The Interplay of Technology Development and Media Convergence: Examples

14

Hermann Hellwagner

Chapter Objectives

- To introduce and illustrate the position that progresses in the media domain is driven by interplaying "technology push" and "application pull" forces
- To give examples of important technology developments in the audio–video area and their impact on the media world, businesses, and consumers
- To give examples of today's user requirements, changing behavior, and growing demands and their impact on research and technology development
- To introduce immersive environments as a potential significant future, interdisciplinary direction in the rich media area

14.1 Introduction

Media convergence is defined by Encyclopaedia Brittanica as a *"phenomenon involving the interlocking of computing and information technology companies, telecommunications networks, and content providers [. . .] Media convergence brings together the 'three Cs' – computing, communications, and content"* (Encyclopædia Britannica Inc. 2012) Considering a modern smartphone as an example, this confluence of the *"three Cs"* is obvious. But what is the driver of this confluence? What did enable the industry to develop such "smart" devices? From the point of view of the author, a researcher in multimedia technology, this confluence is not a monolithic *"phenomenon"* or the result of a specific development; rather, the background of convergence is a sophisticated interplay between "technology push" and "application pull" forces, as depicted in Fig. 14.1.

H. Hellwagner (✉)
Research Group "Multimedia Communication", Institute of Information Technology, Alpen-Adria-Universität Klagenfurt, Universitätsstraße 65-67, 9020 Klagenfurt, Austria
e-mail: hermann.hellwagner@aau.at

S. Diehl and M. Karmasin (eds.), *Media and Convergence Management*,

DOI 10.1007/978-3-642-36163-0_14,

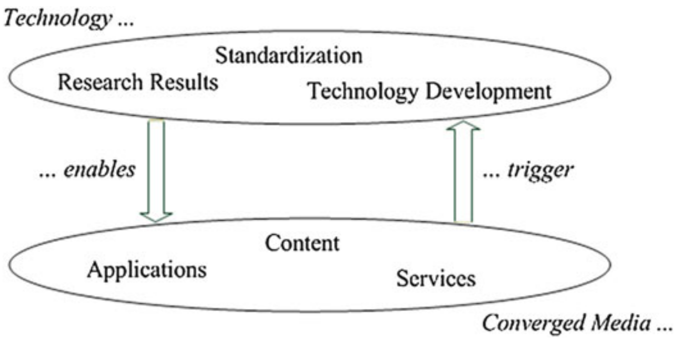

Fig. 14.1 Interplay of technology and media convergence: "technology push" and "application pull"

The figure indicates that, on the one hand, intense research work and technology developments (RTD) as well as standardization efforts in the areas of content processing, computing, and communications have for many years been strong enablers of new content offerings, applications, and services. On the other hand, there is a reverse effect: new converged media applications and visions require and trigger new RTD efforts that again bring about new devices, software, and systems. This constitutes a positive feedback loop that has created remarkable progress in both areas, converged (or rich) media applications and their supporting technologies, and that will continue to foster progress in the years ahead (Jayant (Ed.) 2012).

This contribution discusses these two aspects by means of examples of how RTD and standardization enabled rich media applications (Sect. 14.2) and of how content growth, creativity, and visions in the application space set new requirements for and trigger scientific and technological advances (Sect. 14.3). While Sect. 14.2 looks back into the past and Sect. 14.3 examines recent and current developments, Sect. 14.4 addresses some of the future trends and challenges. The focus will be on audiovisual (A/V) content and communications since these are the author's areas of expertise.

14.2 From the Past: "Technology Push" Aspects . . .

In this section, three examples of developments in the technology arena will be covered that enabled novel media applications and new consumer electronics devices or software, and even facilitated new, or disrupted existing, business models:

- Digital coding (compression) and storage of A/V content
- Standardization and interoperability
- Advances in wired and wireless broadband networks

A major first step toward converged media was undertaken in the 1980s and 1990s by developing effective techniques, tools, and devices for *digital coding* *(compression) and storage of A/V content*. A leading group in that effort was established in 1988, the *Moving Picture Experts Group (MPEG)*, a working group of the International Organization for Standardization (ISO), formally named ISO/IEC JTC 1/SC 29/WG 11. The group's responsibility is *"the development of standards for coded representation of digital audio and video and related data"* (MPEG Home Page 2012).

The first work item and goal of MPEG was *"coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s,"* which in practice meant to enable *"efficient storage and retrieval of audio and video on compact disc"* (MPEG Achievements 2012). This was the first large-scale attempt to develop techniques to store digital audio and video on digital storage media; it succeeded in 1992 in that a first international standard called *MPEG-1* was released.

MPEG-1 provides the means for efficient encoding of digital video at roughly VHS quality and of digital audio at a subjective quality level that is close to the original stereo audio. MPEG-1 deploys *lossy compression* techniques, which for instance for video leads to a compression factor of 25–30, (Sikora 1997), as compared to a "naïve" digital coding format.

But can this be done without substantial quality loss of the media, and how? The answer is apparently, yes, it can be done, by employing sophisticated mathematical-algorithmic techniques that basically *eliminate redundancy and irrelevancy* from the original uncompressed video or audio data stream. In a video, for instance, redundancy takes two forms: *spatial redundancy*, which denotes that adjacent pixels or areas in an individual picture are similar (correlated) to each other, and *temporal redundancy*, which means that successive pictures in a video do not differ too much from each other, enabling pictures to be predicted from previous ones when object or camera motions are taken into account. Original A/V data streams also contain a significant amount of *irrelevant information* in the sense of information that is imperceptible for the visual or auditory human systems and can thus be eliminated or encoded with less detail. In the auditory domain, for example, this led to the development of advanced psychoacoustic models that guide what audio information must be encoded and at what accuracy.

MPEG-1 led to a number of products and applications, among them (Chiariglione 1999; MPEG Achievements 2012): A/V players both in hardware and in software, portable cameras, the Video CD format and associated decoders/playback devices and software, and the use of MPEG-1 technologies in Digital Audio Broadcasting (DAB). Yet, the most prominent outcome is *MP3*, formally MPEG-1 Audio Layer III, which undoubtedly has changed the way we handle and consume music, and has transformed the music industry.

MPEG-2, started only 2 years after MPEG-1 and completed in its most important parts in 1994, was specifically targeted to enable *Digital TV* services. MPEG-2 basically extended and improved MPEG-1 to support interlaced video (required for TV sets of that time), to better encode stereo and multichannel audio (resulting in the well-known Advanced Audio Coding (AAC) standard, for instance), and most importantly to support efficient transmission (broadcasting) and storage of audio and video. MPEG-2 video encoding can achieve compression factors of

30–40 (Sikora 1997), facilitating high-quality digital video comparable to NTSC/PAL TV signals prevalent at that time; in terms of bit rates, 4 Mbit/s and higher bit rates were targeted, significantly higher than for MPEG-1, but still manageable for TV broadcasting systems.

MPEG-2 was a recognized success. It provided the core formats and protocols for digital TV broadcasting over satellite, terrestrial, and cable networks as well as for storage and distribution of movies and other programs on DVDs. The production and consumption of digital A/V material changed drastically, with high-quality digital cameras, DVDs and DVD players, digital TV receivers, "set-top boxes," storage of A/V content on hard disks and initial distribution over the Internet being representative developments enabled by MPEG-2 technologies. New standards adopted and tailored MPEG-2, for example, Digital Video Broadcasting (DVB), and new industries and businesses emerged. MPEG-2 is also widely used for High-Definition TV (HDTV) systems meanwhile.

In recognition of the achievements and the significant impact of the standards MPEG-1, MPEG-2, and the well-known image compression standard JPEG on the media and consumer electronics industry, the ISO/IEC JTC 1/SC 29 received the *1995–1996 Engineering Emmy Award for Outstanding Achievement in Technological Development* (Chiariglione 1996).

MPEG-1 and MPEG-2 devised the basic methods for video and audio compression and storage in digital format. These methods were developed further, refined, and extended in subsequent years, and many of them are still in use today. One further landmark development on this way deserves to be mentioned: *MPEG-4 Advanced Video Coding*. The main goal of this standardization activity, joint work between ISO and the International Telecommunication Union (ITU), which delivered the video coding standard known as *H.264/MPEG-4 AVC* in 2003, was to further increase the compression performance and to provide *"network-friendly video representations"* (Wiegand et al. 2003). More precisely, it was designed to achieve twice the compression efficiency of MPEG-2 (MPEG Achievements 2012) and to support high-quality video "over the Internet," both live services like video conferences and on-demand services such as video-on-demand streaming applications. Since its approval, H.264/MPEG-4 AVC has fully achieved these goals and become the most commonly used video format on the Internet. It is being used by Internet streaming services such as YouTube and the iTunes Store, by Web software such as the Adobe Flash Player and Microsoft Silverlight, and also in various HDTV broadcasting systems, for instance, DVB; moreover, it is being used as a standard format on Blu-Ray Disks (Wikipedia 2012).

The group that developed the H.264/MPEG-4 AVC standard was even presented two Emmy Awards in 2008 and 2009 (ISO/IEC 2009), recognizing the substantial influence that this standard has on business and society.

The presentation so far already gave some insight into the second enabling aspect of media convergence that will be briefly addressed here: the importance of *standardization and interoperability*.

Interoperability denotes *"the ability of two or more systems or components to exchange information in a heterogeneous network and use that information"*

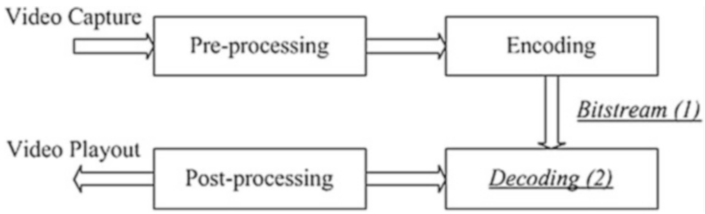

Fig. 14.2 Scope of MPEG-1 and MPEG-2 video standardization (*underlined items*)

(IEEE 2000). In the media world, this concept requires that content must be produced, stored, distributed, and retrieved in a standardized way by diverse devices and/or software components, irrespective of type, vendor, or provenance of the equipment.

The need for standardization in the media arena was judiciously respected from the very beginning in the efforts described above. Research institutions and companies from all over the world, represented by hundreds of individuals, cooperated to create the basic standards that eventually formed the core of the value chains and business opportunities, for instance, in the digital TV field. The scope of the standardization activities was carefully cast, and topics outside this scope were left open for competition among the industry players in the field.

This is illustrated by the following example. Figure 14.2 shows the coarse steps of capturing, preprocessing, digitally encoding, decoding, post-processing, and playing out video data. The scope of standardization in MPEG-1 and MPEG-2 was restricted to *(1) the format of the video bitstream* (data stream), i.e., regulations on the codes and their values to be used in the bitstream (syntax) and their meanings (semantics), and *(2) the process of video decoding* as a reference software decoder, specifying what visual information should be produced for a given encoded video stream. In addition, subsets of the standards for different applications were defined, for example for SD and HD TV, in conjunction with conformance tests. All this allowed extensive testing of whether or not a given video stream or a specific encoder or decoder implementation conformed to the standard.

This represented a first-ever approach to standardization, leaving *freedom to implement* the encoding and decoding steps in diverse ways. In other words, the encoder and decoder could be realized in hardware or software, could be simple or complex, cheap or expensive, support a minimal subset or the full functionality of the standards, and realize low or high compression ratios, as long as they produced conforming encoded bitstreams (encoding) or valid visual information from the bitstreams (decoding), respectively. These concepts left ample room for competition and diversification for the industry and fostered the wide adoption, momentum, and impact of the standards. There were clear benefits for the stakeholders and the awareness that standards and interoperability were the basis of the media business and had to be developed and maintained in a cooperative fashion.

Finally, it is worth noting that consistent *quality* testing was at the core of the standards developments. In case of competing proposals of functionality for the coding standards, the proposals were evaluated and selected according to their figures of merit concerning visual/auditory quality and compression efficiency. The quality evaluations were based on objective quality measures and/or partially extensive subjective tests. This ensured that the best technical proposals made it into the standards, building confidence in the techniques.

The importance of standards for digital media communication and convergence and in particular the processes and achievements of MPEG are summarized in (Chiariglione 2012).

The third element of enabling technology to be briefly addressed here are the continuous *advances in wired and wireless broadband networks* that we have experienced over the past years. Growth in network speeds has been exponential in different technological and usage domains, as indicated by the following data:

- Both Internet traffic and bandwidth grow by approx. 55–60 % per year, as determined by the so-called "Global Internet Geography" analysis for the period 2007–2011 (TeleGeography 2012).
- The speed of high-end users' wired connections to the Internet grows by approx. 50 % per year, as exemplified by "Nielsen's Law of Internet Bandwidth" for the time frame 1998–2010 (Nielsen 2010).
- Wireless connections, both in wide-area cellular networks (e.g., GSM, UMTS, and forthcoming LTE systems) and in short-range WiFi networks (e.g., IEEE 802.11 WLANs), have become faster by almost 50 % per year, as shown for the past 20 years in (Raychaudhuri and Mandayam 2012).

These growth rates mean that network capacities double every 1½–2 years, almost matching Moore's law on advances in computing speed. These improvements have enabled a wide spectrum of applications that involve transmitting high volumes of data over the standard Internet, e.g., Voice over IP (VoIP), video streaming, Internet Protocol TV (IPTV), or rich media embedded in social networks. The growth trends are expected to continue, supporting new classes of applications and services, which will be discussed below.

14.3 ... to the Present: "Application Pull" Aspects

The technologies described above, among others, have become *"true enablers of next-generation services and facilities, specifically diverse rich digital media services,"* as (Jukan and Mambretti 2012) put it. Simultaneously, the new opportunities are creating new demands and challenges for technology. This leads to a situation that is characterized by (Jukan and Mambretti 2012) as follows: *"Driver applications and support technologies have challenged and enabled each other in an unprecedented progression."*

Again, three examples will be explored to illustrate the "application pull" side:

- Increasing demand for, and volume of, multimedia content
- Growing number and diversity of devices
- Higher quality and novel forms of media content.

A first challenge for technology is that the *"appetite" for rich media content* in the Internet is rising sharply. At the time of writing, multimedia content, dominated by digital video, has already become a major component of traffic in the Internet. Moreover, the amount of media content is expected to double every 1–1½ years (Jukan and Mambretti 2012). It is worth noticing that this growth rate is higher than that of the basic network capacities pointed out above.

The networking equipment vendor Cisco, in a periodic forecast of Internet traffic called "Visual Networking Index (VNI)," predicts that Internet video will reach 50 % of consumer Internet traffic by the end of 2012 and 62 % by the end of 2015, not including peer-to-peer (P2P) file sharing traffic. Accounting for P2P traffic as well as other forms of visual information transmission, e.g., IPTV, this percentage is significantly higher (Cisco 2011). In the arena of mobile devices and networks, the predictions are similar, stating for instance that by the end of 2016, more than 70 % of the mobile data traffic will be due to video (Cisco 2012).

The growing demand for, and volume of, media content stems from both the rapidly increasing trend toward user-generated content and sharing this content, and more and more appealing media services and portals on the Internet. Flickr, YouTube, and Facebook are well-known examples of the former; Netflix is representative of the latter. Netflix is a video-rental company mainly active in North America that nowadays typically streams the content to the users over the Internet on demand. With more than 100,000 movies and TV shows available and more than 23 million users (Netflix 2012), Netflix alone accounts for up to 20 % of the downstream Internet traffic in the USA at peak times, according to Cisco's VNI.

The technological responses to these challenges are manifold. Besides the progress being made in improving the core networking technologies as outlined in Sect. 14.2, Internet Service Providers (ISPs) and telecommunication companies worldwide are continuing to heavily invest in their networks at all levels (access, distribution, and backbone networks) and to deliver higher-bandwidth connectivity to their customers, in both the wired and wireless domains.

On the media content distribution side, novel forms of content delivery over the Internet have been devised and are being deployed to cope with the load of serving a potentially vast user community; examples are peer-to-peer (P2P) systems and content delivery networks (CDNs).

Traditionally, for instance in proprietary IPTV systems, content is delivered from a central server (farm) along a tree of sub-servers toward the clients/users. In contrast, in P2P systems all nodes contribute to the delivery of content by assuming both roles, requesting content (pieces) as clients and providing content (pieces) as servers. While P2P systems have a bad reputation as illegal file sharing platforms, they do have technical and economic merits as sophisticated and cost-saving solutions for media distribution, even of live content such as TV channels. For an example, the interested reader is referred to the European project "P2P-Next," the goal of which was to devise and build a next-generation P2P content delivery platform (P2P-Next 2012).

A CDN is typically a large, distributed network of servers, mostly deployed at or near the "edge" of the Internet, in data centers close to the users. Content is replicated and transmitted to the CDN servers and cached there (i.e., stored for a certain time period). The major benefit of a CDN is that even high volumes of media content can be served to many users on demand with high performance (e.g., short download times or streaming startup latencies), high availability (due to fallback options to other CDN servers in case of a server failure), and at low costs (lower than if the content had to traverse the entire Internet from origin to destination for each user). CDNs are highly popular and widespread in today's Internet, with different deployment and business models: large companies running their own CDNs (e.g., Google), specialized enterprises providing CDN services (e.g., Akamai Technologies), or media companies (e.g., Netflix) making use of cloud computing/storage offerings.

Finally, since rich media distribution has become the dominant source of traffic in the Internet today, the basic principles of the Internet are being questioned. Starting more than 40 years ago, the Internet was basically designed to transmit text messages (e.g., e-mails), probably files of moderate size. The number of users and devices on the network, the type and volume of data/content to be transported, the diversity of applications and services, and the role of the network as a crucial worldwide infrastructure were unforeseeable in the initial years and decades when the basic principles and protocols were defined and realized. Thus, partly due to the digital media revolution, the Internet is seen as *"just working,"* not more (Handley 2006). There is wide consensus that the Internet needs to be reworked in order to be able to cope with the future requirements. In recent years, therefore, intense activities on *Future Internet* research and experimentation are being performed, with mainly the USA, Europe, and several Asian countries pursuing their own programs. The European efforts are substantial, for instance with 128 collaborative ongoing projects in 2011 (Domingue et al. (Eds.) 2011). A major thrust of the Future Internet research is to address the growth and diversity of the media content and work out feasible and sustainable concepts for *Content-Centric Networking (CCN)*.

Increasing difficulties arise in today's rich media systems due to the *growing number and diversity of devices* requesting services and content, which will be discussed as the second example of challenges posed to RTD.

The sheer number of devices connecting to the Internet today is astounding. Again according to Cisco's forecasts, *"the number of mobile-connected devices will exceed the number of people on earth"* by the end of 2012, and there will be more than 10 billion such devices in 2016 (Cisco 2012). The main technological measures to keep up with the growing demand emerging from these devices have been discussed above.

The diverse capabilities and constraints of the devices pose even more serious problems. We all enjoy consuming rich media services on a variety of devices nowadays, including (Smart) TV sets, stationary and portable computers, game consoles, tablet computers, and smartphones. However, none of us wants to bother with configuring those devices or with selecting specific content variations and services fitting specific device characteristics like display size, operating system, or the media formats supported. End users just desire (and expect, meanwhile) to access the media content and services anytime, anywhere, and from any device, in the highest quality possible for the device in use.

These expectations have been anticipated and worked on in the multimedia communication community since more than a decade under the term *Universal Multimedia Access (UMA)* (Timmerer and Hellwagner 2005). Again, *interoperable* solutions are highly desirable, which is why MPEG started standardization efforts and addressed the technical issues in this problem domain. This resulted in the *MPEG-21* series of standards, called *Multimedia Framework*, which the author and his group actively contributed to (Timmerer and Hellwagner 2005; Burnett et al. 2006). Together with an earlier family of standards, *MPEG-7*, the *Multimedia Content Description Interface* (Manjunath et al. 2002), the basic means to realize UMA are provided.

At the core of the MPEG-7 and MPEG-21 standards are *descriptions*, also known as *metadata* (data about the multimedia data), that are intended to instruct and steer the delivery of the content to the end user devices. MPEG-7 basically describes the *content* properties, for instance, the coding format and the spatial resolution, frame rate, and bit rate of a video, but also provides general and semantic information like title, description, director, and actors of a movie, in a standardized way. MPEG-21 in contrast provides descriptions for the *usage environment* of the content, which means metadata specifying the device properties (e.g., display size), the characteristics of the networks traversed (e.g., the transmission capacity to be expected), the preferences and constraints of the end user with respect to the content (e.g., genre preferences), and even the natural environment the content is consumed in (e.g., brightness or noise level). In a sense, both the source (content) and the destination (usage environment) of the media consumption chain are captured. (There are many other elements to the MPEG-7 and MPEG-21 standards families, which for the sake of brevity will not be dealt with here.)

Given these descriptions, which basically define the problem of UMA in a technical sense, the media can be adapted for consumption by a specific end user on a specific device to provide the best possible experience. Thus, *multimedia content adaptation* has been a major thrust of RTD for many years, with many hard problems in the details being addressed and solutions being proposed. For example, the decisions on *where* to perform content adaptation and *what* precisely to do in that process and *how* are interesting optimization problems; several other issues are addressed in (Timmerer and Hellwagner 2005).

Unfortunately, the concepts worked out in the MPEG-21 Multimedia Framework have not been widely adopted in practice to date. From the author's point of view, the main reasons are the following. The MPEG-7 and MPEG-21 standards are very complex, amounting to hundreds or even thousands of pages of specification text and requiring complex software to be realized. Also, the solutions envisaged in MPEG-21 seem to be too "static" in hindsight, making it difficult to adapt media content to dynamically fluctuating network conditions, for instance. (This issue will be further addressed below.) Most importantly, apparently the (industry) players in the field do not see immediate benefits in implementing interoperable and principled solutions to the UMA problem; rather, they prefer to provide more or less proprietary platforms that deliver rich media content end to end with high quality and satisfying experience for the user.

A prominent example for the latter behavior is Apple Inc., with a largely closed multimedia infrastructure that delivers excellent services to the users, though. A more ad hoc solution is pursued by Netflix that reportedly are capable of streaming content to more than 700 types of devices (Netflix 2012), which they basically cannot control. To that end, each content item is "transcoded" into dozens of different formats and variants, and the specific version that best fits the user's request and device is selected for delivery—a costly solution in terms of computational, storage, and maintenance effort.

Yet, there are noteworthy developments that originated from the MPEG-21 Multimedia Framework efforts and that are still attractive to the entire media world. One is the concept of practically useful *scalable media content* that directly addresses the urgent need of serving the diversity of devices. This concept denotes the approach to encode media content in *layers*, one base layer and one or several enhancement layers; the base layer contains the content in low quality, the enhancement layers, building on the base layer, enhance the quality progressively, proportionally to the additional amount of data delivered, and along several dimensions. For instance, for video these enhancements can pertain to the spatial domain (higher resolution pictures), the temporal domain (increased frame rate), and the quality domain (fewer or less severe coding artifacts). The obvious benefit of scalable content is that many different device types can be served from a *single* content source, with different (number of) layers being transmitted to, and processed and displayed by, different devices. The most promising example in that area is an extension of H.264/MPEG-4 AVC called *Scalable Video Coding (SVC)*, again a standard jointly developed by the ISO and the ITU (Schwarz et al. 2007; Hellwagner et al. 2011). While SVC approaches were part of earlier standards, SVC is novel and useful in that it provides competitive compression efficiency; the development goal was to incur only about 10 % overhead as compared to single-layer H.264/MPEG-4 AVC, which was basically achieved.

In response to the need to dynamically adjust the media streams to possibly rapidly varying network conditions, e.g., when a user is moving in a car, recent work focused on approaches toward *dynamic adaptive streaming* of multimedia content. Several companies had already introduced their own solutions in that direction, for example Adobe, Apple, and Microsoft. All these solutions make use of the most widespread protocol in the Internet, HTTP, and thus can exploit the existing HTTP Internet infrastructure, notably CDNs, as explained above. A general and interoperable solution was, however, not available until the Third Generation Partnership Project (3GPP), a standardization organization in the mobile broadband communication field, and later on MPEG integrated dynamic adaptive streaming into their portfolios (Stockhammer 2011; Sodagar 2011). MPEG recently released the *Dynamic Adaptive Streaming over HTTP (DASH)* specification, a standard solution for HTTP streaming and adaptation of multimedia content that enables interoperable communication between servers and clients by different vendors (Chiariglione 2012)

Simply put, the basic principle is that content is encoded and stored on the HTTP server(s) in various forms called *Representations* that may differ in terms of coding format, quality, spatial resolution, bit rate, language, and the like. Furthermore, the content is temporally divided into *Segments*, each typically a few seconds long; each segment is accessible by a unique identifier and can be fetched using standard HTTP GET requests. The content structure and the identifiers are described in a standardized XML document called *Media Presentation Description (MPD)*, kind of a directory of the content offering on the server(s).

A client device accessing the content is initially provided by the MPD; it reads and interprets the MPD to find out which content representation is desirable and appropriate for the device under the given network conditions. Then, the segments (or parts thereof) are being fetched one by one using HTTP GET requests. When the network conditions become better or worse, subsequent segments can be requested in higher or lower quality from different representations. This client-driven approach, in contrast to typical earlier server-driven streaming approaches, is highly practical, well proven, and flexible since each client can select, on its own, the best content representation and adjust that decision on a segment-by-segment basis when network bandwidth conditions change, for instance. Moreover, it runs "on top of" the established protocols and HTTP/CDN infrastructure of the Internet, avoiding changes to be made in the Internet. Finally, it can be combined with scalable media coding to form different content representations.

The final challenge that requires technology developments is the quest for *higher quality and novel forms of media content*. Today, a growing number of users are not satisfied with 2D content, not even in HD format. 3D content is regarded as exciting, and the content industry pushes very high-quality and 3D content into cinemas, for example, to increase revenues. There is pressure that even better or richer content be made available for the home environment as well: *"Time for Video to Become 3-D,"* as (Chiariglione 2012) puts it.

Not surprisingly, RTD has been going on in this area as well for several years. One notable development is the specification of *Multiview Video Coding (MVC),* a 2008 extension of the H.264/MPEG-4 AVC standard, which defines efficient coding of multiple camera views (Chiariglione 2012); MVC has been adopted by Blu-Ray 3-D (Tanimoto et al. 2012). Building on MVC, a current project of MPEG is *3D Video (3DV)*, the goal of which is *"to define a 3-D format that enables both advanced stereoscopic display processing and improved support for autostereoscopic N-view displays"* (Chiariglione 2012). These efforts are destined to finally bring about *Free-viewpoint TV (FTV)*, in which the user should be able to freely choose and change viewpoint in 3D space (Tanimoto et al 2012).

Substantially higher quality of 2D content is another direction that is being pursued; improvements in temporal and spatial resolutions, in color fidelity, and in pixel depth (bits per pixel) are being envisaged. Work on *Ultra-HD (UHD)* content and display technology is under way, with 4k $\times$ 2k resolution currently being targeted. Again, a joint team of ISO and ITU is working intensely in this area, currently developing a *High Efficiency Video Coding Standard (HEVC)* (Chiariglione 2012).

The interested reader is recommended to consult the literature cited above and further reading provided therein.

14.4 Future Trends and Challenges

Given the enormous advances of rich media services and supporting technologies as sketched above, the question is exciting, "where is this heading to?"

Several papers in (Jayant (Ed.) 2012) provide highly informed and enlightening answers: *truly immersive, rich media communication* as well as *mixed-reality systems* will be among the next frontiers (Apostolopoulos et al. 2012; Barba et al. 2012; Steinbach et al. 2012).

The field of *truly immersive communication* is driven by the vision *"to enable natural experiences and interactions with remote people and environments"* (Apostolopoulos et al. 2012). Progress has been shown recently by the advent of high-end video-conferencing/telepresence systems; however, these systems are expensive and require all participants to sit in specifically equipped studios.

There are many challenges ahead, both for research and technology development and for understanding, representing, and serving the *users*, the humans who are central to the immersive experience. The notion of *quality of experience* must be understood and quantified along several dimensions (Jayant (Ed.) 2012), which requires interdisciplinary research, at least involving computer scientists, telecommunication experts, and psychologists.

The vision of immersive communication is excellently structured and illustrated in (Apostolopoulos et al. 2012) as shown in Fig. 14.3, which is a simplified version of a diagram of that paper. The important dimensions in immersive communication are to support *(1) natural conversation* among participants and *(2) information sharing* as conveniently as possible.

As an example, consider immersive communication systems. Today's high-end telepresence systems, exemplified by HP Halo and Cisco Telepresence, give participants at different sites the feeling of being in the same room, by using a wall of displays, life-size video feeds of remote users, and high-quality audio. More details and snapshots are given in (Apostolopoulos et al. 2012). One future direction could be toward virtual 3D telepresence, by 3D-capturing the movements of remote participants and the objects in the remote environment, analyzing the data, transmitting it, and locally rendering the remote people and items on stereoscopic displays. This description makes obvious that the quality of experience of the users must be the definite yardstick when designing and realizing such an environment. It is interesting to note that the communication systems mentioned above were *"conceived by veteran storytellers in Hollywood"* (Apostolopoulos et al. 2012).

Haptic communications is regarded as an important part of immersive systems by (Apostolopoulos et al. 2012) and is technically further explored in (Steinbach et al. 2012).

With modern smartphones and cloud systems to back them up in terms of computational power and storage space, *mixed-reality systems* are *"moving out of the lab and into the real world"* (Barba et al. 2012). The "classical" notion of *augmented reality (AR)* enriches the environment or objects therein with computer-generated content, for instance with images or maps, and allows interaction with

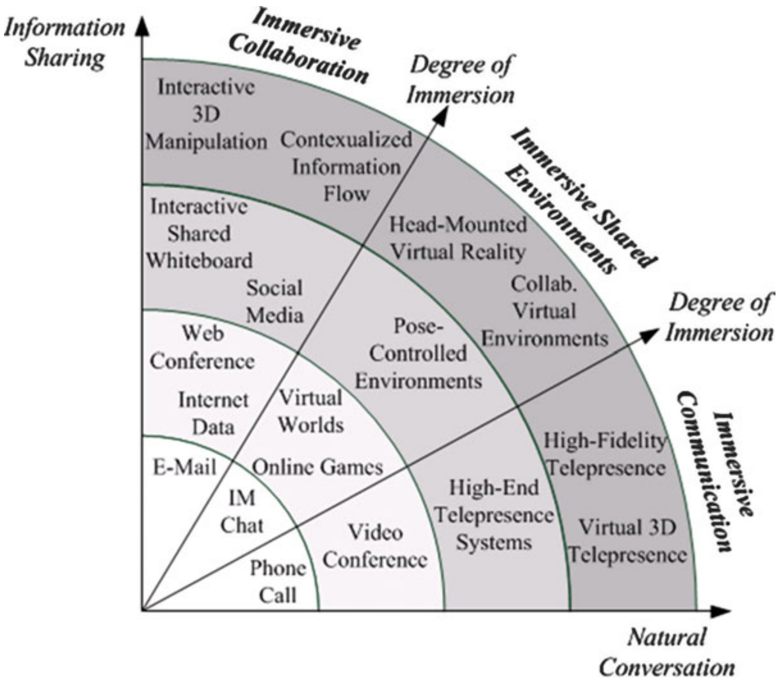

Fig. 14.3 Toward immersive environments [adapted from Apostolopoulos et al. (2012)]

this content. This can be taken further into what Barba et al. (2012) call *mixed reality (MR)*, denoted as integration of *"physical and virtual elements into a new hybrid reality."* Clearly, this new reality should be 3D, posing the challenge to align virtual content with the physical environment as naturally and seamlessly as possible in three dimensions, on our mobile devices that today still by far lack the computational, communication, and storage capabilities to evoke true immersive feeling. But progress is being made, as illustrated in (Barba et al. 2012) by means of several examples.

It will be exciting to monitor the further technological advances that will be made and to experience novel applications and services that will emerge. The interplay of technology and rich media applications will continue to thrive; in the words of Apostolopoulos et al. (2012): *"as technology pushes (i.e., supplies), society pulls (i.e., demands)."*

Questions

1. Why does coding (compression) of audiovisual material work so well? Give an example.
2. Why is standardization and interoperability important in the media world?
3. What was the impact of the MPEG-2 video coding standard?

4. What was specific about the scope of the MPEG-1 and −2 standards and what did this entail?
5. What are the reasons (sources) of the growing demand of multimedia content in the Internet nowadays?
6. What are technological responses to this growing demand?
7. What is a content delivery network (CDN)? Create a rough sketch of a CDN and outline its benefits.
8. Why does the diversity of multimedia-enabled devices pose problems for multimedia communication? What do the MPEG-7 and −21 standards provide to ease these problems?
9. What are the benefits of Dynamic Adaptive Streaming over HTTP (DASH)? What are the advantages of a client controlling the media streaming process?
10. What does immersive communication denote?

Discussions

1. In your opinion, which features should a future immersive communication system (e.g., a 3D video conferencing system) have in order to feel "natural"?
2. Discuss what the concept of quality of experience (QoE) might mean in detail, concretely, possibly quantitatively. Why is this important for immersive communication? Recall your own personal "experiences" with IT devices and software.
3. Do some research and reading about so-called "second-screen applications", i.e., people using two devices when consuming multimedia content or services, e.g., TV and smartphone. Do you think that this will be the future of media and entertainment or information?
4. Can you find counterexamples for the position of the chapter, i.e., that there are both "technology push" and "application pull" forces driving the media domain. In other words, are there examples for notable developments that work(ed) just one way or the other?

References

Apostolopoulos, J. G., Chou, P. A., Culbertson, B., Kalker, T., Trott, M. D., & Wee, S. (2012). The road to immersive communication. *Proceedings of the IEEE, 100*(4), 974–990.

Barba, E., MacIntyre, B., & Mynatt, E. D. (2012). Here we are! Where are we? locating mixed reality in the Age of the Smartphone. *Proceedings of the IEEE, 100*(4), 929–936.

Burnett, I. S., Pereira, F., Van de Walle, R., & Koenen, R. (Eds.). (2006). *The MPEG-21 book*. New York, NY: Wiley.

Chiariglione, L. (1996, October). *Emmy Awarded for the Development of MPEG Standards*. (MPEG, Ed.) Retrieved May 2012, from http://mpeg.chiariglione.org/meetings/chicago/chicago_press.htm

Chiariglione, L. (1999). MPEG: Achievements and future projects. *IEEE International Conference on Multimedia Computing and Systems (ICMCS'99)*. Florence, Italy.

Chiariglione, L. (2012). Multimedia standards: Interfaces to innovation. *Proceedings of the IEEE, 100*(4), 893–904.
Cisco. (2011). *Cisco Visual Networking Index: Forecast and Methodology, 2010–2015 (White Paper)*. Retrieved May 2012, from http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360.pdf
Cisco. (2012). *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2011–2016 (White Paper)*. Retrieved May 2012, from http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/VNI-Forecast_QA.pdf
Domingue, J., et al. (Eds.). (2011). *The Future Internet. Future Internet Assembly 2011: Achievements and technological promises (Vol. LNCS 6656)*. Heidelberg: Springer.
Encyclopædia Britannica Inc. (2012). *Encyclopædia Britannica Online, Entry "Media Convergence"*. Retrieved May 2012, from http://www.britannica.com/EBchecked/topic/1425043/media-convergence
Handley, M. (2006). Why the Internet only just works. *BT Technology Journal, 24*(3), 119–129.
Hellwagner, H., Kofler, I., Eberhard, M., Kuschnig, R., Ransburg, M., & Sablatschan, M. (2011). Scalable video coding: Techniques and applications for adaptive streaming. In C. Zhu, Y. Li, & X. Niu (Eds.), *Streaming media architectures, techniques, and applications: Recent advances* (pp. 1–23). PA: IGI Global.
IEEE. (2000). *IEEE Std-100–2000: The authoritative dictionary of standard terms* (7th ed.). IEEE.
ISO/IEC. (2009, March). *Second Emmy Award for ISO/IEC MPEG-4 AVC Standard*. Retrieved May 2012, from http://www.iso.org/iso/pressrelease.htm?refid=Ref1213
Jayant, N. (2012). Special issue: Frontiers of audiovisual communications: New convergences of broadband communications, computing, and rich media. *Proceedings of the IEEE, 100*(4), 821–823.
Jukan, A., & Mambretti, J. (2012). Evolution of optical networking toward rich digital media services. *Proceedings of the IEEE, 100*(4), 855–871.
Manjunath, B. S., Salembier, P., & Sikora, T. (Eds.). (2002). *Introduction to MPEG-7: Multimedia content description interface*. New York, NY: Wiley.
MPEG Achievements. (2012). Retrieved May 2012, from http://mpeg.chiariglione.org/achievements.php
MPEG Home Page. (2012). Retrieved May 2012, from http://mpeg.chiariglione.org/
Netflix. (2012). *Netflix Home Page*. Retrieved May 2012, from http://www.netflix.com
Nielsen, J. (2010). *Nielsen's Law of Internet Bandwidth*. Retrieved May 2012, from http://www.useit.com/alertbox/980405.html
P2P-Next. (2012). *P2P-Next Home Page*. Retrieved May 2012, from http://www.p2p-next.org/
Raychaudhuri, D., & Mandayam, N. B. (2012). Frontiers of wireless and mobile communications. *Proceedings of the IEEE, 100*(4), 824–840.
Schwarz, H., Marpe, D., & Wiegand, T. (2007). Overview of the scalable video coding extension of the H.264/AVC standard. *IEEE Transactions on Circuits and Systems for Video Technology, 17*(9), 1103–1120.
Sikora, T. (1997). MPEG digital video coding standards. In R. K. Jurgen (Ed.), *Digital consumer electronics handbook*. New York, NY: McGraw-Hill.
Sodagar, I. (2011, October–December). The MPEG-DASH standard for multimedia streaming over the Internet. *IEEE MultiMedia*, pp. 62–67.
Steinbach, E., Hirche, S., Ernst, M., Brandi, F., Chaudhari, R., Kammerl, J., et al. (2012). Haptic communications. *Proceedings of the IEEE, 100*(4), 937–956.
Stockhammer, T. (2011). Dynamic adaptive streaming over HTTP – standards and design principles. *Proceedings of the ACM Multimedia Systems (MMSys) Conference* (pp. 133–143). San Jose, CA, USA.
Tanimoto, M., Tehrani, M. P., Fujii, T., & Yendo, T. (2012). FTV for 3-D spatial communication. *Proceedings of the IEEE, 100*(4), 905–917.
TeleGeography. (2012). *Global Internet Geography*. Retrieved May 2012, from http://www.telegeography.com/research-services/global-internet-geography/index.html

Timmerer, C., & Hellwagner, H. (2005, January–March). Interoperable adaptive multimedia communication. *IEEE MultiMedia*, pp. 74–79.

Wiegand, T., Sullivan, G. J., Bjøntegaard, G., & Luthra, A. (2003). Overview of the H.264/AVC video coding standard. *IEEE Transactions on Circuits and Systems for Video Technology, 13*(7), 560–576.

Wikipedia. (2012). *Wikipedia, Entry "H.264/MPEG-4 AVC"*. Retrieved May 2012, from http://en.wikipedia.org/wiki/H.264_AVC