a useful representation of the original data if decoded separately. There is currently a significant number of scalable coding schemes available, each with different characteristics in terms of the coding technology, the domain in which they provide scalability (e.g., spatial, temporal, quality), the granularity of the scalability provided, and the efficiency cost of providing scalability. MPEG-4 is the content representation standard where the widest range of scalability mechanisms is available, notably in terms of data types, granularities, and scalability domains [13].

Transcoding describes the conversion of a coded (image, video, 3-D, audio, or speech) bit stream to another bit stream representing the same information in a different coding format or with the same format but with reduced quality (less bit rate), spatial resolution, frame rate (for video), or sampling rate (for audio). The fundamental issue in transcoding is to achieve these outcomes without requiring the complete decoding and reencoding of the content [14].

With the explosion of multimedia content, it soon became evident that there was a need for efficient content description (with so-called metadata) to achieve more efficient content management, retrieval, and filtering [15]. Content description is also essential for effective usage environment adaptation if systems are to avoid computationally expensive content analysis at adaptation time. This is particularly critical when content has to be adapted in real time. Rapid navigation and browsing capabilities are prime requirements of user interaction with structured content, and data abstraction techniques enable such facilities through the provision of different types of summaries. The MPEG-7 standard provides a complete and powerful solution for multimedia content description [2].

Content Representation

While there will be a necessity for future content representation tools, it is increasingly desirable to keep formats backward compatible and cross compatible, e.g., the use of common file formats for MPEG-4 and JPEG2000 [13], [16]. This ensures that storage and transport of content can be handled uniformly while maintaining the differing purposes of the actual content streams. Further, there are interesting proposals that explicitly separate the content from the “bit stream format.” In particular, several techniques are suggested for describing a bit stream in terms of XML metadata and manipulating it using XML tools [3], [17]. The idea behind this is that future systems could be implemented quickly with just the provision of high-level mappings between one bit stream format and a second. This opens the possibility of broader flexibility in content representations and transcoding while maintaining standard solutions.

Currently, the work to provide more efficient fine granularity video scalability solutions is still ongoing. For example, there is work under development based on MPEG-4 Parts 2 and 10 (video coding solutions); the latter is also known as advanced video coding (AVC) or ITU-T H.264 [18]. MPEG is also studying wavelet-based video coding scalability, notably considering the relevance that scalable coding assumes in the context of the MPEG-21 multimedia framework (see [1]).

Fine granular scalability for audio remains a significant challenge for researchers. Traditional audio and speech coders (e.g., ITU G.728 [19], ITU G.729 [20]) are bit rate-centric. For instance, audio and speech coders are separated by model dependence, but within each group certain algorithms dominate for certain bit rates and signal bandwidths. This is partly due to the perceptual audio effects exploited by various coders and the necessity of tuning these to bit rates. This has proved to be a significant obstacle in the development of single algorithm scalability. The result has been scalable algorithms that operate in limited bandwidth and quality ranges; e.g., the 3GPP adaptive multirate (AMR) coders [21], [22]. One new trend in audio coding is a shift from a low-rate compression focus to a perceptual quality focus. In particular, lossless audio coding and scalability from lossy to lossless are now considered important areas [23]. Coupled with the move to higher quality, a new focus is the delivery of multichannel audio and the scalability of such experiences to, e.g., users with headphones.

Usage Environment Description

For the user to obtain the best possible multimedia experience, it is essential that the content is adapted taking into account as many relevant usage environment parameters as possible. Typical relevant usage environment dimensions are the network, terminal, natural environment, and user; each of these dimensions may be characterized by multiple parameters. Standard usage environment description tools, just like content description tools, are needed to ensure a high degree of interoperability between terminals and servers. As mentioned in [1], MPEG-21 digital item adaptation (DIA) [3] is a relevant development in this area. In part, it targets the specification of standard description tools for the usage environment.

Since multimedia experiences are centered on the user, the incorporation of the user into the adaptation and delivery processes needs to be at a deeper level than today's typical approach. As an example, the MPEG-21 DIA specification [3] only considers a narrow set of user preferences for multimedia selection and adaptation. With the future emergence of continuously wearable devices (e.g., with the capability of measuring several human parameters such as blood pressure, cardiac rhythm, and temperature), it is possible to take a much wider variety of data into account during the multimedia adaptation process. The result will be the provision of a multimedia experience adapted precisely to the user's current characteristics, e.g., in terms of mental and physical mood or circumstance. For example, the Internet radio's wake-up music could be selected based on these parameters; this would allow an improved multimedia experience since it would not only be based on static (or slowly varying) user preferences but also on instantaneous measures of the user's conditions. This type of capability would require not only sensors to perform the physiological measurements (wireless devices with this type of capability are already available) but also a standard way to adequately describe the features. This standard solution could simply be an extension of the MPEG-21 DIA framework (currently under development) [3].

Besides the selection and standard description of the relevant features, the provision of the capabilities in question would require (nonstandard) solutions to map the physiological sets into psychological dispositions or moods with which certain types of multimedia content are associated. This would likely involve complex psycho-physiological studies so as to enrich the user/human dimension that UMEs require.

One of the big problems with context description is the large set of descriptors that could be generated. It is also clear that many descriptors will only be relevant to a small subset of applications and situations. One possible solution is intelligent agent technologies that could cope with a large, diverse set of context inputs, determine those that are relevant and thus simplify adaptation processing.

Content Adaptation

A central role in adapting multimedia content to different usage environments is played by content adaptation (see Figure 3). This encompasses a wide range of processing mechanisms, namely single object/data type transcoding of video or audio, structured content adaptation (e.g., filtering of MPEG-21 digital items), and cross-modal adaptation (e.g., conversion of video to images or text to speech). Content adaptation may be performed at various locations in the delivery chain or even distributed across several nodes. Moreover, it may be performed in real time or offline providing a set of variations from which to select at adaptation time.

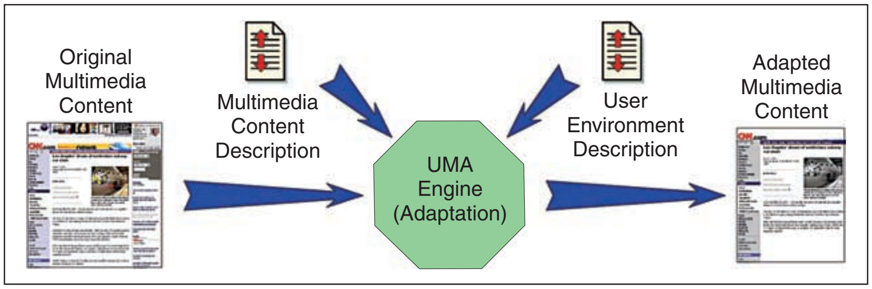
While adaptation through transcoding is well known, cross-modal adaptation is a more challenging process that may range from a simple key-frame extraction process (to transform video into images) to more complex conversions such as text to speech, speech to text, or even video to text or speech. A video-to-speech cross-modal conversion may be as simple as using the information in the textual annotation field of the associated metadata stream (e.g., MPEG-7 stream describing MPEG-2 or MPEG-4 content) or as complex as analyzing the content, recognizing some objects (e.g., for which specific models are available), and synthesizing some speech based on the recognized objects.

The adaptation of structured content such as portals or MPEG-21 digital items may be divided into two stages: first, filtering the content components to give a set suitable for the consumption environment and, second, after a certain content component is selected, adapting that component in terms of transcoding or cross-modal conversion.

Intellectual Property Management and Protection

A key factor in the delivery of multimedia content today is an increasing desire by intellectual property owners to establish, control, and protect their rights. A number of schemes for protecting music and video content have been used, most notably, the protection of DVD content and more recently attempts by the music industry to protect CDs from being copied. Such techniques have generally proved to be limited in their level of protection, and new technologies continue to be introduced. In practice, it seems likely that it is the platforms themselves that must be altered to incorporate mechanisms that protect content from copying.

Currently, various bodies are creating standards (an example is the work in MPEG on intellectual property management and protection, a rights expression language, and a data dictionary [1]) that will provide a framework


Original Multimedia Content
Multimedia Content Description
UMA Engine (Adaptation)
User Environment Description
Adapted Multimedia Content

▲ 3. Content adaptation using a UMA engine.

for digital rights management. The enforcement technologies are likely to be a significant area of research and development for the foreseeable future. For UMEs, the expression of rights to access and alter content metadata, perform (or prevent) certain adaptations, and the enforcement of those rights will be critical if content providers are to be willing to distribute material into new user environments. Further, for UMEs, intellectual property management and protection will be an issue not only for the content but also for usage environment information. Usage environment information could reveal personal information such as location and user's state of health; it is unlikely that users would be happy to have such information freely available on a network.

Presentation Conditions and Devices

In general, universal multimedia involves adaptation of high-quality/functionality content to reduced functionality usage environments, such as mobile terminals. This is a growing trend, as presentation devices become smaller, thinner, and lighter. There are more sophisticated content presentation conditions and devices that may also require adaptation processing to achieve an improved user experience. Presentation devices may be very broad in their capabilities, ranging from small mobile devices to sophisticated immersive rooms. For example, it is possible to imagine a multimedia device that is simply a pair of semitransparent (see through) glasses capable of overlaying stereoscopic visual data, both natural and synthetic; this example illustrates future human-information wearable interfaces that are both network- and signal- (image, video, and 3-D) processing enabled [11].

Presentation conditions and devices also determine the type and level of interactivity provided to the user. While simple, portable devices mostly have limited interaction capabilities, more sophisticated systems such as immersive environments [24], [25] may provide very powerful interaction mechanisms, e.g., tactile gloves to manipulate objects in virtual worlds. The presentation interface itself may also be adapted based on user preferences and skills; today, game interfaces can change depending on the user's skill level, but in the future a multimedia experience may adapt to a consumer's age, past experience, and desired outcomes.

Three-dimensional audiovisual environments provide an interesting example of advanced presentation. In terms of visual information, this may imply the adaptation of available 3-D content to specific types of 3-D or stereoscopic displays or even the processing of two-dimensional (2-D) content to provide 3-D-like visual sensations. The same 3-D content may also have to be adapted to rather simple consumption conditions, such as a personal digital assistant (PDA), targeting the provision of 3-D navigational capabilities, or even 3-D sensory impressions using glasses-based technology.

In terms of audio information, users are now expecting and experiencing near-cinema-grade sound in their living rooms. Already, amplifiers offer several modes of different "acoustic environments," such as concert, stadium, theater, and studio. This is just the beginning of the possibilities, and it is quite feasible to adapt the audio delivery to a room's acoustics to maximize a user's experience and to provide elaborate manipulation of audio objects and streams in space. Just as with video, the consumption of audio, particularly in 3-D (either simulated using stereo systems or actual multispeaker systems), is likely to become less passive and more interactive in the near future. To achieve this, sophisticated adaptation algorithms will be a vital part of ensuring a quality user experience.

Multiple Terminals at Work

While wearable computers [11], [26], [27] will give the ultimate in mobile computing power, there will still be a need for the delivery of content to devices that are impractical to "wear," e.g., large displays and audio systems. If a transparent and universal multimedia experience is to be achieved, however, users should be able to move from one environment to another seamlessly with the content automatically transferring to the relevant and optimal devices. This is an extension of the session mobility available in enhancements to some desktop environments today. It will be substantially more complex if the goal is to ensure seamless mobility of a user's experience.

If the best experience is to be provided, the user's terminal capabilities should be seen as a whole and no longer as a set of independent devices as if they were not all working for the same user. Logically, if the user has multiple terminals, it should be possible to use the terminals in combination to maximize the impact of the user experience. This means that over time and, for a mobile user, the number of terminals used to deliver an experience may vary. In turn, there is a requirement for not only seamless hand-over between terminals but also of the complete experience.

Mobile and Wearable Devices

A key aspect of the desire of users for multimedia content delivery has been the significant uptake in mobile devices both in the cellular phone and PDA markets. While it remains unclear whether the enhanced phone or wireless networked PDA will prevail, the desire for the consumption of content on such devices is clear [12]. Various companies are offering proprietary low-complexity codec solutions for the current limited processing power devices, and several have MPEG-4 decoders available. In the future, a multimedia experience may be provided to a user using multiple mobile terminals (all carried by the user) or even nonpersonal terminals (e.g., the larger screen in a coffee shop).

Beyond simple mobile devices, wearable devices are a new area of research [28]. The prospect of "suits" of high-capability processing and sensor devices offers significant possibilities for future multimedia delivery. The vision of wearable computers is to move computing from being a "primary task" to a situation where the computer is permanently available and augments another activity that the user is performing. For example, when a user walks down a street or rides a bicycle, sensors could make spatiotemporal context information available to the computer nodes to alter user interfaces and multimedia delivery. The embodiment of computers and humans into a "user" removes the typical sharp boundaries between the user and terminal and could have a profound effect on multimedia delivery. An example of this is given in [26], where advertising matter is replaced for a user (via suitable multimedia enhanced glasses) with more useful subject matter. Such systems have the potential to truly offer a transparent, augmented, and universal multimedia experience.

Active and Programmable Networks

Coupled with the increase in mobile terminals, today's passive networks that route traffic from source to destination are evolving. Increasingly, some level of processing is being offered in switches, and the extension of this concept to active or programmable network nodes provides a platform for adaptations that will be required to deliver UMEs. This may be a solution to the limited processing power available on small, mobile devices since a user (or, more likely, their agent) could request a network service provider to perform certain processing (e.g., transcoding) on their behalf. In practice, this is an extension of the current transcoding of HTML WWW content to wireless application protocol (WAP) content in some mobile networks. The potential of programmable networks is much greater, however; the programmable network node (equipped with suitable software) could be the assembly point for a UME. In such a case, an agent with suitable usage environment information would perform the necessary constituent content adaptations before delivering the UME as a set of, for example, low bandwidth content streams.

Peer-to-Peer Content Delivery

The traditional content delivery model is one where users access content from a set of servers. The "master" of the content is thus the administrator of the server and, in general, they control who has access to the content. An alternative vision of content delivery is one where every user is interconnected and any user can act as a server of content for any other user. Such peer-to-peer networks are currently being used heavily for music and file sharing [30] on the Internet. While such use is currently enshrouded in controversy due to the legal issues, there is no doubt about the power of the model.

Peer-to-peer networking has been a long-term form of communication; face-to-face conversation, telegraphy, telephony, and the postal system are all examples. It has thus been natural that peer-to-peer has quickly grown in electronic form on the Internet and in cellular systems. at first, e-mail was the prime example but now we see instant messaging, the file-sharing networks and, among cellular users, short message services (SMS) growing rapidly. The latter is already being improved with multimedia messaging services as users demand an improved peer-to-peer multimedia experience. It thus seems likely that legitimate content delivery (and our current infrastructures) will evolve to one where peer-to-peer transfer is commonplace. This will require changes in the way we consider rights and security of access to content for users.

While peer-to-peer will clearly be very important, and perhaps dominant, server-delivered content is likely to remain significant. One reason for this is the reliability of information and experience gained from legitimate and commercial content providers. Thus, we will see a mixture of the two architectures mirroring today's society where the original news content (rapidly exchanged via e-mail and SMS) likely came from a broadcaster.

The impact of a mixed peer-to-peer- and server-based architecture is that adaptation will be required both in "professional" delivery from servers as well as between individual users. Since the latter may not have the processing capability or software to deliver complex adaptations, a new breed of network services may emerge to provide content adaptation to peer-to-peer users. This further emphasizes the necessity of transparency of the users' wish to exchange content as easily as they can converse in the street (already this is a reality in the text world of SMS), but most will not have or even wish to acquire technical skills. Hence, the use of intelligent agents to ensure consistent user experiences will be a logical and necessary step.

Role of Open Standards

The need for any standard that is basically an agreement between interested parties comes from an essential requirement: interoperability. In a communication context, interoperability expresses the user's dream of exchanging any type of data and experience without any unnecessary technical barriers. There is also a commercial reality that standards create markets and, thus, manufacturers benefit significantly when their products support those standards. Without a standard way to perform some of the operations involved in the processing and communication stages of the data exchanged, easy interoperability between terminals would be impossible.

The key to effective standardization is to create a "minimum" standard that normatively defines a minimum set of tools that will guarantee interoperability. Such specifications provide space for competitive, proprietary, and alternate developments (which will be nonnormative) to be built on top of the standard. This allows for the incorporation of technical advances and thus increases the lifetime of the standard as well as stimulating technical and product competition. The existence of a standard also has important economic implications since it allows the sharing of investment costs and the acceleration of application deployment. Some people believe that open standards are the future since the relevant intellectual property (patents) will be licensable by everybody on fair and reasonable terms and under nondiscriminatory conditions. Thus, users will have access to a large variety of interoperable products from different companies, competition will improve product performance, and costs will decrease facilitating broad access to rich multimedia content. Others prefer de facto, private "standards" that may be subject to simpler licensing.

In the context of UMEs, it is clear that standards will play a central role. Technologies where standardization is and will be essential are: content representation (including scalability capabilities), content and usage environment description, transport protocols, and intellectual property management and protection of the associated descriptions and adaptations. In this context, some of the standards that are particularly important are those developed by ISO/IEC WG 11 (MPEG) and ITU-T, notably MPEG-1/-2, H.261, and H.263 for content representation; MPEG-4 for content representation and intellectual property management and protection (IPMP) [13]; MPEG-7 for content description [2]; and MPEG-21 for usage environment description [3] and more extensive intellectual property management and protection tools [1].

Limitations and Risks

While today multimedia adaptation is possible, we are only beginning the process of adapting the multimedia experience. The adaptation boundaries within which the multimedia experience for a certain content is still worthwhile or valid is a new and complex issue deserving substantial study and research. In fact, a host of questions remain: What would the user experience be like when using terminals with limited screen size and sound capabilities? What is the effect on an experience of having a small keyboard or using a voice interface? Overall, it appears that the more challenging the experience consumption environment becomes, the more critical it will be to establish the most effective forms of presenting the various multimedia elements.

Today, we have few mechanisms to measure the quality of an "experience." In image, video, speech, and audio compression, we have struggled to find objective quality measures and the success of subjective testing remains controversial in many areas; e.g., digital cinema. Yet we must now measure the quality of an experience consisting of all these elements and more. This is essential as in some cases there may need to be a quality threshold under which it is declared that no acceptable experience can be provided. Alternatively, providers could always deliver the best possible experience, but given its variability it is likely that rights holders will demand that applications incorporate some quality thresholding, or, perhaps more likely, agent technology, to make such decisions. This dilemma is particularly relevant in mobile environments where typically the consumption conditions are more challenging, and to date most mobile devices have just tried to reproduce the experiences (and applications) already available for fixed networks.

One of the key aspects, particularly for educational applications [29], is to ensure that adaptation of the multimedia content maintains the information content so that the essential knowledge can be gleaned by a user. In adapting experiences that are intended to impart precise knowledge, educational issues, such as pedagogy and how the adaptation will contribute/detract from the learning process, need consideration. This is an area where smart algorithms will be important, adapting the experience intelligently (according to a set of pedagogical as well as technical multimedia adaptation rules) such that a user will still gain the essential knowledge from the presentation.

Commercial value is also an important issue. In general, the boundaries within which users will be willing to pay for an experience must be determined to avoid negative reactions to poor-quality adaptations, below expectations. The definition of experience boundaries may be determined in advance through exhaustive testing for each piece of content, e.g., by an author who wants to explicitly prevent poor adaptations. Alternatively, they may be based on visual and aural perception as well as on psychological factors that are algorithmically expressed and, very likely, derived from empirical knowledge. These two approaches are in stark contrast; the first is cumbersome but defined while the second is risky and quite subjective. Unless the derived rules are kept conservative, which would prevent many users from getting still-useful experiences, the results will be difficult to guarantee for all types of content.

If adapting multimedia content to different usage environments involves some degree of simplification of that content, this could imply that the content will be easier to "steal" since IPMP mechanisms have been removed, simplified, or reduced in number. It may also be more appealing to steal simpler content since it is generally easier to copy and suitable for a greater variety of terminals, although the content itself may be less attractive. The key will be to persuade consumers that the quality is worth paying for, a battle that is currently being won in the DVD marketplace. Overall, these scenarios again illustrate the need for content protection and rights management tools that can control the copying and presentation of the content. The same type of protection is required for usage environment descriptions if privacy is to be guaranteed.

Final Remarks

The significant developments of recent years in terms of communications, and, more specifically, multimedia access, have been dominated by the explosion of the Internet and mobile communication services. Within the latter, deployment of digital multimedia to an increasing variety of terminals and conditions has grown rapidly. While universal multimedia delivery is still in its infancy it has already become clear that, as delivery technology evolves, the human factors associated with multimedia consumption increase in importance. In particular, the importance of the user and not the terminal as the final point in the multimedia consumption chain is becoming clear. To emphasize this fact, we have discussed UMEs as opposed to UMA, which, in our view, weakens the user's role in favor of delivery to a terminal. In the context of UMEs, the most relevant emerging and future trends and technologies have been reviewed with special emphasis on signal processing-related developments. The vision of mass delivery of identical content is being replaced by one of mass customization of content centered on the user. In considering the accomplishment of this goal, we have described a landscape for the future of multimedia creation, delivery, and consumption. It is now time to work on the development of the missing technologies and interfaces required to bring the future universal multimedia world to reality.

Acknowledgments

The authors would like to thank the anonymous reviewers for their valuable comments. Special thanks go to Rob Koenen for his very detailed and useful reviewing and to João Magalhães for the figures.

Fernando Pereira received his B.S. in electrical and computers engineering from the Instituto Superior Técnico (IST), Universidade Técnica de Lisboa, Portugal, in 1985. He received his M.Sc. and Ph.D. degrees in electrical and computers engineering from IST in 1988 and 1991, respectively. He is currently a professor at the Electrical and Computers Engineering Department of IST, where he is responsible for the participation of IST in many national and international research projects. He has contributed more than 100 papers to journals and international conferences. He has been participating in the work of ISO/MPEG for many years, where he is currently the chairman of the MPEG Requirements group. His current areas of interest are video analysis, processing, coding and description, and multimedia interactive services.

Ian Burnett received his B.Sc., M.Eng., and Ph.D. from the University of Bath, U.K., in 1987, 1988, and 1992, respectively. Since 1993, he has been with the University of Wollongong, NSW, Australia, where he is currently an associate professor and director of the Telecommuncations Research Centre. He is also an active participant and project leader in the Cooperative Research Centre for Smart Internet Technology. He has published more than 60 papers, primarily in the areas of speech and audio coding. He has been involved in the ISO/MPEG MPEG-21 activities since 2000 and has played a significant role in the development of a number of parts of the MPEG-21 standard. His current research interests lie in speech and audio coding, 3-D audio, audio separation, multimedia delivery, and structural sound synthesis.

References

[1] J. Bormans, J. Gelissen, and A. Perkis, "MPEG-21: The 21st century multimedia framework," *IEEE Signal Processing Mag.,* vol. 20, pp. **xx-xx**, Mar. 2003.

[2] B.S. Manjunath, P. Salembier, and T. Sikora, Eds., *Introduction to MPEG-7: Multimedia Content Description Language*. New York: Wiley, 2002.

[3] MPEG MDS Subgroup, "Multimedia framework—Part 7: Digital item adaptation," presented at the Awaji MPEG Meeting, Japan, Dec. 2002.

[4] D. Adams, *The Hitchhiker's Guide to the Galaxy*. London, U.K.: Pan Books Ltd., 1979.

[5] Trisenx Home Page. Available: http://www.trisenx.com

[6] AromaJet Home Page. Available: http://www.aromajet.com

[7] M. Schwarz, L. Ewe, N. Hijazi, B.J. Hosticka, J. Huppertz, S. Kolnsberg, W. Mokwa, and H.K. Trieu, "Micro implantable visual prostheses," in *Proc. 1st Annu. Int. Conf. Microtechnologies in Medicine and Biology*, 2000, pp. 461-465.

[8] M. Gross, R. Buss, K. Kohler, J. Schaub, and D. Jager, "Optical signal and energy transmission for a retina implant," in *Proc. 1st Joint BMES/EMBS Conf.*, 1999, vol. 1, pp. 476.

[9] I.F. Akyildiz, S. Weilian, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Commun. Mag.*, vol. 40, pp. 102-114, Aug. 2002.

[10] *Habitat Monitoring on Great Duck Island*. Available: http://www.greatduckisland.net/

[11] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre, "Recent advances in augmented reality," *IEEE Comput. Graph. Appl.*, vol. 21, pp. 34-47, Nov.-Dec. 2001.

[12] M. Frodigh, S. Parkvall, C. Roobol, P. Johansson, and P. Larsson, "Future generation wireless networks," *IEEE Pers. Commun*., vol. 8, pp. 10-17, Oct. 2001.

[13] F. Pereira and T. Ebrahimi, Eds., *The MPEG-4 Book*. Englewood Cliffs, NJ: Prentice-Hall, 2002.

[14] A. Vetro, C. Christopoulos, and H. Sun, "Video transcoding architectures and techniques: An overview," *IEEE Signal Processing Mag.,* vol. 20, pp. 18-29, Mar. 2003.

[15] P. van Beek, J.R. Smith, T. Ebrahimi, T. Suzuki, and J. Askelof, "Metadata driven multimedia access," *IEEE Signal Processing Mag.,* vol. 20, pp. 40-52, Mar. 2003.

[16] D.S. Taubman and M.W. Marcellin, *JPEG2000: Image Compression Fundamentals, Standards, and Practice*. Norwell, MA: Kluwer, 2001.

[17] M. Amielh and S. Devillers, "Bitstream syntax description language: Application of XML-schema to multimedia content," in *Proc. 11th Int. World Wide Web Conf. (WWW 2002)*, Honolulu, HI, May 2002.

[18] Joint Video Team, "Coding of audio-visual objects—Part 10: Advanced video coding," presented at Awaji MPEG Meeting, Japan, Dec. 2002.

[19] *Coding of Speech at 16 kbit/s Using Low-Delay Code Excited Linear Prediction*, Recommendation ITU-T G.728, Sept. 1992.

[20] *Coding of Speech at 8 kbit/s Using Conjugate-Structure Algebraic-Code-Excited Linear Prediction (CS-ACELP)*, Recommendation ITU-T G.729, Mar. 1996.

[21] *Adaptive Multi-Rate Speech Codec; Transcoding Functions,* 3GPP TS 26.090 v.4.0.0, 3GPP Technical Specification, 2001.

[22] *Adaptive Multi-Rate Wideband Speech Transcoding*, 3GPP TS 26.190, 3GPP Technical Specification, 2002.

[23] MPEG Audio Subgroup, "Final call for proposals on MPEG-4 lossless audio coding," Doc. ISO/MPEG N5208, presented at the Shanghai MPEG Meeting, Shanghai, China, Oct. 2002.

[24] *IEEE Comput. Graph. Appl. (Special Issue on Virtual Reality)*, vol. 21, Nov.-Dec. 2001.

[25] *IEEE Comput. Graph. Appl. (Special Issue on Virtual Worlds, Real Sounds),* vol. 22, July-Aug. 2002.

[25] S. Mann, "Wearable computing: Toward humanistic intelligence," *IEEE Intell. Syst*., vol. 16, pp. 10-15, May/June 2001.

[27] S. Mann, "Wearable intelligent signal processing," *Proc. IEEE*, vol. 86, pp. 2123-2151, Nov. 1998.

[28] MIThril Home Page (MIT Media Lab). Available: http://www.media.mit.edu/wearables/mithril

[29] *IEEE Multimedia (Special Issue on Distance Learning)*, vol. 8, July-Sept. 2001.

[30] Kazaa Home Page. Available http://www.kazaa.com