Beyond Multimedia Adaptation: Quality of Experience-aware Multi-sensorial Media Delivery

Zhenhui Yuan, *Member IEEE*, Gheorghita Ghinea, *Member IEEE*, and Gabriel-Miro Muntean, *Member IEEE*

***Abstract*—There is an increasing user interest in rich media content, including multiple sensorial media (mulsemedia). Mulsemedia combines multiple media elements which engage three or more of human senses, and as most other media content, requires support for delivery over the existing networks. This paper proposes an ADAptive MulSemedia framework (ADAMS) for delivering scalable video and sensorial data to users. Unlike existing two dimensional joint source-channel adaptation solutions for video streaming, the ADAMS framework includes three joint adaptation dimensions: video source - sensorial source - network optimization. Using an MPEG-7 description scheme, ADAMS recommends the integration of multiple sensorial effects (i.e. haptic, olfaction, air motion, etc.) as metadata into multimedia streams. The MPEG-7 description is compatible with MPEG codecs (i.e. MPEG-4) and therefore enables ADAMS deployment with existing video codecs. ADAMS designs both coarse- and fine-grained adaptation modules on the server side: 1) mulsemedia flow adaptation module; and 2) packet priority scheduling module. Feedback from subjective quality evaluation and network conditions are used to develop the two modules. Subjective evaluation investigated users' enjoyment levels when exposed to mulsemedia and multimedia sequences, respectively and to study users' preference levels of some sensorial effects in the context of mulsemedia sequences with video components at different quality levels. Results of the subjective study inform guidelines for an adaptive strategy that selects the optimal combination for video segments and sensorial data for a given bandwidth constraint and user requirement. User perceptual tests show how ADAMS outperforms existing multimedia delivery solutions in terms of both user perceived quality and user enjoyment during adaptive streaming of various mulsemedia content. In doing so, it highlights the case for tailored, adaptive mulsemedia delivery over traditional multimedia adaptive transport mechanisms.**

***Index Terms*—mulsemedia, quality of experience, subjective testing.**

I. Introduction

The latest rich media services including video streaming, voice over IP, video conferencing, on-line gaming, social networking, etc, require high bandwidth networks for their distribution to users. At the same time, the current network infrastructure has evolved towards a heterogeneous network environment in which wired, wireless, satellite, optical, etc. networks co-exist and support network content delivery based on various technologies and protocol families including the IEEE 802.3 (Ethernet), IEEE 802.11 (WiFi), IEEE 802.16 (WiMax), UMTS, LTE, etc. Noteworthy is also the very large diversity of devices, many of them mobile, which enable an increasing number of users to access the latest services over these networks. Despite the increased bandwidth availability, the exponential growth in the number of users, coupled with growing network resource requirements of the most popular applications, makes for an uphill battle to support high quality for these services. This is especially true for multimedia-based services, more sensitive to network delivery factors. Against this background, diverse solutions have been proposed to increase user perceived quality, including adaptive multimedia delivery schemes [1][2][3].

These solutions, however, in line with traditional multimedia applications, have only engaged two human senses: visual and audio. Existing multimedia services are limited in their ability to fully imitate the immersive scenarios and cannot provide an immersed sense of reality, which would have the potential to increase their perceived quality levels. For instance, when delivering traditional multimedia content, users cannot feel real environmental/ambiental elements such as scent of the flowers, air motion of the ocean wind, haptic effect of a push, etc.

Thanks to advanced computational technologies, it is now possible to deliver applications that engage other human senses, such as olfaction, touch, gustatory, etc. A new paradigm has been introduced to extend the traditional multimedia streams with additional components and is referred to as *mulsemedia* – multiple sensorial media – engaging more human senses than the two involved in multimedia [4][5]. As such, mulsemedia content consists of both traditional media objects (e.g. audio and video) and non-traditional ones such as olfaction, gustatory, haptic, temperature, humidity, and air motion, all of which target supplementary human sensorial inputs.

This paper proposes an **ADAptive MulSemedia delivery solution (ADAMS)** for end-user quality of experience enhancement. ADAMS recommends using MPEG-7-based coding [7] to integrate multiple sensorial effects (i.e. haptic, olfaction, air motion) into multimedia streams. Novel subjective tests are conducted to analyze users' enjoyment levels when exposed to mulsemedia and multimedia sequences, respectively and to study users' preference levels of some sensorial effects in the context of mulsemedia sequences with video components at different quality levels. By utilizing the results from these subjective tests, ADAMS was designed to perform adaptive mulsemedia streaming according to the user preferences in variable network conditions. A mulsemedia presentation tool was developed to present audiovisual media synchronized with olfaction, haptic, and air motion data. This system can be extended by including more human sensory-related media objects such as humidity, temperature, etc. Making use of this mulsemedia presentation tool, subjective experimental tests were performed and their results indicate how ADAMS provides high levels of user experience, especially in terms of enjoyment of sensorial effects, under highly loaded network conditions.

The paper is organized as follows. Section II reviews research on adaptive delivery of multimedia streams and existing mulsemedia work. Section III presents the subjective tests, including the test-bed setup, media sequences, scenarios and results analyses of the perceptual mulsemedia service delivery. Sections IV and V introduce ADAMS, the proposed adaptive mulsemedia delivery solution and the system design issues. Performance evaluation of the proposed scheme is presented in section VI, while section VII concludes the paper.

II. Related Work

A. Multimedia Applications

Multimedia data, unlike traditional media content that uses text only, refers to a combination of text, still images, animation, audio, and video. Most multimedia streaming protocols have been designed at different OSI layers in order to improve streaming performance and end user experience. An adaptive client-server multimedia streaming mechanism, the Quality-Oriented Adaptation Scheme (QOAS) was designed for the application layer [1]. The QOAS client application uses a Quality of Delivery Grading Scheme to evaluate the delivery quality by monitoring the transmission related parameters (such as packet loss, delay, jitter, late packet for play out rate) and estimate end user perceived quality. The QOAS server uses a Server Arbitration Scheme to analyze the received feedback reports and adjust the delivery of video stream by varying its quality. In [8], the Internet Engineering Task Force (IETF) developed a novel transport layer protocol referred to as Partial Reliable-Stream Control Transmission Protocol (PR-SCTP). It is an unreliable service mode extension of SCTP which differentiates retransmissions based on a reliability level that could be set dynamically. By using PR-SCTP, users can specify rules for data transmission. When a certain pre-defined threshold is reached, the sender abandons packet retransmission and sends the next incoming packet from the application layer. The reliability level is set based on different data types or the stream requirements. Other advanced multimedia streaming solutions have been also been proposed such as [9], [10], [11], [12], etc.

B. Olfaction

Olfaction - or smell – is one of the last challenges which multimedia applications have to conquer. Enhancing such applications with olfactory stimuli has the potential to create a more complex and richer user *mulsemedia* experience, by heightening the sense of reality and diversifying user interaction modalities. Nonetheless, olfaction-enhanced multimedia is a challenging research area, and this is reflected by the relative paucity of research.

Pioneering efforts were first carried out by Kaye [13][14]. His work played a significant role in creating an awareness of the issues, problems and limitations associated with the use of olfactory data, incidentally also serving as a good summary of olfaction incorporation in various applications and industries across the years. His work revealed that olfactory data are better suited for ambient displays of slowly changing, continuous information and that its use should rely on differences between smell rather than the intensity of a particular smell. He also distinguishes between different types of olfactory data output and thus discriminates between smell output to convey information, where the smell released is related to the information to be conveyed, which he calls *olfactory icons*, and smell output to provide an abstract relationship with the data it expresses, which he calls *smicons*.

One benefit of having information displays that are multi-modal and interactive in nature is to share attention and information processing demands between our different senses. Applications used to gain the users attention, more popularly known as notification or alerting systems, represent one of the areas in which olfactory data output has shown great potential. Kaye designed two such applications, *Smell Reminder*, which allows users to use smicons to create personal, notification alarms, and *Honey, I'm home,* an application shared between two people which ensures that out of sight, is not out of mind where smicons are used to alert the other that you are thinking of him/her [14]. Unfortunately, he does not report any detailed evaluation of these applications.

Bodnar et al. [15] also created a notification system that uses olfactory data. They conducted an experimental study to compare the effect of the use of visual, audio or olfactory displays to deliver notifications on a user's engagement of a cognitive task. Participants were given an arithmetic task to complete and at various intervals two types of notifications were triggered, one where the participants had to immediately stop what they were doing and record some data before returning to the completion of their task, and the other they were to ignore. With their experiment, they found that while olfactory notifications were the least effective in delivering notifications to end users, they had the advantage of producing the least disruptive effect on a user's engagement of a task.

In the realm of information processing, we mention the study carried out by Brewster et al. [16] in which they use olfactory data for multimedia content searching, browsing and retrieval, more specifically to aid in the search of digital photo collections. In their experiment, they compare the effects of using text-based tagging and smell-based tagging of digital photos by users to search and retrieve photos from a digital library. To achieve this, they developed an olfactory photo browsing and searching tool, which they called *Olfoto*. Smell and text tags from participants' description of photos (personal photographs of participants were used) were created and participants had to use these tags to put a tag on their photos. At a later date, participants then had to use the same tags to search and answer questions about the previously tagged photographs. The results of their experiment show that although the performance with the text-based tags was better, smell (and its ability to trigger memories in individuals) does have potential for being used as a querying method for multimedia content search.

Whereas the work presented so far has focused on the use of olfaction as an alternative to traditional output modalities, it must be said that relatively little work has explored the impact of olfactory data when integrated with other media objects. Of such efforts, most have been undertaken in the virtual reality field (VR), in applications ranging from education and training systems [17][18], to gaming [19] and have shown the potential success of olfaction-enhanced multimedia applications.

C. Haptics

Haptic user interfaces are relatively new, but have been actively applied to the domain of human-computer interaction in virtual environments since the early 1990s [20][32]. As such, haptic technology is widely used across a variety of domains, including medical, automotive, mobile phone, entertainment, controls, education, training, rehabilitation, assistive technology, and the scientific study of touch [20][21]. For example, Immersion Corporation, a company recognized worldwide for developing, licensing, and marketing haptic technology, reported that 2,000 medical simulators with haptic technology have been sold worldwide to hospitals and teaching institutions to train clinicians [21][22]. Haptic technology is, for instance, also embedded in mobile phones to enhance users' communication experience related to ringtones, games, messaging, alerts, dialing cues, and user interfaces for touch screen presses.

Today, haptic technology has become an important component of effectively accessing information systems. A haptic device interacts with virtual reality interfaces in which users are allowed to manipulate and obtain mechanical feedback (e.g., vibration) from three-dimensional objects (e.g., images and graphs). The haptic interface could be supported by a real-time display of a virtual environment where users explore by pushing, pulling, feeling, and manipulating the virtual objects with a device (e.g., a mouse or stylus) [23][22]. Users are thus able to experience simulations of various characteristics of the objects and the environment, such as mass, hardness, texture, and gravitational fields.

D. Mulsemedia

Incipient efforts in mulsemedia research have been forthcoming. For instance, there have been a few studies carried out to investigate the user-perceived experience associated with the use of the newer media objects such as tactile (touch) and olfactory media objects. However, because the use of these media objects is relatively new in the multimedia field, most of these perceptual studies have concentrated their efforts on the practicality and possibility of incorporating these media objects into these applications.

One such research effort is a virtual reality (VR) learning system called VIREPSE which provides both olfactory and haptic feedback [24]. An earlier mulsemedia VR learning environment from the same group of researchers was one in which research investigated the effect of olfaction on learning, retention, and recall of complex 3D structures such as organic molecules in chemical structures [25]. However, neither of the two studies report on any detailed evaluation of either of these applications, but rather focus their research efforts on discussing the significance of developing such mulsemedia virtual environments for education.

In related work, [26] describes an investigative study which explored the possibility of using a vibro-tactile device on the whole body for simulating collision between the user and a VR environment. Here, the effects of using a vibration feedback model (for simulating collision with different object materials), saltation, and simultaneous use of 3D sound toward spatial presence and perceptual realism, are tested. The results from their study revealed that their proposed vibro-tactile interface did enhance the sense of presence, especially when combined with 3D sound. It was, however, also discovered that the vibration feedback model was not significantly effective, and sometimes even hindered the correct sense of collision, but this was attributed to the limitation of the vibro-tactile device itself.

It is of little surprise that, because of the relative novelty of the mulsemedia combinations involved, the studies reviewed so far also explore user acceptance of these new media objects, a theme carried forward in more recent research [27], which looked at user perception and acceptance of olfactory media combined with the more traditional audio and video.

The researchers of the study reported in [28] present strategies and algorithms to model context in haptic applications that allow users to explore haptically objects in virtual reality/augmented reality environments. The results from their study show significant improvement in accuracy and efficiency of haptic perception in augmented reality environments when compared to conventional approaches that do not model context in haptic rendering. Indeed, the use of haptics in mulsemedia VR environments has very recently also been the subject of the research reported in [29].

In related work [29], researchers reported on a perceptual study carried out to establish an algorithm to provide high quality inter-media stream synchronization between haptic and audio (voice) media objects in a virtual environment. Indeed, synchronization seems to be a common theme across mulsemedia research. Thus, recent work has explored synchronization of olfactory media with audio-visual content [30], whilst [31] investigated synchronisation issues between different modalities, as well as the integration of video and haptics in resource constrained communication networks – a topic closely related to the work described in this paper.

Concluding, there is important interest in mulsemedia and its delivery to go beyond the state of the art. There is a need to propose an adaptive mulsemedia delivery scheme to improve user quality of experience levels when transmitting mulsemedia content over heterogeneous networks and such a solution has not been proposed so far.

III. Effect of Multi-sensorial Inputs on User Perception

A. Overview

This section investigates the effect of multi-sensorial inputs on user perception. Three types of sensorial effects (i.e. haptic, air, and olfaction) were integrated into sequences selected from two movies, creating mulsemedia content. User perception on played back movies and user enjoyment of the mulsemedia was studied with the help of a specially built test-bed and subjective tests.

B. Test-bed Description

The subjective tests were conducted in the Performance Engineering Lab at Dublin City University, Ireland (PEL@DCU) in a separate room with no outside disturbance. Testing conditions suggested in ITU-T R. P.910 [32] and ITU-T R. P.911 [33] were complied with and the single stimulus method was employed. The tests involved 16 users which included 9 males and 7 females. The subjective test was arranged according to a matrix shown in *Appendix* II [34].

TABLE I
ENCODING CHARACTERISTICS OF THE MULTIMEDIA CONTENT

Video sequence	Quality	Codec	Frame rate	Resolution	Bitrate (Kbps)
Jurassic Park	High	MPEG-4	30 fps	1280x720	2500
	Avg	MPEG-4	24 fps	853x480	1100
Back To The Future	High	MPEG-4	30 fps	1280x720	2500
	Avg	MPEG-4	24 fps	853x480	1100

TABLE II
SENSORIAL EFFECT DESCRIPTIONS ATTACHED TO THE MULTIMEDIA CONTENT FROM "JURASSIC PARK"

Motion	Video clip	Effects	Movie scenario	Olfaction Aroma
High	1	None	Animal attack	
	2	Haptic	Animal attack	
	3	Air	Wind as car moving fast	
	4	Olfaction	tear gas	Burnt
	5	Haptic, Air	Vehicle vibration and wind	
	6	Haptic, Olfaction	Animal attack and smoke	Burnt
	7	Air, Olfaction	Wind and fire	Burnt
	8	Haptic, Air, Olfaction	Vehicle vibration, wind and forest	Forest
Low	1	None	None	
	2	Haptic	Animal attack	
	3	Air	Subway train comes	
	4	Olfaction	Decomposed animal odor	Rubbish
	5	Haptic, Air	Pull by parasail and wind	
	6	Haptic, Olfaction	Air plane and crash	Methane
	7	Air, olfaction	Ocean wind and wine	Rock pools, Mulled wine
	8	Haptic, Olfaction, Air	Movement, gas and wind	Methane

TABLE III
SENSORIAL EFFECT DESCRIPTIONS ATTACHED TO THE MULTIMEDIA CONTENT FROM "BACK TO THE FUTURE"

Motion	Video clip	Effects	Movie scenario	Olfaction Aroma
High	1	None	None	
	2	Haptic	Car crash	
	3	Air	Wind	
	4	Olfaction	Smoke	Burnt
	5	Haptic, Air	Crash and wind	
	6	Haptic, Olfaction	Car crash and manure	Rubbish acrid
	7	Air, Olfaction	Wind and smoke	Burnt
	8	Haptic, Air, Olfaction	Car movement, wind and smoke	Burnt
Low	1	None	None	
	2	Haptic	Car crash	
	3	Air	Wind	
	4	Olfaction	Burning bread	Burnt
	5	Haptic, Air	Falling down and wind	
	6	Haptic, Olfaction	Sound waves and smoke	Burnt
	7	Air, olfaction	Smoke and wind	Burnt
	8	Haptic, Olfaction, Air	Car movement, fire and wind	Methane

The participants were from different backgrounds, e.g. engineering, education, finance, etc., in the 20-36 age range, with a mean of 26. All users initially took part in a pilot test in order to be familiar with the test operations. The instructions given to the participants and the personal information form to be filled are provided in *Appendix I* [34].

Each user was asked to watch 16 unique multimedia sequences taken from the movies "*Jurassic Park*" and "*Back To The Future*".

Fig. 1 Mulsemedia content including video, haptic, air, and olfaction (images from "*Back To The Future"* Courtesy of Universal Studios Licensing LLC)

(a) USB fan (b) Olfaction dispenser

(c) Haptic vest

Fig. 2 Multi-sensorial devices

Fig. 3 Mulsemedia perceptual test-bed

Each sequence was 30s long and was encoded at two different quality levels, namely 2.5 Mbps and 1.1 Mbps, which differed in terms of both frame rate and resolution and were labeled "*High*" and "*Avg*". The encoding characteristics of the movie sequences are shown in Table I. MPEG-4 AVC video and AAC audio compression are used in conjunction with an MP4 container. Three sensorial effects (haptic, air, and olfaction) were integrated into the 16 multimedia clips according to the sequence content scenarios, as given in Table II and Table III. Fig. 1 illustrates the video content and lists the sensorial affect added to different sequences from the two movies. For each of the two movies, there were four video clips with high motion content, and four video clips with low motion content; further each video clip was encoded at both high and low quality levels. In this paper, high and low motion refers to video content which changes rapidly or slowly, affecting the process of motion prediction and motion vector computation. For instance, action and sports movies are typical high motion content videos, whereas talk shows and

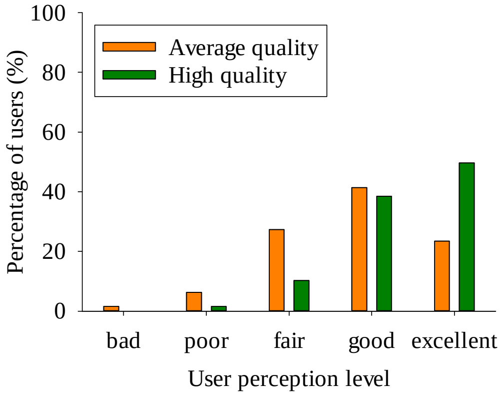

Fig. 4 User quality perception of the multimedia content

Fig. 5 User enjoyment of the mulsemedia content

news are typical low motion videos. These mulsemedia clips were shown to users in a random order according to the algorithm presented in *Appendix II* [34].

Fig. 2 presents the devices which provide the three sensorial effects: a USB fan for air, an olfaction dispenser for smell, and a haptic vest for vibrations. The duration of haptic and olfaction effects were determined based on the actual movie content and ranges from 1s to 3s. Fig. 3 shows the picture of the mulsemedia delivery test-bed. Users were asked to complete a paper questionnaire (presented in *Appendix III* [34]) which was given to them before the tests. The time interval between every two users was around one hour in order to fully refresh the test room (i.e. windows were opened), as otherwise the scent lingered in the air.

C. Result Analysis

In this section, the responses to the questionnaires received from the 16 users are summarized. The most relevant statistics for each questionnaire item are as follows.

1. *The sensorial effects enhance the video content.*
   53.6% of users agree and 31.4% strongly agree, therefore, **85%** of users tend to agree.
2. *The sensorial effects are annoying.*
   39.7% and 41.5% of users strongly and slightly disagree, respectively; therefore, **81.2%** of users tend to disagree.
3. *The sensorial effects improve the sense of reality when watching the video.*
   47.5% of users agree and 36.1% of users strongly agree, therefore, **83.6%** of users tend to agree.
4. *The sensorial effects are distracting.*
   37.1% and 35.5% of users strongly and slightly disagree, respectively; therefore, **72.6%** of users tend to disagree.
5. *I enjoyed the experience.*
   41.9% of users agree and 45.9% of users strongly agree, therefore, **87.8%** of users tend to enjoy.
6. *Which sensorial effect do you prefer (or you like the best)?*
   62.5% of users prefer haptic, 31.25% of users prefer air, 6.25% of users prefer olfaction.

User perceptions on both high and average quality multimedia traffic are summarized in Fig. 4. It is shown that the large majority of users rates “avg” and “high” quality multimedia sequences good (41.4%/38.5%) and excellent (23.4%/49.7%), respectively. Additionally, user enjoyment levels for the mulsemedia content are shown in Fig. 5. The results demonstrate that the majority of users (76.3%/84.4%) agree that regardless of the video quality level, mulsemedia content increases user enjoyment.

D. Test Conclusions

The following conclusions can be drawn by looking at the results from Fig. 4 and Fig. 5:

1. The higher quality multimedia sequences result in higher overall user quality of perception levels.
2. When delivering mulsemedia content, there is no statistical difference between user enjoyment levels when exposed to “avg” and “high” quality sequences, respectively.
3. There is a definite user preferred degree in terms of multi-sensorial effects: haptic effects are preferred by the majority of users; this is followed by air effects, whilst olfactory effects are least popular.

Additionally, by analysing the questionnaires note that:

1. Synchronization between sensorial effects and multimedia content needs to be precise, especially when olfaction is included.
2. Unpleasant smells such as methane and rubbish annoy the users and result in reduced user enjoyment levels.

In conclusion, in the absence of multi-sensorial inputs, the large majority of users noticed the difference in multimedia quality. However, user enjoyment levels were maintained high when lower multimedia quality sequences were used in conjunction with multiple sensorial effects. Fig.6 illustrates the principle of the ADAMS system; its adaptation strategy takes advantage of the fact that multi-sensorial effects partly mask decreases in video quality. In terms of sensorial stimuli, there is a clear preference for **haptic, air and olfaction in this order**. This work does not consider, but future work can focus on, the effects of either multi-sensorial input synchronization and/or other pleasant olfaction stimuli on the user perceptual quality, extending earlier work on the subject [26] [30].

IV. Adaptive Mulsemedia Delivery Solution (ADAMS)

A. Solution Overview

In the context of an increasing amount of data traffic communication networks are often subject to very high loads. These affect the service quality of the delivered multimedia content. Existing content adaptation solutions have considered making multimedia content adjustments dynamically [1] (these mostly affect video, the largest component) to match the transferred content bitrate to the available bandwidth and

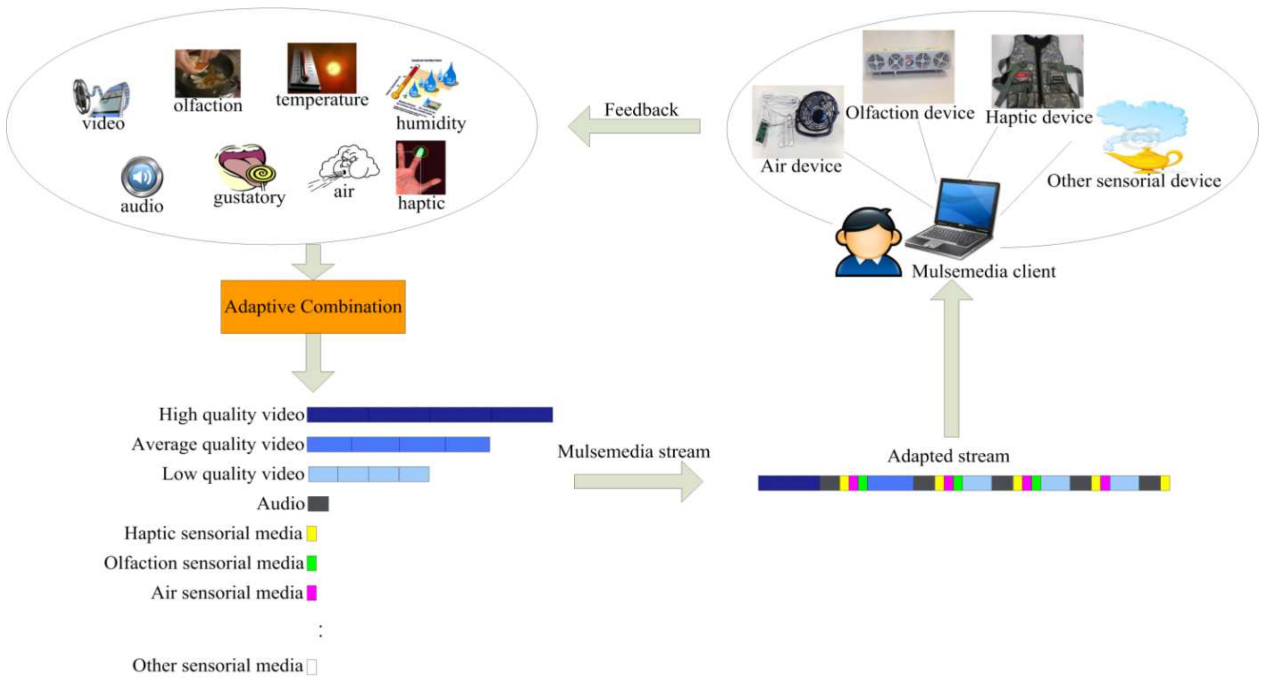

Fig.6 Illustration of the principle of ADAMS system

decrease the loss rate. Despite the adaptation efforts, the reduction in encoding multimedia quality is observed and the end-user perceived quality decreases. However, mulsemedia perceptual tests described in section III have shown that in the presence of additional sensorial inputs, the overall user quality experience is higher than in their absence during adaptive multimedia content delivery. Consequently this section introduces a novel ADAptive MulSemedia delivery solution (ADAMS) for end-user quality of experience enhancement, which considers multi-sensorial content in the network-based content delivery adaptation process.

Fig. 6 illustrates a scenario in which ADAMS performs adaptive mulsemedia content delivery to an end user. On the left side the ADAMS server selects content and/or metadata related to a number of sensorial media types. These include video, audio, olfaction, haptic, air, temperature, humidity, etc. In general these media types are meant to excite various components of the human sensory system (e.g. sight, smell, touch, etc.). Following the adaptive selection process, the adapted content is delivered to the ADAMS client at the remote multi-sensorial user in chunks. Feedback informs the server about both network delivery conditions and user preferences (if any) and ADAMS adjusts the multi-sensorial content delivery process accordingly. The illustration shows that following negative feedback, the video component is sent at lower and then the lowest quality levels available, without any alteration in the other sensorial components. When feedback information continues to suggest loaded delivery conditions, sensorial content is dropped in inverse order of user preference (i.e. olfaction, air and haptic), before the video is eventually dropped and audio only is delivered.

Thus, the ADAMS adaptation algorithm extends the Quality-Oriented Adaptive Scheme (QOAS)'s [1] classic video quality adjustment process with a second stage of adaptation of the sensorial components according to user interest levels. In this manner ADAMS's mulsemedia-aware adaptation truly benefits from the multidimensionality of the solution space and improves the user multisensorial experience. This multidimensionality was not taken into consideration when QOAS was proposed, as QOAS has performed linear adaptation of the video content only.

B. *ADAMS Architecture*

Fig. 7 presents the block-level architectural of the proposed scheme ADAMS, which involves a feedback-based client-server approach. During the content delivery sessions, the ADAMS server exchanges multi-sensorial data in the server ~~to~~ with the ADAMS client, which, in turn passes feedback information back to the server. ADAMS specific information processing is performed in the hashed blocks, whereas the other blocks employ already existing solutions.

The **ADAMS server** is composed of five major blocks. The *ADAMS Adaptation Module* gets regular feedback information from the ADAMS client and based on the received quality of delivery scores, takes multi-sensorial media adaptation decisions according to the ADAMS adaptation algorithm. The ADAMS adaptation algorithm is implemented in two sub-modules: **Mulsemedia Flow Adaptation (MFA)** and **Packet Priority Scheduling (PPS)**. The Multi-sensorial Data and Metadata block stores the relevant content and associated information in order to be able to perform the delivery. The MPEG-7-enabled encoder puts together the selected and transcoded multi-sensorial components into a mulsemedia presentation ready for delivery. The delivery to the client is performed by the *Packet Delivery* Unit.

The *MFA* module provides flow-based coarse-grained adaptation which transmits proper multi-sensorial content and performs video content transcoding if required, according to client feedback. The feedback includes both network conditions and user profile (i.e. priority level of sensorial effects). The network conditions are indicated using off-the-shelf bandwidth estimation techniques, such as the Model-

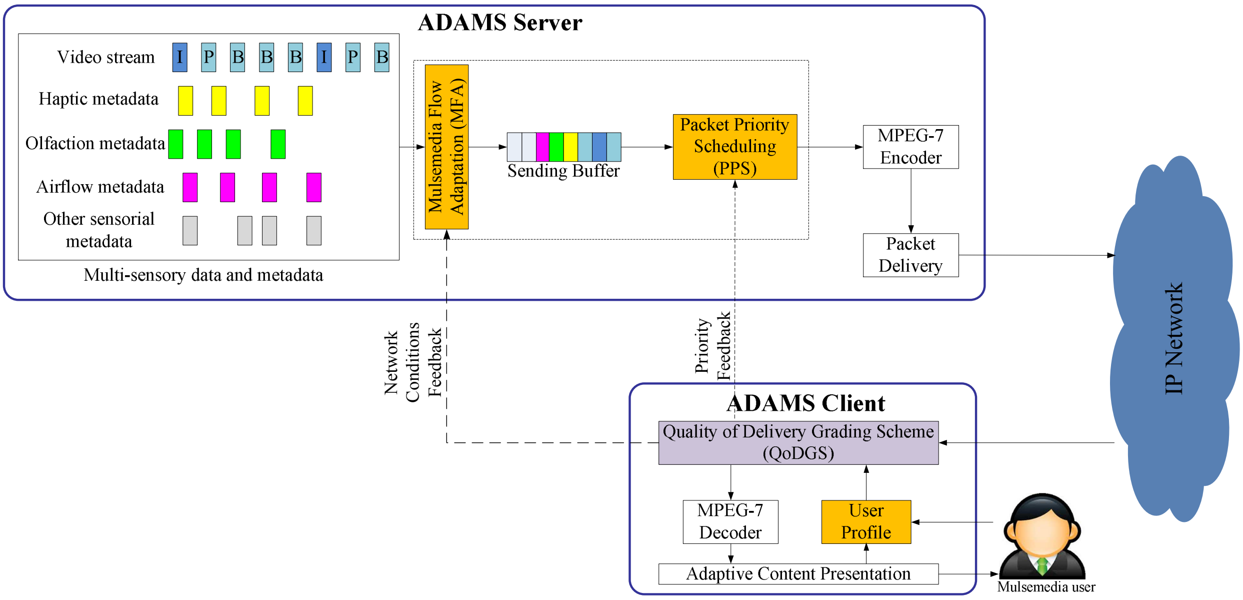

Fig. 7 ADAMS block-level architecture (QoDGS was introduced in QOAS [1])

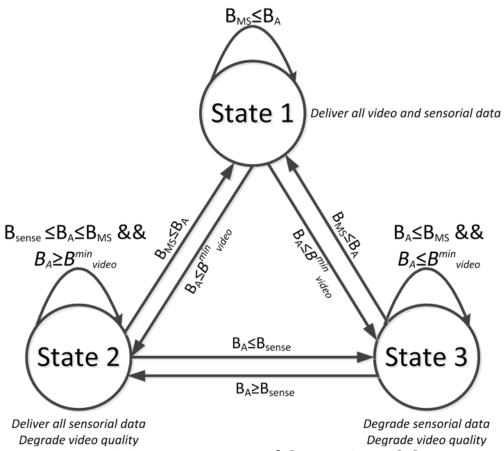

Fig. 8 States transition of the MFA module

based Bandwidth Estimation (MBE), introduced in our previous paper [36]. Other bandwidth estimation techniques are also reported that are outside the scope of this paper (e.g., as in [37] [38]). MBE computes the estimated bandwidth using two three parameters: number of mobile stations, packet loss, and packet size. Equation (1) gives the computation of estimated available bandwidth (B_A) for TCP flows based on MBE. The parameter *b* is the number of packets acknowledged by a received ACK, P_{retr} denotes the probability of packet retransmission, *MRTT* is the transport layer round-trip time between sender and receiver, and *MSS* is the maximum segment size. T_o is the timeout value used by the congestion control. The estimated available bandwidth for UDP flows used in this paper is also given in [36].

$$B_A = \frac{MSS}{MRTT \times \sqrt{\frac{2bP_{retr}}{3}} + T_o \times \min(1, 3\sqrt{\frac{3bP_{retr}}{8}}) \times P_{retr} \times (1+32P_{retr}^{\ 2})} \quad (1)$$

User profiles are configured and updated by the ADAMS client. *MFA* involves three states to perform mulsemedia flow adaptation. Let B_{MS}, B_{sense}, and B_{video} represent the bitrate of mulsemedia flow, sensorial data flow and video flow, respectively. B_{MS} is defined in equation (2).

$$B_{MS} = B_{sense} + B_{video} \quad (2)$$

MFA maintains a *state* parameter to dynamically control the *MFA* process according to network conditions. Three states are considered in the design of *MFA*, as illustrated in Fig .8.

1) The first state (*State 1*) is active if $B_{MS} \le B_A$. *State 1* indicates that the available bandwidth is enough to deliver both video and sensorial data flows and there is no need to perform any content quality adaptation.

2) The second state (*State 2*) is active if $B_{sense} \le B_A \le B_{MS}$ and $B_A \ge B^{min}_{video}$ where B^{min}_{video} is the bandwidth threshold associated with good video quality level. In *State 2* the available bandwidth is between the bitrate of the sensorial data flow and the bitrate of the video flow and therefore the video flow is adapted (i.e. involves quality reduction and therefore bitrate decrease) while all the sensorial data flows are still transmitted. ADAMS adjusts the video bitrate to meet the available network bandwidth following the feedback reports. This is based on an additive increase-multiplicative decrease policy and on N granularity quality levels defined in inverse order of video quality. Each such quality level is defined in terms of a triplet <*resolution, frame rate, color depth*>, directly related to a video bitrate value. When increased traffic in the network affects the client-reported QoDGS grades - QoDGS, the *Quality of Delivery Grading Scheme* will be described in more detail when presenting the ADAMS client later in this section - ADAMS switches fast to a lower quality level and accordingly adjusts the values of some of the triplet's components. This action results in a reduction in the bitrate of the video sent, easing the pressure on the network and helping it to recover from congestion. This eventually determines lower loss rates and consequently better end-user perceived quality. In improved delivery conditions as reported in terms of QoDGS scores, ADAMS cautiously and gradually increases the transmitted video quality level and therefore

improves the values of some of the triplet's components. In the absence of loss this determines an increase in end-user perceived quality.

3) The third state (*State 3*) is active if $B_A \leq B_{MS}$ and $B_A \leq B^{min}_{video}$. *State 3* indicates that the available bandwidth has reached very low values and therefore, the video flow is degraded as indicated in *State 2*. Additionally, following delivery quality feedback reports, ADAMS removes sensorial media components from the mulsemedia stream, in inverse order of user interest in their corresponding sensorial effects. This decision is taken based on user profile information if it includes user preference for some sensorial media objects, or explicit user feedback. When such information is not available, a default preference order is assumed. Section III has shown a definite preference of the test subjects for haptic, air motion and olfaction effects, respectively, in this order.

The *PPS* module provides packet-based fine-grained adaptation using a priority model which specifies that packets with higher priority are scheduled earlier than those with lower priority. The priority model is derived based on the results of the previously described subjective tests, detailed in section III C. It was hypothesized that a lower quality video sequence integrated with mulsemedia effects is capable of producing as good a user experience as that of a higher quality video sequence.

Let W_v, W_h, W_o, and W_a denote the weight factors associated with the priority levels of video, haptic, olfaction, and air-flow data packets, respectively. According to results from the subjective tests, on average 63%, 31%, and 6% of users prefer haptic, air-flow, and olfaction sensorial effects, respectively. Let W_v, W_h, W_o, and W_a denote the weight factors associated with the priority levels of video, haptic, olfaction, and air-flow data packets, respectively. The importance or priority of each sensorial effect is normalized according to the ratio in equation (3). This might not be the perfect model for the priority levels of these sensorial data packets, but it initializes the mulsemedia adaptive system using low complexity computation and based on the average opinions of the subjects tested. To the best of our knowledge, this is the first equation that models the relationship between haptic, olfaction, and air-flow in terms of human preferences, and is incorporated in the ADAMS adaptation strategy. Future work will extend equation (3) to improve the solution in terms of flexibility and scalability.

$$W_h : W_a : W_o = 0.63 : 0.31 : 0.06 \qquad (3)$$

Additionally, the subjective tests in section III show that the user enjoyment levels were maintained high when lower multimedia quality sequences were used in conjunction with mulsemedia effects. Naturally, we assign sensorial data packets an equal or higher priority level than that of the video packets in terms of the user-perceived experience. Based on the results from the subjective tests, it can be concluded that sensorial data packets have equal or higher priority level (in terms of the impact on user perception) than that of the video packets. Therefore, equation (4) is derived to describe the priority relationship between these sensorial data packets.

$$\{W_h, W_o, W_a\}_{\min} \geq W_v \qquad (4)$$

Equation (4) is a general approximation of the priority model between sensorial packets (i.e. haptic, olfaction, air-flowing) and video packets. In order to obtain the initial values of the weighted factors for different packet types, it is assumed that olfaction packets have the same priority with the video packet, which results in W_o equals W_v. This assumption is supported by the fact that, in terms of user perception, olfaction data has lower priority than both haptic and air-flow data. According to equation (4), by normalization, W_h, W_a, W_o, *and* W_v values are 0.595, 0.293, 0.056, and 0.056, respectively.

The probability of scheduling the next packet in the queue is computed by equation (5), which takes into account both packet priority and flow bitrate. Parameters *i* and *j* refer to the i^{th} packet of flow *j* in the queue and *N* is the number of queued packets. $Bitrate_j$ denotes the bitrate of the j^{th} flow. The value of packet weight factor W_i^j is set based on the packet type (i.e. video, haptic, olfaction, air-flow). For instance, if the i^{th} packet is a haptic packet, then W_i^j equals W_h which is 0.293 according to the previously described default configuration.

$$P_i = \frac{W_i^j \times Bitrate_j}{\sum_i^N W_i^j \times Bitrate_j} \qquad (5)$$

The **ADAMS client** consists of four major blocks. When it receives the multi-sensorial content via the network, the *MPEG-7 Decoder* gets the multi-sensorial media components and passes them to the *Adaptive Content Presentation* which performs synchronized presentation of the various content items. Apart from the regular screen and speakers necessary to present multimedia content, this unit makes use of various devices, such as haptic vests, fans, smell releasing devices, heaters, etc. for presentation of other sensorial effects. The client maintains a *User Profile* in order to enable both automatic feedback gathering and explicit (if users desire to provide) in terms of user multi-sensorial adaptive preferences. The performance of network delivery is assessed by the *Quality of Delivery Grading Scheme (QoDGS)*, which has been implemented in QOAS [1]. QoDGS maps quality of service related parameters such as loss, delay and jitter and their variations and estimations of viewer perceived quality on application level scores that describe the quality of the delivery session. This delivery quality is monitored over both short-term and long-term. Short-term monitoring is important for learning quickly about transient effects, such as sudden traffic changes, and for quickly reacting to them. Long-term variations are monitored in order to track slow changes in the overall delivery environment, such as new flows over the network. These short-term and long-term periods are set to be an order and two orders of magnitude (respectively) greater than the feedback-reporting interval (e.g. for 100 ms inter-feedback intervals, these would be 1s and 10s, respectively).

V. ADAMS: Tailored for Mulsemedia Delivery

This section follows the ADAMS architecture introduction in section III and presents several key issues in designing the ADAMS system. Besides the adaptation algorithm itself, ADAMS concerns three critical aspects, which are addressed next:

1) Mulsemedia data packet header;
2) Data combination of diverse types of sensorial data;
3) Mulsemedia components synchronisation.

TABLE IV
PACKET HEADER DESCRIPTION OF SENSORIAL DATA

Name	Size	Description
sequence number	2 byte	Used for identify the packet. The sequence number increases by one for each sensorial data packet sent.
type	1 byte	Type of the sensorial effects, e.g. olfaction, air, haptic, etc.
intensity	1 byte	The intensity of the sensorial effects, e.g. strong, medium, weak, etc.
start time	4 bytes	Start time of the sensorial effect when synchronized with video
duration	4 bytes	The duration of the sensorial effects. Example: how long the olfaction last.
option	4 bytes	Extensible by users

A. *Mulsemedia Data Packet Header*

In order to create and deliver the sensorial packets in IP-based networks, a new packet header for sensorial data needs to be defined. A typical way to transmit mulsemedia data is to first create mulsemedia packets using the mulsemedia data packet header and then encapsulate these mulsemedia data packets into an existing codec (e.g. MPEG-4/7). MPEG packets can then be multiplexed and streamed over the IP networks. The new packet header for mulsemedia data is designed and the description of each header filed is given in Table IV. The header size is 16bytes.

B. *Mulsemedia Data Combination*

Fig. 9 illustrates the hierarchical organization of diverse media components employed for mulsemedia data combination including video, audio, olfaction, gustatory, haptic, etc. These components are represented in terms of metadata only or both metadata and content data. Metadata representation is enough to describe most sensorial effects which will be reproduced at remotely located devices, following mulsemedia network delivery. The metadata associated with the different sensorial media components have similar entries which, for each sensorial effect, identify its *start time*, *duration*, and *intensity*. Some metadata differs due to some specific sensorial characteristics such as *flavor* for the gustatory effect, *direction* for air motion, and *scent type* for olfaction. Video and audio components are very well known and require, apart from the metadata, also the presence of the actual content data, which will be decoded and presented remotely.

The sensorial media data is sent as metadata separately to the client. In parallel with the video stream, the sensorial data stream delivers the control command information (e.g. duration, strength, types, etc) to manage the end user's sensorial devices. The client then synchronizes the sensorial effects to the actual content by activating/stopping the associated sensorial devices.

All of the sensorial meta data can be conveniently expressed using already accepted standards such as MPEG-7 [7] and Fig. 10 shows how one can use the MPEG-7 framework to define Mulsemedia Description Schemes and (in this particular case) olfactory media types.

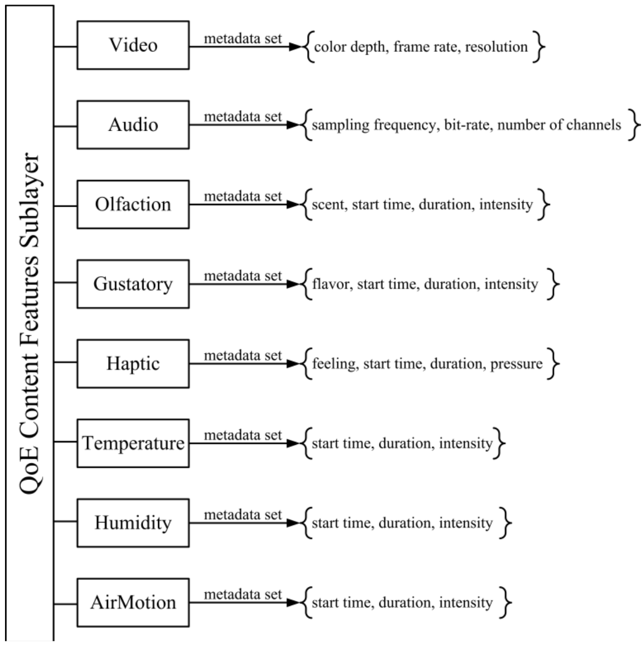

Fig. 9 Mulsemedia components description

C. *Sensorial Media Synchronization*

The purpose of synchronization is to achieve the desired temporal relationship between the various sensorial media objects, all part of the mulsemedia stream. There is the natural desire to record zero ***intermedia skew*** between different mulsemedia components for best user quality of experience levels. For instance, a zero skew between the visual stream and haptic stream would indicate a perfect temporal relationship.

In order to help achieve this excellent inter-media synchronization, the metadata associated with all human sensing-related media objects considered part of the mulsemedia stream (i.e. olfaction, gustatory, haptic, temperature, humidity, and air motion) includes two independent features, start time and duration, which help control the synchronization during presentation. However, unlike the traditional multimedia components (i.e. audio and video), sensing-related media objects might cause unexpected user perception effects. For instance, the perceived duration of olfaction, gustatory, temperature, and humidity-data may be less or greater than the intended duration, mostly due to effects such as propagation and lingering. Employing solutions such as adding constant offsets, allowing larger inter-media time intervals solves some of these problems as demonstrated by olfaction-video synchronization research [30] [35].

D. *Conclusions*

This section has introduced ADAMS as an adaptive scheme for mulsemedia content delivery, which adjusts the various sensorial media content according to feedback-reported network delivery conditions. As opposed to traditional multimedia adaptive delivery solutions, the design of ADAMS's adaptation algorithm was informed by both user mulsemedia subjective tests and delivered video quality in loaded network conditions.

```xml
<complexType name="MulsemediaSegmentMediaSourceDecompositionType">
    <complexContent>
        <extension base="mpeg7:MediaSourceSegmentDecompositionType">
        <choice maxOccurs="unbounded">
            <element name="VideoSegment" type="mpeg7:VideoSegmentType"/>
            <element name="VideoSegmentRef" type="mpeg7:ReferenceType"/>
            <element name="AudioSegment" type="mpeg7:AudioSegmentType"/>
            <element name="AudioSegmentRef" type="mpeg7:ReferenceType"/>
            <element name="OlfactionSegment" type="mpeg7:OlfactionSegmentType"/>
            <element name="OlfactionSegmentRef" type="mpeg7:ReferenceType"/>
            <element name="HapticSegment" type="mpeg7: HapticSegmentType"/>
            <element name="HapticSegmentRef" type="mpeg7:ReferenceType"/>
            <element name="AirSegment" type="mpeg7: AirSegmentType"/>
            <element name="AirSegmentRef" type="mpeg7:ReferenceType"/>
            <element name="GustatorySegment" type="mpeg7: GustatorySegmentType"/>
            <element name="GustatorySegmentRef" type="mpeg7:ReferenceType"/>
            <element name="TemperatureSegment" type="mpeg7: TemperatureSegmentType"/>
            <element name="TemperatureSegmentRef" type="mpeg7:ReferenceType"/>
            <element name="HumiditySegment" type="mpeg7: Humidity        SegmentType"/>
            <element name="HumiditySegmentRef" type="mpeg7:ReferenceType"/>
        </choice>
        </extension>
    </complexContent>
</complexType>
<complexType name="OlfactionSegmentType">
    <complexContent>
        <extension base="mpeg7:SegmentType">
            <sequence>
                <choice minOccurs="0">
                    <element name="MediaTime" type="mpeg7:MediaTimeType"/>
                    <element name="TemporalMask" type="mpeg7:TemporalMaskType"/>
                </choice>
                <choice minOccurs="0" maxOccurs="unbounded">
                    <complexContent>
                        <extension base="mpeg7:MultimediaContentType">
                        <element name= "Component">
                            <simpleType>
                                <restriction base="string">
                                    <enumeration value="burnt"/>
                                    <enumeration value="forest"/>
                                    <enumeration value="rubbish"/>
                                    <enumeration value="mulled wine"/>
                                    <enumeration value="methane"/>
                                    <enumeration value="strawberry"/>
                                    <enumeration value="wallflower"/>
                                </restriction>
                            </simpleType>
                        </element>
                        </extension>
                        <element name=”Intensity”>
                            <simpleType>
                                <restriction base="string">
                                    <enumeration value="low"/>
                                    <enumeration value="medium"/>
                                    <enumeration value="high"/>
                                </restriction>
                            </simpleType>
                        </element>
                    </complexContent>
                </choice>
            </sequence>
        </extension>
    </complexContent>
</complexType>
```

Fig. 10 MPEG-7-based Mulsemedia Description Scheme

VI. ADAMS - User Perceptual Testing

ADAMS – the adaptive delivery solution for mulsemedia content is evaluated in this section via user perceptual tests. Multimedia content was delivered to a mobile user over a wireless LAN experiencing growing congestion levels. The congestion was simulated by increasing the number of transmission flows of the two typical transport layer protocols UDP and TCP. The performance of ADAMS is evaluated in comparison with a classic adaptive multimedia delivery scheme QOAS [1] and a non-adaptive scheme in typical network delivery conditions. The test-bed, test conditions, subjects and scenarios are presented and the test results are analyzed next.

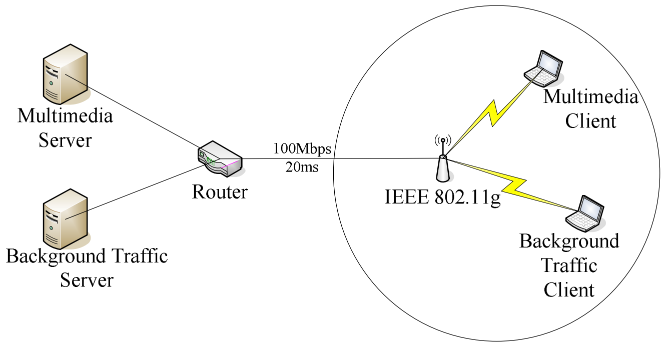

Fig. 11 Test topology (simulation)

TABLE V
DOWNLINK AND UPLINK TRAFFIC CHARACTERISTICS

	TCP		UDP	
	Packet size (bytes)	Transmission rate (Kbps)	Packet size (bytes)	Transmission rate (Kbps)
downlink	1000	512-1024	1200	1024
uplink	300	256-512	500	512

TABLE VI
NUMBER OF BACKGROUND TRAFFIC FLOWS

Time (s)	Number of flows	Number of downlink flows		Number of uplink flows	
		TCP	UDP	TCP	UDP
0-20	0	0	0	0	0
20-80	6	2	2	1	1
80-140	12	4	4	2	2
140-200	18	6	6	3	3
200-260	24	8	8	4	4
260-320	30	10	10	5	5

A. Test-bed, Test Conditions and Test Subjects

In order to enable fair reference to the subjective tests shown in section III, the same test-bed was used in these tests. The tests took place in the same location under the same conditions, already described. None of the users had participated in the initial round of tests described in section III, so they were not familiar with mulsemedia testing. Each user was asked to watch multiple mulsemedia clips and experience the associated integrated sensorial effects. The mulsemedia content was composed of multimedia objects (audio and video) and other sensorial components.

The impact of network congestion on video quality levels was studied using the Network Simulator version 2 (NS2). The simulation test-bed used the wired-cum-wireless "dumbbell" topology illustrated in Fig. 11. The scenarios involved a wireless client receiving video traffic from a video server over WLAN via an IEEE 802.11g access point (AP). Background traffic was delivered from a dedicated server to a background traffic client, in order to increase the load on the wireless network. The video server and background traffic server were connected to the AP through one router and the wired link between the router and the AP was overprovisioned (100Mbps bandwidth and 20ms propagation delay), so that the IEEE 802.11g WLAN remained the only bottleneck link on the end-to-end path.

The background traffic was generated according to Table V and Table VI. The video transmission time is set to 320s. During the first 20s, there was no background traffic. From 20s to 320s, the number of background flows was gradually increased from 6 to 30, with 6 new flows added every 60s. The background traffic consisted of UDP and TCP flows which were implemented by model agents provided by NS2. The UDP agents carried traffic generated by Constant Bit Rate (CBR) applications and the TCP agents transported File Transfer Protocol (FTP) application traffic. CBR and FTP models were also provided by NS2. As the transmission bit-rate of TCP flows was variable, as detailed in Table II, the TCP sending rate was adjusted by changing the size of the receiving window.

B. Mulsemedia Synchronization

For proof of concept testing purposes, synchronization between the sensorial components and the multimedia content was achieved manually according to the sensorial events timeline (i.e. times when the sensorial effects occur in the video stream). This is an ideal implementation in which the sensorial events were perfectly synchronized with the video scenario, namely, the inter-media skew was zero. The equipment and software used to synchronize the media objects are shown in Figure 2 and include three devices which generate the sensorial effects: a haptic vest, an air fan and an olfaction dispenser. A C++ software developed to control these devices uses as input the multi-sensorial timeline. The haptic effects were generated by the vest which provided fully programmable control of the haptic effects in terms of intensity levels, types, and duration. The USB fan provided the air-flow effects and can be controlled to generate strong, medium, and weak levels of air-flow and be turned on/off via a program. Olfactory stimuli were released from the dispenser, which uses four miniature fans to respectively emit the scents contained in its four cartridges. There is a wide variety of scents to choose from and each fan was programmable by a dedicated on/off control.

C. Scenarios and Assessment

In order to evaluate the performance of ADAMS, three separate test scenarios were designed, as shown in Table VII. The multimedia clips have both high/low motion intensity and high/medium/low quality levels from two movies: <<Jurassic Park>> and <<Back To The Future>>. Similar to the previous subjective tests, the performance is assessed in terms of: 1) user perception of the multimedia content; 2) user enjoyment experience for the mulsemedia clip.

The three test scenarios are described next: 1) *Scenario A - Non-adaptivity*. High quality multimedia clips are shown to users as a result of their delivery using a non-adaptive scheme in various network conditions. The quality level of each clip was affected by the increased network congestion. The sensorial effects were maintained unchanged; 2) *Scenario B - Multimedia adaptivity*. High/medium/low quality multimedia clips were presented to users following their adaptive delivery using QOAS in increasingly loaded network conditions. The sensorial effects were unmodified; 3) *Scenario C - Mulsemedia adaptivity*. The default order for sensorial effects adaptation was employed, as given by the results detailed in

TABLE VII
SUBJECTIVE TEST SCENARIOS

Number of background traffic flows	Scenario A (Non-Adaptivity)		Scenario B (Multimedia Adaptivity)		Scenario C (Mulsemedia Adaptivity)	
	Video quality	Sensorial effects	Video quality	Sensorial effects	Video quality	Sensorial effects
0	High	Haptic, Air, Olfaction	High	Haptic, Air, Olfaction	High	Haptic, Air, Olfaction
6	High	Haptic, Air, Olfaction	High	Haptic, Air, Olfaction	High	Haptic, Air
12	High	Haptic, Air, Olfaction	Medium	Haptic, Air, Olfaction	Medium	Haptic, Air, Olfaction
18	High	Haptic, Air, Olfaction	Medium	Haptic, Air, Olfaction	Medium	Haptic, Air
24	High	Haptic, Air, Olfaction	Low	Haptic, Air, Olfaction	Low	Haptic, Air
30	High	Haptic, Air, Olfaction	Low	Haptic, Air, Olfaction	Low	Haptic

TABLE VIII
USER GROUPS SETUP FOR SUBJECTIVE TEST SCENARIOS

	User Group 1				User Group 2				User Group 3			
	Case 1	Case 2	Case 3	Case 4	Case 1	Case 2	Case 3	Case 4	Case 1	Case 2	Case 3	Case 4
Movie	JP	JP	BF	BF	JP	JP	BF	BF	JP	JP	BF	BF
Motion	High	Low	High	Low	High	Low	High	Low	High	Low	High	Low
Scenario	A	B	A	B	B	C	B	C	C	A	C	A

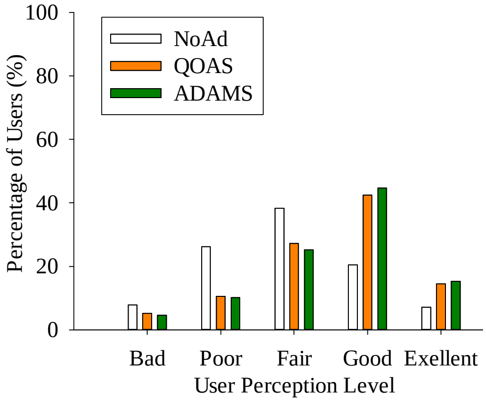

Fig. 12 User perception level for the multimedia delivery affected by network condition

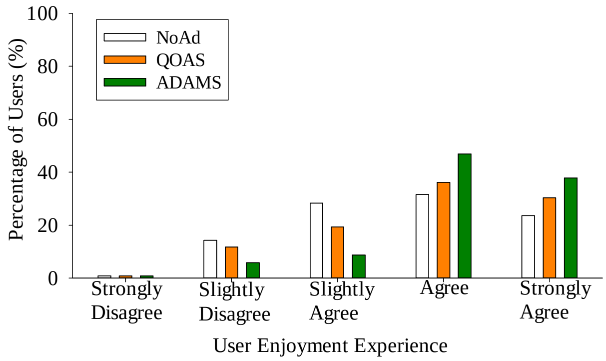

Fig. 13 User enjoyment experience for the mulsemedia delivery affected by network condition

section III. For instance, when high network congestion is noted, olfaction is removed before other sensorial effects.

Although ADAMS allows users to indicate their adaptation preference, for simplicity, the users were not asked to specify a preference for certain sensorial effects.

The 16 users were divided into three groups as shown in Table VIII. Each user group includes four test cases involving different combinations of movie type, motion intensity, and scenario. A user belonging to a certain user group was asked to complete all the four test cases. The test time duration for each user was roughly 15 minutes.

D. Result Analysis

Fig. 12 presents the user perception for multimedia components when the increased background traffic was delivered. It is clear that, by using QOAS, the percentage of “Good” and “Excellent” levels increases by 22% and 7.4%, respectively, in comparison with that of the non-adaptive scheme. QOAS results in less video distortion and therefore better received video quality. This is consistent with the results from the objective tests in section IV. Additionally, comparing usage of ADAMS and QOAS, no statistically significant difference between the two adaptive schemes in terms of user perception levels was noted. This indicates that the reduction of some sensorial effects in the mulsemedia delivery has no negative impact on user perception of the multimedia component.

Fig. 13 presents user enjoyment results when the mulsemedia content was delivered in increased background traffic conditions. The results demonstrate that both multimedia and mulsemedia adaptive schemes improve the user enjoyment experience. For instance, in comparison with the non-adaptive scheme, QOAS and ADAMS increase the percentage of “Strongly Agree” answers by 6.7% and 14.2%, respectively. Additionally, ADAMS outperforms QOAS, as the percentage of users enjoying their experience in the “Agree” and “Strongly Agree” categories has increased by 10.7% and 7.5%, respectively. This is because ADAMS reduces both the number of sensorial effects and multimedia quality level, saving bandwidth. Additionally, the amount of sensorial effects was decreased according to the users’ preference level, as determined through the user subjective mulsemedia tests described in section III, which gracefully reduced the negative impact on user enjoyment levels.

E. Conclusions

Following mulsemedia adaptivity testing, it can be stated that ADAMS, the proposed mulsemedia adaptive scheme, improves both user perception levels and user enjoyment experience in variable network delivery conditions. Additionally, ADAMS does not sacrifice user enjoyment experience despite the reduction of multimedia quality and number of sensorial effects, as the latter is performed in inverse order to user interest levels.

VI. Conclusions

In the quest to further enhance user quality of experience mulsemedia combines multiple media elements which engage an increased number of human senses. As any other type of rich media content, mulsemedia delivery over limited bandwidth networks is challenging. This paper has proposed ADAMS, an ADAptive MulSemedia delivery solution in order to increase end-user quality of experience in loaded network delivery conditions. ADAMS's design was informed by extensive subjective tests conducted to study users' preference of various sensorial effects in the context of mulsemedia sequences. Perceptual user tests have been organized to assess ADAMS in comparison with existing state of the art delivery solutions. ADAMS outperforms these solutions in terms of both perceived quality and user enjoyment during adaptive streaming of different multi-sensorial content. In so doing, ADAMS makes the case for having tailored adaptive delivery solutions for mulsemedia content, as traditional multimedia techniques will deliver a lower user quality of experience. Accordingly, one valuable future direction that our work opens up is that of mulsemedia-aware adaptation - we have shown that multimedia adaptation is not enough for mulsemedia applications - for these one has to use mulsemedia-aware adaptation, and we hope that future work shall further explore this area.

Acknowledgements

The authors acknowledge Universal Studios Inc., the distributer of movies *Back to the Future* and *Jurassic Park*. and the support from Enhancing Performance Initiative at Dublin City University Ireland, Enterprise Ireland Innovation Partnership with Ericsson (IP/2011/0135) and China Scholarship Council.

References

[1] G.-M. Muntean, P. Perry and L. Murphy, "A New Adaptive Multimedia Streaming System for All-IP Multi-Service Networks," *IEEE Transactions on Broadcasting*, vol. 50, no. 1, pp. 1-10, Mar. 2004.

[2] N. Mastronarde, K. Kanoun, D. Atienza, P. Frossard, and M. van der Schaar, "Markov Decision Process Based Energy-Efficient On-Line Scheduling for Slice-Parallel Video Decoders on Multicore Systems," *IEEE Transactions on Multimedia*, vol. 15, no. 2, pp. 268–278, 2013.

[3] M. Kennedy, H. Venkataraman, G.-M. Muntean, "Battery and Stream-aware Adaptive Multimedia Delivery for Wireless Devices", *IEEE Local Computer Networks Conference (LCN)*, pp. 843-846, 2010

[4] Z. Yuan, S. Chen, G. Ghinea, G.-M. Muntean, "User Quality of Experience of Mulsemedia Applications", *ACM Trans. on Multimedia Computing, Communications, and Applications (TOMM)*, 2014.

[5] Z. Yuan, G. Ghinea, and G.-M. Muntean, "Quality of Experience Study for Multiple Sensorial Media Delivery", *IEEE Int. Wireless Comms. and Mobile Computing Conf. (IWCMC)*, Nicosia, Cyprus, Aug. 2014.

[6] G. Ghinea, S. R. Gulliver and F. Andres (eds.), "Multiple Sensorial Media Advances and Applications: New Developments in MulSeMedia", *IGI Global*, 2011.

[7] T. Sikora, "The MPEG-7 visual standard for content description: An overview," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 6, pp. 696–702, Jun. 2001.

[8] R. Stewart, M. Ramalho, Q. Xie, M. Tuexen, and P. Conrad, "SCTP Partial Reliability Extension," *IETF RFC 3758*, May. 2004.

[9] J. Dai, F. Liu, B. Li, B. Li, and J. Liu, "Collaborative Caching in Wireless Video Streaming Through Resource Auctions," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 2, pp. 458–466, 2012.

[10] L. Zhou, H.-C. Chao, and A. V. Vasilakos, "Joint Forensics-Scheduling Strategy for Delay-Sensitive Multimedia Applications over Heterogeneous Networks," *IEEE Journal on Selected Areas in Comms.*, vol. 29, no. 7, pp. 1358–1367, 2011.

[11] T. Melodia and I. F. Akyildiz, "Cross-layer QoS-aware Communication for Ultra Wide Band Wireless Multimedia Sensor Networks," *IEEE Journal on Selected Areas in Comms.*, vol. 28, no. 5, pp. 653–663, 2010.

[12] B. Ciubotariu, G-M. Muntean and G. Ghinea, "*Objective Assessment of Region of Interest-Aware Adaptive Multimedia Streaming Quality*", *IEEE Trans. Broadcasting*, vol.55, no.2, pp.202-212, 2009.

[13] J. N. Kaye, 'Symbolic Olfactory Display', Master's thesis, Massachusetts Institute of Technology, 2001. Available: http://www.media.mit.edu/~jofish/thesis/.

[14] J. N. Kaye, "Making Scents: Aromatic Output for HCI", *Interactions*, vol. 11, no. 1, pp. 48 – 61, 2004.

[15] A. Bodnar, R. Corbett, and D. Nekrasovski, "AROMA: Ambient Awareness Through Olfaction in a Messaging Application", *Int. Conf. on Multimodal Interfaces*, PA, USA, 2004, pp. 183 – 190.

[16] S.A. Brewster, D. K. McGookin, and C.A. Miller, "Olfoto: Designing a Smell-Based Interaction", in *Proc. of Conference on Human Factors in Computing Systems*, Montréal, Québec, Canada, 2006, pp. 653 – 662.

[17] S. Issanchou, D. Valentin, C. Sulmont, J. Degel, and E.P. Köster, "Testing Odor Memory: Incidental versus Intentional Learning, Implicit versus Explicit Memory" in Olfaction, Taste, and Cognition, Cambridge University Press, New York, USA, pp. 211- 230, 2002.

[18] H.Q. Dinh, N. Walker, L.F. Hodges C. Song, and A. Kobayashi, , "Evaluating the importance of multi-sensory input on memory and the sense of presence in virtual environments", in *Proceedings - Virtual Reality Annual International Symposium*, 1999, pp. 222-228.

[19] S. Boyd Davis, G. Davies, R. Haddad, and M. Lai, "Smell Me: Engaging with an Interactive Olfactory Game", in *Proceedings of the Human Factors and Ergonomics Society 25th Annual Meeting*, 2006, pp. 25-40

[20] P. Kortum, "HCI beyond the GUI: Design for haptic, speech, olfactory, and other nontraditional interfaces.", Morgan Kaufmann, 2008.

[21] V. Hayward, O.R. Astley, M. Cruz-Hernandez, D. Grant, and G. Robles-De-La-Torre, "Haptic interfaces and devices", *Sensor Review*, Vol. 24, pp. 16-29, 2004.

[22] Immersion (2008). Fact Sheet. http://www.immersion.com/corporate/fact_sheet.php [On-line].

[23] K. Kostopoulos, K. Moustakas, D. Tzovaras, G. Nikolakis, C. Thillou, and B. Gosselin, "Haptic Access to Conventional 2D Maps for the Visually Impaired", *Journal on Multimodal User Interfaces*, vol. 1, no.2, pp. 13-19, 2007.

[24] E. Richard, A. Tijou, P. Richard and J. Ferrier.-. 2006, "Multi-modal virtual environments for education with haptic and olfactory feedback", *Virtual Reality*, vol. 10, no. 3-4, pp. 207-225, 2006.

[25] A. Tijou, E. Richard, and P. Richard, "Using Olfactive Virtual Environments for Learning Organic Molecules", *Int. Conf. on Technologies for ELearning and Digital Entertainment, Edutainment*, 2006, pp. 1223-1233.

[26] J. Ryu, and G.J. Kim, "Using a vibro-tactile display for enhanced collision perception and presence", *ACM Symposium on Virtual Reality Software and Technology (VRST)*, New York, USA, 2004, pp. 89-96.

[27] G. Ghinea and O. A. Ademoye, "The sweet smell of success: Enhancing multimedia applications with olfaction", *ACM Trans. on Multimedia Computing, Communications and Applications* vol. 8, no.1, 2, 2012

[28] K. Kahol, P.Tripathi, T. Mcdaniel, L. Bratton, and S. Panchanathan, "Modeling context in haptic perception, rendering, and visualization", *ACM Trans. on Multimedia Computing, Communications and Applications*, vol. 2, no. 3, pp. 219-240, 2006.

[29] J. G. Apostolopoulos, P. A. Chou, B. Culbertson, T. Kalker, M. D. Trott, and S. Wee, "The Road to Immersive Communication," *Proceedings of the IEEE*, vol. 100, no. 4, pp. 974 –990, Apr. 2012.

[30] Y. Ishibashi, T. Kanbara, and S. Tasaka, "Inter-stream synchronization between haptic media and voice in collaborative virtual environments", in *Proceedings of the 12th Annual ACM Int. Conf. on Multimedia* ACM, New York, NY, USA, 2004, pp. 604-611.

[31] G. Ghinea and O. A. Ademoye, "Perceived Synchronization of Olfactory Multimedia", *IEEE Trans. on Systems, Man, and Cybernetics-Part A: Systems and Humans* , Vol. 40, No. 4,pp. 657 – 663, 2010.

[32] E. Steinbach, S. Hirche, M. Ernst, F. Brandi, R. Chaudhari, J. Kammerl, and I. Vittorias, "Haptic Communications," *Proceedings of the IEEE*, vol. 100, no. 4, pp. 937 –956, Apr. 2012.

[33] ITU-T Recommendation P.910, "Subjective Video Quality Assessment Methods for Multimedia Applications," Apr. 2008.

[34] ITU-T Recommendation P.911, "Subjective Audiovisual Quality Assessment Methods for Multimedia Applications," Dec. 1998.

[35] Z. Yuan, G. Ghinea and G.-M. Muntean, "Appendix Docs", Technical Report, Performance Engineering Lab, Dublin City University, Ireland, Available: http://elm.eeng.dcu.ie/~yuanzh/papers/appendix.pdf [Accessed: Feb 18, 2014].

[36] N. Murray, Y. Qiao, B. Lee, Karunakar A. K, G.-M. Muntean, "Subjective Evaluation of Olfactory and Visual Media Synchronization", *ACM Multimedia Systems Conf.*, Oslo, Norway, 2013.

[37] Z. Yuan, H. Venkataraman, and G.-M. Muntean, "MBE: Model-Based Available Bandwidth Estimation for IEEE 802.11 Data Communications", IEEE Transactions on Vehicular Technology, no.5, vol.61, pp.2158-2171, Jun.2012.

[38] A. Capone, L. Fratta and F. Martignon, "Bandwidth Estimation Schemes for TCP over Wireless Networks," IEEE Trans. on Mobile Computing, vol.3, no.2, pp.129-143, Apr. 2004.

[39] M. Li, M. Claypool, and R. Kinicki, "WBest: A Bandwidth Estimation Tool for IEEE 802.11 Wireless Networks", IEEE Int. Conf. Local Computer Networks (LCN), Montreal, Canada, Oct.2008, pp. 374-381.

Dr. Zhenhui Yuan is a Postdoctoral Researcher with the Performance Engineering Laboratory, School of Electronic Engineering, Dublin City University, Ireland, which has also awarded him the Ph.D degree for research in video delivery differentiation in wireless network environments in 2012. In 2008, Zhenhui received his B.Eng from Wuhan University China. His research interests include wireless mobile networks, quality-oriented and battery-aware multimedia streaming, and multisensorial services. He is a member of IEEE.

Dr. Gheorghita Ghinea is a Reader in the School of Information Systems and Computing at Brunel University, United Kingdom. He received the B.Sc. and B.Sc. (Hons) degrees in Computer Science and Mathematics, in 1993 and 1994, respectively, and the M.Sc. degree in Computer Science, in 1996, from the University of the Witwatersrand, Johannesburg, South Africa; he then received the Ph.D. degree in Computer Science from the University of Reading, United Kingdom, in 2000. His research activities lie at the confluence of Computer Science, Media and Psychology. In particular, his work focuses on the area of perceptual multimedia quality and building end-to-end communication systems incorporating user perceptual requirement. He has over 200 publications in leading international conferences and journals; Dr. Ghinea has co-edited two books on Digital Multimedia Perception and Design, and Multiple Sensorial Media Advance and Applications. He is a member of the IEEE and the British Computer Society.

Dr. Gabriel-Miro Muntean is a Lecturer with the School of Electronic Engineering, co-Director of the Performance Engineering Laboratory at Dublin City University, Ireland and Consultant Professor with Beijing University of Posts and Telecommunications, China. He obtained his Ph.D. degree from Dublin City University, Ireland for research in quality-oriented adaptive multimedia streaming over wired networks in 2003. He was awarded the B.Eng. and M.Sc. degrees in Software Engineering from the Computer Science Department, "Politehnica" University of Timisoara, Romania in 1996 and 1997, respectively. Dr. Muntean's research interests include quality and performance-related issues of adaptive multimedia streaming, performance of content delivery over wired and wireless networks and with various devices, and energy-aware networking. Dr. Muntean has published over 200 papers in top-level international conferences and journals and has authored three books and sixteen book chapters and edited six other books. Dr. Muntean is Associate Editor with the IEEE Transactions on Broadcasting, Associate Editor with the IEEE Communications Survey and Tutorials and reviewer for important international journals, conferences and funding agencies. He is a member of ACM, IEEE and IEE Broadcast Technology Society.