



SALERO

Digital Media Technology, Research, and Production Trends Report

SALERO Deliverable 10.3.1



Digital Media Technology, Research, and Production Trends Report

SALERO Deliverable 10.3.1

SALERO identifier: SALERO-D1031-BLITZ-DigitalMediaTechnology-v4.doc

Deliverable number: D10.3.1

Author(s) and company: M. Mathews, R. Hackett (BLITZ), P. Stansfield (WCR)

Work package / task: WP10

Document status: Final

Confidentiality: Confidential / Public

DOCUMENT HISTORY

Version	Date	Reason of change
1	2006-07-19	Document created
2	2006-12-06	Revisions after Blitz meeting
3	2006-12-24	Additions of ELSPA material and updates of MPEG sections
4	2006-12-28	Final Version

The work presented in this document was partially supported by the European Community under the Information Society Technologies (IST) priority of the 6th framework programme for R&D.

This document does not represent the opinion of the European Community, and the European Community is not responsible for any use that might be made of its content.

This document contains material, which is the copyright of certain SALERO consortium parties, and may not be reproduced or copied without permission. All SALERO consortium parties have agreed to full publication of this document. The commercial use of any information contained in this document may require a license from the proprietor of that information.

Neither the SALERO consortium as a whole, nor a certain party of the SALERO consortium warrant that the information contained in this document is capable of use, nor that use of the information is free from risk, and does not accept any liability for loss or damage suffered by any person using this information.

Table of Contents

1	Executive Summary	1
2	Introduction	2
2.1	Purpose of this document.....	2
2.2	Scope of this document.....	2
2.3	Status of this document.....	2
3	Digital Media Technology Today	3
3.1	Domestic Digital Media Technology	3
3.1.1	<i>HD-DVD and Blu-Ray</i>	<i>3</i>
3.1.2	<i>Games.....</i>	<i>5</i>
3.2	Professional Digital Media Technology	8
3.2.1	<i>Introduction</i>	<i>8</i>
3.2.2	<i>Acquisition.....</i>	<i>8</i>
3.2.3	<i>Storage.....</i>	<i>8</i>
3.2.4	<i>Editing.....</i>	<i>9</i>
3.2.5	<i>Special Effects</i>	<i>9</i>
3.2.6	<i>Post Production.....</i>	<i>9</i>
3.2.7	<i>Data Logging.....</i>	<i>10</i>
3.3	MPEG Implementations.....	10
4	Production Trends	13
4.1	Audiovisual Media Production Trends.....	13
4.2	Games Media Production Trends.....	13
4.3	Cross Media Production Trends.....	15
4.4	IPTV and moves towards 'Intelligent Media'	15
5	Digital Media Research	17
5.1	Research in Intelligent Media Content Technologies.....	17
5.1.1	<i>Introduction: European research projects</i>	<i>17</i>
5.1.2	<i>Image analysis, description and reconstruction</i>	<i>17</i>
5.1.3	<i>Analysis and generation of sound and acoustic content.....</i>	<i>18</i>
5.1.4	<i>Analysis, translation and generation of speech</i>	<i>18</i>
5.1.5	<i>Intelligent content semantics and workflow technologies</i>	<i>19</i>
5.1.6	<i>Audiovisual information retrieval</i>	<i>20</i>
5.2	Trends in Industrial RTD and Patent Activity.....	21
5.3	EU RTD trends, policy, and opportunities for Intelligent Media Research.....	21
5.3.1	<i>Objective IST-2007.4.2 (IST-2007.4.4): Intelligent Content and Semantics</i>	<i>21</i>
6	Conclusions.....	24
6.1	Possible Future Directions of Digital Media.....	24

6.2	SWOT analysis of SALERO, in the light of current trends	25
6.3	Recommendations for SALERO	25
7	References.....	26
8	Glossary.....	29

1 Executive Summary

This Deliverable contains the research findings on the 'State of the Art' today on Digital Media. It identifies many trends, some of which are now established, and others which are just beginning.

Through both the Professional and Domestic sectors, we can see convergence taking place. We can see the start of 'Intelligent Content' happening.

The overall reasons for the emergence and growth of Intelligent Content will benefit Europe as a whole. Intelligent content can be used to satisfy these needs

Principal Highlights of the findings of this report are

- Domestic Technology is now becoming available that will implement Intelligent Content. This is seen in the forthcoming HD DVD players and Blu-Ray Players. These devices contain XML layers and Internet connections, so that media can be 'tagged' and re-used or re-purposed.
- The Games market continues to grow. Of particular interest is the MMORPG market
- IPTV distribution is growing rapidly, and could easily be a \$20 Billion market within five years.
- The consumer demand for compelling 'content' continues to grow
- Many new forms of content are starting to appear – the so called 'User Generated' content, in forms such as Blogs, Podcasts etc
- The barriers between 'professional' creators of media and 'home / amateur creators are being slowly eroded.
- The first signs of Intelligent content 'products' are beginning to appear from the UK company Volantis
- Significant research effort is being expended on Intelligent Content
- European Policy continues to support Intelligent Content; as can be seen in the Framework 7 Workprogram
- Standards continue to emerge, with the MPEG Family of standards starting to support Intelligent media (the MPEG-4 standards can support facial reference sites)
- It is now recognised from many sectors that Media Search systems are a 'key' to the success of intelligent content.
- None of the trends discovered or reported on in this report contain anything that will prevent the SALERO project from completing its goals. In fact, many of the changes identified here will positively support the SALERO project.

2 Introduction

2.1 Purpose of this document

The purpose of this document is to record and present trends in media production. These trends are used to measure the desirability of the SALERO project by. Ideally, we would see that the market for Intelligent and / or reusable content was emerging, and that no-one had managed to satisfy this niche. However, the findings here will be presented in as fair and neutral way for the reader to determine his own conclusions.

2.2 Scope of this document

This document covers the first eleven months of the Project, from 1 January 2006 to 30 November 2006.

2.3 Status of this document

This is the final version of D10.3.1.

3 Digital Media Technology Today

3.1 Domestic Digital Media Technology

Domestic digital media technology is still developing rapidly, with no potential end in sight. The iPod is now capable of video replay, smartphones incorporate radios and MP3 and video players; the take up of digital home cinema systems is accelerating, and HD-DVD and Blu-Ray systems are reaching the European market. Each one of these items increases the expectations for media. The iPod has pioneered the portability of media, while the Home Cinema has increased expectations of the 'wow' factor of media on a screen of two to three metres diagonal. Home Cinema has made multi-channel sound the norm, relegating stereo to merely 'old fashioned'.

3.1.1 HD-DVD and Blu-Ray

HD-DVD and Blu-Ray are more than new storage media offering higher capacity. The combination of very large disc capacity with an XML layer, a player built around a Pentium 4 processor with 1GB of RAM, and an Ethernet port for connection to broadband Internet produces a platform that suggests a pathway for the implementation of intelligent media. HD-DVD and Blu-Ray both offer more 'interaction' (and possible repurposing) than previous DVD implementations.

The 'interactivity layer' for HD-DVD is called iHD¹. iHD was developed by Microsoft and Toshiba as an open specification, with inputs and support from Hollywood movie studios, especially Disney. Microsoft has made available a free download to let users author and debug iHD content under Windows XP. iHD allows interactivity to be authored into a data format using XML and ECMAScript (a version of JavaScript) as its interpreted scripting engine. The iHD model includes an XML subset for formatting images, buttons, and video objects; cascading style sheets (CSS) for layout, colour, and fonts; SMIL for timing and synchronization; and ECMA scripts for programmability.

iHD is not limited to interaction with the optical media but can also be used on media delivered or streamed over the Internet or any other network. iHD allows menus to be displayed during video playback, persistent user-defined bookmarks, picture-in-picture display, and network access to download new features and HD trailers. It is even possible to watch a movie while searching for another scene, taking advantage of the 'picture in picture' function.

The competing Blu-ray format uses either HDMV protocol or BD-J for authoring interactive features. HDMV was designed to offer new features, while keeping the authoring process as simple as possible. The production of Blu-ray discs and DVD-Video titles incorporates many identical phases. HDMV offers improved navigational and menu features, improved graphics and animation, better subtitling support and new features like browsable slideshows.

Unlike DVD-Video, the Blu-ray Disc format allows data to be read from a different location on the disc, while uninterruptedly decoding and playing back video. This allows the system to call up menus, overlay graphics, pictures, button sounds, etc. at user request without stopping playback. Two individual, full HD resolution (1920x1080) graphics planes are available, on top of the HD video plane. One plane is assigned to video-related, frame accurate graphics (like subtitles), and the other plane is assigned to interactive graphic elements, such as buttons or menus. For both planes, various wipes, fades and scroll effects are available, for example to present a menu.

Blu-ray button graphics are implemented in HDMV. Menu buttons support 256 colour full-resolution graphics and animation. Buttons can be called and removed during video playback, so there is no need to return to a menu screen. Button sounds can be loaded into memory of the Blu-ray Disc player. When a user highlights or selects a menu option, the sound can be played (such as a voice-over explaining the highlighted menu choice, or button clicks). These button sounds can even be mixed with the running audio from the movie or menu. HDMV facilitates multi-page menus, which do not interrupt playback each time a new menu screen is called so users can browse the menu pages, while the audio and video remain playing in the background.

HDMV supports the creation of user browsable slide shows with continuous audio. Browsing includes forward and backward scrolling, and selections from thumbnail images. HDMV also facilitates many

advances in subtitling. DVD-Video stores subtitles in the video stream, limiting the number of languages and display styles. In Blu-ray, subtitles can be stored independently on the disc. A user may select different font styles, sizes and colours for the subtitles, or location on screen. Subtitles can be animated, scrolled or faded in and out.

Every Blu-ray Disc player will be equipped with a Java interpreter, so that it is capable of running discs authored in BD-J (Blu-ray Disc Java) mode. BD-J was designed to offer the content provider enhanced functionality when creating interactive titles. BD-J supports network access (e.g. for downloading updated trailers or accessing live studio events), picture-in-picture and access to local storage. In BD-J mode, the author has complete freedom in designing the user interface. The interface is controlled by the navigation buttons on the remote and can display up to 32-bit dynamically generated graphics and pictures in standard file formats like JPEG, PNG, etc.

The BD-J environment offers all of the playback features of HDMV mode, including the selection of subtitle, trick play modes, angles, etc. Video can be scaled dynamically, so that it can be played in a small size in the corner of a menu, and resume full screen when a selection is made. Blu-ray Disc players will contain a small amount of non-flash memory for storing game scores, bookmarks, favourites from a disc, training course results, etc. As a manufacturer's option, a Blu-ray Disc player may also be equipped with a hard disk, to store large amounts of data like audio/video.

BD-J supports the TCP/IP and HTTP Internet protocols like. The player may connect to the disc publisher's web site to unlock special content on the disc (after certain conditions, like payment, are met), or dynamically display extra information (like theatre schedules for a movie) on the screen. The disc's program may be extended with JPEG pictures or audio fragments downloaded from the Internet, or stream full new audio/visual content to a hard drive.

BD-J is based on a packaged media profile of Globally Executable MHP, or GEM. GEM, in turn, forms the basis of most global digital television application standards, including Multimedia Home Platform ("DVB-MHP") for broadcast, satellite, and cable worldwide, OpenCable Application Platform ("OCAP") for North American cable, and Advanced Common Application Platform ("ACAP") for US broadcast. All GEM-based standards are built on top of Java Micro Edition ("Java ME"), invented by Sun Microsystems and maintained and enhanced by the Java Community Process. Because BD-J, MHP, OCAP, and ACAP are all based on a common Java platform, there is a high degree of content interoperability, making it easier, for example, to deliver Blu-ray Disc content features on cable Video on Demand, or to repackage for Blu-Ray on Blu-ray Disc.

HD-DVD and Blu-Ray offer not only increased picture quality and interactivity but also substantially better sound, since the extra storage capacity requires the use of less compression to accommodate the soundtrack.

Both the major sound companies in cinema and home entertainment, Dolby and DTS, have produced new audio standards for HD. Dolby® TrueHD is a lossless technology developed for high-definition disc-based media. Dolby TrueHD delivers sound that is bit-for-bit identical to the studio master and supports

- 100 percent lossless coding
- Up to 18 Mbps bit rate.
- Up to eight full-range channels of 24-bit/96 kHz audio
- High-Definition Media Interface (HDMI™), single-cable digital connection for audio and video.

Dolby TrueHD introduces dialogue normalization, which maintains the same volume level when you change to other Dolby programming, and dynamic range control (night mode) which enables the viewer to customize audio playback to reduce peak volume levels (no loud surprises) while experiencing all the details in the soundtrack. This is designed for late-night viewing of high-energy surround sound without disturbing others. True HD has been selected as the mandatory format for HD DVD and as an optional format for Blu-ray Disc.

Meanwhile DTS has developed DTS-HD, which is implemented at two levels. DTS-HD Master Audio offers a 'bit for bit' version of the studio master; DTS-HD High Resolution Audio offers compressed sound, but at typically twice the bandwidth of standard Definition DVD.

3.1.2 Games

The Games market continues to develop and grow. The exact figures here are hard to come by, as although the value of games sold 'over the counter' is available from publishers, the total value of 'on line' gaming is only estimated.

In the UK, the trade body ELSPA (Entertainment and Leisure Software Publishers Association) reported that in 2004 the UK Computer and Video games industry recorded sales of in excess of £2 Billion (€3 Billion) In the United States, more than 35 percent of all homes have one or more video game consoles, or a computer that is used to play games. Competition continues between the two major console manufacturers, Sony and Microsoft, but whichever succeeds it is without doubt that games platforms available for the public will continue to increase significantly in performance.

Feeling you are a part of certain types of games is extremely important and therefore the more real a game can look and feel obviously the better it is as the person taking the controls. It helps you feel more as if you are the character whereas before you were merely in control of the character on screen.

Most of this is due to the manner in which the graphics have gone from a totally two dimensional game to the games we have now, with proper 3D images, lighting and shadow effects. An example of this is a game released both on the Microsoft Xbox360 and on the PC called 'Oblivion – The Elder Scrolls IV'. The look of the water and grass were much better than seen before. While some may think the most important thing about a new game is its playability, in this day and age that isn't enough. Participants want more realism in games and although 'Oblivion' may be in a fantasy world with elves and goblins the whole appearance and feel of the game somehow gives the game that certain #real' touch.

When the first Grand Theft Auto game was released by Rockstar on the first Sony Playstation, the gamers perspective was from above (like a helicopter view) with everything laid out like a map on the screen as a totally 2 dimensional image. The game however was still a massive success as the 3D aspect was not around to add any competition from other games and the playability of the game was almost second to none. The first version was released in 1997/98 and the second version GTA 2 was released only a year or two later. However the Grand Theft Auto III released in 2001 suddenly had 3D graphics and no longer were you stuck with a view from the 'sky' but you were down on the ground amongst the action. Since then GTA Vice City, GTA San Andreas, GTA Liberty City Stories and GTA Vice City Stories have all been released and all go straight to the top of the game charts. What also adds to this game and 'Oblivion' is that both games leave the user free to explore and choose how to play the game. Whether it is following a storyline or not, the user has the option of total freedom of movement. Early games gave the user no option as to the route to take or the way to play a game – you just had to follow their storyline or buy another game. Having this total freedom when playing a game increases the possibilities of what can be achieved as a player.

'Black and White' by Lionhead was a game released in 2001 solely for the PC market which brought an extremely new aspect to game playing with characters that learnt from your behaviour and reacted to their environment like no other game had done before. In the game the user played a 'God' and you controlled a giant creature which only behaved as you instructed or 'taught'. When you first meet the creature it was very much like that of a baby, it was hungry and emotional and it was your objective to teach the creature how to behave and what it could or couldn't eat. The user had to make sure it learnt not to destroy the villagers' homes – as it was big enough to trample whole buildings under foot – and not to eat the villagers if it was hungry. The user could show the creature where to go to collect fish and fruit and cattle to eat if necessary and to protect the villagers against natural or unnatural disasters. Of course that's only if the user decided to play the game that way. He might have a cruel streak and decide he wanted the creature to misbehave, but whatever you did in time the creature could be left alone in the knowledge that it would behave mostly as you had instructed.

Significant changes have taken place in User interfaces – a game called 'Dance Dance Revolution' (or DDR) was released in the arcades about 8 years ago as a game that played music while you dance on a mat with symbols on it. The idea was to hit the symbols when the screen instructed you to. Therefore effectively to end up dancing to the music played. This was introduced later into the homes when released on PS2, Xbox and the Nintendo Gamecube. Young girls entered what had previously been a male dominated hobby and other games were released to encourage these new gamers into the stores. DDR added aerobic exercise to what was previously something you only did sitting down in front of your TV at home. In fact certain options on this game are specifically for people to use as a keep fit tool and not as a game as such – showing a multifaceted approach to exercise and gaming.

There is also a game out now called 'Guitar Hero' where the user can buy a special guitar peripheral (similar to a Gibson guitar) to plug in to the PS2 to again give it a more 'playable' feel. This wasn't the first game to do this but it is a lot more successful game than previous attempts and is already about to release the sequel Guitar Hero II only a year later. There are other games on the market like golf, fishing and more obviously the driving games with a steering wheel available that are turning gaming into a more physical entertainment than ever before.

This has been taken on board by Nintendo with the release of its Nintendo Wii in Q4 2006. They have created a console whose peripheral controls have motion sensors that suddenly open up a whole new world where any movement made is transferred onto the screen. So now you can play tennis using only the motion of your arm or drive or fish and really it creates a whole new style of games for us to enjoy.

Finally there is the addition of the internet and broadband enabling you to not just play like you can with friends on the net via PCs but the new consoles let the user play with people in any continent (if they have the same game) and pit your wits against a whole new competition. Before they were restricted to friends playing against each other at home but now the user can wanted play football against someone in Paris or race a car against another opponent in New York or maybe even try to shoot someone they've never met somewhere in Sydney! The world in that sense has become the Gamer's Oyster.

MMORPG (massive multiplayer online role-playing game) are online computer role-playing games (RPG) in which a large number of players interact with one another in a virtual world. As in all Role Playing Games, players assume the role of a fictional character (most commonly in a fantasy setting) and take control over many of that character's actions. MMORPGs are distinguished from single-player or small multi-player RPGs by the number of players, and by the game's persistent world, usually hosted by the game's publisher, which continues to exist and evolve while the player is away from the game.

MMORPGs are very popular throughout the world, with combined global memberships in subscription and non-subscription games exceeding 15 million as of 2006². Overall, revenues for MMORPGs exceeded half a billion dollars in 2005 and are expected to reach over a billion dollars by 2009³.

Meridian 59, launched by 3DO in late 1996, was one of the first Internet MMORPGs. It was the first Internet game from a major publisher, the first to be covered in the major game magazines, and the first to introduce the flat monthly subscription fee. Most significantly, however, and another first for online RPGs, was its 3D engine, allowing players to experience the game world through the eyes of their characters. A cult following quickly grew for *Meridian 59* that still exists today.

Commercial online gaming has become extraordinarily popular in South Korea. *Nexus: The Kingdom of the Winds*, designed by Jake Song, was commercially released in 1996 and eventually gained over one million subscribers. Song's next game, *Lineage* (1998), enjoyed even greater success gaining millions of subscribers in Korea and Taiwan. This helped to secure developer NCsoft's dominance in the global MMORPG market for several years.

Final Fantasy XI The first MMO to allow console and PC gamers to play together

In 2002, *Final Fantasy XI* was released, which hoped to bring in fans of the extremely popular (but mostly single-player) *Final Fantasy* series, which was particularly popular in Japan. Like the previous *Final Fantasy* games, this one was playable on a console, using a handheld controller, but bore little resemblance to gameplay in prior *Final Fantasy* titles.

World of Warcraft is currently the most widely played MMORPG in the world.

The most recent generation of MMORPGs, based on arbitrary standards of graphics, gameplay, and popularity, is said to have launched in November 2004 with Sony Online Entertainment's *EverQuest II* and Blizzard Entertainment's *World of Warcraft* (WoW). At the time, Sony expected to dominate the market, based on the success of the first *Everquest*, and decided to offer a flat monthly rate to play all of their MMORPGs including *EverQuest*, *EverQuest II*, and *Star Wars Galaxies*, to keep from competing with itself. However, *World of Warcraft* immediately overtook all of these games upon release, and indeed became so popular that it dwarfed all previous monthly-fee MMORPGs. At present, *WoW* is one of the most played games in North America, and the most played MMORPG worldwide, with a total of over 7.5 million customers. A total of approximately 97.5 million dollars a month is being made by blizzard in subscription costs.

In April 2005, *Guild Wars* launched successfully, introducing a new financial model which might have been partly responsible for its success. Though definitely an online RPG, and technically having a

persistent world (despite most of the game's content being instanced), it required only a one-time purchasing fee. It was also designed to be "winnable", more or less, as developers wouldn't profit from customers' prolonged playtime. Other differences compared to traditional MMORPGs included strictly PvP-only areas, a relatively short playtime requirement to access end-game content, instant world travel, and strategic PvP. For these differences it was termed instead a "Competitive/Cooperative Online Role-Playing Game" (CORPG) by its developers, and is now seen by some as occupying a new niche in the MMORPG market. With two million players as of June 19, 2006^[5], Guild Wars is still continuously profitable (due to several expansions) and can be seen as a serious competitor to WoW.

Finally, despite WoW's domination in the pay-MMORPG market, there still has been significant competition (and potential for profit) among free MMORPGs. A good example of a profitable free MMORPG is the Korean *MapleStory*, a 2D side-scroller developed by *Wizet*, which features purchasable in-game "enhancements". Due to its free nature, the game is said to have more than 50 million players in all of its many versions, with the majority of them from East Asia. Another good example would be the popular Java-based online RPG *RuneScape*, developed by British developer *Jagex*. It allows for players to sign up for a member's account for extra content, in addition to the ad-supported free accounts. It is also extremely popular, and claims almost ten million members.

Psychology of MMORPGs

Since the interactions between MMORPG players are real, even if the environments are not, psychologists and sociologists are also able to use MMORPGs as tools for academic research.

Sherry Turkle, a clinical psychologist, has conducted interviews with computer users including game-players. Turkle found that many people have expanded their emotional range by exploring the many different roles (including gender identities) that MMORPGs allow a person to explore.^[13]

Nick Yee, a Ph.D student, has surveyed more than 35,000 MMORPG players over the past few years, focusing on psychological and sociological aspects of these games. His research can be found at The Daedalus Project⁴. Recent findings included that 15% of players become a guild-leader at one time or another, but most generally find the job tough and thankless. Other researchers have found that the enjoyment of a game is directly related to the social organization of a game, ranging from brief encounters between players to highly organized play in structured groups.

User-Generated Content is a term that has come into the mainstream during 2005 in web publishing and new media content production circles. It refers to on-line content that is produced by users of websites as opposed to traditional media producers such as broadcasters and production companies. It reflects the democratisation of media production through new technologies that are accessible and affordable. These include digital video, blogging, podcasting, mobile phone photography and, of course, wikis. Prominent examples of websites based on User-Generated Content include Flickr, Friends Reunited, FourDocs, YouTube, Second Life, Top Ten Media and Wikipedia. The advent of User-Generated Content marks a shift among media organisations from creating on-line content to creating the facilities and framework for non-media professionals (i.e. 'ordinary people') to publish their own content in prominent places.

Microsoft has launched the XNA toolset to let users generate their own games for the Xbox 360. It comes as a set of tools, complete with a managed runtime environment, provided by Microsoft that facilitates computer game design, development and management. XNA does this by freeing game designers from writing "repetitive boilerplate code,"^[2] and brings all aspects of game production into a single system

Using XNA it is not only possible to generate games for home use, but it is possible to upload them to a website where these games can be played by other players. This even includes a mechanism for the owner of the game to enjoy a royalty stream from his game. We believe that this indicates another important trend – that the differences between 'professional' and 'amateur' media creators (in this example, games) are becoming more indistinguishable than ever. It will be interesting to see whether SONY mimics this model with the Playstation 3.

One interesting parameter of user generated content is that it allows the bypassing of the traditional 'gatekeepers' of media – the Cinema distributors, the broadcasters, Games publishers, advertising agencies etc.

The British Broadcasting Corporation (BBC), for example, set up a User-Generated Content (UGC) team as a pilot in April 2005 with 3 staff. In the wake of the 7 July 2005 London bombings and the UK's

Buncefield oil depot fire, the team was made permanent and was expanded, reflecting the arrival in the mainstream of the 'citizen journalist'. During the Buncefield disaster the BBC received over 5,000 photos from viewers. The debate over the publication of the Mohammed cartoons (January 2006) prompted over 22,000 emails from viewers in the first couple of weeks. The BBC does not normally pay for content generated by its viewers.

In recent years User-Generated Content has been a path to mainstream success for some. Afroman's novelty song Because I Got High became the most requested song on American radio in July 2002 due to its success on Napster. In the summer of 2006, Brooke Brodack was signed into a talent development deal based on her short skits on YouTube.

3.2 Professional Digital Media Technology

3.2.1 Introduction

There have been many changes in Media production technology in the Professional sector. A lot of these changes have been implemented to reduce costs, and increase flexibility in production. Looking at these changes in detail

3.2.2 Acquisition

The capture medium for feature movies is still almost always motion picture film, although the number of exceptions is growing. The 2006 *Miami Vice* movie, was shot by Michael Mann mostly on Grass Valley VIPER digital cameras and in March 2006 the IP-RACINE website listed 19 movies on which the VIPER camera was used. In the large majority of cases where film is used, the film is now normally linked into the production process by software techniques. One common feature is to use 'Keycode' readers on the Telecine or film scanning equipment that transfers the film images into electronic representations. Keycode works by using film stock that has unique identifiers in the margin of the film, and which are read optically by a reader. This unique code is effectively used as an 'asset number' in the later stages of post production. Another common technique is to shoot with 'Video Assist'. This works by using a small video camera attached to the film camera. The Video camera produces an exact parallel (or nearly exact – the two can never be on the same axis) image stream, which can even be used for editing, sometimes even on site of the shoot. The film frames are then related to the generated timecode of the video material.

Film Cameras are now thought to be at their peak. Cameras such as the ARRI 535 are well refined machines, and are unlikely to be substantially developed further. One trend in Cinematography is the use of Super 16mm film stock. It is argued that the emulsions used today are so much better than those of twenty years ago, and that the smaller and much cheaper 16mm film stock can be made to give results that can be comparable to old 35mm stock. One such high profile movie shot on Super 16 is "The Last King of Scotland" by Kevin Macdonald, to be released in early 2007.

Digital Cinematography Cameras are now appearing from several manufacturers. Apart from the market leading Grass Valley VIPER mentioned above, there is the GENESIS camera from Panavision (which is a modified SONY Camera), the Dalsa ORIGIN Camera, and a proposed 'RED' camera from the manufacturers of the Oakley brand of sunglasses, which in 2006 was only shown in concept form.

Acquisition for television is almost exclusively electronic. The quality of broadcast cameras is now so good that the expense of shooting on film is considered to be a waste with the exception of some drama material, and commercials. The arguments are that drama has a long shelf life – sometimes being sold to new broadcasters after ten years or more – and video formats change. For commercials the logic seems to be that the advertisers have always used film, can afford to use film, so they do.

3.2.3 Storage

Although magnetic tape cassettes are still in use, nearly all operations on data are done on workstations connected to image servers. Some post production facilities have in excess of 100 Terabytes of data 'on line', with many times that stored off line on backup tapes. Tape still has the edge for archiving purposes, although difficulties still exist in print through and humidity susceptibility. Note that whilst DigiBeta cassettes are often used as a delivery or transport medium, they are not usually used as an archive medium.

The bandwidth of data generated by digital cinematography cameras creates different problems: the digital data from cameras such as the Grass Valley VIPER or the Dalsa ORIGIN has to be stored on purpose designed storage systems such as those from the S2 company in Nevada, USA. Codex in London supplies a similar digital data acquisition unit.

3.2.4 Editing

Film editing evolved from the process of physically cutting and taping together pieces of film, using a viewer such as a Moviola or Steenbeck to look at the results. The assembly cut is the first of four stages of the film editing process. An assembly cut generally includes the clapboard at the beginning of each shot and does not flow very well. It is normally a very rough approximation of what the film will be when it reaches picture lock. All initial editing is done with a positive copy of the negative called a workprint. This allows the editor to do as much experimenting as he or she wishes, without the risk of damaging the original. When the workprint has been cut to a satisfactory state, it is then used to make a negative cutting list. The negative cutter refers to this list while processing the negative, splitting the shots into A and B rolls, which are then optically printed to produce the final film print.

Since the film was physically cut and pasted, a 'nonlinear' style of editing evolved. At the workprint stage, strips of film could be placed in any order. This approach is generally considered superior to the strictly linear approach that was necessary in video editing through the 1970s. A video 'cut' is really the copying of scenes from various camera tapes onto a master. Before the development of powerful computer systems that could store large amounts of visual data for transfer, it was necessary to make the transfer in strictly linear order. Trying to insert a shot between two shots already on the master tape would create noise. A system such as Avid allows the creation of a workprint.

In recent years, 'film editing' has come to mean 'what a film editor does', even though the work involved is now generally performed on a computer-based non-linear editing system, such as Avid, Lightworks or Apple's Final Cut Pro and, at the semi-professional level, on programs such as Adobe Premiere Pro, Pinnacle Edition or Sony Vegas.

If the end product is to be a traditional movie, the final negative cutting list is produced from the software, and the negative cutting process occurs as before. In other cases, an edit decision list may be generated for a video editing system.

3.2.5 Special Effects

Special effects (abbreviated SPFX or SFX) are used in the film, television, and entertainment industry to realize scenes that cannot be achieved by normal means, such as space travel. They are also used when creating the effect by normal means is prohibitively expensive such as building a mediaeval castle or sinking a real ocean liner. With the advent of computer graphics imaging, special effects are also used to improve previously filmed elements, by adding, removing or enhancing objects within the scene.

Many different special effects techniques exist, ranging from classic film techniques invented in the early 20th century (such as aerial image photography and optical printers) to modern computer graphics imagery (CGI). Often several different techniques are used together in a single scene or shot to achieve the desired effect.

Special effects are traditionally divided into two types. The first is optical effects (also called visual or photographic effects), which rely on manipulation of a photographed image. Optical effects can be produced with either photographic (i.e. optical printer) or visual (i.e. CGI) technology. A good example of an optical effect would be a scene in *Star Trek* depicting the USS Enterprise flying through space.

The second type is mechanical effects (also called practical or physical effects), which are accomplished during the live-action shooting. These include mechanized props, scenery, and pyrotechnics. Good examples would be the ejector seat of James Bond's Aston Martin, R2D2 in the *Star Wars* films, or the zero-gravity effects employed in *2001: A Space Odyssey*.

3.2.6 Post Production

Since 2005 it has become normal practice for postproduction to be carried out on material for cinema release through the Digital Intermediate process. In this mode, all frames of relevant shot film material are digitised, and stored. These frames are manipulated through workstations to edit, composite, and

grade material, until the director is happy with the end result. The new digital files are then recorded to new film material, which then becomes the new 'master'.

The master is optically duplicated, through a cascade process of Inter-negatives and Inter-positives, to produce the cinema release prints. For cinemas that can exhibit digitally, the postproduction facilities make a 'Digital Master', which is used to digitally make Digital release copies, which are sometimes called 'Virtual prints'

Commonly used postproduction systems include the Quantel iQ system, the Pandora International 'Revolution', the Nucoda Film Master, and the 'Lustre' system from Autodesk.

In these systems, a skilled operator will perform scene-by-scene grading to get the overall portrayal of the image as desired by the director or his delegate.

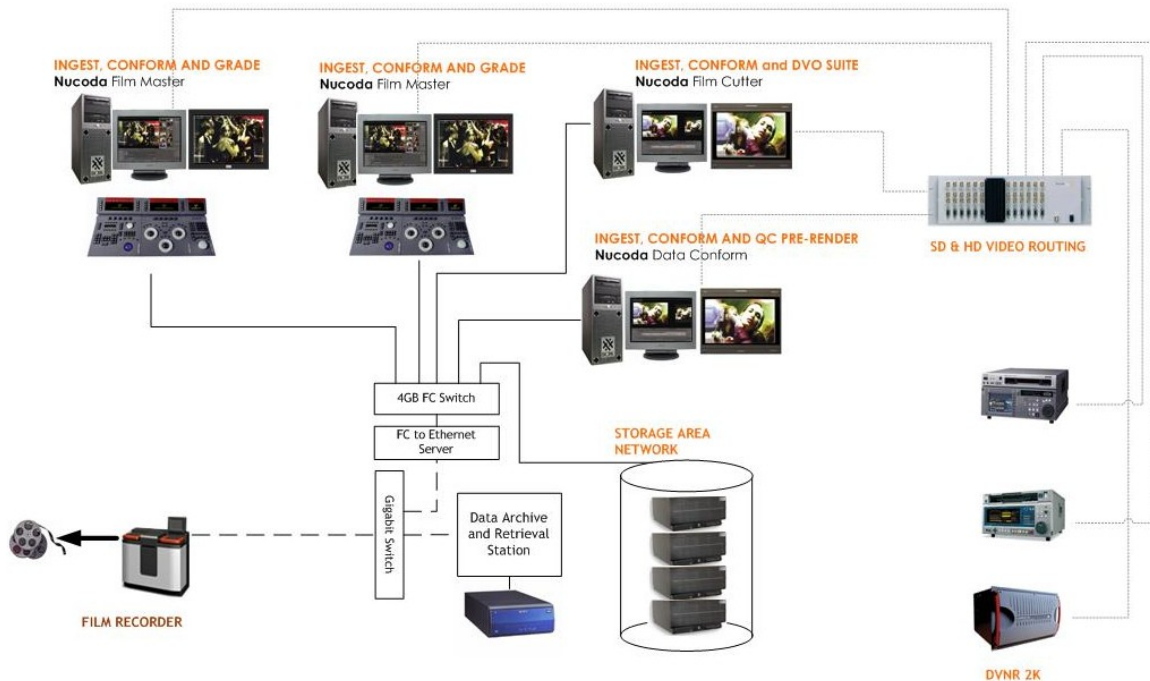


Figure 1: Diagram showing Digital Intermediate workflow (Nucoda)

Up until recently, all previewing of images used to be done by viewing high quality Cathode Ray Tube monitors. These days, for cinema work, the grading is quite often carried out by viewing projected images, similar to the way that the final work will be projected. In Q4 2006, more than 120 post production facilities had the capability to grade images in this way.

3.2.7 Data Logging

It is obviously important to be able to locate media, and to retrieve media from archives (an archive without the ability to locate and withdraw items would be totally pointless). Usually each organisation develops its own methodology for cataloguing material. These catalogues are quite often dependent on one technology (eg a certain software platform or application) and often require significant maintenance. No standard cataloguing protocols have been adopted, and therefore there is little interoperability with post production facilities who collaborate on visual effects work for

3.3 MPEG Implementations

The **Moving Picture Experts Group** or **MPEG** is a working group of ISO/IEC charged with the development of video and audio encoding standards. Its first meeting was in May of 1988 in Ottawa, Canada. As of late 2005, MPEG has grown to include approximately 350 members per meeting from various industries, universities, and research institutions. MPEG's official designation is ISO/IEC JTC1/SC29 WG11.

MPEG has standardized the following compression formats and ancillary standards:

MPEG-1⁵ video was originally designed with a goal of achieving acceptable video quality at 1.5M bit/second data rates and 352x240 resolution. While MPEG-1 applications are often low resolution and low bitrate, the standard allows any resolution less than 4095x4095. Nevertheless, most implementations were designed with the Constrained Parameter Bitstream specification in mind.

At present MPEG-1 is the most compatible format in the MPEG family; it is playable in almost all computers and VCD/DVD players.

One big disadvantage of MPEG-1 video is that it supports only progressive pictures. This deficiency helped prompt development of the more advanced MPEG-2.

MPEG-2⁶ is the designation for a group of coding and compression standards for Audio and Video (AV), agreed upon by MPEG (Moving Picture Experts Group), and published as the ISO/IEC 13818 international standard. MPEG-2 is typically used to encode audio and video for broadcast signals, including direct broadcast satellite and Cable TV. MPEG-2, with some modifications, is also the coding format used by standard commercial DVD movies. Where software patentability is upheld, the use of MPEG-2 requires the payment of licensing fees to the patent holders via the MPEG Licensing Association.

The Video part (part 2) of MPEG-2 is similar to MPEG-1, but also provides support for interlaced video (the format used by analog broadcast TV systems). MPEG-2 video is not optimized for low bit-rates (less than 1 Mbit/s), but outperforms MPEG-1 at 3 Mbit/s and above. All standards-conforming MPEG-2 Video decoders are fully capable of playing back MPEG-1 Video streams.

MPEG-3⁷ is the designation for a group of audio and video coding standards agreed upon by MPEG (Moving Picture Experts Group). MPEG-3 was designed to handle HDTV signals in the range of 20 to 40 Mbit/s.

It was soon discovered that similar results could be obtained through slight modifications to the MPEG-2 standard. Shortly thereafter, work on MPEG-3 was discontinued.

MPEG-4⁸ is a standard used primarily to compress audio and video (AV) digital data. Introduced in late 1998, it is the designation for a group of audio and video coding standards and related technology agreed upon by the ISO/IEC Moving Picture Experts Group (MPEG). The uses for the MPEG-4 standard are web (streaming media) and CD distribution, conversation (videophone), and broadcast television, all of which benefit from compressing the AV stream.

MPEG-7⁹ is a multimedia content description standard. Thus, it is *not* a standard which deals with the actual encoding of moving pictures and audio, like MPEG-1, MPEG-2 and MPEG-4. It uses XML to store metadata, and can be attached to timecode in order to tag particular events, or synchronise lyrics to a song, for example.

It was designed to standardise:

- a set of description schemes and descriptors
- a language to specify these schemes, called the Description Definition Language (DDL)
- a scheme for coding the description

The combination of MPEG-4 and MPEG-7 has been referred to as MPEG-47.

There are many software implementations and Demonstrators for MPEG-7. These include:

- **Caliph & Emir:** Creation and Retrieval of images based on MPEG-7 (GPL).
- **Frameline 47 Video Notation:** Frameline 47 from Versatile Delivery Systems. The first commercial MPEG-7 application, Frameline 47 uses an advanced content schema based on MPEG-7 so as to be able to notate entire video files, or segments and groups of segments from within that video file according to the MPEG-7 convention (commercial tool)
- **Eptascape ADS100** uses a real-time MPEG 7 encoder on an analog camera video signal to identify interesting events, especially in surveillance applications, check the [demos](#) to see MPEG-7 in action (commercial tool)

- **IBM VideoAnnEx Annotation Tool:** Creating MPEG-7 documents for video streams describing structure and giving keywords from a controlled vocabulary (binary release, restrictive license)
- **iFinder Medienanalyse- und Retrievalsystem:** Metadata extraction and search engine based on MPEG-7 (commercial tool)
- **MPEG-7 Audio Encoder:** Creating MPEG-7 documents for audio documents describing low level audio characteristics (binary & source release, Java, GPL)
- **XM Feature Extraction Web Service:** The functionalities of the eXperimentation Model(XM) are made available via web service interface to enable automatic MPEG-7 low-level visual description characterization of images.
- **TU Berlin MPEG-7 Audio Analyzer (Web-Demo):** Creating MPEG-7 documents (XML) for audio documents (WAV, MP3). All 17 MPEG-7 low level audio descriptors are implemented (commercial)
- **TU Berlin MPEG-7 Spoken Content Demonstrator (Web-Demo):** Creating MPEG-7 documents (XML) with SpokenContent description from an input speech signal (WAV, MP3) (commercial)
- **MP7JRS C++ Library** Complete MPEG-7 implementation of part 3, 4 and 5 (visual, audio and MDS) by IIS, JOANNEUM RESEARCH Institute of Information Systems and Information Management (JRS).

MPEG-21¹⁰ ISO 21000 defines an open framework for multimedia applications. MPEG-21 defines a "Rights Expression Language" standard as means of sharing digital rights/permissions/restrictions for digital content from content creator to content consumer. As an XML-based standard, MPEG-21 is designed to communicate machine-readable license information and do so in a "ubiquitous, unambiguous and secure" manner.

4 Production Trends

4.1 Audiovisual Media Production Trends

Audiovisual Media has hardly changed in the past ten years. Probably the 'only' new format is the 'DVD Extras'. These usually contain additional information relating to the finished movie or program. Typical content of these may be an alternative 'commentary' by the Film's Director, out takes from the movie making, interviews with the performers, producers, set designers, Visual Effects producers, historical material about the period or location, and so on. All of these are believed to enhance the viewer's enjoyment of the medium.

It is noted that this material also exists in the new emerging 'HD-DVD and 'Blu-Ray' formats, but starts to become more intertwined with the production itself. The technical replay capability of these media include facilities to be able to view the 'rough' special effect taking place in a window in the finished production, so as to compare the two, whilst listening to the special effects producer's description.

There is a trend to the making of more 'Documentary' style feature films. 'Bowling for Colombine' (about the Colorado shootings) was probably the turning point here, followed by 'Smartest Guys in the Room' (about the Enron fraud), 'The high cost of low price' (about Enron), 'Super Size me' (about McDonalds) and 'An Inconvenient Truth' (about Global warming).

Another trend is the so called 'Movie-cal' (being the movie of the Musical). Recent evidence of this is the 'Chicago' movie, and more recently, the 'Dreamgirls' movie, based on the 1980's Broadway show.

There is also a developing market for material made in new territories. It is noticeable that countries that are not known for their media productions to become visible. For example, the 'Cinema du Monde' village at the Cannes Film festival contained a 'Thai Cinema day', which would have been unthinkable a few years ago. Such countries as Bahrain are now making films that are being given limited distribution in other countries, and so on. The 'Global Village' concept is truly being applied to Media.

4.2 Games Media Production Trends

Many changes are taking place in the production of Games media. Many of these are easily observed, but some are more subtle.

One trend that is immediately visible is the size and complexity of Games. Typically, about five years ago, in the 'PlayStation 1' days, a game could take around 8-10 Man years of effort. Current Playstation 2 games take typically 40-60 Man years of effort. It is anticipated that Playstation 3 games are likely to consume 60-100 Man years of effort. This begins to approach the effort taken for High quality animated features.

For Commercial reasons, it is still only possible to produce a game over a one to one and a half year timeframe. This means that the only way of producing games today and in the near future is to enlist significantly larger teams, which in itself poses new challenges.

Similarly, budgets have had to increase significantly, from £1Million five years ago, through current budgets of £3-5 Million, to Playstation 3 games with a likely development cost of around £10 Million. This means that with distribution, marketing, and publicity, the total cost of developing and publishing a game may be up to £12 Million (\$25 Million)

Games are getting better because of several factors

- The availability of High resolution displays in the home (the 'HD ready' displays)
- The take-up of Parallel processing in games platforms and development of games
- Better 'shader' technology
- Improved Audio (often 5.1 now, and 7.1 capability coming on PS3)

The skillsets employed for games productions have radically changed over the past five years, and are anticipated to continue changing. The biggest observed difference is that the ratio of 'Programmer to digital artist' is changing. The considerably bigger games now being developed require a few more programmers than previous games, whilst requiring a lot more 'digital artists'. Another trend is the specialisation of digital artists – previously an artist did everything – backgrounds, characters, props, effects, animation, model making. Now there is more likely to be a specialism in the Artist's activities. This is getting to the point where many artists are forming external companies for the provision of specific items. Examples of the websites, which often contain some 'free' models and some models to purchase are www.exchange3d.com and www.3dextra.cygad.net. Other specialist companies exist to provide either 'Digital Cities' or Digital Horticulture' Some doubts exist in the business and legal position should two rival games developers both license, for example, a given type of Getaway car for the villains in a game. The 'non – uniqueness' of this item between games could seriously damage the potential success of each, and therefore the use of such libraries is generally for lesser items.

It could be seen that the 'programming' side of games is now maturing, and that the fastest growing element of modern and future games is the creation of the 'artwork' or 'asset creation'. Note that at present, with the exception of the occasional use of purchased library items, most games 'artwork' – be it backgrounds, props, characters, animation, or effects, are still usually created from 'scratch' for every production or game. Typically a 50 person game development team these days will contain between five to ten programmers, with the rest of the team being 'digital artists' of one form or another. We estimate that whilst the programming effort required for a game has probably increased by 50% over the past three years, the animation effort has increased by 200%. Yet another way of looking at this trend in Games development is that the programming tools are reusable from game to game, so the programmer teams are small, whilst the digital assets aren't – and that is why large teams are common.

Globalisation is also been seen to affect the development of games. Gone is the possibility of developing a game for one country or territory only – For example, the Japanese games developers used to develop games for their domestic market. Now, this is financially not viable, and since no other major games playing country really understands Japanese culture, then this sort of game will no longer be made.. Another trend that is based on this is a 'shift' of games development from Japan to the West. This seems to be based on the logic that the Japanese can understand Western based games, as can obviously the west, whilst only Japan can understand Japanese culture.

Looking at the 'Games of the Film' sector, it seems that the number of such tie in games has decreased over the past five years, and is currently likely to decrease further. Underlying logic is that games development is so expensive, it is very risky to produce games of films of unknown success, and therefore publishers concentrate on 'near certainties' such as Superman, Batman, Harry Potter, and so on. A further trend is the game which involves a genre rather than a particular movie. The latest 'James Bond' game is based partly on a number of James Bond films, giving it a longer 'shelf life' and more chance to break even!

If we examine how 'Games Assets' are produced from 'Film Assets', then we notice that there are now some signs of element re-use. For example, when BLITZ was developing the 'Mummy 2' game, no digital elements whatsoever were reused. The 'Mummy 2' game was totally developed from the movie script, and set photos. The game wasn't really representative of the movie's 'feel'. In production of the 'Bratz' games, there was real use of some CGI assets from the 'Bratz' animation series. There are many reasons why assets aren't more often used. One is that the models are designed and created for renderers with special or 'unusual' plug-ins that aren't appropriate for Games development, and also there are resolution issues – film assets are created with higher resolutions in mind.

Another observation is that there is an emerging trend in games not to just 'recycle' animation. Whilst a game may require a person to walk from A to B several times, it is not necessary to exactly repeat that walk. In real life if a person walks to a door, there are many ways in which he can complete that walk. All are 'walking', but vary. By storing the information 'man walks to door' and generating it each time it is needed, with a different starting 'seed', much more 'reality' is obtained

Further games trends detected are seen as a 'backlash' against the very large games. These are typified in the downloadable 'Geometry Wars' game downloadable from 'X-Box Live' at around 7 Euros. It is obvious that such games cannot give the extended experience of 'major' games costing 70 Euros – but still there is a growing market here. Downloadable games must currently be 'small' in file size, otherwise download times, even with current broadband, become intolerable. The beauty of downloadable games is that no-one needs to carry 'stock'.

Our limited survey of games developers indicates that a 'Utopian vision' would be one where it was possible to 'drag and drop' between media would be VERY welcome!

4.3 Cross Media Production Trends

There has always been a 'Cross Media' aspect to Media, ever since it was invented. Hollywood, as long ago as 1906 was producing 'the film of the book', or 'the film of the play'. These days, the producers see themselves as 'rights holders' of 'properties'. (An example here for the non-media person is that 'Superman' or 'Batman' are 'properties' this sense).

Media Downloading

XBOX live offers a download service for TV and film, choose from ordinary or HD and very cheap (\$6 for a full movie), which is downloaded to your Hard Drive and then erased after a set period. The SONY PS3 is expected to do the same – Microsoft and Sony are now not just battling with each other for console domination but to be the entertainment platform of choice.

It is noted that the BBC are using IP address identification to prevent downloading material of cultural value to territories that they cannot enforce their copyright in easily. Whilst it is believed that sophisticated network pirates could defeat this, it shows how media download control can be compatible with Digital Rights Management.

It is noted that some Open Source platforms are available for home gateways (consisting of PCs using LINUX) – apart from the two major manufacturers, Sony and Microsoft's offerings. It remains to be seen how much take-up there is of these platforms.

Recently, British Telecom announced their new service of Movie downloads.¹¹

PSP and UMD Format

Another sign of convergence and Cross Media is the establishment of the UMD format for movies. This format is used by the SONY PSP portable games console, and has had a reasonable uptake in the production of 'Hollywood' films in this format. The very idea of watching Hollywood 'blockbusters' on your games console would have been unthinkable a few years ago!

4.4 IPTV and moves towards 'Intelligent Media'

IPTV is of particular interest to this report, as it is uniquely a media designed for one consumer at a time (as opposed to a 'shared experience' as seen in the typical domestic living room). IPTV was well in evidence at the recent IBC show in Amsterdam (IBC is the International Broadcasting Convention – the largest show for Media technology in Europe). Market size estimates for IPTV vary widely, but typical estimates are around \$17 Billion by 2010 (see <http://informatv.com/articles/2005/01/17/globaliptvrevenue/>)

The architecture of IPTV systems is usually quite complex and includes a diverse set of network elements that must be brought together into a single operational network for the service to operate correctly. An IP TV network typically includes:

- Access Systems
- Video Headend Systems
- Video On Demand Servers
- Middleware
- Content Protection and Digital Rights Management Software
- The underlying transmission system

The complexity of these systems has created a market for IP TV System Integration and Professional services to insure that service provider deployments will be successful. Each of these areas includes its own set of complications. They can be organized into natural groupings that reflect how these networks are integrated and installed:

- **The Access Network** – includes the DSL or fiber systems and includes the evaluation and grooming of the local loop plant.
- **The Video Network** – includes the Video Headend, the Video On Demand servers, the satellite and off the air capture systems, and the Content Protection and Digital Rights Systems.
- **The Middleware and Backend Software** – includes the customization and installation of the Middleware as well as the integration of it with the service provider's legacy backend OSS/BSS software systems.

Moves towards 'Intelligent Media' are starting to become visible. The well known Media retailer, Amazon, is in the USA selling 'on line' versions of movies. These movies are naturally compressed files, designed to work on a specified number of MPEG4 based Media players. However, rather than the user have to remember the make and model number of the media player, it is possible for him to connect the media player to his Internet enabled PC via a USB connection, and the Media downloading software (actually part of Amazon.com) interrogates the device, and downloads to the user's PC the right file version optimised for that player. Whilst this is not a very large step forward, it is an example of the 'media configuring itself for the output device'.

IPTV 'Piracy'

Providing users with Digital content makes piracy of this material easy. Much IPTV Piracy is reported in Asia, where the piracy business is estimated to be a Billion Dollar Industry¹²

This in turn makes the 'Anti-piracy' of IPTV turn into a viable industry. The market leader is probably Widevine¹³

5 Digital Media Research

5.1 Research in Intelligent Media Content Technologies

5.1.1 *Introduction: European research projects*

Intelligent Media Content Technologies represent a very new field for research, whose parameters are comparatively undefined, although it builds on a very large body of well-established component disciplines.

The FP6 IST Programme launched the first ever funding call for research into Intelligent Content in 2005 (Call 4), which established SALERO as well as the MESH project, which is researching multimedia semantic syndication for enhanced news services, to find ways of extracting, comparing and combining news stories from multiple sources and creating personalised summaries.

The Intelligent Content research topic combines semantic technologies with various aspects of media, communications and knowledge technologies. One approach has been to apply semantics to the management of multimedia content. AXMEDIS (IST FP6 Integrated Project) started in 2004, with the aim of developing a framework of semantic methods and tools to speed up and optimise content production and distribution. AceMedia is another FP6 IP, which also started in 2004, researching knowledge discovery and embedded self-adaptability to enable content to be self-organising, self-annotating, and self-associating, based on the concept of the Autonomous Content Entity; and the CONTENT4ALL IST STREP is developing intelligent media management methodologies for publishing over heterogeneous devices, with an open-source platform for leisure, entertainment and communication.

Another category of project is researching novel technologies for creating media: IP-RACINE¹⁴ (started 2004) is researching Digital Cinema 'from scene to screen' and NM2 'New Media for a New Millennium' (FP6 IST Integrated Project) is researching new production tools for non-linear broadband media that are personalised for individual users. POLYMNIA (IST STREP) is developing an intelligent cross-media platform for personalised entertainment in thematic parks. A cluster of projects for audiovisual search and retrieval, which draw on semantic techniques and have relevance to intelligent content, is coming on stream early in 2007 as a result of IST FP6 Call 6, including SEMEDIA, which has strong ties to SALERO and extends research in media technologies, web semantics, AI, CBIR and interface design.

We are aware of much research currently being undertaken in contributory fields of Digital Media Technology that will feed into the future of intelligent media. While it is beyond the scope of this chapter to provide a complete 'state of the art' review, the following summaries give an overview of relevant work.

5.1.2 *Image analysis, description and reconstruction*

Digital photographs, video and film capture the projected appearance of real-world scenes in real time, but do not let us re-visualise the scene from new viewpoints, edit shape or appearance, or combine the scene with virtual models. During the past ten years, several Image-Based Rendering techniques have proposed solutions by using multiple images from different viewpoints. With a dense sampling of viewpoints, new images can be synthesised from intermediate or extrapolated viewpoints without any knowledge of 3D scene geometry. In the light field approach, tens or even hundreds of images from different viewpoints are combined.

Where IBR approaches are invoked for the rendering of static natural objects and scenery, more recent work Video-Based approaches extend to the rendering and animation of dynamic scenes (objects, scenery, animals and humans) and facial animation. Very recent work in video-based character animation uses a multi-camera studio to capture multiple motions (walk, run, jump, kick etc.). Photo-hull reconstruction is applied to obtain unstructured mesh sequences, which are automatically converted to geometry videos using spherical mapping. A motion graph of geometry videos is used to achieve real-time rendering of multiple characters with control of both movement and viewpoint¹⁵. Various approaches to model extraction have been proposed, using multiple views to reconstruct animated models of people. The latest work at the University of Surrey reconstructs shape and texture from 13

camera views in a chromakey studio¹⁶. A hierarchical approach has been developed to reconstruct the movements of multiple people¹⁷. This work is, for the most part, at the laboratory stage but Surrey has been working with the BBC toward practical media applications and the University of Hasselt Expertise Centre for Digital Media¹⁸ is working (in the IP-RACINE project) on applications of VBR to digital cinema postproduction. The ability automatically to generate new camera viewpoints and movements in postproduction, or even at playout, is a major research challenge for the next generation of intelligent content technologies.

The point of view of the camera is an essential part of film language. Existing work deals essentially with intelligent positioning of cameras, but is not related to the story being told. BMCI is currently developing an ontology of camera movements, related to cinematic language, to support more intelligent camera movements for programme production.

Turning from the camera to the character, despite the long history of efforts to generate convincing and potentially autonomous virtual characters, there is still little published work on the automatic generation of movement involving a significant part of the body beyond the traditional procedural, kinematic approaches. Movement generated by kinematic procedures does not yet look sufficiently natural for media applications. One promising approach to the generation of more natural movements involves the interpolation of concatenated movements and intelligent enlargement. This approach now produces output for TV broadcasting, Internet, and mobile phones. Further work will be needed to enable characters to be processed at the scene representation level. The situated robotics work pioneered by Brooks et al. at MIT AI lab¹⁹ and work on socially intelligent agents are relevant for creating expressive movement and work on dance expression developed, by DIST of the University of Genoa²⁰ may be useful supports for the description of body language.

5.1.3 Analysis and generation of sound and acoustic content

One current research strategy in sound and music computing is to complete the understanding of human perception and cognition processes, in order to provide technology for meaningful semantic interaction. Many research areas are included on the improvement of this knowledge, as for instance computational scene analysis (identification of sound sources), auditory attention memory & learning or music structural analysis. Another strategy addresses augmented perception, which provides technology for intelligent interaction between devices and users.

The simulation of acoustic effects in 3D virtual environments builds on the techniques for the acoustic design of concert halls²¹. It involves the study of algorithms for the propagation of sound in virtual spaces to create methods for designing virtual spaces that provide a sense of acoustic presence and immersion.²² An interesting approach, which is being developed in the IP-RACINE project, involves the detailed analysis and modelling of the acoustics of real spaces, to derive an acoustic model that can be applied at will²³ or allow the semi-automatic creation of audio ambience²⁴. The problem of modelling the acoustic space is complicated by the computational difficulty of solving the necessary partial differential equations. BMCI has been working on a hybrid method that uses finite difference methods for low frequencies and ray tracing for high frequencies, as the basis for a potential software solution to the problem of rendering the virtual acoustic space in (or close to) real time²⁵.

The problem of modelling and rendering acoustic spaces is complementary to problems of modifying the voice, either to change its timbre²⁶ or modify the speed and rhythm²⁷. Some of the most advanced work in relation to voice modelling²⁸ and morphing²⁹ has been carried out in relation to the singing voice and for 'virtual karaoke' applications.³⁰

5.1.4 Analysis, translation and generation of speech

General-purpose speech recognition systems such as Autonomy's 'Virage' distinguish between the data on which they operate and the algorithms they work with. Decoding is usually based on neural network technology (to train phoneme recognition modules) or hidden Markov models (HMMs) and use statistical methods to process the digital audio signal for speech recognition, speaker recognition, language recognition and speech enhancement. Digitised speech is decomposed into its phonetic constructs, the phonetic sequence is then analysed in conjunction with acoustic model and statistical improbabilities to calculate the most probable sequence of words, which are presented on screen as text in real time. Systems require very considerable training, with very large numbers of voices and

vocabularies: English language systems (which have been trained on corpora of billions of words) are the most advanced. The basic problem remains to find a way to build a system that is not only speaker independent but also domain independent.

Work in Machine Translation (MT) started as long ago as 1950, and machine translation technology is widely used to obtain rough-and-ready translations of web pages. There is currently a focus, for security reasons, on R&D into MT from languages such as Arabic, Pashto and Dari. Machine translation systems increasingly use statistical methods based on bilingual text corpora, and achieve the best results for pairs of languages with similar structures³¹. The Periódico de Catalunya introduced MT from Spanish to Catalan in 1997³². A consortium of Spanish universities is currently developing open-source MT systems for the languages of Spain³³ and has produced an open source system 'Apertium'.³⁴ Statistical systems improve considerably if trained for specific input (such as weather reports), and can be used in conjunction with Translation Memories or TM, large databases of humanly translated sentences that may occur again in successive source texts. The database provides suggestions for sentences that have already been translated in other texts and avoids the typical errors and noisy output associated with MT systems. Although designed for technical and repetitive texts, Translation Memories can be used for all types of texts and can be fed directly with aligned corpora. Recent approaches look to integrate complementary TM and AT systems by allowing recovery from TMs of sub-sentence segments (phrases) and also building the MT system at sub-sentence level³⁵.

Although speech synthesis systems have improved considerably over the last 20 years, they rarely sound properly human, with problems of timing, prosody and intonation³⁶. Digital speech synthesis systems contain an engine to normalise the input text and perform a text-to-phoneme conversion, with a synthesiser to generate synthetic speech waveforms. In Spanish it is possible to predict the pronunciation of words based on the spelling in nearly all instances, so speech synthesis can be largely rule-based, whereas English has an irregular spelling system and relies more on dictionaries. Current systems include Festival from Edinburgh University, Loquendo TTS³⁷, AT&T Natural Voices³⁸, and ScanSoft RealSpeak³⁹. Although speech synthesis is increasingly widely used in many areas of activity, its application in media fields is just emerging from the research level. One of the most important problems is to make the voice expressive, with a convincing suggestion of emotion. Approaches are classified into three main groups: rule-based synthesis, such as HAMLET⁴⁰ and the Affect Editor⁴¹; diphone concatenation synthesis⁴²; and unit selection or corpus-based synthesis⁴³. Promising approaches integrate diphone and corpus-based analysis: if a larger unit (full words or even phrases) is found it is taken as such; if not the diphone sequences are used.⁴⁴

Unlike rule-based methods, concatenative and corpus-based methods build on a human recorded voice. The speech resulting from recorded methods is claimed to be more natural, but rule-based methods are often preferred for small devices because of their reduced data storage requirements. This is becoming less important as device memory increases. Synthetic speech quality decreases when the target domain or idiom diverges from the style covered by the speech database⁴⁵, so the desired domain, emotion, or style should be included in the corpus. Intelligent content semantics, workflows and technologies for managing content

In the audiovisual media industries, data types are quite often 'bundled' under a wrapper, intended to keep all of the necessary files together. Typical file wrappers such as MXF and AAF⁴⁶ are designed to facilitate the transfer of content data between applications, by associating the necessary metadata. Wrapper formats are limited by the lack of formal semantic representations, means of automatically generating metadata from content, or indeed of storing metadata that goes beyond the simplest descriptors.

5.1.5 Intelligent content semantics and workflow technologies

Ontology modelling and ontology-based metadata creation has focused on textual content for the past decade and multimedia ontologies have been mostly used to assist manual content annotation⁴⁷. The creation of media semantics and ontologies, which underpin several aspects of intelligent content technology approaches, is a very new field of research, which has yet to reach industry application, although its potential value is widely acknowledged. FP6 IST research projects have now started to address multimedia ontologies seriously, although no accepted standard for multimedia ontologies exists. AceMedia⁴⁸ has developed the Visual Descriptor Ontology (VDO) and the Multimedia Structure

Ontology (MSO) for semantic multimedia content analysis and reasoning with representations of MPEG-7 visual descriptors and models.

The University of Reading Intelligent Media Systems and Services Research Laboratory (IMSS)⁴⁹ is leading the workflow area of Axmedis⁵⁰, which is applying Artificial Intelligence techniques and semantic technologies to content composition, representation (format) and workflow control, and exploiting and extending MPEG-21 for DRM. The toolsets now emerging from Axmedis are beginning to attract serious industry interest.

5.1.6 Audiovisual information retrieval

The large body of research into non-text-based media information indexing and retrieval has so far made little impact on the media industries. Object recognition in general image collections remains a seemingly insurmountable challenge and the semantic gap⁵¹ between low-level features and the user is a major stumbling block for CBIR. However, automatic image understanding is not critical for retrieval systems that exploit the user's knowledge and learn to infer semantics from low-level feature representations. An intuitive interactive environment can improve the system's effectiveness. Techniques are mostly based on machine learning or pattern recognition, involving either semi-automatic annotation⁵² or image classification⁵³.

Relevance feedback is invaluable for CBIR systems. Low-level features are used to estimate the relevance of the images and the initial estimation is progressively improved, with prompts to steer the user. Relevance feedback may be formulated by a vector space model⁵⁴, probabilistically as belief propagation⁵⁵, or a classification task⁵⁶. Recent approaches revolve around the user, as opposed to the user having to 'fit in' with the data types and structures. Innovative interfaces replace the traditional linear result display, ranked by similarity to the query, with two- or three-dimensional maps of the returned images⁵⁷. Multidimensional displays aim to reveal relationships by visualising similarities between images in a global view⁵⁸.

When faced with large amounts of multimedia data, a compact representation of content is required: media summarisation clusters and filters information to generate a pictorial summary or a visual skim. Content segmentation (as opposed to structural segmentation) expresses elements that have a semantic meaning for the author or viewer: the task of audiovisual content analysis is to extract metadata from audiovisual material, reversing the authoring processes⁵⁹. One approach uses semantic content segments such as a topic (e.g. a news report), a genre or sub-genre (horror movie), logical units (dialogue), or named events (a car crash, an explosion). Low-level feature extraction may be based on colour, shape, and motion, using MPEG-7. Summarisation as an image strip is another approach, used for example in the FP6 STREP GMF4iTV project⁶⁰. Personalisation and semantic technologies can be exploited to deliver personal summaries or predict whether a user would be interested in a particular sequence or scene. Techniques from video content analysis may be combined with user logs and shot ranking to generate video summaries⁶¹. Logs may also consider the way a user consumes a video or the transition between the viewing of several video sequences⁶² and the user's trail through the data set⁶³.

Audio retrieval systems are in general more advanced than image systems and there has been significant work on the use of MPEG-7 for content-based audio retrieval and transformation.⁶⁴

None of the current media production and management systems employs context-based retrieval methods. Industrial practice relies on meticulous human created and entered metadata, sometimes supported by script or extracted text linked accurately to timecode (as produced by Autonomy's Virage⁶⁵ system). The lack of proper information retrieval methods is now seen as a major problem by the technology suppliers and we expect to see numerous competing systems emerge in the next few years as a result of research supported by the IST and national projects (for which this has become a major research priority).

5.2 Trends in Industrial RTD and Patent Activity

Intelligent content technologies are just starting to make an impact on the content-based industries. In the area of content tailoring, Volantis⁶⁶ provides a server-based means of recognising registered user devices and automatically reformatting web pages or content streams to fit them. In film and TV, behavioural animation techniques and software such as Massive⁶⁷ are widely used for generating crowd-related visual effects. The reactions of the characters determine what they do and how they do it and can simulate basic emotive qualities such as bravery, weariness, or joy (as in the *Lord of the Rings* 'Helm's Deep' sequence). Crowd characters are mute, unlike the script-driven virtual weather presenter SAM being further developed by CCRTV and BMCI in SALERO, which shows a high degree of verbal as well as visual behaviour.

The state of the art of IPTV is represented by Telefónica's 'imagenio' platform⁶⁸, which allows the distribution of video channels and VoD through DSL networks and has acquired 300.000 users in Spain alone in the past year. IPTV is expected to stimulate the demand for intelligent content, which will build on the various constituent technologies.

The media industries apply advanced image processing algorithms for motion vector analysis to assist in functions such as automatic matte pulling, degrading, regrading, and wire removal as part of plug-in toolsets (such as The Foundry's Furnace and Tinderbox, which was extensively used on *Lord of the Rings* and *King Kong*)⁶⁹. In Europe, many media software companies have not previously protected their IP, arguing that the pace of change made this unnecessary. However, increasing US patent filing in the field along with industry consolidation have

A significant amount of Industrial R&D is going into 3-D Display systems. We are aware of the legal battles in this field between 'In-Three' and IMAX. Both companies are busy filing new 3-D applications in an attempt to monopolise this area. Over 50 patents have been granted by the USPTO in 3-D display systems in the past three years.

Media industry sound creation and acoustic alteration is still a craft process, using tools such as 'Protools'⁷⁰.

5.3 EU RTD trends, policy, and opportunities for Intelligent Media Research

The area of Intelligent Content continues to be recognised by the European Commission as an important one. This can be seen in Framework 7, which has the following areas identified in its Workprogram for 2007-8

5.3.1 Objective IST-2007.4.2 (IST-2007.4.4): Intelligent Content and Semantics

Target outcome

Medium term:

- a) Advanced **authoring** environments for the creation of **novel forms** of interactive and expressive content enabling multimodal experimentation and non-linear story-telling. These environments will ease content sharing and remixing, also by non-expert users, by automatically tagging content with semantic metadata and by using open standards to store it in networked repositories supporting symbolic and similarity-based indexing and search capabilities, for all content types.
- b) Collaborative automated **workflow** environments to manage the **lifecycle** of novel and legacy media and enterprise content assets, from the acquisition of reference materials to the versioning, packaging and repurposing of complex products, including their linguistic and cultural adaptation to target markets and user groups. Empirical results from the psychology of human perception and attention will be used to identify salient multimedia segments and apply summarisation and encoding schemes that will improve content storage and transmission without affecting its perceptual properties.

- c) Architectures and technologies for **personalised distribution, presentation and consumption** of self-aware, adaptive content. Detecting and exploiting emergent ambient intelligence they will use features embedded in content objects and rendering equipment to enable dynamic device adaptation, immersive multimodal experiences and contextual support of user goals and linguistic preferences. Privacy preserving learning algorithms will analyse user interactions with devices and other users so as to update and effectively serve those goals and preferences..
- d) Actions geared towards **community building**, intended to stimulate cross-disciplinary approaches and a more effective user/supplier dialogue, and other measures, including field validation and standards, aimed at a faster **uptake** of research results. Usability and technology assessment studies, economic analyses and roadmaps to chart the democratisation of personal and community based multimedia production and management tools.

Longer term:

- a) **Semantic foundations**: probabilistic, temporal and modal modelling and approximate reasoning through objective-driven research moving **beyond current formalisms**. Theoretical results will be matched by robust and scalable reference implementations. Usability and performance will be tested through large scale ontology mediated **Web integration** of heterogeneous, evolving and noisy or inconsistent data sources ranging from distributed multimedia repositories to data streams originating from ambient devices and sensors, supporting real time resolution of massive numbers of queries and the induction of scientific hypotheses or other forms of learning.
- b) **Advanced knowledge management systems** for information-bound organisations and communities, capable of extracting **actionable meaning** from structured and unstructured information and social interaction patterns, and of making it available for activities ranging from information search through conceptual mapping to decision making. Such systems will exploit semantics embedded in multimedia objects, data streams and ICT-based processes, and rely on formal policies to manage user access as well as audit trails in support of dynamic virtual organisations. Research advances will be embedded within end-to-end systems using **computer-tractable knowledge** in support of dynamic data and application integration, automation and interoperation of business processes, automated diagnosis and problem-solving in a variety of domains. Robustness, scalability and flexibility will be tested in real-life settings, together with interworking with legacy systems.

Expected impact

These activities will make digital resources that embody creativity and semantics easier and more cost-effective to produce, organize, search, personalise, distribute and (re)use, across the value chain.

- Creators will be able to design more participative and communicative forms of content.
- Publishers in creative industries, enterprises and professional sectors will increase their productivity with innovative content of greater complexity and ease of repurposing.
- Organisations will be able to automate the collection and distribution of digital content and machine-tractable knowledge and share them with partner organisations in trusted collaborative environments.
- Scientists will operate more efficiently by automating the link between data analysis, theory and experimental validation.

Funding schemes

CP, NoE, CSA

Indicative budget distribution

101 M€

- CP 91 M€ of which a minimum of 39 M€ to IP and a minimum of 25 M€ to STREP; NoE 5 M€
- CSA 5 M€

Calls:

FP7-ICT-2007-1 [51 M€], FP7-ICT-2007-3 [50 M€]

6 Conclusions

6.1 Possible Future Directions of Digital Media

We can envisage the media in the year 2020.

Audiovisual media in 2020 will offer a much richer experience than today, with wider possibilities for creators and audience. Entertainment media will be primarily 3D, in image and sound, but intelligent technologies will make the content seamlessly self-adaptive to different platforms and contexts. Audiovisual content objects with intelligent functionality will customise programmes and advertising for individual users, and automate the production of synthetic characters with convincing behaviour and appearance. The functional distinction between content 'creators' and 'users' will get blurred – although the professionals will usually make more polished media and more complex programmes than the amateurs. It will be possible for people (*auteurs* or consumers) to change the viewpoint of a scene, or change a storyline, automatically populating new scenes with characters or objects. It will be possible to transform a voice to impersonate a different character or to change the language of a programme. Media users and creators will be able to retrieve and navigate vast quantities of content simply, accurately and intuitively and repurpose content across domains such as healthcare, assisted living or learning.

Media Capture

By 2020, we believe that nearly all image capture will be carried out using digital devices. However, unlike now, it will be possible to capture reasonable 3-D with a single device (whether it uses twin lens stereo vision, combines optical capture with depth sensing, or is a mobile monocular device that moves around the shooting area).

Digital data will be pre-processed on the acquisition device, which will export scenes composed of intelligent media objects. These are likely to consist of 3-D models with associated behavioural and relational metadata. It is anticipated that human characters will also be captured in this way. The models will be resolution independent, so that they can be rendered at any resolution and any frame rate, for any purpose or delivery platform. The capture device will have a large amount of on-board storage and secure wireless links to even larger external data stores will be available.

Post Production

In 2020 postproduction will, we believe, be a 3D process, taking the captured 3D models and stock elements (also available as 3D models) and combining them in new ways. Again, this process will be vector based, implying no resolution limits. The 3D postproduction paradigm will make a very sophisticated use of metadata. Low-level features (extracted from 3D motion vector analysis for example) will combine with automatically deduced high-level interpretations of behaviour and emotion.

Characters, actions and sets will respond to changes in the script. While high quality media will still rely on the expressive interaction and nuance of human actors, virtual characters will be driven from the script, with expressive synthetic voices (modelled on human actors) and realistic emotional responses both to the script and to each other.

Intelligent content technologies and rich metadata will make it possible, for example, to rearrange the set after shooting. If we want to move a character in front of the sofa rather than behind, not only will it be easy to move the figure but also the intelligent 'sofa' object will be editable as a single object that responds coherently to a different lighting model. If we wish to change the local colour of the sofa, from red to green, and turn on another lamp, all the colours will change correctly (including the shadows and reflected colour) without skilled manual intervention. If we move the actor to sit on the sofa's arm, it will 'know' how sofas behave and automatically deform correctly. It will also be possible to specify new arbitrary points of view for the camera. Furthermore, although the resulting image stream may be presented from the Director's chosen point of view, sufficient information will be available for the viewer / consumer to change the point of view to his liking – unless the Director has disabled this function or restricted the permitted points of view. Characters, being Intelligent Media Objects, will be able to be replaced, or complemented with other characters. It will be possible to change 'parts' of characters, so that we can replace a stunt man's face with the face of a famous actor. It will be also possible to change

attributes of a character – for example to copy the ‘flying’ motion of Superman to another chosen character, or the ‘run’ of a Famous Olympic runner to another selected character. Audio Postproduction will complement the visual material: the 3D soundscape will be under the Director’s control, automatically matched to the acoustics of the visual environment and similarly navigable by the user.

Exhibition / Consumption

The universal availability of broadband networks will not kill public media display, but the cinema of 2020 will, we believe, offer 3D autostereo, which can be viewed from a very wide range of viewpoints with no special glasses or headsets. Every member of the audience will hear the audio optimally balanced for ‘their’ sitting position, with correct audio positioning.

Many more homes will have ‘Home Cinema rooms’, extending the trend started in the early 2000s, and 3D autostereo will be coming to home cinemas and domestic TV. Of course, it is always easy to produce a 2D representation from 3D material, just as it is possible to produce monochrome from colour images, and intelligent media objects will be able to adapt themselves to any sort of ploy, in the home, the office, the school or on the move. The linear movie or TV drama, as we know it today, will be one form of entertainment genre that the consumer can construct from a series of intelligent media objects that are assembled on the fly to give a choice between the longer ‘Directors’ Cut’ or the consumers’ ‘optimal cut’, containing the deleted scenes, and with the ‘political statement’ end. The children get to watch the happier shorter, less violent version. The ancillary material will be far greater than today’s ‘DVD extras’ and be seamlessly presented by high-speed computer networks. This material will be dynamic, presenting new information every time it is accessed, adding substantially to the entertainment value.

We believe that there will be no perceptible difference in visual quality between the film and the ‘game of the film’, which will in 2020 be made from exactly the same elements. By repurposing the film content objects in real time in response to the consumer’s commands, in conjunction with an underlying script or ‘plot engine’, the game and film become visually (and audibly) identical. 2020 technology and intelligent media will also support new forms of information and entertainment, which will include immersive, interactive multi-sensory experiences, driven by the participants, where elements of drama, game and exploration merge. Although it’s hard to imagine the experience, we may even begin to navigate intelligent content freely not just in the three dimensions of space but the fourth dimension of time.

6.2 SWOT analysis of SALERO, in the light of current trends

The *strength* of SALERO is that the reuse of media components will reduce the cost of media production, thus benefiting the European Media production Industry.

The *weakness* of SALERO is that it is important to establish standards for Intelligent Media. If this is not done, then the chances of SALERO having a successful uptake are small.

The *opportunities* afforded by SALERO are almost boundless. The potential savings of Intelligent Media that is easily ported across platforms and media types is likely to be in the Billions of Euros per year savings.

The *threat* to SALERO is that if it does not remain totally in touch with trends, then it could produce unusable results.

6.3 Recommendations for SALERO

The recommendations from this deliverable are that

- SALERO should continue in its present form and direction.
- SALERO should continue to monitor the market, and continue to analyse trends etc to ensure commercial relevance
- SALERO should continue to liaise with its User Group – the views of the Users are important!

7 References

- ¹ www.ihd.org
- ² Chart of Subscriber Growth, <http://www.mmogchart.com>
- ³ Parks Associates (2005), Online Gaming Revenues to Triple by 2009
http://www.parksassociates.com/press/press_releases/2005/gaming-1.html
- ⁴ <http://www.nickyee.com/daedalus/>
- ⁵ <http://en.wikipedia.org/wiki/MPEG-1>
- ⁶ <http://en.wikipedia.org/wiki/MPEG-2>
- ⁷ <http://en.wikipedia.org/wiki/MPEG-3>
- ⁸ <http://en.wikipedia.org/wiki/MPEG-4>
- ⁹ <http://en.wikipedia.org/wiki/MPEG-7>
- ¹⁰ <http://en.wikipedia.org/wiki/MPEG-21>
- ¹¹ <http://news.bbc.co.uk/1/hi/entertainment/5217212.stm>
- ¹² <http://www.convergedigest.com/bp-ttp/bp1.asp?ID=211&ctgy=>
- ¹³ www.widevine.com
- ¹⁴ <http://www.ipracine.org/overview/overview.html>
- ¹⁵ J. Starck, G. Miller and A. Hilton. Video-Based Character Animation ACM Symposium on Computer Animation, 2000
- ¹⁶ J. Starck and A. Hilton. Model-based multiple view reconstruction of people. IEEE International Conference on Computer Vision, pages 915-922, 2003
- ¹⁷ J. Starck and A. Hilton. Model-based multiple view reconstruction of people. IEEE International Conference on Computer Vision, pages 915-922, 2003
- ¹⁸ http://www.edm.uhasselt.be/about/the_institute
- ¹⁹ <http://people.csail.mit.edu/~brooks>
- ²⁰ <http://www.infomus.dist.unige.it/Lab/People/Camurri.html>
- ²¹ M. R. Schroeder and, B. Atal, "Computer simulation of sound transmission in rooms," IEEE Conv. Rec. Vol. 11, pp. 150-155 (1963)
- ²² Udo Zölzer Ed. "DAFX – Digital Audio Effects", John Wiley & Sons, LTD (2002)
- ²³ D. Stanzial, D. Bonsi, G. Schiffrer, "Four-dimensional Treatment of Linear Theory of Acoustic Fields and Radiation Pressure", Acustica-Acta Acustica, Vol. 89 (2003), pp.213-244.
- ²⁴ Cano, P. Fabig, L. Gouyon, F. Koppenberger, M. Loscos, A. Barbosa, A. 2004.-'Semi-Automatic Ambiance Generation'- Proceedings of 7th International Conference on Digital Audio Effects Naples, Italy
- ²⁵ A. Garriga, C. Spa and V. López, "Computation of the Complete Acoustic Field with Finite-Differences Algorithms". (Invited paper) Forum Acusticum, Budapest, (2005).
- ²⁶ Loscos, A. Bonada, J. 2004.- 'Emulating Rough And Growl Voice In Spectral Domain'-Proceedings of 7th International Conference on Digital Audio Effects, Naples, Italy
- ²⁷ Bonada, J. 2000.- 'Automatic Technique in Frequency Domain for Near-Lossless Time-Scale Modification of Audio' Proceedings of International Computer Music Conference 2000 Berlin, Germany

- ²⁸ Bonada, J. Loscos, A. Cano, P. Serra, X. 2001.- 'Spectral Approach to the Modeling of the Singing Voice'-Proceedings of 111th AES Convention New York, USA
- ²⁹ Bonada, J. 2004.-'High Quality Voice Transformations Based On Modeling Radiated Voice Pulses In Frequency Domain' -Proceedings of 7th International Conference on Digital Audio Effects Naples, Italy
- ³⁰ Cano, P. Loscos, A. Bonada, J. de Boer, M. Serra, X. 2000. -'Voice Morphing System for Impersonating in Karaoke Applications' Proceedings of International Computer Music Conference 2000 Berlin, Germany
- ³¹ Jan Hajic, Jan Hric & Vladislav Kubon: *CESILKO – an MT system closely related languages*. ACL 2000 Companion Volume, Hong Kong, October 1-8, 2000
- ³² Montoliu, C. (1998): *Informe sobre el sistema de traducción automática del Periódico de Catalunya* <http://europa.eu.int/comm/sdt/bulletins/puntoycoma/51/periodico.htm>. 12 November 1999.
- ³³ 'Open Source Machine Translation for the Romance languages of Spain' PROFIT FIT340101-2004-0003.
- ³⁴ <http://apertium.sourceforge.net/>
- ³⁵ Hutchins, W.J. Current commercial machine translation systems and computer-based translation tools: system types and their uses. *International Journal of Translation*. vol.17, no.1-2, Jan-Dec 2005
- ³⁶ *Improvements in Speech Synthesis: Cost 258: The Naturalness of Synthetic Speech* ed E. Keller, G. Bailly, A. Monaghan, J. Terken, M. Huckvale
- ³⁷ <http://www.loquendo.com/>
- ³⁸ <http://www.naturalvoices.att.com/>
- ³⁹ <http://www.scansoft.com/>
- ⁴⁰ http://www.computing.dundee.ac.uk/ac_research/themedetails.asp?id=11
- ⁴¹ Cahn, J. E., *Generating Expression in Synthesized Speech*, Master's Thesis, MIT, 1989. <http://www.media.mit.edu/~cahn/masters-thesis.html>
- ⁴² Bulut, M., Narayanan, S. and Syrdal, A., Expressive speech synthesis using a concatenative synthesizer, *Proc. of ICSLP 2002*.
- ⁴³ Iida, A., Campbell, N., Higuchi, F. and Yasumura, M., *A corpus-based speech synthesis system with emotion*, *Speech Communication*, Vol. 40, Issue 1-2, 161-187, April 2003.
- ⁴⁴ Hunt, A., Black, A. W., Unit selection in a concatenative speech synthesis system using a large speech database, in *ICASSP*, vol. 1, pp. 373 - 376, Atlanta, USA, 1996.
- ⁴⁵ Black, A.W., Lenzo, K, *Limited Domain Synthesis*, in *ICSLP*, Beijing, China, 2000.
- ⁴⁶ <http://www.mxf.org.uk>; <http://www.aafassociation.org>
- ⁴⁷ A.T. Schreiber, B. Dubbeldam, J. Wielemaker, and B.J. Wielinga. "Ontology-based photo annotation". *IEEE Intelligent Systems*, 16:66-74, May/June 2001
- ⁴⁸ AceMedia (EC-project), URL: <http://www.acemedia.org/aceMedia>
- ⁴⁹ <http://www.imss.reading.ac.uk/>
- ⁵⁰ <http://www.axmedis.org/>
- ⁵¹ A. W. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, Dec. 2000
- ⁵² E. Chang, K. Goh, G. Sychay, and G. Wu. CBSA: Content-based soft annotation for multimodal image retrieval using bayes point machines. *IEEE Trans. Circuits Syst. Video Technol. (Special Issue on Conceptual and Dynamical Aspects of Multimedia Content Description)*, 13(1):26–38, Jan. 2003.

-
- ⁵³ A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *Int. Journal of Computer Vision (Kluwer Academic Publishers)*, 2001.
- ⁵⁴ G. Salton and M. J. McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill, Tokio, 1983.
- ⁵⁵ N. Vasconcelos and A. Lippman. Bayesian relevance feedback for content-based image retrieval. In *IEEE Proc. of Workshop on Content-based Access of Image and Video Libraries*, pages 63–67, 2000
- ⁵⁶ M. E. J. Wood, B. T. Thomas, and N. W. Campbell. Iterative refinement by relevance feedback in content-based digital image retrieval. In *ACM Multimedia 98*, pages 13–20, Bristol, UK, Sept. 1998. ACM Press.
- ⁵⁷ J.-Y. Chen, C. A. Bouman, and J. C. Dalton. Hierarchical browsing and search of large image databases. *IEEE Trans. Image Processing*, 9, 2000
- ⁵⁸ K. Rodden, W. Basalaj, D. Sinclair, and K. Wood. Does organisation by similarity assist image browsing? In *Proc. of the ACM Int. Conf. on Human Factors in Computing Systems, Sensable Navigation Search*, pages 190–197, 2001
- ⁵⁹ C.G.M. Snoek, “The Authoring Metaphor to Machine Understanding of Multimedia”, Ph.D. Thesis, University of Amsterdam 2005
- ⁶⁰ http://www.joanneum.at/en/informatik/projekte_detail.php?p_iid=IIS&p_pid=41
- ⁶¹ Bin Yu, Wei-Ying Ma, Klara Nahrstedt, Hong-Jiang Zhang, “Video Summarization based on User Log Enhanced Link Analysis” , Technical Report MSR-TR-2003-51, Microsoft Research, Microsoft Corporation. Aug 14, 2003
- ⁶² Sylvain Mongy. Fatma Bouali. Chabane Djeraba. “Video usage mining. Multimedia Dictionary”, CRC Press, 2005
- ⁶³ Ullas Gargi and Pere Obrador. “Multimedia links and metadata channels for distributed shared browsing and authoring”. Proceedings of SPIE 2004 Volume 5307
- ⁶⁴ Celma, O. Gómez, E. Janer, J. Gouyon, F. Herrera, P. Garcia, D. 2004. -'Tools for Content-Based Retrieval and Transformation of Audio Using MPEG-7: The SPOffline and the MDTools' Proceedings of 25th International AES Conference London, UK
- ⁶⁵ <http://www.autonomy.com/>
- ⁶⁶ <http://www.volantis.com/>
- ⁶⁷ www.massivesoftware.com
- ⁶⁸ www.telefonica.es/tol/imagenio.html
- ⁶⁹ <http://www.thefoundry.co.uk/>
- ⁷⁰ www.digidesign.com

8 Glossary

Partner Acronyms

AM	Activa Multimedia, ES
BLITZ	Blitz Games, UK
CINESITE	Cinesite Europe Ltd., UK
DIT	Dublin Institute of Technology, IE
DTS	Digital Theatre Systems, UK
FBM-UPF	Fundació Universitat Pompeu Fabra, ES
GVG	Grass Valley Germany, DE
JRS	JOANNEUM RESEARCH Forschungsgesellschaft mbH, AT
LFUI	Leopold-Franzenzs Universtät Innsbruck, AT
PGP	Pepper's Ghost Productions Ltd., UK
TAIK	Taideteollinen Korkeakoulu, FI
UG	University of Glasgow, UK
URL	Universitat Ramon Llull, ES