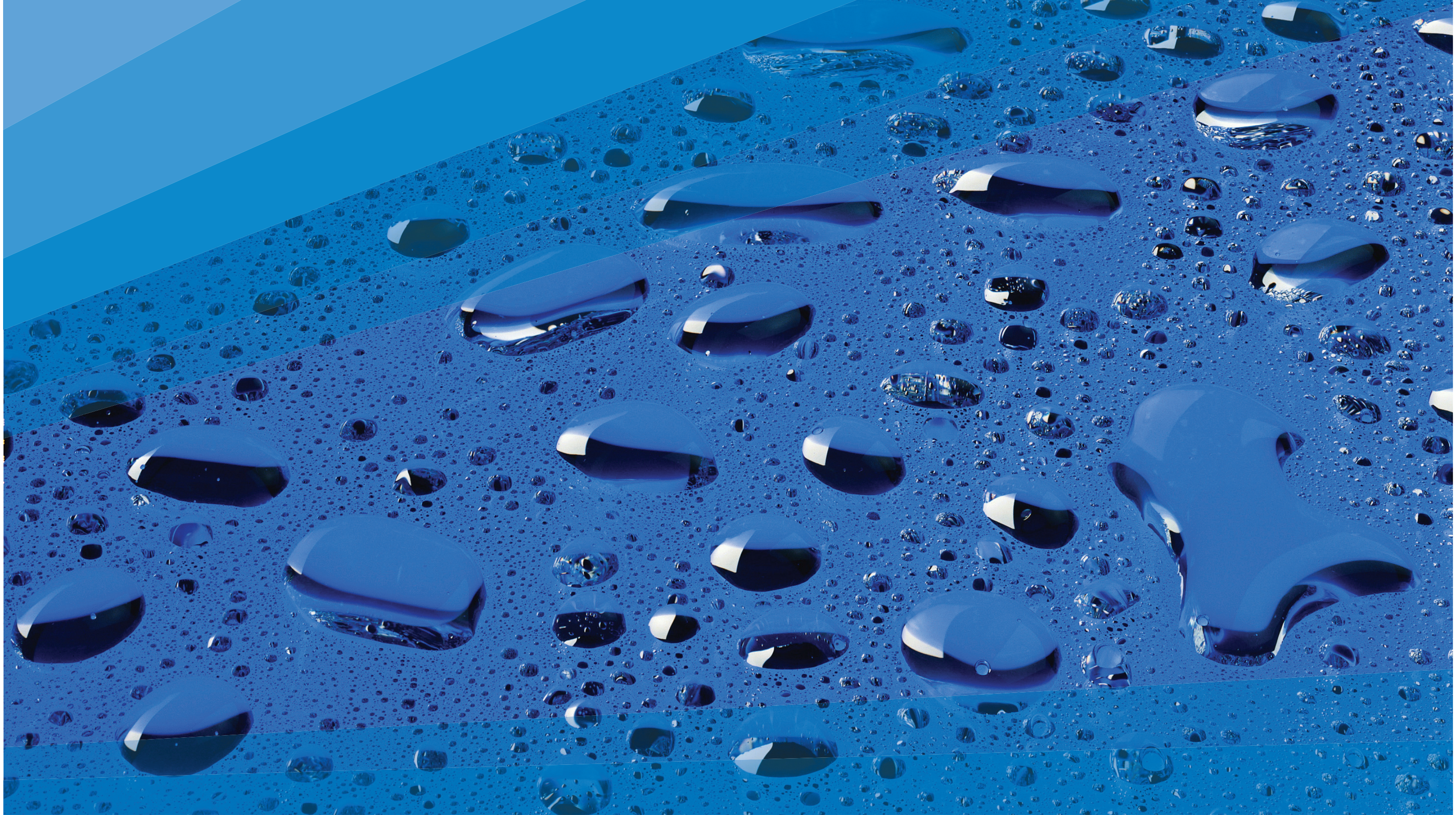
# Research on Future Media Internet



**January 2009**

Created by "Future Media Internet Task Force"

Coordinated by five FP7 Networks of Excellence supported by the Networked Media Unit of the DG Information Society and Media of the European Commission

European Commission
Information Society and Media

This work was coordinated by 5 NoEs:

- *Content (http://www.ist-content.org)*
- *Emanics http://(www.emanics.org)*
- *Intermedia (http:// intermedia.miralab.unige.ch)*
- *Petamedia (http:// www.petamedia.eu)*
- *Visnet II (http:// www.visnet-noe.org)*

And was coordinated by Isidro Laso Ballesteros. Supported by the Networked Media Unit, DG Information Society and Media of the European Commission.

# 1 Executive Summary

This White Paper reflects the consolidated opinion of 36 experts from the EU and the USA on aspects related to the Future Media Internet (FMI), under the guidance of the Networked Media Systems Unit of the Information Society and Media Directorate General of the European Commission. More specifically, this paper describes the challenges provisioned by the experts for the upcoming years, concerning the FMI, along with potential scenarios, paying specific attention to its provisional network and 3D content characteristics.

The vision of the FMI Task Force is that the current Internet is evolving towards providing gully immersive experiences, which give rise to innovative applications notably in gaming technologies, virtual worlds (as mixed world's) and communications.

Due to the double nature of the Future Media Internet (network and media/content) different, but definitely cooperative, aspects are exhaustively discussed and analyzed. These aspects concern both potential ingredients of next generation networks and new forms of media. With respect to **network-related characteristics** the FMI-TF believes that the characteristics of the FMI should involve:

- higher bandwidth needs to be coupled to new traffic patterns;
- content adaptation in the network and the terminals that enable the availability of media for a range of heterogeneous devices;
- new models of content distribution that consider not only the protocols involved but also the social characterization of the nodes and the data;
- new network management modules and mechanisms to provide and monitor Quality of Experience (QoE), trust and privacy.

The aforementioned network characteristics will allow for **content-related characteristic of the FMI**, such as:

- interactive/proactive autonomous characters, where a character is any object in a 3D scene with its own opinions and suggestions, able to understand its environment and take decisions;
- 3D social communities which allow people to use 3D environments to communicate and interact with each other using rich communication means similar to those used in face-to-face meetings (gaze awareness, gestures, facial expressions, correct sound direction, manipulation of social signals);
- personalised entertainment supporting interactive, non-linear story-telling, senses to be engaged in an immersive experience (participation in - or bringing theatre, movies, games);

- capturing and reproduction of the real world in 3D, involving novel characteristics like smell, haptics, etc;
- coexistence of virtual and real worlds maintaining perceptual coherence.

Based on the abovementioned network and content related characteristics of the FMI, the Task Force identified the following major research challenges, the fulfillment of which is expected to lead to the new era:

- *Embedded intelligence for search and retrieval in virtual words;* how can we develop intelligent systems able to search for multimodal objects using as queries multimodal objects?
- *Physical-based worlds (including behaviour modelling and understanding);* how can we achieve interaction realism able to allow performing new experiences in the virtual world, improving training and execution of tasks in the virtual environment?
- *Compression, encoding, transmission & Portability and adaptation;* how can we enable face to face audio and video interactions to be held between various different types of fixed and mobile terminals connected by a heterogeneous network?
- *Media driven network;* how can we achieve higher end-to-end QoE for 3D content streaming in distributed environments?

The timeline for the aforementioned research challenges can be seen in the following table.

| RTD Challenge | Timeline |
|---|---|
| 4.1: Embedded intelligence for search and retrieval in Virtual Worlds | Medium (5-10 yrs) |
| 4.2: Physical-based worlds | Medium to Long (>10 yrs) |
| 4.3: Compression, encoding, transmission & Portability and adaptation | Short (3-5 yrs) to Medium (5-10 yrs) |

| | |
|---|---|
| *4.4: Media Driven Networks* | Medium to Long (clean slate - Revolutionary technologies to deeper embed media into the network design (media aware networks and infrastructure devices etc.).<br><br>Short to Medium (Evolutionary technologies for making the Internet media-enabled (protocols, media-distribution architectures and cost metrics etc.)) |

***Table 1***: Timeline for the Common Research Challenges

Another challenge is to move from the aforementioned components to harmoniously combined and integrated solutions exploring new possibilities for boosting individual and social creativity and productivity in a Future Media Internet environment. The ultimate result of such combinations is expected to have a fundamental impact on society in the FMI on how we communicate, how we work and how we live.

## 2 Introduction

The FMI Task Force (TF) engaged a brainstorming session in order to provide experiences and the vision to pave the way towards the FMI. It is a common belief that nowadays **the Current Internet is evolving towards providing richer and immersive experiences**. This is expected to be a reality soon due to the advances in 3D processing software tools along with the technological innovations in 3D graphics and computational equipment, **which give rise to innovative applications notably in gaming technologies, virtual worlds and communications**. However, these new applications place new types of traffic demands and constraints on network architectures. Thus, there is a worldwide concern in the networking research community about the future of Current Internet and its manageability, security, openness to innovation and scalability.

Towards finding solutions to the aforementioned issues, many research initiatives in the world are running in parallel: the EU Future Internet Research and Experimentation (FIRE)[1], the NSF Future Internet Design (FIND) and GENI programmes[2], and similar efforts in the China Science and Technology Network[3], in Japan through the AKARI Architecture Design Project[4] and in Korea[5].

We can classify the current initiatives of the worldwide research community in two broad categories: *clean slate* vs. *evolution*[6]. The first one proposes to redesign from scratch a new Internet from a long term perspective, while the latter proposes to respect the current architecture and also maintain backward compatibility. The vision of the Future Media Internet Task Force (FMI-TF) is not to bet on one of these orthogonal approaches, but to identify the problems of the current Internet and propose solutions for providing media and 3D content via the Internet in the next decade.

The future needs of networks will be inconsistent with the principles on which the Internet of today is based. It is not only a question of redesigning a protocol stack and

---

[1] www.cordis.europa.eu/fp7/ict/fire

[2] www.nets-find.net & http://www.geni.net/office/office.html

[3] www.cstnet.net.cn

[4] akari-project.nict.go.jp/eng/overview.htm

[5] mmlab.snu.ac.kr/fiw2007/presentations/architecture_tschoi.pdf

[6] Constantine Dovrolis. *What would Darwin Think about Clean-Slate Architectures?.* ACM SIGCOMM Computer Communication Review. Volume 38, Issue 1, January 2008.

packet system based network. It is also a matter of finding new ways of using the Internet in different domains such as work and entertainment.

Moreover, the dynamism of content is moving from a pull based on local interest to a push based on a social network distribution mechanism. The new economics of the Internet identify new business in the long tail[7] of content available to a large number of small communities. This new method of content consumption introduces new roles for the media business that allow us just to discern new changes in the economics of Internet. The reduction of intermediaries in the distribution of content allows the end users to introduce new sources of content at multiple places, in multiple formats and for multiple methods of consumption. This represents just the start of a new era of network traffic.

[7] Chris Anderson. *The Long Tail: How Endless Choice Is Creating Unlimited Demand.* Random House, New York, 2006.

# 3 Vision of the Future Media Internet

The Internet was initially designed and primarily used by scientists for networking research and for exchanging information between each other. However, due to the explosion of the World Wide Web (which started as a document repository) and its successful descendants (Web 2.0), along with the dramatic increase of net-based audiovisual material (networked media) that has been produced by professional and most recently by amateur users, the Internet is rapidly transforming into a fully fledged virtual environment that facilitates services, interaction and communication. Therefore, the vision that the Future Internet will be an Internet of Media is about to be a reality.

The vision of the FMI Task Force is that the current Internet is evolving towards providing gully immersive experiences, which give rise to innovative applications notably in gaming technologies, virtual worlds (as mixed world's) and communications

Towards this aim, specific attention should be given to both main pillars of the FMI, namely the content and the network. The long term vision of the task force is that the future content will have the following characteristics:

- the interactive/proactive autonomous characters, where a character is any object in a 3D scene with its own opinions and suggestions, able to understand its environment and take decisions;
- 3D social communities which allow people to use 3D environments to communicate and interact with each other using rich communication means similar to those used in face-to-face meetings (gaze awareness, gestures, facial expressions, correct sound direction, manipulation of social signals);
- personalised entertainment supporting interactive, non-linear story-telling, senses to be engaged in an immersive experience (participation in - or bringing theatre, movies, games);
- the ability to capture and reproduce the real world in 3D, multi-view, smell, haptics, etc;
- the coexistence of virtual and real worlds maintaining perceptual coherence.

From the network-related point of view, the characteristics of the Future Media are expected to involve:

- higher bandwidth needs to be coupled to new traffic patterns;
- content adaptation in the network and the terminals that enable the availability of media for a range of heterogeneous devices;

- new models of content distribution that consider not only the protocols involved but also the social characterization of the nodes and the data;
- new network management modules and mechanisms to provide and monitor Quality of Experience (QoE), trust and privacy.

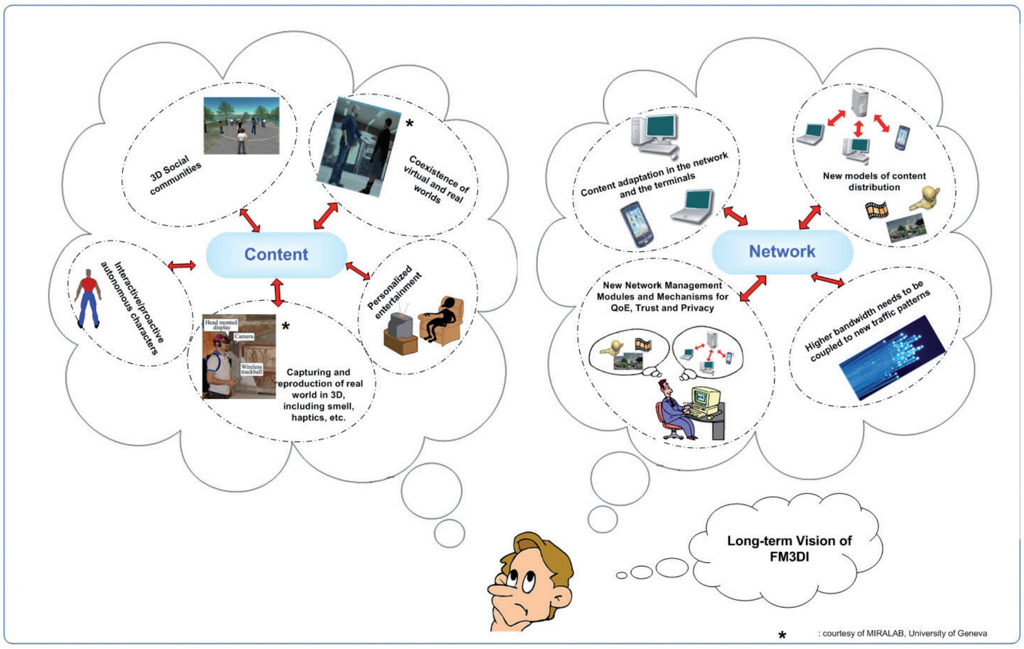



***Figure 1**:* Characteristics of the FMI

The long term vision of the FMI-TF defined above (Figure1) requires the following long term research endeavours:

1. research on the optimum methods for capturing the visual appearance of the real world, including 3D/multiview, high dynamic range and high frame rate;
2. research on the development of rich interfaces allowing for multimodal interaction 3D navigation and strong personalization;
3. research on the extensibility, scalability, distribution and availability of the content anywhere, anytime and in any terminal;
4. research on new mechanisms for native searching.

In the next sub-sections the TF sheds light on the FMI characteristics and explains which research challenges are envisioned to lead to the new era.

# 4 Major Common Challenges

This subsection presents major challenges in the areas of embedded intelligence for search and retrieval in Virtual Worlds, physical-based worlds (including behaviour modelling and understanding), compression, encoding, transmission, portability, adaptation as well as media driven networks.

## 4.1 Embedded intelligence for search and retrieval in Virtual Worlds

The embedded geometry of 3D objects provides an effective tool for representation-independent search and retrieval in virtual worlds. Indeed, it seems to be very unlikely that different objects present the same geometric structure. This suggests identifying a 3D object with its coordinates in a suitable shape space, whose geometry and topology should play an essential role in 3D matching, indexing, and retrieval. From this point of view, it is worth noticing that natural shape spaces turn out to be highly non-linear, hence standard descriptions exploiting linear feature vectors seem to be intrinsically inadequate.

Further intelligence can be added by providing real objects (and their reproduction in virtual worlds) with an own intelligence, connected with identification codes (IDs, for instance obtained from barcodes or RFIDs), annotation (tags or meta-tags associated with the description of the object and describing it), learning (from past history/use of the object itself or similarity with other objects). Such characteristics can be connected with the intrinsic properties of the object (static features such as shape, colour, texture, dimension, etc.) and/or with its dynamic properties (how it behaves, in which activities it is normally used, who uses it, etc.).

Based on the above information, a future retrieval system may simply query the surrounding world: the smart objects present in the scene will answer according to the information embedded.

Growth of popularity of media is not accompanied by the rapid development of media search technologies. The most popular media services in the Web are typically limited to textual search. This method of searching is not entirely natural for humans. If we are performing a search in the real life, we do not only query by providing a spoken or textual description of the searched item or service. We provide examples: “Do you have a jacket, like this one, but dark green, rather than orange?”; “Could you please give me a haircut, like the one in the catalogue?” This allows us to omit long and unnecessary description of the searched item. Such services are nowadays available in the Web only in form of low scale client-server test-beds and are not present in P2P.

One of the difficulties in designing good mechanisms for distributed autonomous systems is a lack of a unified process for evaluating the efficiency of mechanisms, both in the research community and in the industry. As content distribution systems have widely recognised standard overlays to compare with, in search there is no such base.

Another problem affecting the state of the art media search solutions is the distinction between the relevancy of the answer to the query and the relation of such relevancy to the users' satisfaction with the performance of the given service. This issue is well known to anyone who tried to use a multimedia search service such as Google Image Search. The answers to a given query are often relevant, but not satisfactory.

The solution to those problems is probably the creation of a search benchmarking system for the distributed media delivery system, which would cover all critical aspects, including the user perceived quality. The proposed benchmarking mechanisms should allow, on one hand, to evaluate the performance of the search algorithms, and, on the other hand to make the research and tuning processes easier.

## 4.2 Physical-based worlds (including behaviour modelling and understanding)

The evolution of 3D media will continue the trend in the increase of realism of virtual environments. The focus of FMI will be on realistic interaction, in which the physics of the environment is correctly simulated and transformed depending on the application. This aspect will be challenging in the context of mixed reality setups, coordinating the physics of the real and the virtual world. The interaction realism will allow performing new experiences in the virtual world, improving training and execution of tasks in the virtual environment. However, the entities in the virtual environment will not be limited to visualization and physical simulation. Objects will need to be associated to behaviours, allowing reaction to the user action and the creation of new interaction paradigms between virtual and real worlds. To this end, it requires a tremendous amount of calculation from lots of physical parameters which give effects to the change of virtual object states. On the contrary, any state change of a virtual entity needs to give a feeling to a user as if he/she is interacting with a real physical object and environment. Such a feeling of realism can be realized by various force feedback mechanisms using a virtual environment equipped with wearable hardware. The problem is that current devices are still too uncomfortable to be used due to their complicated, heavy and sensitive characteristics. Therefore, it is a big challenge to harmonize both virtual and real worlds altogether where virtual and real entities naturally interact with each other.

An example of the fundamental model of this harmonization is the physical based cloth simulation. Because of its applications in the textile industry, its benefits for the artistic garment design process and its use within the movie and entertainment industry, cloth simulation is a popular topic in the computer graphics research community, and the cloth animation and rendering techniques have dramatically evolved during the last two decades. However, while there have been spectacular advances in the visual realism of simulated clothes, there has not been a similar progress in the related interaction modalities. The ability to manipulate and modify virtual textiles intuitively using dedicated ergonomic devices has been definitely neglected. Humans are used to handle clothing materials with their hands since prehistoric ages. Nowadays, comparable interaction methods for handling virtual textiles do not exist, and the limits imposed by the use of mouse and keyboard make them a poor solution decrease many a potentiality. Using the sense of touch to interact with computer-simulated clothes would significantly increase the realism and believability of the user experience and ease the creation process of virtual textiles. Moreover, being able to render the feeling of touching digital clothes would introduce a completely new way of communicating 3D products, e.g. during the online purchase of garments or when assessing the fabric hand of specific materials.

### 4.3 Network-driven media engineering

Content- and context-aware media engineering plays a key role in the success of Future Media Internet. This fundamental technology relates to the processing of content aiming at overcoming network limitations and at satisfying the needs of future network and Internet based multimedia systems and applications. It is termed network-driven media engineering and includes several important aspects related to efficient, flexible and scalable storage, transmission and streaming of multimedia content, as well as high quality dynamic visualization and display of new n-dimensional and multi-modal content categories. A summary of the most critical requirements and related research challenges is given in the remaining of this section.

*Fine granularity scalable media compression:* To achieve real-time two-way realistic 3D communication an important end-to-end coding research challenge requires focused attention. Furthermore, standard multimedia distribution frameworks such as MPEG-21 for content formatting and content adaptation should enable face to face audio and video interactions to be held between different types of fixed and mobile terminals connected by heterogeneous networks. To achieve this, the media streams must be encoded with very fine-grained random access points and these must remain accessible when the streams are placed in digital item containers and transport and network packets.

Error resilient coding is also required to ensure that media quality at the receiving terminal degrades gracefully with packet loss, whether from congestion in the wired sections of the network or link imperfections in wireless sections. Multiple description coding and multi-view coding are two example techniques that might be useful for 3D video content. It would seem appropriate to focus on video content for compression improvements, given the lead which has already been established for voice with current VoIP systems.

*Encoding, 3D representations and transcoding:* Any 3D object, up to geometric transformations preserving its shape, could be encoded by its coordinates in a suitable shape space. Despite the non-linearity of such a parameter space, higher mathematics show that such coordinates can be explicitly computed and efficiently managed. Key challenges in this framework are related to the effective representation of 3D worlds either in their full-3D format, or in the different 3D-to-2D conversions, i.e., 3D to stereo, 3D to multi-angle and 3D to 2D. In this sense, transcoding not only concerns the conversion of the coded video among different standards, but also the adaptation to different user terminals, e.g., a full-3D virtual world should be adapted to the players of a game, each of whom has a different viewpoint and a different device, stereo or monocular, single-view or multiview, handheld or wearable, etc.

Additional important aspects concern scalability and adaptability. Concepts such as quality- or resolution-scalability and regions-of-interest have non-trivial extensions in the 3D world. For instance, the concept of quality in the reproduction of a three-dimensional shape deformed by a compression process is hard to define, in particular from a perceptual viewpoint. Well-known concepts such as the rate-distortion function need to be suitably redefined in this context.

All of these aspects need to find appropriate solutions in existing and incoming international standards to allow real interoperability and openness of future 3D media.

*Scalable 3D multimedia streaming:* It includes strategies to consider the number of views and view/stream dependencies, as well the specific 3D display technology of the receiving peer. Here, packet scheduling can make a significant impact on the end-to-end QoE. The main current research challenges in 3D multi-view video streaming are:

- determination of the best video encoding configuration for each streaming strategy. Multi-view video encoding methods provide some compression efficiency gain at the expense of creating dependencies between views that hinder random access to views.
- determination of the best rate adaptation method – adaptation refers to adaptation of the rate of each view and inter-view rate allocation depending on the available network rate and video content, and adaptation of the number and quality of views transmitted depending on available network rate and user display technology and desired viewpoint.
- packet-loss resilient video encoding and streaming strategies as well as better error concealment methods at the receiver.
- best P2P multicasting design methods, including topology discovery, topology maintenance, forwarding techniques, exploitation of path diversity, methods for enticing peers to send data and to stay connected, and the use of dedicated nodes as relays.

# 5 Content-related characteristics of the FMI

According to the content-related FMI-TF members, the following characteristics were identified to play a significant role in FMI:

## 5.1 Co-existence of real and virtual worlds including interactive and proactive autonomous characters

To bridge the gap between real and virtual worlds it is necessary to reproduce the environmental conditions of the target “world” involving as much as possible the entire human senses. Users should feel the virtual environment as a natural one from every possible viewpoint (e.g., the feeling of the temperature should be adequate with the reproduced environment). Furthermore, the appearance of the objects should be realistic in all respects (e.g., correct lighting, colour and texture, and possibly tactile effect, weight, smell, etc.). In the mixing of real and synthetic objects, the imperfection of the real world should be taken into account (perfection does not look realistic: the addition of noise, distortion, deformations may make virtual worlds more convincing). Furthermore, the interaction with virtual objects plays a major role. Multiple users should be able to manipulate and interact with objects and actuators simultaneously, as it would happen in real life.

It is no longer sufficient for the characters to look like dumb imitations of humans or other objects. They have to behave like humans as they are supposed to interact with real users. This is the sole motivation behind the current research going on in the area of proactive and autonomous virtual characters. The vision is to create an intelligent virtual character that cannot only display verbal and nonverbal behavior, but also has a personality much like a human being has, and the character is also able to emotionally respond to the user during a conversation.

This has led to the need of computational modeling of emotions, behaviour, mood and personality for virtual characters. It requires a systematic description of how all of these models can be used to get an emotionally responsive character. However, a critical component that has been missing from these models is a concept of memory – not just a memory of events, but a memory of past emotional interaction. Thus, we need a memory-based emotion model that uses the memory of past interactions to build long-term relationships between the virtual characters and users.

## 5.2 3D Social communities including rich communication aspects

Collaboration systems are already well established in the Current Internet at the level of asynchronous shared workspaces (slow-time) and Internet Messaging applications (near real-time). Voice over Internet Protocol and low quality face to face video communications are also available provided that reliability, video frame rate and ease of call setup are not considered critical issues. However, high quality, high reliability and easy to use Telepresence systems are still the preserve of dedicated network connections. This is a result of the lack of end-to-end Quality of Service (QoS) provision and high bandwidth in the general Internet.

Moreover, as audio and speech are important for face-to-face communications, spatial audio systems will form an essential part of 3D environments in the FMI. People will be able to hold virtual meetings with anyone, anywhere and anytime. The virtual acoustic environment will accompany the virtual visual environment to improve realism with perfectly matching sound source positions and acoustics. The compatibility issues between various recording techniques and reproduction systems will be overcome. Audio recording and reproduction systems will become ubiquitous and high-quality spatial audio will become the norm. In addition to reproduction, recorded audio will be used in analysis for automatic transcription of speech and other audio events that will help maintain a record of interactions between people.

Another important issue in face-to-face communication concerns the remote rendering of the appropriate auditory and visual cues responsible of expressive communication of the empathy, the resonance and synchronization that characterise physical face-to-face situation. The "embodiment" in the communication process is an important issue that the future internet is expected to support and exploit.

Furthermore, the user acceptance of social community models strongly based on technological intermediation strictly depends on the capability of creating realistic and immersive virtual worlds, able to mimic reality, while providing richer experiences and wider accessibility. This includes the possibility of communicating and interacting in a natural and intuitive way with the rest of the world, in a multicultural, multi-lingual, multi-modal, multi-sensorial environment. The Quality of Experience (QoE) should be as much as possible independent of physical location, time, user characteristics (age, skills, etc.), available technological resources, and is closely related to the support of non-verbal components including embodiment, affection, and empathy. Relevant applications range from interactive social games, to cultural events, active music listening, debates, meetings, distributed cooperative work, etc.

## 5.3 Personalised entertainment

One of the most notable features of personalised entertainment services and applications supported over the FMI is envisaged to be its real-time support for interactivity of its users with services and applications. This one feature will enable the wide proliferation of personalised entertainment among numerous individual users as well as users in virtual social communities and groups that are geographically dispersed. In this way, these various groups of users will get in touch, stay connected and entertain each other as they please whenever they want and in whatever way they wish.

Bringing the dynamic atmosphere of a 3D digital cinema/movie or the heat of a football action (or any other live event, such as a sports game, concert, clubbing night etc) to a preferred display is merely one of the future personalised entertainment services and applications that can be provided to users. Another potential application area could be interacting within a theatre play or a computer game, whose script would be interactively written in collaboration with remote users as well as played by a few or all of the collaborators. This kind of an entertainment application would involve non-linear story-telling features, in which the main story would be shaped by various users with some mechanism for voting to choose the best ending and/or providing means to realize all of the possible endings. All of these future applications are believed to encompass technologies that will provide immersive experiences for engaging most (if not all) of the human senses, which in return will influence the degree of key personalization factor, and thus the main acceptance and selling point of these technologies as well as personalised entertainment services as a whole.

Yet another personalised entertainment application or service can be realized in the area of education, as particularly the trend for self-learning is increasing notably nowadays. This is due to the fact that people are making great efforts to use their scarce, yet highly valued leisure time more effectively. They have an entirely new attitude towards spending their leisure time, which results in a totally new application environment for multimedia content access/consumption, namely “edutainment”, which encompasses both education through self-learning and entertainment while learning new things. Studying/learning the wild animals of Africa, archaeology to reveal past lives and cultures of ancient civilizations, sea-life under the oceans, planets in the outer space, atmospheric events when different weather conditions are simulated or purely for “do-it-yourself (DIY)” purposes are just a few examples, which would provide an environment for both learning and entertainment.

Personalised entertainment also includes music industry, where there is a strong potential for novel applications and services, in a broad market perspective: for example, active music listening, fully exploiting context-aware and social music fruition.

### 5.4 Vision of capturing and reproducing immersive content

The emergence of a very high bandwidth communication network should allow very high fidelity communication, supporting both conventional high-quality media such as film & HDTV, and new forms of immersive media that provide a real sense of being present at a remote location. This would have many applications, ranging from high-quality versions of conventional media such as film, through to remote teleconferencing (allowing a user to make a ‘virtual visit’ to distant family members). To provide such a feeling of immersion through vision, fundamental image characteristics such as very high spatial definition, wide viewing angle, high frame rate, wide colour gamut and high dynamic range will be important. 3D aspects of the image and sound will also be crucial, ideally allowing a viewer to change their location to experience the scene from a different angle, thus implying multi-view capture to provide true 3D, rather than just conventional stereoscopy. The video should be accompanied by multi-channel audio to provide surround sound. Even with very high fidelity audio-visual reproduction, the feeling of presence will not be complete without considering other aspects such as haptics and smell, which should be included in the long-term vision. Further, expressive qualities in voice, physiological, and motoric activities, should be tracked in a multimodal synchronized framework and rendered remotely in appropriate ways, in order to be able to enhance the effectiveness of the communication.

# 6 Network-related characteristics of the FMI

According to the network-related FMI-TF members, the following topics were identified to play a significant role in FMI:

## 6.1 Higher bandwidth needs

The Internet is growing and advancing in several dimensions and one of them is the number of users. Currently there are over a billion Internet users and with machines and appliances also connecting to the Internet (the "Internet of Things") the number of users will increase rapidly. A natural consequence of the increase in the number of users is an increase in the amount of content transported over the Internet. Moreover, with more and more demanding Peer-to-Peer (P2P) and video applications appearing, the increase in traffic will grow more than linearly with the amount of users. Within the core, such bandwidth can only be provided with new, optical approaches, such as lambda, burst or optical packet switching

P2P application developers (we can think of BitTorrent as the clearest example) use several ways of achieving higher bandwidth needs. For example, they continually probe for peers with more upload bandwidth, they run TCP sessions for long periods so that their congestion windows grow, and they open multiple TCP flows simultaneously. One of the main consequences of this is that interactive traffic either fights to find sufficient bandwidth or suffers from unpredictable delay and jitter that degrades applications like VoIP.

This means that P2P applications eliminate the assumption on which statistical multiplexing and consequently today's Internet is built. Nowadays there are some users who download (and upload) content 24 hours a day, 7 days a week. With the growth in P2P based TV this can only get worse. Operators could just give the P2P flows dedicated circuits but the end user would suffer the cost of this provisioning. It is clear that a new approach to solve the increase in the demand of bandwidth is needed. But the problem is not only the bandwidth demand, but also the asymmetrical bandwidth characteristics of the current access networks. DSL, cable and wireless network design philosophies no longer reflect current bandwidth usage. The asymmetrical design of these networks, which dictates that downstream traffic is faster than upstream traffic, was originally based on usage patterns from early content-consuming applications like e-mail and web-browsing. However, the continual evolution of applications from content consuming to always-on content distribution has meant that current traffic patterns no longer fit asymmetrical bandwidth assumptions.

This trend will continue with high bandwidth video communications applications, which will eventually become the most demanding traffic. Some video applications run as open ended streams, and do not permit a long start-up delay to fill the buffering that allows delay variations to be smoothed in download or broadcast video applications

## 6.2 Content adaptation

We need to cope with a highly heterogeneous environment in terms of diversity of user devices. The mobile telephone market is quite far from witnessing a consolidation in terms of operating systems and of terminal capabilities. Different capabilities ask for different formats and therefore it is expected that content adaptation will become even more of a challenging problem in the design of future media services. To ensure a real seamless access to such services, it is desirable that such services could be implemented automatically inside the network. Scalable coding will surely help in implementing these features, but we do need to consider also scenarios where more complex functions such as transcoding can be efficiently performed within network elements. This calls for more complex software tasks to be executed inside routers and other devices.

In the specific area of content distribution towards mobile terminals, more efficient and dynamic techniques to store and distribute complex media are required. We expect to move towards a distributed caching architecture with improved management features, again based on the availability of smarter network nodes equipped with high capacity repositories.

The most notable beneficiaries of such context-aware content adaptation functionalities are the media companies that provide entertainment content, such as live media coverage for popular sporting events, musical concerts, etc and their customers, i.e., the general public. Availability of QoE-enhanced services in return is expected to accelerate service providers' efforts to expand their customer base beyond the conventional living room framework and offer a personalised service to their customers over any access network and terminal device of their choice for extended home environments. At the same time, customers will have the liberty of choosing when, where and how to enjoy rich multimedia content and select from whom to get them when multiple content and service providers are available. Most anticipating multi-view and free-view video services will also benefit from context-aware content adaptation that could deliver a user-selected view from an encoded multi-view sequence to a terminal device with only conventional single-view compatibility and thereby winning an even larger audience for the service. Moreover, user-generated content, produced with a restricted set of resources, will be able to penetrate to a wider audience.

In general, context-aware content adaptation is mostly discussed with respect to plain (i.e., unencrypted) content. However, there is a substantial demand, mainly from the business domain, for adaptation of encrypted and trusted content, which may pose a significant challenge for the FMI. Applications that demand end-to-end content security, such as virtual collaboration, security and surveillance, e-health, and e-education, can also reap the benefits of content adaptation, should the FMI be capable of adapting encrypted content seamlessly.

## 6.3 Media driven networks

Content-aware real-time transmission of future media means that the relative importance of each packet towards increasing the end-to-end utility function is established. That is, the more important packets should be better protected (by allocating appropriately network resources) or should be transmitted first in a scheduling scenario. All layers in the communication protocols should be content aware. There are many examples in the literature of current research on media driven wireless networks, i.e. the application of streaming pre-coded video over wireless access networks to multiple users. For this application, many cross-layer scheduling and resource allocation methods have been proposed that exploit the time-varying nature of the channel to maximize the throughput of the network while maintaining fairness across multiple users. A content aware utility function needs therefore to be defined, which enables optimizing over the actual quality of the received video and which is even better suited for multimedia streaming applications. In addition, error robust data packetization at the encoder and realistic error concealment schemes at the decoder need to be determined.

Routers should be allowed to make routing decisions based on content characteristics. Such content information most typically will be provided by the encoder (as part of the header information), but scenarios can be envisioned according to which the router could also extract such information (if also transcoding is taking place). In the latter case encryption issues need to be addressed. DiffServ provides an example of a system where the routers are taking content (priority class) into account. Intelligence could also migrate from edges into the network (active networks, active routers).

There are also issues specifically related to streaming of 3D video. Many alternative technologies for 3D video representation, including holographic, volumetric, geometric (3D mesh models), and multi-view (light-field) exist. Stereoscopic/multi-view 3D video seems to be the most mature technology at the moment. Promising approaches for encoding stereoscopic and multi-view video have been standardized in the form of MPEG “video-plus-depth” and the JVT MVC standards. A multitude of strategies have been considered for streaming such encoded multi-view video using RTP/UDP/IP or RTP/DCCP/IP.

## 6.4 Content distribution

Today's user experience of the network is all about media. Thanks to high speed connections, the Web pages are full of (2D and 3D) multimedia. Video and audio sharing services are attracting a large number of users. Podcasts are becoming a popular source of information. Multimedia is now rapidly moving into the every-day life. Users are not only able to access media via radio, television and the Internet but now also create their own media content. Digital cameras, video recorders, media-enabled mobile phones and the wide spread availability of content creation and editing software, which were previously available only to professionals, all contribute to this data volume growth. The traditional division of users has to be extended by a new kind – the "prosumers". This trend is caused by a fact, that the users are not only able to consume ("consumers") the content, but also are able to become producers ("prosumers") themselves.

The future Internet will therefore be characterized by an explosion of information that can be retrieved (and represented correctly) anywhere and anytime. In order to deal in an efficient way with huge amounts of content, networks of the FMI are made content-aware such that content can be retrieved and delivered with a proper Quality of Service (QoS). Searching - from a user perspective - will become easier through the deployment of semantic search and user-tailored search (e.g., user-based tagging) techniques. The search processes will allow users to search efficiently for content among content, but will also extend the search towards finding content within content. For instance, when we are looking for a particular clip from a large video, this clip is extracted, possibly adapted, and presented to the user conforming to his/her expected QoE.

Unfortunately, with the increasing trend of sharing information over the Internet, misbehaving users also get more opportunities to attack others, especially in decentralized environments such as P2P or ad hoc networks. Future networks need therefore be designed to be robust against such attacks and must have an incorporated ability to recognize malicious traffic.

## 6.5 Overlay networks for content distribution

The main research challenge in overlay networks is the creation of overlay network infrastructures to support the provisioning of media services to end-user communities with little or no support from centralized computing facilities. The existence and reliability of such infrastructures will mainly depend on its members, and on the self-configuration and autonomicity characteristics of robust P2P networks.

Overlay networks in the context of FMI have to suitably support the provision of content services to user communities. The creation of an overlay is mainly justified by the necessity of creating an infrastructure to support the provision of these services among the community members without relying on centralised computing facilities. As such, designing the end-to-end infrastructure involves the creation of new supporting functions that can be implemented either in user terminals or in access networking facilities (e.g. the so called 'home platform'). However, since media traffic is of great importance for Internet Service Providers, and due to the revenue-generating capability of content services research work has to take into account the role of Telecom operators. In particular it has to be investigated how Telecom operators and Internet Service Providers can exploit Traffic Engineering techniques to match their own management needs with the user's expectations in terms of QoS.

Some of the issues related to overlay networks have a wider impact and span, in fact, multiple areas. For instance the specification and measurement of QoS parameters and other metrics that can be used to (1) assess the underlying communication technologies, (2) organize efficient and resilient overlays and (3) suggest most suitable service instances for the end-user should be co-ordinated at the overlay layer. A possible approach to deal with these cross-layer issues is to gather information from the underlying networks and combine it with the higher-level quality assessment and requirements of applications to adjust the overlay networks. In this sense, the use of overlay network technologies is the instrument that makes it possible for user communities to exert some form of control over QoS when no direct interaction with the network is allowed. However, considerable research in this area is still required to come up with an optimal strategy

Collaboration tools fall into a number of categories based on their QoS requirements. These classes match quite closely the broad types of content currently found in the internet and on telecoms networks.

| Collaboration tool | Examples | Characteristic requirements |
|---|---|---|
| Asynchronous collaboration tools | Shared websites and team work spaces | Not delay sensitive<br>May need guaranteed or receipted transactions<br>Large data volumes<br>High integrity |
| Instant messaging | Two way text comms. | <10s one way latency<br>Medium integrity<br>Small volumes |
| Shared synchronous workspaces | Shared white-boarding, document sharing and shared applications | <1s one way latency<br>High integrity<br>Medium/Large volumes |
| VoIP | Two way voice comms | <100ms one way latency<br>Medium integrity<br>Medium volumes |

| | | |
|---|---|---|
| Video conferencing | Two way AV comms | <100ms one way latency<br>Medium integrity<br>High volumes |
| Telepresence | HD quality AV comms with immersion features such as life size video | <100ms one way latency<br>Medium integrity<br>Very high volumes |

***Table 2:*** Collaboration tools and their QoS-based requirements

Virtualisation of Internet can be envisaged as the way forward to support QoS by allowing multiple virtual meta-networks on top of a base heterogeneous multi-domain global network for achieving service differentiation. To achieve this, a multi-topology cross-domain routing mechanism is also required as a suitable platform for supporting service differentiation.

Through the use of virtual networks or programmable network elements, the FMI allows for accelerated innovation and service deployment. This technology can be applied at the application layer or at the networking layer. Virtual networks at the application layer allow users to create their own network with some specific characteristics tuned to their service needs. Different virtual networks or services, which are run over one infrastructure, try to improve on different performance measures. For instance, there could be a virtual network that improves on the end-to-end delay where another focuses on robustness of the communication. Also social behaviour or user similarity can be taken into account when forming the distributed overlay (virtual network), so that management of end users, and performance of content searching can be improved.

## 6.6 Green Media Internet

Although estimates of the carbon footprint of the Internet vary, it is indisputably a big energy consumer. With the increase in high-capacity network equipment and the amount of data centres to accommodate media, energy consumption and costs are rising fast. Fortunately, the future Internet provides two significant opportunities to reduce the amount of energy used:

*The deployment of energy-aware network equipment and routing protocols to reduce the energy consumption of the Internet itself.* One can think of energy scavenging, more energy-efficient equipment (e.g., optical switches), server provisioning techniques that dynamically allocate some servers while placing unnecessary servers in sleeping mode, the alteration of network traffic patterns to facilitate packet processing, etc. Naturally, all such measures should not compromise (significantly) on the QoE perceived by the users.

*The use of ICT and media applications that enable a more efficient use of energy in other areas.* This opportunity is possibly even more pronounced than the first.

Many applications can be envisioned here, for instance 3D videoconferencing applications that provide an alternative to travelling to conferences/meetings, or low-power sensor networks that steer (boot up or shut down) high-power equipment (e.g., smart streetlights).

These opportunities hold for wired as well as wireless scenarios, but in a wireless mobile environment, saving energy has even more benefits (e.g., increased lifetime of nodes and consequently better connected networks). The FMI therefore holds the promise of being an enabler of energy efficient use.

### 6.7 Mechanism to monitor and provide QoE

At the transport layer we currently use UDP for real-time audio and video communications assuming that data communications and other non-delay critical traffic will use TCP. This situation is recognized as undesirable and unsustainable because UDP is insensitive to congestion and will swamp any TCP traffic once congestion arises. As a result, a range of Data Congestion Control Protocols (DCCP variants) have been developed and are approaching the performance of UDP when bandwidth is available, while protecting TCP when it isn't. DCCP with fast restart and small packet modifications is still not as good as UDP or even TCP (for G.711 and G.729 voice). Recommendations have been made to improve these variants to near UDP quality[8] but are apparently still not as good as modern TCP with the correct features enabled. Further modifications to DCCP for video streaming over wireless links and to support multiple QoS (QoE) levels are currently under study.

Skype manages an acceptable solution for low rate, delay critical audio (VoIP) but usage is still orders of magnitude below the capacity provided by the telecoms operators. It is achieved by putting intelligence in the application (out of necessity as this is the only way to provide QoS for a current application[9]). We should be able to do better in a future Internet.

There aren't any clever solutions for video out there yet as their bandwidth demands are much higher. HP and Cisco are still using dedicated networks on the order of 45Mbit/s for their telepresence offerings.

Quantification of QoE using objective and subjective measures, especially for 3D video, remains as a challenging research problem. While measures to predict and

[8] Horia Vlad Balan, Lars Eggert, Saverio Niccolini, Marcus Brunner*. An Experimental Evaluation of Voice Quality over the Datagram Congestion Control Protocol.* INFOCOM 2007

[9] Tobias Hoβfeld, Andreas Binzenhöfer*. Analysis of Skype VoIP traffic in UMTS: End-to-end QoS and QoE measurements.* International Council for Computer Communications – Computer Networks 2008 Volume 52 part 3

evaluate the visual quality of uncompressed 3D video is itself an unsolved problem, developing such measures in the presence of packet losses and other compression artifacts is even more challenging and dependent on the specific display technology. Furthermore, the relationship between network level QoS measures and overall QoE must be studied.

Deployment of QoS-based services at large scale across the Internet requires a large set of providers to co-operate. Appropriate means are required to enable providers to extend their QoS over multiple domains, thus enabling the providers to offer connectivity and high-level services to networks beyond their own domains with QoS performance levels suitable for delivery of these end-to-end QoS-enabled services. New approaches need to be devised for co-operative and distributed quality monitoring of services across many domains.

## 6.8 Identity management

Recent years have seen an explosion in the number of content providers, and this trend is set to continue. More and more content is being produced and distributed by so-called microproviders, for example on YouTube, MySpace or blogs. The creation of multimedia content is moving into the hands of the ordinary user, in addition to the more traditional major distributors of content.

A consequence of this trend is that access to content by consumers is spanning more and more distribution sites, each often requiring identity and personal information to be provided in order to tailor access to the user's preferences, for purchasing or for other reasons. This currently causes a major usability headache for users, as they need to re-enter their details to register on each site, remember lots of credentials and passwords etc. which makes the system unwieldy, unmanageable and slow. Alternatively, users simply re-use their usernames/passwords, what can be considered significantly risky from the security point of view. In addition, having to manage user identities causes a significant overhead to content providers, and especially microproviders, potentially discouraging or even preventing them from being involved. What is needed is a seamless and interoperable way of managing users' identities across multiple providers of content. Ideally, this should be focused on the users and content providers and their ease-of-use, as well as speed of set up at new sites. This should also enable multiple identities to be used and to allow what information is provided to be controlled according to the users' preferences.

Similar problems arise in the area of virtual presence and collaboration technologies. More and more, organisations and companies will collaborate virtually, requiring dynamic access to each other's networks, systems and data. At the same time, this access needs to be controlled based on the users' identities or roles. A particular issue here is that different organisations are likely to use different mechanisms to control access. In addition they will wish to remain in control of who has access to their own resources. To avoid the case of users having to register at each organisation they wish to collaborate with, with the usability problems this gives rise to, mechanisms to provide interoperable and rapid access in such situations will be required, while still enabling each organisation to maintain control of its own assets.

## 6.9 Trust, Privacy and Security

Users expect to be able to take advantage of the future widespread availability of multimedia and access to virtual worlds. At the same time, they need to feel confident that their security and privacy is being protected. Natural questions to ask include: How can I tell that this content has come from where it claims to have? Can I trust the provider of this content? How do I know that my personal information will not be disclosed to or used by those I don't want it to? How do I know that the message and cues I receive from the system are genuine and not the result of malware in one of the intervening layers of software? How can I ensure that my actions are not tracked and traced against my wishes?

The increasing complexity and scale of future media systems will make all of these problems harder to solve. Users wish to be kept secure, but do not want to have to be security experts in order to use and understand future media systems. They expect to be able to make use of such systems and data, with their security preferences and the current context being taken care of automatically and transparently. In addition, cues as to the current security situation need to be provided to them in a way that they can understand. Users do not want to have to put blind trust in systems, but wish to be made aware in a language they understand of the current situation. Any security solutions should work with the user rather than getting in their way, with the consequent risk that users will try to get around security protection, or alternatively switch it off and leave themselves vulnerable. Users need to be assured about the quality of content and who has produced it.

On the other hand, content providers wish to ensure that their content only goes to those who are authorised to access it. Users depend on the correct operation of the underlying network, which implies that also the network should be protected against potential mis-use and attacks.

# 7 RTD Challenges with a major content component

The computer industry and with that the Internet is fueled by Moore's law. This simply means that time will ensure that memory sizes increase exponentially, cycle budgets also, enabling computing tasks that were impossible only years ago. These possibilities in turn enable functionality that is offered to consumers by means of applications of all sorts. The simple example of browsing through millions of video's that are online to a billion users worldwide was absolutely impossible until a few years ago (YouTube). The usefulness of this website was not the only cause of its success. Part of the success of this application is the user interface that was simply embedded into a medium that a large part of consumers had seen before: a web page with a search engine. This meant that the threshold to try this new application was low and therefore appealing.
However, when designing applications that will provide a wealth of new uses combined with multiple information streams (touch, 3D audio, 3D video as inputs for the user and voice, gestures, facial expressions, keyboards, data entry and environment as input for the application) it becomes crucial that applications remain intuitive and simple to use. The first prerequisite of a new application is of course to have a function, to be useful, but all the functionality and usefulness in the world will not save an application that has the wrong UI.

The objective of tomorrow's applications is to unlock a new virtual world with new technology, but it needs to be accessible to the majority of people. This could mean that the applications are either self explanatory, but it might also mean that applications offer excellent low threshold multisensory help. Much development is needed in user interactions for multiple IO's.

Having in mind the aforementioned usability, simplicity and functionality aspects, the content-related FMI-TF members, identified the following RTD challenges:

## 7.1 Multi-sensory capturing/rendering (including object understanding/recognition)

<u>Research is needed on the optimum methods for capturing the visual appearance of the real world, including 3D/multiview, high dynamic range and high frame rate.</u>

In terms of 2D images, we are already at the point where images can be captured at a spatial resolution above that of the human eye, although the interplay between spatial resolution and frame rate still needs exploring. Wide colour gamut and high dynamic range still present challenges. 3D capture techniques (including multi-view capture) are still being developed, and more work is needed both on the capture methods themselves and methods of data representation and compression to allow transmission over different kinds of networks and reproduction on different display devices. To allow meaningful interaction with the captured world, research is needed on robust object recognition and tracking technologies, for example to allow metadata to be attached to real scene elements in an automated way; this will support a wide range of applications beyond conventional passive viewing of content. Optimum methods of surround sound capture need to be determined, to produce a sound field well-matched to the immersive visual experience. Research is also needed to explore the benefit that may be achieved by adding new modalities such as smell and haptics.

There are various challenges that are required from a successful and widely used multi-sensory capturing and rendering system. These can be listed as small size, simple data acquisition interface, low cost, reduced power requirements, low computational complexity of accompanying algorithms and robustness to varying environmental conditions. While various multi-sensory systems exist that meet some of these requirements, a novel system that is easily applicable needs to satisfy all of them.

The size of a multi-sensory system, such as a microphone array, needs to be small, so that it is ubiquitous and easily integrated with another system. The number of sensor elements has to be optimum, so as to capture as many channels of information as possible without a complicated data acquisition interface. As a general requirement, the cost of the overall system has to be low. Power consumption of the system should be as little as possible, especially for systems that are employed in mobile devices, such as laptops. Alternatively, the multi-sensory system should have its own mechanical or chemical power supply that can be replenished easily. The computational complexity of the accompanying analysis algorithms should be low, so that the real-time operation can be considered. Finally, the system should not be affected much by changing environmental conditions or should automatically adapt to it in order to eliminate the need for customization of settings for different environments and to minimize the need for servicing.

## 7.2 Rich interfaces, users' acceptance

<u>Future rich multimedia interfaces will be needed to support the novel input modalities and the envisioned interactive applications.</u>

The development of user interfaces for FMI will witness tremendous advancements in the coming years thanks to the incorporation of novel input modalities (e.g., 3D position and orientation sensors, tangible interfaces, body haptics etc), display and presentation modalities (e.g., 3D displays, micro projectors, augmented/virtual reality etc) and increased richness and complexity of interactive applications (e.g., 3D audio/video, multi-modal support, speech recognition etc).

As we witness a shift from graphical interfaces to multimedia interfaces, we believe that developing richer and more intelligent interfaces is a key challenge to the adoption of the envisioned applications. Such interfaces give more flexibility to the use by offering a choice of modalities depending on the nature of the information conveyed. They also better accommodate a broader range of users with different capabilities.

Finally, multimedia user interfaces provide the adaptability needed to effectively operate in changing operating conditions.

Future user interfaces need to replace current I/O devices enabling natural and seamless interaction with the computer. Joint exploitation of multimodal interaction means (gesture recognition, head/eye tracking) and strong personalization might introduce novel paradigms of interaction. Crucial aspects concern the design, both from graphical and semantic point of view, to make the user feel comfortable while interacting. Rear-projection, multi-touch screens, beamer-projection, 3D projection can play a major role in this scenario, depending on the target application. Particular attention should be paid to users with reduced skills and capabilities (e.g., elderly, disabled) for which the interface may become the access to crucial services such as medical care or home assistance.

### 7.3 Extensibility, scalability, availability

The access to the FMI content should be supported anywhere, anytime and using any terminal.

This is a notion referred to as Universal Multimedia Access (UMA) in the relevant literature. Thus, the 2D/3D media content should be extensible and scalable, in other terms it should be adaptable to various usage environments and available to a wide range of users. For instance, it should be possible to visualize the sophisticated 3D material using light-weight devices, which only have processing and visualization support for 2D rendering and displaying. Similarly, 3D audio content should also be rendered in the best possible way to give the user the sensation of 3D immersive experience although his/her terminal may not be capable of rendering the 3D content. Therefore, it is important to generate the FMI content in a scalable manner during either its capture or compression stage. This may in return require the development of brand-new media coding algorithms or utilization of state-of-the-art tools with added functionalities to cater for 3D content. Where scalability is not inherently supported, then intelligent content adaptation mechanisms (context-aware as well as content-aware adaptation) are needed to enable on-the-fly availability of the FMI content to various usage contexts. Cross-modality (e.g., video, audio, text etc) and multi-sensory (i.e., new modalities for sensory, olfactory information) aspects of the original content should also be provided interchangeably, so as to make the FMI services extensible to various network conditions, terminal features, user preferences etc, and thus available to all.

All of the aforementioned discussions fall into the category of providing UMA in FMI. There is also another significant dimension to be considered here, which is the support for universal multimedia experience Universal Multimedia Experience (UME), so as to provide similar user experiences while accessing FMI content anywhere, anytime while using any device. This point will be discussed further under the topic of QoE later.

As previously highlighted, the delivery of future 3D media requires the availability of sophisticated scalability features, in order to fulfil the requirements of different interfaces, network resources and applications. Scalability and extensibility in particular refer to the video resolution, framerate, colour depth, post-processing and post-filtering, but also 3D appearance (full-3D vs. stereo or monocular vision). Mechanisms such as layered coding, fine-grain scalability, multiple-description coding, etc. need to be suitably redefined to be applied to the more complex 3D context. Quality criteria need to be revised as well.

## 7.4 Security (individual, community, including biometrics)

New solutions will be required to ensure identity management and personal information protection.

3D environments allow interaction with avatars and objects contained in it and all content can be accessed anywhere and at any time. It is therefore crucial to find appropriate solutions for user identity control and personal information protection. Secure access to sensitive information like private biometric and behavioural data should be granted. Active solutions such as cryptography, watermarking and fingerprinting could support different types of control: confidentiality, integrity, authenticity, but also illegal copy prevention and misuse detection. On the other hand, digital forensic should reveal tampering and manipulations. Also in this case, a major problem is in the non-trivial extension of concepts such as imperceptible watermarking to more complex data such as a three-dimensional media.

## 7.5 Embody personality in the virtual characters

Personality and emotions will be embodied to virtual characters to increase their believability and make them distinguishable.

Personification means giving human properties to non-human objects and this issue is becoming important in the world of virtual characters in recent years. While looking at the perspective of traditional intelligent systems such as expert systems or decision support system, having emotions can be seen as a non-desirable property. However, this is not the case for the domain of believable agents since we prefer them to behave as human as possible. Social behaviour of computer characters with emotion and personality increases the realism and quality of interaction such as in games, story-telling systems, interactive dramas, training systems and therapy systems. They also can replace many of the service areas in which real people are employed such as a museum guide or a receptionist. When we talk about personification, we usually consider two factors: Personality and Emotion. Personality is a phenomenon that makes it possible to distinguish between different people. This is also the case in our interaction with virtual characters. When we are immersed into a virtual environment populated with virtual characters in FMI, experiencing that they all behave differently under same conditions increases their believability. Emotion is another major component of personification since in real life emotions affect all a person's cognitive processes, their perceptions, beliefs and the way they behave.

### 7.6 3D navigation with physical and emotional involvement of the user

Novel 3D navigation methods will be needed able to take into account not only the audiovisual but also new types of haptic feedback and user's emotional involvement.

The navigation in real and in virtual worlds will be transformed by the FMI. During real world navigation the user is primarily guided by the audio and visual channel, but in the future the advancement in portable haptic interfaces will allow receiving new types of feedbacks. Haptic feedback will be more transparent and will reduce the mental load of the navigation process. These devices will be wearable and lightweight, providing a variety of stimuli both during walking and on vehicles. In addition hand held augmented reality interfaces will give the possibility to explore the real world interactively projecting information over real entities using hand held projectors; visualization will be possible over flexible surfaces or in general over any surface at the user's disposal. The interaction of the user with the navigation support will become hands free, with the possibility of tracking the gestures performed around the user's space, customizing or modifying the navigation process.

Emotional involvement will imply the reconsideration of the 3D rendering process, which will integrate realism with non-real rendering of objects, users cues, and actions, in order to enhance the expressive and emotional communication within a community sharing the virtual/mixed environment.

### 7.7 Quality of Experience

Novel techniques for ensuring QoE will be needed taking into account 3D media streaming, user generated content and two-way communications.

The realisation of end-to-end Quality of Experience (QoE) relies on the generation of content at the application layer based on the requested QoE at the service layer and network capability. The content then needs to be adapted in line with the network constraints to provide acceptable QoE.

Future users of Media Internet will generate a vast amount of 2D and 3D multimedia content and expect satisfactory QoE. Nowadays, QoE provisioning and monitoring techniques are dedicated either for audio or video (audio-visual content is not considered as a whole), restricted to 2D content, and being still under development. Future Media Internet will involve multimedia streaming in both uplink and downlink directions as user-generated content and two-way communications become more and more popular. IPTV-like services (downlink transmission only) will coexist with user-generated video streaming services in both wired and wireless/fixed and mobile networks. Future multimedia services will focus around user-generated content: YouTube-like services for live video streaming, Assisted Ambient Living (Personal Health

Systems), or intelligent urban environment observation systems. This will cause huge QoE assurance concerns and will require advanced QoE monitoring and assurance techniques in uplink and downlink direction. Diversity of future services (2D, 3D, plethora of audio and video codecs, and bitrates) will require a wide range of QoE-related techniques, being complementary to each other. Future users will not accept low QoE and Telcos would have to provide services that live up to the expectations at all costs.

### 7.8 Human factors

<u>There is a number of human factors issues that need to be tackled for media generation and consumption in general</u>

Issues of the effect of increased internet capacity on demand should be examined in more detail as well as the effects of cost implications on individual services.

The extent to which security functions should be visible to the user also needs further investigation to establish guidelines as to when cues are required and what feedback should be given about progress towards the user's desired content or network access.

QoE can be considered as a human factors issue. With a reasonably simple simulation, the effects of transmission delay, packet loss and reduced bandwidth can be assessed for small groups of game players or collaborators. The large groups typically found in multiparty and shared network collaboration and massively multiplayer online role playing games present more of a challenge.

An additional issue is specific to real time face-to-face collaboration: The round trip delay requirements for multiparty discussions where multiple parties are eager to speak and need rapid feedback of real time cues to judge their interventions ("real-time multi-party arguments" in short.

### 7.9 User Generated (3D) Content

<u>The trend created by UGC is expected to involve also high-quality 3D content able to be used in applications such as 3DTV, immersive environments and so on.</u>

Since the combination of media and networking moves in parallel with the user's drive to acquire more control of their media, the individual "freedom of creation", understood as a generic ability to act either as a consumer, creator or distributor, is constantly increased. In this context, it is expected that in the near future the User Generated Content (UGC) concept, which implies that the user will become an active member of the overall media chain by generating, distributing and experiencing high-quality media content, will flourish[10].

[10] *"The majority of the 7 billion online videos streamed each month are user generated"* (Source Future Exploration Network)

Within a broadband world with social networks, high-speed mobile networks and digital TV, the increased offering of content and related technologies have revolutionised the user experience. On one hand the media industry is finding an audience with an ever growing demand for better content and a better and completely personalised experience and on the other hand the user community suddenly appeared and assumed the role of producer.

We have entered the world of integrated user experience – users who are not only watching and listening to new forms of content on new platforms of technology but also creating 3D content of their own. Thus, the media landscape is undergoing a revolution driven by the more active participation of users. In this context, the UGC requires overcoming of numerous long-term challenges.

The trend created by UGC is expected to speed up the seamless access to content, regardless of its location / repository or terminal device, leading to pervasive media. This will bring new solutions, such as the new distribution models for user generated content. These new solutions will enhance human creativity by enabling a rich digital media-world which, while being associated with our physical existence, is revolutionary in terms of human creativity. UGC will support inventive and creative practices in arts, science, engineering, education, and business by enabling entirely new types of creative media production.

### 7.10 In-network content enrichment

Novel methods for in-network content enrichment and cross-network adaptation will be needed to allow for optimal use of available resources and enriched QoE.

The Future Internet may consider the evolving H.264 SVC (Scalable Video Coding), H.264 MVC (Multi View Coding) and their emerging standards, as the major foreseen content delivery technologies over heterogeneous networks, multiple terminals and large audiences:

- The SVC coding technique may offer layered temporal/spatial and quality content scalability, enabling the user and/or the terminal to select the basic layer and the enhancement layers that it requires in order to decode/reconstruct the video sequence.
- The MVC coding technique may introduce a personalised video delivery experience by allowing the user to select among the different views embedded in a single video stream or by adding depth information to have a basic 3D video presentation.

On the other hand, advanced and innovative technologies involving multi-source and multi-networks streaming like P2P and Media Description Coding (MDC) are also foreseen.

By dynamically combining the inherited content scalability (different content layers, views and representations) of the same resource (video stream), transmitted from multiple sources (different servers or peers in case of P2P streaming) and/or received over multiple diverse paths or networks (utilise the MDC features), on-the fly content adaptation, inherited resiliency and enriched QoE may be achieved. Reconstruction of the content segments may take place either within the network or at the edge of the network (at content aware edge routers) offering transparent streaming to low-end terminals or at the terminal side in case multi-network connectivity is available.

Cross-network adaptation and in-network content enrichment especially in P2P overlay topologies, will offer traffic adaptation (load balancing to avoid network flooding), optimal use of available resources (bandwidth), and enriched QoE.

The timeline for the aforementioned research challenges can be seen in the following table.

| RTD Challenge | Timeline |
|---|---|
| 7.1: Multi-sensory capturing/rendering | Short (3-5 yrs) |
| 7.2\|: Rich interfaces, users' acceptance | Short (3-5 yrs) to Long (>10 yrs) |
| 7.3: Extensibility, scalability, availability | Short (3-5 yrs) to Medium (5-10 yrs) |
| 7.4: Security | Short (3-5 yrs) to Long (>10 yrs) |
| 7.5: Embody personality in the virtual characters | Medium 5-10 yrs) to Long (>10 yrs) |
| 7.6: 3D navigation with physical and emotional involvement of the user | Medium 5-10 yrs) to Long (>10 yrs) |
| 7.7: Quality of Experience | Short (3-5 yrs) to Medium (5-10 yrs) |
| 7.8: Human Factors | Short (3-5 yrs) to Medium (5-10 yrs) |
| 7.9: User Generated (3D) content | Short to Medium w.r.t infrastructure availability<br>Medium to Long for enabling the user to create 3D content) |

| | |
|---|---|
| 7.10: In-network content enrichment | Short (3-5 yrs) to Medium (5-10 yrs) |

***Table 3***: Timeline for the Content-related Research Challenges

# 8 RTD Challenges with a major network component

The RTD challenges to be tackled to realize the vision from the perspective of the networks are the following:

## 8.1 New access technologies

New access technologies will be needed to cope with future increases in traffic volumes and more symmetric bandwidth usage patterns.

Where there is traffic, inevitably there will be congestion. This is a reality in today's broadband networks. Bandwidth saturation and the resulting network congestion are appearing with a greater frequency due primarily to the surge in use of bandwidth intensive applications that have exasperated network infrastructure, and confounded over-subscription models. With goal bandwidth use estimated to double by 2010, service providers must find a balance between maximizing the value of the network for all involved and maintaining the network neutrality expected by its users.

Traffic management offers the best chance of reducing the impacts of congestion control on subscribers and ensuring fair usage of shared and limited resources. No single traffic management technique can fully address all problems relating to network congestion and bandwidth availability. Instead, service providers need to use a layered approach that combines several techniques. Bandwidth intensive applications have changed how the game is played for service providers. The minority of users employing these applications is impacting the overall user experience for the majority of subscribers.

Obviously, new access technologies (e.g. Fibre To The Home - FTTH) will be needed to cope with future increases in traffic volumes. However, due to higher access capacity, the transmission bottleneck might shift from the residential Internet to the core of the Internet. We need a high-speed infrastructure (both access and core) that can support the many bandwidth-hungry media applications. This pertains to fixed (optical) networks as well as wireless networks. Wireless mesh networks is another example of networking technology that provides a number of benefits such as cheap (aggregated) community internet access, self-healing internet connectivity and distributed data storage for backups. Work is still in progress to provide QoS across wireless mesh networks so that they can be used for multimedia applications including real-time communications.

At the access points individual fixed users will buy the bandwidth they require for the services they use whether it is provided by fibre, copper or fixed wireless links. Mobile users will always be at a delay, bandwidth and packet loss disadvantage because of the inherent problems of providing dynamic wireless connectivity such as access protocol delays, limited spectrum and unpredictable channels. This points to an inevitable requirement for content adaptation somewhere in the wired network.

At the edge routers, simple fast high-bandwidth core switching needs to give way to devices that perform high levels of aggregation from thousands of users with different service level agreements to a few high capacity streams for feeding into the core. This requires intelligent routers that can agree, assign and bill a wide range of different services for a large number of customers. This is where QoS will have to be resolved in a flexible and high capacity fashion.

In the core it is relatively easy (although probably expensive) to provision network capacity accurately (and therefore economically) because of the high traffic aggregation levels which result in predictable load. As a result, the QoS mechanisms required for reliable transmission of communications are less important and more likely to have been installed already.

Although necessary, adding capacity will not be sufficient. Adding capacity is often costly, not flexible to shifting traffic demands, may lead to increased bandwidth consumption, and cannot give QoS guarantees. The infrastructure therefore needs to be made QoS-aware.

We list several research directions:

- Research in optical networks relates to setting up lightpaths (the routing and wavelength assignment problem). Similar problems, as currently under investigation for the “normal” Internet need to be tackled, such as path selection, rerouting, protection schemes, constraint-based QoS routing, and multi-domain routing. However, optical constraints (such as wavelength continuity) may make these problems more challenging. An advantage is that the field of optical networking is fairly new and has many opportunities to create new architectures and protocols, while similar new radical approaches to improve the current Internet (the much discussed clean-slate debate) seem more difficult. Research is therefore needed to extend the above-mentioned topics to a multi-domain setting in which lightpaths can be set-up dynamically, reliably, and completely tailored to the QoS constraints of the user/service.

- When designing our networks we also need to make them robust. However, what is robust? There does not seem to be a common consensus on what is robust. We need proper protection schemes to assure that QoS is maintained even under network failures.

- WLAN (with its unlicensed spectrum) may increasingly be affected by interference if the number of users grows. Can systems with a licensed spectrum (e.g., Wi-MAX, Femto cells, etc.) provide a better solution?

- With P2P-based content distribution, the performance of the P2P network is not optimized in terms of routing efficiency, from the perspective of the network providers. Techniques used to better couple the performance of the two layers (AL and IP layer) are needed.

- A future content distribution network is envisioned as a heterogeneous network with wired or wireless devices. How to guarantee smooth multimedia communications and content distribution between different networks with different bandwidths?

### 8.2 New mechanisms for native searching

New mechanisms for native searching (i.e., analyze objects as opposed to only their metadata), including real-time mechanisms, will be needed.

In the FMI the vast majority of transported data will be media objects. Irrelevant of whether this happens on a file or on a stream basis, media objects will take over the role of today's text-based HTML/XML files.
In addition to the severe consequences this shift in character of the transported data has (and which are described in Sections (6) and (7)) one of the most important applications of the internet – the content search – will also have to undergo dramatic changes and developments:

Media objects often are described by so called metadata (data describing data). This on the one hand enables a dramatic reduction in the amount of data required to describe a scene and on the other hand paves the way to using the classical text-based search engines (Google, Lycos, Yahoo, etc).

There is, however, an inherent restriction of this approach: The description either is "free-form" (which means it is edited manually by a human being or by a community of human beings (social tagging)) or the description does follow a rigid metadata scheme (e.g. MPEG 7) (then can include semantic representations based on machine-readable ontologies). In the first case there is a restricted semantic content in the metadata (at least the semantic content cannot be extracted by a search engine) while in the latter the expressible semantic is finite and restricted to the ontology. Metadata per definition can only represent a subset of the underlying object's content!

The desire to achieve search results like the ones user's are used to today (formulating whatever combination of textual objects ("strings")) is by searching the media objects directly: Input a chord progression and the search engine should give results of music pieces following the same or a very similar chord progression; input a photo of your favorite pet and the search engine should retrieve results of the same breed.

Several activities have been started on native media search engines, but the change to one of the most essential applications of the Internet – the content search and retrieval – is so fundamental that it is seen as one of the RTD challenges with a major network component.

### 8.3 New mechanisms to optimize content distribution

New mechanisms to optimize content distribution from the users' (e.g. faster transfers), network providers' (e.g., local transfers), and content providers' points of view as well as from an energy efficiency point of view (i.e., to support mobile devices and to be more environ-mentally friendly) will be needed.

High intensity traffic applications and in particular P2P contribute a significant portion of today's Internet traffic. By building overlay networks that are oblivious to the underlying Internet topology and routing, these systems have become one of the greatest traffic-engineering challenges for Internet Service Providers (ISPs) and the source of costly data traffic flows. Recently, some ISPs have attempted to reduce p P2P traffic by placing caches at the ISP's gateway to the Internet or by using network appliances. The legality of these approaches is questionable. By caching content, ISPs may become participants in illegal distribution of copyrighted material, while interfering with P2P flows in a non-transparent way may not only break the law but also lead to a significant problem.

New solutions are coming from a close collaboration between ISPs and P2P applications where the ISP can reduce the cost impact of high demand traffic applications in the core and peering agreement and the end users can improve their accessibility to the content.

P2P technology, as an alternative to the traditional client-server model, has provided a scalable framework for content distribution. Another technology for content distribution is provided by using the dedicated broadband networks of service providers. In this case, content is often distributed via multicast on the networking (IP) layer.

With the increase in capacity of mobile devices, the popularity of video-based services and applications will extend into the wireless realm as well. However, to date, it is not clear which technology (IP-based, P2P-based, or a combination of the two) is best suited to deliver streaming media to wireless (and wired) devices. It is clear that we need to have a notion of QoS to support streaming media.

Some questions that need to be addressed are:

- Can virtual or programmable networks provide a flexible and future-safe architecture?
- How to compose, configure and provision networks in virtualized environments?
- Can we make our routing protocols energy-aware?
- Where should we store content, such that it can be retrieved by the user fast and without loss of quality?
- How to spread information of particular content as fast as possible through the network while reducing the overhead used for information exchange.

## 8.4 New mechanisms for content identity management

New mechanisms for content identity management, content protection (content integrity and user privacy), and to ensure content veracity (integrity and using reputation systems) and the right of use will be needed.

More work is needed in the area of user-centric security, to provide easy-to-use security solutions that naturally fit in with the way users behave and present the current security situation in a way they understand. To achieve this, the complexities of security management need to be hidden from the user (transparent). This requires systems to have built-in and automated trust, security and privacy management technologies. This also requires automated context aware security, in order to ensure that the right security is applied at the right time.

At the same time, security must not be completely transparent if users are to trust the system, and in fact this is unlikely to be possible due to the complexities involved unless numerous incorrect decisions are acceptable. Users need to be adequately informed about the trust, security and privacy risks. Multidisciplinary research including the social sciences is needed to ensure that technologies that are applied are easy-to-use and fit into the way people behave in a natural way. Such research is also needed to ensure that the current security situation is presented in a way users can understand and make informed decisions about.

In identity management, user-centric technologies are required that place control and ease-of-use of the end user at the forefront. Key challenges are to enable single sign-on across multiple content providers, and allow users to control access to their private information in a simple and easy to use way. Federated identity management technologies are needed to provide rapid and seamless access across multiple organizations in virtual collaborations. These need to cope with heterogeneous security technologies and systems and thus provide interoperability between the partners in the collaboration.

Technologies for data provenance and integrity are required in order to reassure users as to the veracity of the content provided to them. On the other hand, content providers require technologies to enable them to enforce access control on content. Ideally this will enable "sticky policies" to be applied to content which last for its lifetime.

Content integrity can be achieved by exploiting either cryptography or suitable data hiding methods. A posteriori analysis via digital forensic techniques could also support tampering detection and therefore integrity verification.

## 8.5 New mechanisms to evaluate and provide the appropriate QoE

New mechanisms for evaluating and providing the appropriate QoE will be needed.

The *QoE* is the quality that a customer or end-user subjectively perceives/experiences when s/he consumes the service at his/her terminal. The *QoE* monitoring is normally performed at the level of an individual application stream perceived by a user. In contrast and for scalability reasons, the NQoS is measured for a specific QoS class and at an aggregated level. In large scale networks, individual QoE monitoring can be costly. For scalability reasons it might be necessary to find out about the QoE perceived by a collection of users consuming a number of application streams delivered at an NQoS-class level. Using this *NQoS*, an approximation of QoE could be deduced. Each application has certain characteristics and can use different encoding schemes. Therefore, various prediction/mapping models need to be devised to derive the QoE for each application (e.g., in case of a certain network level packet loss, delay, jitter and throughput).

As presented in Figure 2 services related to future Internet would involve QoE measurement, monitoring and optimization in centralized (network servers) and distributed architecture (QoE measurement in end users' terminals). What is more, QoE assurance concerns will affect both uplink and downlink transmission direction in wireless and wired/fixed and mobile networks.

The most appealing challenges related to QoE monitoring and optimizations in FMI are:

- Implementation of accurate QoE metrics dedicated to emerging Future Internet services, being light-weight in terms of computation and resources consumption, applicable in-service in end user's devices (mobile phones, set top boxes, etc.) and network devices.

- Implementation of QoE optimization mechanisms, based on information sent from active probes or end users' devices in feedback channel. QoE optimization should be possible in 'per user' and 'per service' mode at the same time. QoE optimization can be achieved due to traffic shaping and prioritization techniques, IP packets prioritization, end-to-end congestion control schemes and resource allocation protocols, and other techniques.

- Additional mechanisms should allow for long-term QoE optimization (related to network architecture rearrangement), based on QoE statistics.

- Application level techniques for QoE assurance, like Forward Error Coding (FEC) or Automatic Repeat Request (ARQ).

- All mentioned features should be applicable in uplink and downlink transmission, applicable in wired and wireless/fixed and mobile networks.

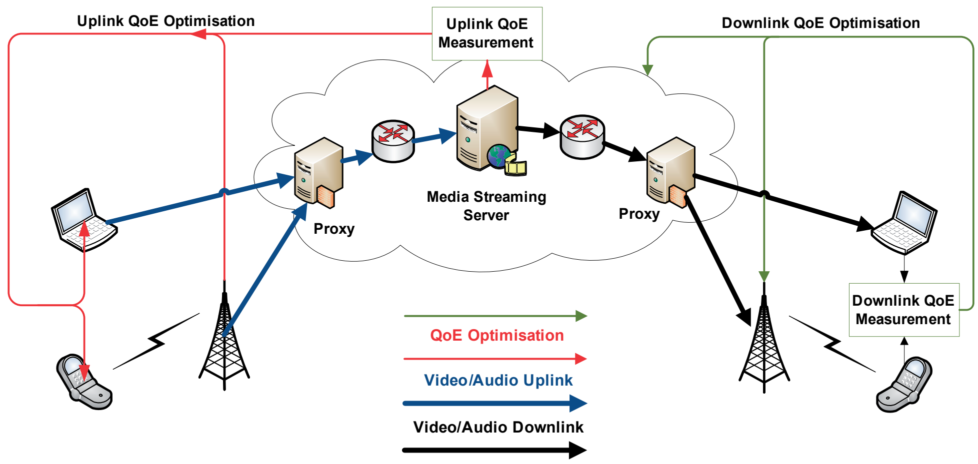



***Figure 2**:* Assuring and optimizing QoE for FMI

Currently there are no clear guidelines on how to dimension and tune the network and how to configure the streaming parameters in order to achieve the desired QoE level. We need tools to assess the QoE and find new ways to improve upon the QoS if the QoE was perceived as unsatisfactory.

Questions that deserve attention are:

- QoS routing is the process of finding paths through the Internet that satisfy certain performance requirements. The value of QoS routing is undisputed, yet its complexity (or fear thereof) has proven a big hurdle on the road towards realization. QoS routing will only be adopted en masse, once its complexity is contained. Therefore, we need a framework for QoS routing that avoids (NP-)hard scenarios and enforces a feasible complexity. Can we overcome the complexity issues of QoS routing by properly designing and managing our networks?

- How to search for content among content, but also within content? Once the desired content is found, it should also be delivered to the user with a satisfactory QoE. Content may need to be processed or adapted before being presented to the user. This requires a certain awareness of the resources and demands of the user.

- In a wireless environment, the extra challenges created by device mobility, link instability, and energy constraints, require cross-layer design. Research is needed to ascertain a high QoE in a wireless scenario, possibly through the aid of distributed computing and storage.

- Search engines can be extended to semantic query or visualized search query, i.e. image, music, segment of a movie as the search reference. Distributed searching is also a technique to be expected in future Internet.

### 8.6 New network models to guarantee scalability

New network paradigms have to be devised in order to cope with substantial scalability needs in different dimensions of the network.

We are witnessing an enormous increase in the need for better applications, more efficient services and improved capabilities of the networks, for example:

- Number of connected terminals
- Packet-switching capacity at network nodes
- Number, and connectivity degree, of Autonomous Systems
- Number of multi-homed sites,

Current technologies are using mostly brute-force approaches to address the above mentioned challenges and critically needed improvements in features and dimensions of the network. This in turn is leading to either unsustainable growth rates or to extremely inefficient solutions that severely harm the cost-effectiveness of the service. Indeed, the brute force approach is usually reflected in largely increased capex or opex costs, reduced functionality and reliability, or a combination of them.

What is required is fundamental research to provide solutions that address the root of the scalability problems, by modifying key elements of the network in order to allow scalability without significant drawbacks. Research focus should be on new routing and addressing architectures, as well as congestion control models and traffic engineering techniques.

### 8.7 New networking techniques to cope with network diversity

New networking techniques able to cope with network diversity will be needed.

The Internet started as a fairly homogeneous technology in terms of terminal capabilities, network node capabilities, network service (i.e. unicast packet delivery), network control, and protocol architecture. However, the network growth has triggered an increasing system diversity that is stressing its current design. We may see an increased diversity in terms of:

- *Capabilities of terminals:* Terminals are not any more basically servers and PCs, but now are a wide range of systems, from supercomputers, to the simplest sensors and actuators, passing through conventional servers, personal computers, PDAs, presentation devices, moving networks, middle-boxes, transcoding elements, and so on.

- *Types of System Protocols:* A system is now required to have installed a huge number of interrelated protocols (DHCP, MIP6, IPv4, IPv6, ARP/RARP, TCP, UDP, RTP, RTCP, SCTP, SIP,ICMP NEMO, ISIS, BGP, etc), installed over a range of different, and simultaneously present, physical network interfaces (Ethernet, Wifi, WiMax, GSM, GPRS, HSDPA, LTE, Bluetooth, infrared, USB, etc).

- *Protocol relationships:* Not only the number and types of protocols and physical interfaces is becoming increasingly diverse, but the interrelation between them is also becoming increasingly complex. The traditional separation of the protocol architecture in three planes (user, control and management) is no longer enough to cope with this diversity. We see that systems include larger and larger of instances of so-called “cross-layer issues”, that are in fact violations of the user-provider protocol layering model. This is not wrong in itself, unless the violations are so numerous that have de-facto destroyed the actual system architecture, making the system a monolithical device extremely difficult to maintain, configure, and evolve.

- *Service models:* The traditional end-to-end best-effort unicast packet delivery model of the internet has been largely overwhelmed by a large range of different services that are considered essential today. These new services include terminal mobility, network mobility, multicast, anycast, NAT, security middleboxes, VPNs, multihoming, multihpath, multiprotocol, delay-tolerant, delay sensitive, network caching, media specific and others.

This already large, and constantly increasing, diversity poses a research challenge on defining new fundamental models that can cope with diversity without rendering the system a chaotic composition of unrelated patches. Fundamental research on network virtualization, media-driven networks, autonomic content networks, new non-layered protocol architectures, and modular network service should lead to systems that can cope with the diversity needs that we are observing.

## 8.8 New Mechanisms to Improve Dependability

The increased usage and importance of the future 3D Internet makes dependability an important challenge.

A system is called dependable, whenever reliance can justifiably be placed on the services it delivers. We thereby interpret the term dependability in a broad way, as encompassing availability, reliability, performance (quality of experience/service), security and privacy. Tailored to the 3D Internet, this means that research is needed to make the various components of the 3D Internet more dependable; these components include (amongst others) the core, access, wireless, overlay and sensor network. Research should focus on design as well as on operational (management) aspects. Although the design should strive for autonomic operation of the network, human intervention by the managers will still be needed to handle unexpected cases and deal with calamities.

The timeline for the aforementioned research challenges can be seen in the following table.

| RTD Challenge | Timeline |
|---|---|
| 8.1: New Access Technologies | Short (3-5 yrs) to Medium (5-10 yrs) |
| 8.2: New Mechanisms for Native Searching | Long (>10 yrs) |
| 8.3: New Mechanisms to Optimjse Content Distribution | Short (3-5 yrs) to Medium (5-10 yrs) |
| 8.4: New Mechanisms for Content Identity Management | Short (3-5 yrs) |
| 8.5: New Mechanisms to Evaluate and Provide the Appropriate QoE | Short (3-5 yrs) to Medium (5-10 yrs) |
| 8.6: New Network Models to Guarantee Scalability | Short (3-5 yrs) to Medium (5-10 yrs) |
| 8.7: New networking techniques to Cope with Network diversity | Medium (5-10 yrs) |
| 8.8: New Mechanisms to Improve Dependability | Medium (5-10 yrs) |

***Table 4***: Timeline for the Network-related Research Challenges

# 9 Scenarios

This chapter introduces three possible usage scenarios: the immersive entertainment scenario, the remote presence scenario and the learning/training/rehabilitation scenario which demonstrate how the characteristics of the FMI can be combined to provide the users the best possible experiences.

## 9.1 Scenario 1: Immersive Entertainment

John, a 32 year-old chemistry teacher is playing his favourite “alternate reality” massively multiplayer online game in the evening. His avatar, in the game, is an alchemist doing his quest to discover the philosopher’s stone. The game, however, has many connections to his real life. During the game he is being called by one of his students and is asked for help in solving a particularly difficult chemistry homework.

John is able to instantly transfer his avatar from the reality of the game to his private 3D meeting space. The transfer is seamless for the user’s avatar – it’s just like being transferred to another world – from the medieval alchemical laboratory to a modern surrounding. John invites the student to his 3D meeting space and arranges the meeting with the student in his virtual laboratory. There, he searches the repositories of items and ingredients and quickly prepares a virtual presentation for his student. The presentation goes well and explains the problem to the student. John wants to recreate the experiment in reality, in front of his class. He is now able to instantly check whether the ingredients and items he used for the experiment are available in stock of his school’s chemical laboratory – and if not – he is able to order them from the internet store just with few clicks.

Later in the evening, while still playing his favourite game and while exploring the game’s world he visits a virtual house of his friend (whom he never met in the reality). There, he sees a particularly interesting book on his friend’s shelf. While browsing through the book he is informed that the book is an avatar of a real book being in possession of the friend. Moreover – the book is offered for sale. Directly from the game John is able to enter the auction of the book at a popular online auction website and place his offer.

This way John’s favourite game mixes in a seamless way with his real life offering him a huge variety of possibilities.

Novel entertainment paradigms will also need to face the problems of increasing blurring between real life and virtual worlds, of potential causes of disorders in personality in users exposed to such novel scenarios.

The above scenario copes with the following challenges:

- Challenge 4.2 & 5.1: in the scenario, visual realism of simulated entities could be realized to satisfy users to feel virtual world as real world. It is one of key technical challenges to accomplish visual realism of physical based simulated entities on the virtual worlds.

- Challenge 5.2: people can enhance their social relationships in the real world through ad-hoc, instant, and immediate social network on the virtual worlds.

## 9.2 Scenario 2: Remote Presence

The ability to give a true feeling of remote presence will, at simplest, support existing forms of media such as film and HDTV. Cinemas are already striving to give an improved sense of immersion, through stereoscopic films and surround sound including both the localization of voice, sound, noise sources and the remote re-creation of acoustic environments. It should be possible to deliver these experiences over a network in very high quality. The availability of a system giving a vivid feeling of presence should enable new forms of content, building on the fact that the viewer will feel a part of the scene, rather than acting as a detatched observer. The ability for the viewer to move their viewpoint and view direction (so-called free-viewpoint video) will be significant. Novel systems that a consumer could use to capture their environment and experience would allow 'virtual visits' to friends and relatives, and new ways of recording events in their lives. Such systems may finally make the videoconference a practical alternative to a face-to-face meeting.

The many meanings of the term "Presence" are very much debated nowadays, and many definitions are equally valid and recognized. Surely relevant to the present document is an interpretation of "remote Presence" that does not imply sending in a remote location an amount of data sufficient to the reconstruction of a virtual-self as it is commonly intended for advanced video-conference applications: in the context of the future networked virtual and augmented reality remote presence could also rightfully be described as the ability to capture the essence of a remote location - with respect to the user -, and streaming high-bandwidth video, audio and haptic data over an advanced network framework, so that such streams could be fully rendered at a remote location. Both capturing and rendering would be obviously compound activities, and in the context of mixed reality environments should be considered intrinsically multimodal. For example, capturing is likely to involve advanced video technology, stereoscopic and/or panoramic filming, spatialized audio sampling, as well as various forms of force measurements that could then later be used for haptic rendering, and more. Why should novel network services be necessary? It is likely that such a large amount of heterogeneous data will need to be conveyed on the network in the form of interleaved streams, combining existing data compression techniques to novel approaches that could take into account the additional burden and restrictions imposed by the required level of immersion. A large amount of data will need to flow over the network infrastructure, requiring novel approaches to Quality of Service assessment and dynamic handling. At the other end of the network, an ad-hoc decoder and de-multiplexer will need to be formalized to allow for proper immersive rendering that could take place both in Virtual Reality or in a Mixed Reality where AR technologies are employed to augment or fuse the perception that the user has of the remote location.

A first, typical kind of scenarios would be the one of an expert industrial technician performing a virtual inspection of a real production plant from a remote location, with a level of immersion and capture/rendering fidelity that could allow for him to properly exert his/her diagnostic skills. Capture technology would be placed in a real plant, producing streams of complex data to be sent over the network to a remote location, where advanced rendering techniques and apparatus will represent, following a fully immersive and multimodal paradigm the remote reality. Provided with an adequate level of fidelity, the technician will be able to inspect the remote location as if "s/he were there", exerting the related diagnosis ability and maybe even interacting physically, by means of tele-operated equipment with the machines he is checking out.

In another kind of scenarios the objective shifts from an as-most-as-possible exact replication of the physical world to the remote communication among users collaborating in high-skill tasks. An exemplary case of this kind of scenarios is networked music performance: it is a skill including different layers of social interaction, strong physical embodiment, social emotional entrainment among users. Further it is a consolidated task which may be measured and compared with the local case, where professional users are source of detailed information on social cues and on the effectiveness of the task. Music performance is an ideal testbed to investigate and measuring aspects on “being in tune” each other, “agreement”, “conflict”, “empathy”, “leadership”, etc. which are core issues in remote presence. Furthermore, networked music  performance requires, from one hand, a detailed, fine-grain re-creation of some views of the remote physical world, including the rendering of the remote acoustic environment. From the other hand, it does not necessarily require realism in the visual rendering of the remote user or of all the objects involved in the scene (e.g. the musical instrument): rather, a remote musician needs for detailed 3D embodied representations of the fine-grain expressive and emotional cues responsible for the social empathic engagement, to guarantee a collaborative performance of the task comparable to the local performance. An optimized network communication able to guarantee the best multimodal synchronization (and minimum, constant, predictable network latency) is also a fundamental issue.

Further, the networked music performance scenario enables to develop metaphors and models of the computer network: as part of the acoustic environment, e.g. the consequences of latency in the network become part of an overall shared acoustic environment, in terms of resonances and a of specific virtual acoustic environments emerging from network features.

Challenges involved in Remote Presence scenarios include the following:

- Challenge 4.2 on physical-based worlds: at least part of the physical world should be reproduced carefully in the remote site, including acoustic environments.
- Challenges 4.3, & 6.1: in scenarios where a fine-grain networked social interaction occurs, it is important to guarantee the best multimodal synchronization, as well as constant, predictable latency.
- Challenge 5.1: involving as much as possible the entire human senses; development of computational models of emotions and of novel approaches to model avatars of remote users.
- Challenge 5.2 on 3D social communities is fundamental as for novel face-to-face remote rendering of appropriate auditory and visual cues responsible of expressive communication of the empathy.
- Challenge 7.1 & 7.2 are important in both kinds of scenarios of Remote Presence, in order to provide the incorporation of the needed novel input modalities, of display and presentation modalities.
- Challenge 7.7, on novel techniques for measuring quality of experience, is also fundamental in order to measure the fine-grain social cues, e.g., level of empathy, social engagement.

## 9.3 Scenario 3: Learning, training, rehabilitation

In a new world, where physical environments are augmented, thus enhanced with virtual ones, one can be easily immersed into a setting s/he wishes for. These settings can be various, and thus be utilized for numerous learning and training purposes, where users are spoilt for choice of topics based on their wide range of interests and/or needs. Given the availability of very high-bandwidth connections, low system response delays for reacting to user interactions in real time with a provision for fast and efficient processing mechanisms, all of which are envisaged to be supported by the FMI technologies, any material for learning or training can be made available for access in this new world. A typical scenario will be in the area of self-learning, where with the use of multi-sensory rendering of 3D context-aware and adaptable material, users will delve into topics that they prefer to learn or often find it hard to learn through conventional methods. In return, this will result in significantly enhanced learning experiences with the help of reproducing the immersive learning content. Another scenario will involve training of professional as well as non-professional individuals in several disciplines. Similar to enhanced learning experiences via virtual augmentation, more realistic scenarios will be reproduced with the co-existence of real and virtual worlds through the use of technologies on offer by the FMI. These scenarios can be various, such as training of emergency service personnel, practicing a range of operations (e.g., medical, military, social etc), exercising precautionary procedures for public safety situations, unprecedented events, natural or self-inflicted disasters etc. Another potential scenario may involve the deployment of immersive real, virtual and mixed environments together with their rich yet adaptable and accessible media content for the purposes of rehabilitation of individuals who need physical or psychological help, assistance and/or treatment. These environments can thus be custom designed with the help of specialists in relevant disciplines while making this particular FMI application scenario very much interdisciplinary based on the targeted focus groups.

Rich 3D media and interfaces open new opportunities in the development of advanced tools for medical rehabilitation and training. Physical and cognitive rehabilitation can be accomplished through vision-based systems, providing important additional capabilities such as the possibility of storing a history of the exercises, record the progresses of the patients, measuring in a quantitative way the exercise quality. This will ease the work of the medical personnel in evaluating patients' conditions and defining personalised care. Visual interfaces can be based on both a single-camera in the case of simpler 2D exercises, or multi-camera systems that might be meaningful in analyzing for example walking, complex 3D movements of articulations. Different types of interfaces could be considered such as tools for rehabilitation within medical institutions (high accuracy) and tools for home rehabilitation (less accuracy, affordable price, communication with remote hospitals for evaluation). Furthermore, easy interfaces can be built upon these technologies that could allow persons with reduced capabilities (elderly, disabled) to improve their interaction with other persons and the environment.

In the rehabilitation scenario the FMI will be able to transform the current local rehabilitation service into a distributed service, in which the patient can receive rehabilitation directly at home. The advancement in immersive technologies allows a direct link between the local patient and the remote therapist, in which mixed reality environments enhance the capacities of the therapist. In this way the expertise of the physiotherapist will be made available to a larger number of patients, reducing geographical barriers. At the same time such systems will allow the online measurement of patient performance, adapting, in real-time, the strategy of the therapy. The advancements in 3D content encoding will allow to personalise the scenarios for a given patient, based on new authoring tools that take into account the specific needs of the patients and the on-line assessment of the therapy session. The immersive capabilities of FMI and their networked nature, will give the possibility to create realistic training session able to transfer new skills to novices or to improve the level of experts. The users will be able to select specific scenarios in the areas of professional activity, sport, entertainment or also medicine, and perform training session, in which the requested skill is measured and enhanced. This scenario will take advantage of new encodings of human skills that capture the enactive knowledge behind human actions.

Elderly monitoring, therapy, and rehabilitation is one of the main challenges of our society. The use of FMI to implement services for fall prevention and cognitive decline measurement, e.g. by means of monitoring motoric activities, will be an important component that FMI should support: for example, future mobile systems might include sensors for multimodal monitoring and communication of user data for therapy and rehabilitation.

Another typical scenario could involve on-the-job training of industrial large mechanical devices. Such machines require periodical maintenance to be performed by trained personnel and it is common practice in the field to provide technicians frequent training courses on how to operate, maintain and repair the very large number of machines used in the various industrial plants. Immersive mixed environments could be used to provide “on-the-job” training directly at the physical location of the trainee. Using wearable AR and VR equipment, and by means of large-bandwidth, reliable and low-lag network communication a remote trainer could establish a multi-modal communication with the in-situ technician (for instance by means of visual, audio and haptic technology), guiding his movements while teaching how to perform the various operations involved in the servicing of the machines, as well as monitoring the trainee progresses. The ability to provide multi-modal, immersive “on-the-job” training from a remote location on any type of existing machines will be a significant breakthrough in the Industry practice, compared to the traditional training approaches based on oral presentations and the use of textbooks and video material. These are, in fact, not very efficient means to train technicians because practical aspects are of much more importance and sensorimotor skills are central elements of many training processes. Internetworked mixed environments also remove the limitations of traditional courses caused by the need of centralized training centres and the scarce availability (and functionality) of machinery mock-ups.

This scenario poses several challenges to the research having strong demands both in terms of content and networking:

Challenge 4.2 & 4.3: The starting point for obtaining enhanced learning, rehabilitation and training systems based on FMI is represented by the creation of Physical-based Worlds and the capacity of transmitting multiple sensory channels on the Network.

Challenge 7.1 and Challenge 7.6: The cornerstones of this scenario are the immersive rendering and the capturing of user's motions and actions in real-time integrated with the generation of physical feedback, for guidance and interactivity.

Challenge 7.9: In parallel to these technological advancements, the learning process mediated by FMI requires the generation of new enhanced content produced directly by the users supporting intelligent adaptation to final user characteristics.

Challenge 8.2, 8.3 & 8.4: The distribution of content for learning and training requires also an improvement in the search techniques adapted to the new type of content, while at the same time the complexity of enhanced training material poses challenges to the distribution on the network. The last challenge posed by this scenario is related to the protection of the content and the safe identification of the users.

# 10 Conclusions

This White Paper is the result of several brainstorming sessions of a group of experts from the EU and the USA on new research trends on the Future Media Internet with respect to both network and content. During the discuss sessions, research challenges were identified and highlighted with the goal of determining the current trend and future perspective of the research community on the Future Media Internet. More specifically, it is envisioned that the FMI is expected to involve the following characteristics:

- *interactive/proactive autonomous characters;*
- *3D social communities;*
- *personalised entertainment;*
- *capturing and reproduction of the real world in 3D;*
- *coexistence of virtual and real worlds maintaining perceptual coherence.*

These characteristics should be supported by the future network which should be able to handle higher bandwidth needs and to provide content adaptation in the network and the terminals, new models of content distribution and new network management modules and mechanisms to provide and monitor QoE, trust and privacy.

Based on the abovementioned characteristics of the FMI, the Task Force identified the following major research challenges:

- *Embedded intelligence for search and retrieval in virtual words;*
- *Physical-based worlds (including behaviour modelling and understanding);*
- *Compression, encoding, transmission & Portability and adaptation;*
- *Media driven network;*

Taking into account the above, the experts created common scenarios where the major challenges of the FMI are expected to flourish. Next steps towards highlighting and better defining the ingredients of such an evolving field are the continuation of the research discussion under the umbrella of this Task Force that constitutes the field for envisioning the Future Internet from the perspective of both the media and networks.

## Editors

**Isidro Laso - Ballesteros**
European Commission
Networked Media Syst. Unit
Isidro.Laso@ec.europa.eu

**Petros Daras**
Centre for Research and Technology Hellas - Informatics and Telematics Institute
daras@iti.gr

**Carmen Guerrero**
Universidad Carlos III de Madrid
carmen.guerrero@uc3m.es

## Annex I - List of contributors

| No | Expert's name | Organization | e-mail |
|---|---|---|---|
| 1 | Isidro Laso-Ballesteros | EC – DG – INFSO – D2 | Isidro.LASO@ec.europa.eu |
| 2 | Martin Hahn | EC – DG – INFSO – D2 | Martin.HAHN@ec.europa.eu |
| 3 | Petros Daras | CERTH-ITI | daras@iti.gr |
| 4 | Nadia Magnenat-Thalmann | University of Geneva, Miralab | thalmann@miralab.unige.ch |
| 5 | Safak Dogan | University of Surrey | S.Dogan@surrey.ac.uk |
| 6 | Francesco De Natale | Trento University | denatale@ing.unitn.it |
| 7 | Ebroul Izquierdo | Queen Mary, University of London | ebroul.izquierdo@elec.qmul.ac.uk |
| 8 | Graham Thomas | BBC R&I | graham.thomas@bbc.co.uk |
| 9 | Massimo Bergamasco | Scuola Superiore S.Anna. University of Pisa | bergamasco@sssup.it |
| 10 | Mikołaj Leszczuk | AGH University of Science and Technology | leszczuk@agh.edu.pl |
| 11 | Robert Paul Berretty | PHILIPS | robert-paul.berretty@philips.com |
| 12 | Radu State | INRIA | Radu.State@loria.fr |

| No | Expert's name | Organization | e-mail |
|---|---|---|---|
| | | | |
| 13 | Anton Nijholt | Universiteit Twente | anijholt@cs.utwente.nl |
| 14 | Gonzalo Camarillo | Ericsson Finland | gonzalo.camarillo@ericsson.com |
| 15 | Jaime Delgado | UPC | jaime.delgado@ac.upc.edu |
| 16 | Olivier Festor | INRIA | Olivier.festor@inria.fr |
| 17 | Carmen Guerrero | UC3M | carmen.guerrero@uc3m.es |
| 18 | Thorsten Herfet | Saarland U.& Intel | Thorsten.herfet@intel.com |
| 19 | Fernando Kuipers | Delft Univ. of Technology | f.a.kuipers@tudelft.nl |
| 20 | Tim Masterton | Thales R&T | tim.masterton@thalesgroup.com |
| 21 | Norbert Niebert | Ericsson | Norbert.niebert@ericsson.com |
| 22 | Xabiel Pañeda | Univ. of Oviedo | xabiel@uniovi.es |
| 23 | A. Murat Tekalp | Koc University – Turkey | mtekalp@ku.edu.tr |
| 24 | Giorgio Ventre | Univ. of Napoli Federico II | Giorgio@unina.it |
| 25 | Theodore Zahariadis | Synelixis | zahariad@synelixis.com |
| 26 | Piotr Romaniak | AGH University of Sci- | Romaniak@kt.agh.edu.pl |

| No | Expert's name | Organization | e-mail |
|---|---|---|---|
| | | ence and Technology | |
| 27 | Michał Grega | AGH University of Science and Technology | Grega@kt.agh.edu.pl |
| 28 | Aggelos Katsaggelos | Northwestern University | aggk@ece.northwestern.edu |
| 29 | Franco Tecchia | Scuola Superiore S.Anna | f.tecchia@sssup.it |
| 30 | Emanuele Ruffaldi | Scuola Superiore S.Anna | e.ruffaldi@sssup.it |
| 31 | Ahmet Kondoz | University of Surrey | A.Kondoz@surrey.ac.uk |
| 32 | Banu Gunel | University of Surrey | B.Gunel@surrey.ac.uk |
| 33 | Hemantha Kodikara Arachchi | University of Surrey | H.kodikaraarachchi@surrey.ac.uk |
| 34 | Antonio Camurri | University of Genova | Antonio.camurri@unige.it |
| 35 | Karim Yaici | University of Surrey | K.Yaici@surrey.ac.uk |
| 36 | Arturo Azcorra | UC3M | azcorra@it.uc3m.es |

European Commission

Luxembourg: Office for Official Publications of the European Communities

2009 — 60 pp. — 21 x 29,7 cm

ISBN 978-92-79-12810-3

KK-78-09-890-EN-C

# Research on Future Media Internet

**This work was coordinated by 5 NoEs:**

- *Content* (http://www.ist-content.org)
- *Emanics* (http://www.emanics.org)
- *Intermedia* (http:// intermedia.miralab.unige.ch)
- *Petamedia* (http:// www.petamedia.eu)
- *Visnet II* (http:// www.visnet-noe.org)

**and it was supported by the Networked Media Unit, DG Information Society and Media of the European Commission.**

http://cordis.europa.eu/fp7/ict/netmedia/home_en.html

ISBN 978-92-79-12810-3

Publications Office
Publications.europa.eu

doi 10.2759/11972

9 789279 128103

**European Commission**
Information Society and Media