Hindawi Publishing Corporation
International Journal of Digital Multimedia Broadcasting
Volume 2008, Article ID 250654, 18 pages
doi:10.1155/2008/250654

Research Article

Live Broadcasting of High Definition Audiovisual Content Using HDTV over Broadband IP Networks

C. E. Vegiris,[1,2] K. A. Avdelidis,[1,2] C. A. Dimoulas,[1,2] and G. V. Papanikolaou[1,2]

[1] *Telecommunications Laboratory, Electroacoustics Unit, Department of Electrical & Computer Engineering, Aristotle University of Thessaloniki, Thessaloniki 54124, Greece*
[2] *Electronic Media Laboratory, Department of Journalism & Mass Communication, Aristotle University of Thessaloniki, Thessaloniki 54006, Greece*

Correspondence should be addressed to C. E. Vegiris, cvegiris@yahoo.gr

Received 8 April 2008; Revised 7 August 2008; Accepted 5 November 2008

Recommended by Thomas Magedanz

The current paper focuses on validating an implementation of a state-of-the art audiovisual (AV) technologies setup for live broadcasting of cultural shows, via broadband Internet. The main objective of the work was to study, configure, and setup dedicated audio-video equipment for the processes of capturing, processing, and transmission of extended resolution and high fidelity AV content in order to increase realism and achieve maximum audience sensation. Internet2 and GEANT broadband telecommunication networks were selected as the most applicable technology to deliver such traffic workloads. Validation procedures were conducted in combination with metric-based quality of service (QoS) and quality of experience (QoE) evaluation experiments for the quantification and the perceptual interpretation of the quality achieved during content reproduction. The implemented system was successfully applied in real-world applications, such as the transmission of cultural events from Thessaloniki Concert Hall throughout Greece as well as the reproduction of Philadelphia Orchestra performances (USA) via Internet2 and GEANT backbones.

1. INTRODUCTION

It is unquestionable that the rapid evolution of next generation networks and broadband access emerging nowadays has an increased impact on traditional information and communication technologies (ICT) services and applications. Among others, digital multimedia production and broadcasting is mostly influenced by these changes, allowing taking full advantage of the contemporary technological advances. Novel, user-oriented, and on-demand services are currently deployed for browsing, searching, and retrieval of AV content, including news, multimedia e-learning, AV streams of cultural events, entertaining shows, and other applications. Web-based video on demand (VoD) services, digital/interactive TV (DTV/ITV), IP-based TV (IPTV) programs over Internet and mobile systems are typical examples where AV content is usually delivered via IP-based topologies [1–7]. This also complies with the persistent demand for continuous extension of the available AV content resolution and fidelity, in an effort to achieve better experience, creating a sense of realism or telepresence [8–19]. High definition (HD) AV technologies [5–7, 9, 10], Ultra High Definition Video (UHDV or Super Hi-Vision) [17–19], and digital cinema (D-cinema) [20] projects are currently focusing on these objectives, increasing the capacity of the related video streams and raising a compulsory demand for even higher data transfer rates. Besides the utilities that have been recently launched in next generation networks architectures [1–3], the combination of broadband services and high definition multimedia broadcasting is a very challenging technological/research field that the current paper aims to discuss in depth.

The widespread of the World Wide Web and the know-how obtained through the successful implementation of the Internet have lead to the global adoption of the IP technology and the IP-based communications over a broad area of contemporary ICT services. There is a worldwide effort to construct broadband network backbones, like the Internet2 [21] and the Geant [22] initiatives that have been implemented during the last decade in USA and

Europe, respectively. It is obvious that digital multimedia broadcasting inherently belongs to demanding broadband services such as the above as well as similar state-of-the art technological approaches [8–10, 21, 22]. The purpose of the current paper is to analyze, implement, and evaluate the use of state-of-the art digital multimedia broadcasting technologies in combination with broadband services for the transmission of demanding AV streams, captured by means of live performance HDTV shots. The proposed AV configuration setup has been successfully deployed in real-world applications, such as the transmission of cultural events from Thessaloniki Concert Hall throughout Greece [23] as well as the reproduction of Philadelphia Orchestra performances (USA) [24] via Internet2 and GEANT backbone. However, the presented methodology can be further utilized as guidance in related IPTV applications, especially those dealing with HD video/multimedia streaming.

The paper is organized as follows. The problem definition is described in Section 2, where state of research and related works are also discussed. Section 3 presents the proposed system configuration, providing detailed information about the development steps of the work, all the physical and the technical aspects faced during the implementation phases, while metrics' statistics/relation and their utilization are also considered. Experimental results are analyzed in Section 4, where evaluation of the adopted system configuration is carried out in combination with conclusion and future work remarks.

2. PROBLEM DEFINITION AND BACKGROUND WORK

The increased popularity of networked multimedia applications has created new demands for reliable and secure video transmission, so that AV content is expected to account for a large portion of the traffic in the future Internet and in next-generation wireless systems [8, 25, 26]. This creates further necessities for broadband networks due to the fact that digital audio and video hold a large amount of information, especially in critical/demanding cases, whereas high resolution is required, while quality compromises are not acceptable. To meet such demands, advanced compression techniques are continuously evolving in combination with novel routing architectures and algorithms, in an effort to guarantee the required QoS via the currently available Internet data transfer rates [8–10, 25, 27, 28]. Classical examples of this category are the IPTV and VoD projects that have been lanced during the last years [1–7]. As a consequence, various studies have been appeared focusing on the evaluation of the network and the compression parameters, in combination with various content types, using quantitative metrics (often perceptually adapted) as well as subjects that incorporate functional, perceptual, and aesthetical mean opinion scores (MOSs) [8, 27–35]. Recently, research effort has began focusing on HDTV-related approaches [5–7, 9, 10, 33, 34], including D-cinema [20] and UHDV [16–19, 36], aiming at evaluating the advantages, as well as the technical difficulties of these technologies, toward the implementation of future, AV broadcasting services.

The scope of the current paper is twofold: first, to present a solid system layout for HD video content production and transmission; second, to evaluate various parameters of the system configuration setup for their effect to the achieved quality. It is important to mention that the current paper deals with technical issues both at the production (physical theatre) and reproduction (remote amphitheatre) stages as well as all the intermediary phases related with content packaging and streaming (AV formats, compression algorithms, routing architectures, etc.) Hence, unlike the previously mentioned research works, it aims at providing end-to-end solutions and their evaluation for successful broadcasting of live shows in auditoriums. In addition, audio and surround sound techniques are essential toward the successful accomplishment of the "high fidelity" target. Thus, audio-related issues are equally important with video and deserve, respectively, careful treatment at all the recording, mixing, multiplexing, coding, and reproduction procedures. Specifically, the problem under study is best described with commonly raised questions that should be answered. *How many cameras and what formats should be employed? How many microphones, what types, and where should be placed? Which AV compression, multiplexing, and coding algorithms should be selected? Which are the most applicable AV content packaging/streaming techniques that should be implemented in combination with the available network topologies and the corresponding routing architectures? Is there dedicated AV equipment to fulfill the reproduction demands, and how should be used? How the above parameters influence the achieved quality and the perception of the transmitted shows?* The outermost target of the current work is to provide guidance for live AV capturing/IP-broadcasting/reproduction as well as an integrated methodology for the prediction/estimation of the achieved, end-to-end QoE ($QoE_{e\text{-}e}$), using QoS metrics (i.e., PSNR) related to both application-demands (QoS_A) and at broadcasting-network level (QoS_B).

The majority of the corresponding research works are mainly focused on the video-related issues, which are the most technically demanding, since they are related to the larger portion of the AV data stream. For example, CCD-camera noise and lighting-related degradation may appear in cases where capturing conditions cannot be adjusted based on the broadcasting demands [8, 37], as the performed show is mainly designated for the audience at the physical auditorium. In general, video signals can be corrupted by noise during acquisition, recording, digitization, processing, and transmission [37]. On the other hand, audio information requires careful treatment in order to be able to create high fidelity sound-field reproduction conditions. The importance of this task has been exhaustively analyzed in D-cinema and UHDV initiatives [13–16, 36], while various stereo sound recording techniques have been proposed in combination with the corresponding surround sound systems for its effective implementation [13–16, 36, 38–43]. From a practical point of view, 5.1 and 7.1 surround systems have prevailed in real-world application, such as movie theatres and cinema industry. In addition, subjective tests have been applied to provide best format

selection guidance, in combination with the corresponding content type and the available reproduction HDTV sets [33, 34], while future trends of HDTV and the third generation of HDTV formats, like the 1080p, have been studied [9, 10, 34]. However, in the case of HDTV, little attention was given to the accompanied surround sound in combination with the viewing distance conditions [9, 11].

Considering that the video compression-related degradation has been studied during the development of the corresponding algorithms, many researchers are exclusively focused on the evaluation of the video broadcasting procedures, studying the generated traffic demands in relation to the characteristics of the involved network technologies [8, 25, 26]. In general, there are three major research approaches in evaluating broadcasting networks video-wise: (a) actual video bitrate-based techniques, (b) video traces-based approaches, and (c) model-based techniques [26, 29–35]. Actual video streams exhibit originality and provide accurate results during their transmission evaluation over lossy networks, but raise difficulties connected with their availability and copyright permissions. In order to overcome such issues, video traces use only the video content number of bits and the related timing information instead the video content itself [26, 29, 30]. These techniques are focusing on the study of the transmission characteristics but fail to estimate their effect on the decoding process being unable to predict how the occurred errors reflect on the perception of the received image sequences [26, 29, 30]. An evolution of the above methods include the design of advanced video traces that incorporate various video-related features [30], enabling the study of lost packets influence in compressed sequences [28, 29]. Finally, model-based video traffic techniques use mathematical models to describe the video streaming propagation over the network, depending strongly on the validity of the selected model in a real-world application [26, 29, 30]. These methods represent the simplest evaluation approaches that provide lost-data statistics at the level of pixel, frame, and groups of pictures (GOPs), over various network bandwidths and topologies, different routing architectures, and in general, variable QoS settings [8, 25, 26].

As already stated, packet-loss information is inadequate to estimate the actual video degradation, related to the received and perceived image quality. In most cases, reference content is available (full reference (FR) methods [26, 29, 30]), therefore comparisons between the broadcasted (reference) and the received AV streams are employed to form the corresponding metrics. Peak signal-to-noise ratio (PSNR) and mean square error (MSE) are the most common arithmetic metrics that are usually employed for objective evaluation of "processed" images and videos (i.e., processes of compression, enhancement/denoising, transmission, etc.) [8, 25–32, 37],

$$\mathrm{mse}(n) = \frac{1}{N_H \cdot N_V} \sum_{i,j} [x(i,j,n) - y(i,j,n)]^2,$$

$$\begin{aligned} \mathrm{PSNR}(n) &= 10 \cdot \log_{10}\left(\frac{\max(x)^2}{\mathrm{mse}(n)}\right) \\ &= 20 \cdot \log_{10}\left(\frac{2^{n_B} - 1}{\sqrt{\mathrm{mse}(n)}}\right), \end{aligned} \tag{1}$$

where $x(i,j,n)$ and $y(i,j,n)$ are the reference and the impaired video frames for the corresponding video components (i.e., luminance component Y, color components $C_R = Y - R$, $C_B = Y - B$, R, G, B, etc.), with i, j the Cartesian image-coordinates, n the frame number, $\max(x) \equiv \max(y)$ is the maximum value of the 2D signals that is related to the number of quantization bits n_B (equal to 255 for an 8-bit image signal), N_H and N_V the horizontal and vertical image resolution, respectively [8, 32, 37].

Although PSNR metric is quite acceptable for evaluation of integrity of final signal, it is not capable of evaluating video degradations caused by structural dissimilarities. A quite simple and "computationally affordable" metric that applies and is widely used for this task is the perceptually-adapted structural similarity (SSIM) metric [31] which is extracted based on the general principle that human vision system (HVS) is highly sensitive to the structural information of the received optical stimulus. Thus, considering again the reference and the impaired signals (video-frames) $x(i,j,n)$ and $y(i,j,n)$ a simplified form to estimate SSIM is given as follows [31]:

$$\begin{aligned} \mathrm{SSIM}(x,y,n) &= \frac{(2 \cdot \mu_x \cdot \mu_y) \cdot (2 + C_1) \cdot (2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1) \cdot (\sigma_x^2 + \sigma_y^2 + C_2)}, \\ \mu_x &= \mu_x(n) = \frac{1}{N_H \cdot N_V} \sum_{i,j} x(i,j,n), \\ \sigma_x &= \sigma_x(n) = \sqrt{\frac{1}{N_H \cdot N_V} \sum_{i,j} [x(i,j,n) - \mu_x(n)]^2}, \\ \sigma_{xy} &= \sigma_{xy}(n) \\ &= \frac{1}{N_H \cdot N_V} \sum_{i,j} [x(i,j,n) - \mu_x(n)] \cdot [y(i,j,n) - \mu_y(n)] \end{aligned} \tag{2}$$

where μ_x, σ_x, σ_{xy} are known statistics (mean, standard deviation, cross-variance), and C_1, C_2 are small positive constants that are introduced to prevent computational overflow and instability when values are very close to zero.

Besides FR methods and their corresponding metrics (PSNR, MSE, SSIM), reduced-reference (RR) and no reference (NR) methods are employed, when source (reference) content is partially available or unavailable during the evaluation process [29–31]. These two approaches are commonly used in combination with subjective tests which are relying on MOS statistics in order to fully incorporate perceptual attributes of the HVS during evaluation. In addition, subjective tests are very useful in cases that the influence of various parameters is not predetermined, or

depends on the characteristics of the AV content itself (i.e., degradations at the acquisition/digitization phases, impact of packet losses at the decompression phase, appearance of errors and concealed errors with respect to image position and motion activity of the content, impact of the selected resolution—scanning to the perceived quality with regard to content type, etc.) [25–31].

To the best of our knowledge, there are not sufficient guidance and related information on real-world, end-to-end implementations using state-of-the art audiovisual HD equipment with the above characteristics, besides the research efforts and demos discussed in the previous paragraphs that are partially focusing on some of the discussed objectives. In the current paper scope, statistical analysis was performed using both FR and NR methods, using standard quality metrics and testing hybrid evaluation functions on real-world broadcasting and simulated transmission on HD streams.

3. THE PROPOSED SYSTEM CONFIGURATION

HD video and surround sound were the least viable choices to meet the minimum configuration requirements of the application in question. Additionally, for the specific demands of the current work in order to be employed in real-world demanding applications, various practical issues should be considered.

3.1. *Physical and technical issues: application demands and technology requirements*

In order to address the above case, there is a necessity of the replication of the conditions present in the actual event venue to a potential virtual event venue. Considering the proportions, a suitably configured projection hall in a remote site could act as a virtual auditorium following certain specifications. Fortunately, the high-speed Internet and digital technology growth during the last years present us with a unique opportunity to combine all of the above in a digital form following suitable standards and transmit the content in real time to even the remotest of audiences, satisfying the uniqueness and motivation needs of an actual cultural event. In the above frame, our research was targeted to the investigation of the conditions of the actual venue, the transmission infrastructure, and the virtual venue in terms of various factors such as actual spectator perception, capturing, transport, reproduction, and so on. An overview of the architecture used is shown in Figure 1(a), while AV capturing/reproduction setups are presented in Figures 1(b) and 1(c), respectively (detailed information about the selected architecture setups are provided in the following paragraphs).

3.1.1. *Capture of the performances at the physical theater: the transmission site*

The first question that emerges in the design of such a system is the study of perception, thus the experience of a spectator in a performance hall (auditorium, concert, opera, etc.). Given the fact that the actual performance is being held in an organized hall, we take for granted its acoustical and visual integrity. As a result, audience "sweet spots" exist and are known in each different case. Thus, if one was to decide a spectator position which would be the experience reference, a suitable selection would be the choice of one of these spots in terms of both audio and video stimuli [23, 42, 43]. Our effort was targeted to the transfer of a selected actual spectator position experience to all the members of the remote audience [23].

In this point, it is useful to discriminate the audio and video capturing properties. As far as audio is concerned, the perceived result in an actual hall is produced from the combination of the following factors: (a) the direct acoustic field, (b) the reverberant acoustic field (hall acoustics), and (c) the field created by the PA system, if any installed [42, 43]. The reproduction system in the remote hall was specified to be a standard 5.1 surround system which, nowadays, is the typical sound system installed in the majority of public projection halls (i.e., cinemas). In order to be able to reproduce effectively the original audio conditions, it is crucial to acquire the above fields as isolated as possible so that the final 5.1 mix corresponds to the experience reference position. For that reason two kinds of microphone setups were used concurrently: (a) a gun microphone array pointing the stage for the direct field and (b) a soundfield microphone in the reference position that holds spatial sound information by means of the four X, Y, Z and W audio components [40–43] and is used to extract the actual spectator surround perception (Figure 1(b)).

Since the direct field arrives to the spectator with a delay proportional to the distance from the stage, the above setup is able to provide the direct field and the reverberant/PA field in the position in question [42, 43]. According to [42, 43], such a hybrid system is capable of providing soundfield localization, virtual source positioning-panning and signal enhancement, by means of amplitude weighting and time delay compensation, especially for the cases of small-sized sources (point source model) or even still sources. Although these conditions were valid for most cultural performances dealt in the current project (i.e., jazz, theatrical acts, recitals, etc.), we decided not to involve sound-source localization for practical reasons as well as in order to propose a universal recording layout that could be applied in every cultural show. However, we adopted the amplitude—delay weighting approach, aiming at capturing and reproducing the audience experience related with the hall acoustic properties of the physical theatre. Based on the above remarks, the gun microphone array signals and the corresponding soundfield components were mixed to a 5.1 setup based on sound propagation criteria (amplitude-phase weighting) according to their positioning and the coordinates of the capturing (camera) spot [44].

As far as the encoding of the six audio channels that had to be transmitted is concerned, an AC3 encoder which created a stream of 320 kbits/s, achieving an easily decodable, high-quality audio form for the consumer side [23]. This is the encoding that DVB format uses and so it was not further examined because of the proven quality that provides [45].

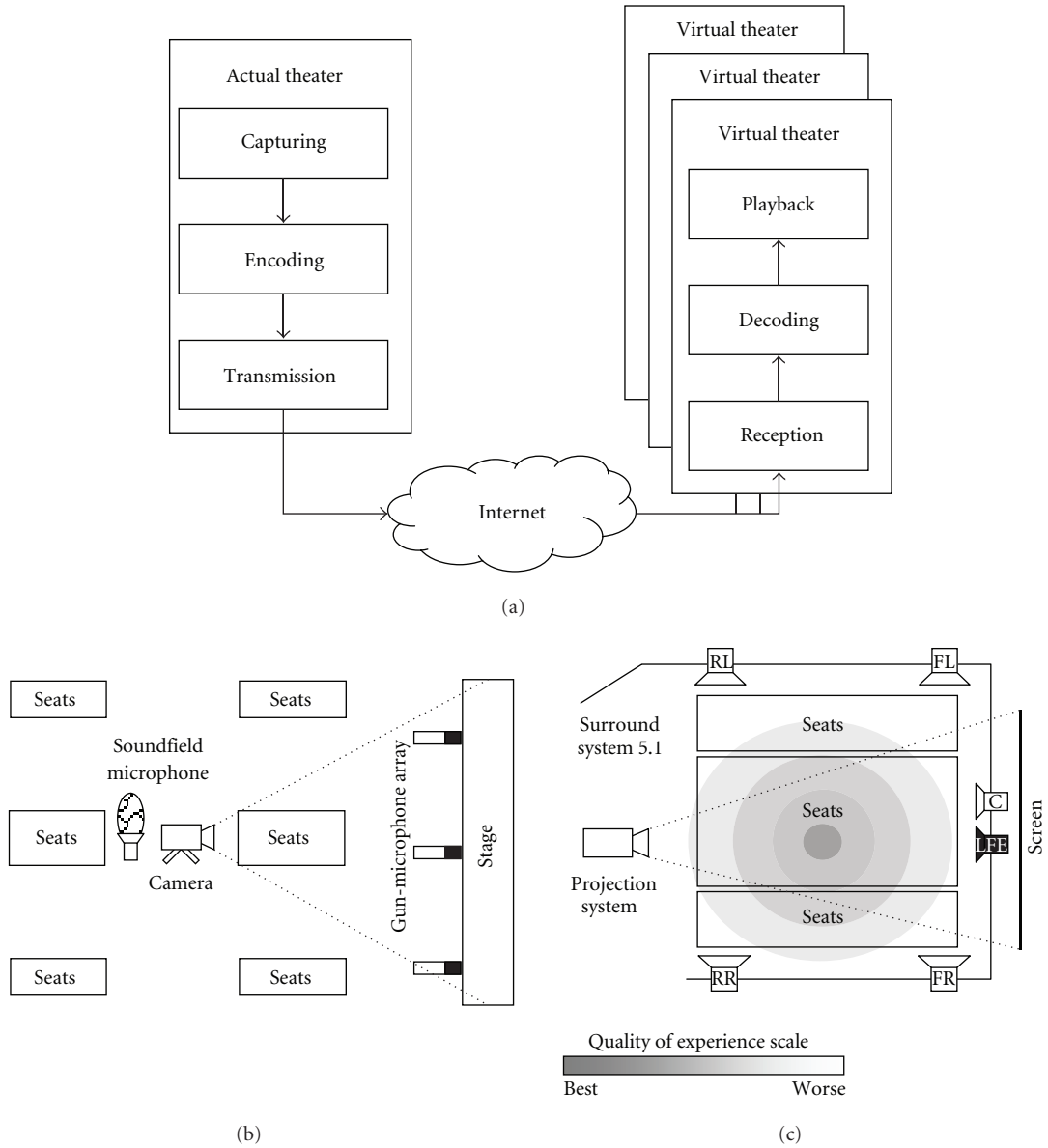

Figure 1: Configuration system setup: (a) architecture overview, (b) actual venue capture setup (transmission site), (c) indicative virtual hall configuration (reproduction site).

The case of video is more straightforward as the view of a spectator in such an environment is limited by the visible field viewing angle and the boundaries of the stage from the selected position. The most critical point in this case is that a spectator is able to focus his/her attention to any point on the stage at any time of the performance. In order to achieve that feeling in the remote hall, it was necessary to provide the virtual spectators with full stage view, in adequate quality and at least near-to-real dimensions. The chosen video capture device (HD-camera) was covering the stage area in a steady position in order to transfer an unbiased stage aspect (without the intervention of a director) on the virtual audience [23]. The setup used in the actual venue is depicted in Figure 1(b).

The exact capturing position, meaning the distance from stage and the height of camera, had to be decided. Specifically, the determination of the acquisition point strongly depends on the stage's physical dimensions (width and height). In addition, special care had to be taken for technical issues, such as camera lens's focus length and zoom in order to avoid deformities due to zoom lenses. In any case, the former set of parameters should be configured according to the restrictions and limitations of the theater stage and seats' layout and so it is quite different in each different physical theater.

An alternative strategy, regarding live-transmission of audiovisual (cultural) shows, is to employ a multicamera director-based setup for video and a 5.1 audio mix

independent of the hall acoustic properties, which is being currently held by the Philadelphia Orchestra with the Global Concert Series [24]. Both strategies can share identical configuration setups regarding AV broadcasting, however there are major differences related to the capturing or even the reproduction layout, issues that have an impact on the achieved QoE, as it will be further commented in Section 4.

The decision of the audiovisual framework was in fact the most crucial part of the design. The video format finally used was 1080i; however, the 720p format was also tested. The main reasons for this choice involved the cost and the consumer side projection capability. The native protocol in the production side used was the HD-SDI standard in order to eliminate the quality loss before the transmission [23]. The visual content was MPEG-2 compressed based on the fact that it is easily decoded, providing sufficient video quality in the scenario under question. At the time of investigation, there was a variety of MPEG-2 enabled devices (encoders, decoders) on the market for a significant amount of time which ensured the reliability of the method. Moreover, the relatively low bandwidth demands for this kind of task were able to ensure reliable transmission (given the Internet framework discussed below). The bitrate chosen was based on subjective, objective, and empirical criteria and was closely related to the nature of the transmitted material (a more thorough analysis is presented in Section 3.1.2) [5, 7, 9, 10].

Other implementation choices included decisions related to the forward error correction (ProMPEG FEC), variable/constant bitrate, encoding profile, and GOP structure. Most of these parameters are under examination by the ITU [46] in relation to the processed content of past research works [33, 34], taking also into account the limitations of the transmitting networks. Based on the low-motion nature of the transmitted content, the final decisions included the use of Main Profile at High Level (MP@HL) [23] format of MPEG-2 encoding, CBR mode, and lack of error correction as the less costly and most easily implemented.

3.1.2. Reproduction at the remote auditorium: the reception site

According to the audio-recording layout already discussed, it is easy to describe the corresponding necessities at the remote auditorium. In fact, a standard 5.1 surround system setup is only required (Figure 1(c)), while slight variation should be occurred related to the dimensions of the remote auditorium and the adequate number of loudspeakers. In any case, neutral acoustical hall behavior (low reverberation) is preferred, allowing transfer of the physical theatre acoustical experience with more fidelity. The sweet spot, in this kind of environment, is also defined mostly based on the audio quality superimposed to the conditions of each projection hall.

For the video part, the constraint posed is relative to the projection size. For the biggest degree of realism to be succeeded, the projection of the event should display the projected objects by their physical dimensions. Ideally, the size of the projection screen should match the dimensions of physical theatre stage, but this is not probable to be the case for the remote auditoriums. Thus, the screen size should be as close as possible to the physical stage, and of course should not exceed these physical dimensions. The screen size and the projection resolution affect the desirable viewing angle to convey the full sensation of presence which for a wide-field video system is 80–100 arcdeg [11, 12, 17]. In order to achieve best audience viewing experience, the viewing distance should be bigger than the shortest distance at which a person with normal vision of 1.0 is unable to recognize the pixel structure on a screen [12]. However, the sensation of telepresence is decreased as distance increase. A compromise between the two above conflicting criteria, giving priority to the first, was adopted and recommended for best viewing experience [44].

In the remote side the standards used for projection varied according to the projection equipment among HD-SDI, DVI, XGA, and HDMI due to the versatility provided by MPEG-2 decoding devices. Although there have been studies focusing on the subjective evaluation of the HD format, in combination with the size of the projection screen in flat TV-sets [33, 34], there are no available published works focusing on large-auditorium projection. A thorough technology-market research pointed out as most applicable the use of DLP technology projector, in combination with over five meter wide, electronic projection screens. As a result, three alternative projection formats were examined: 720p50, 1080i25, and 1080i29.94, with the last one applied during reproduction of Philadelphia Orchestra transmissions [23, 24].

3.2. Transport of AV data and network conformation demands

Most of the attributes describing the transport framework can be derived from the audiovisual framework specifications. The finally selected choices included MPEG TS encapsulation using the RTP protocol at the rate of 30 Mbps combined stream bitrate. This choice was based on the fact that the MPEG TS is capable of encapsulating simultaneous audio stream in AC3 format, eliminating the synchronization problems appeared in past efforts, and allowing the delivery of multichannel surround sound. Finally, incorporation of RTP introduced an amount of jitter immunity to the consumers who could take advantage of this capability. Also, the usage of ProMPEG FEC was tested [23].

As previously mentioned, the present Internet attributes motivated the implementation of the effort in discussion. Following a close collaboration with GRNET, the concluding setup was capable of offering our service reliably to a considerable number of Greek provincial and urban universities, thus covering a potentially large number of spectators. In order to service as much interested consumers as possible, the transmissions were decided to be multicasted through a single group [23]. The connection provided on the transmission site was established through an M-BGP enabled switch with an end-to-end optical Gigabit Ethernet uplink to the GRNET backbone switch and copper wire Gigabit Ethernet downlinks for the internal connections.

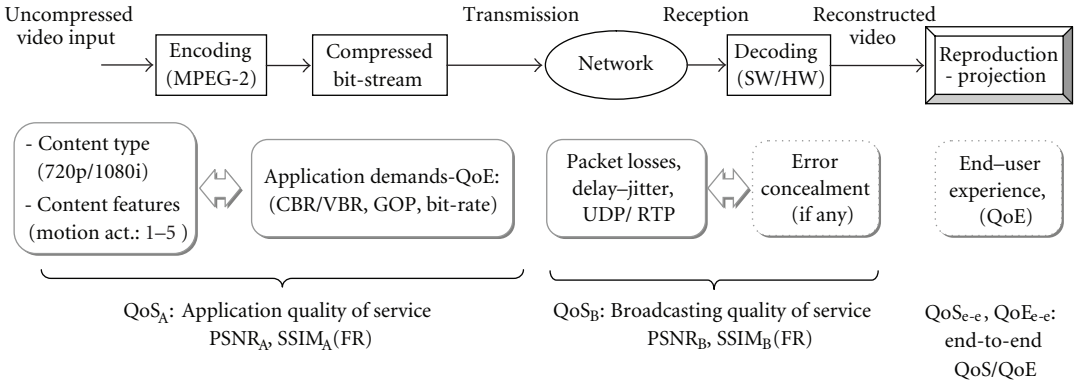

Figure 2: System simulation setup for performance evaluation by means of QoS and QoE metrics.

The reception side setup inside the university campuses relied on the already established network infrastructure consisting of copper wire Fast Ethernet wall sockets of certified functionality and efficiency. The links were tested for transmission/reception errors and were required to appear high quality link state which was a relatively easy task since we were addressing to locally well-organized networks (universities) and not the open public (i.e., home users). For the transmission case, a stand-alone IGMP enabled MPEG-2 encoder/streamer was used while the reception case required a respective receiver/decoder which could be consisted of a stand-alone or a pc-based setup. The multicast traffic was in both cases handled by the Multicast Backbone (MBONE) infrastructure of the providers involved (GRNET and University NOCs). A session announcement through SAP was also statically available during the transmission period in order to encourage participation to our tests. The above setup ensured the minimization of the physical layer related corruptions.

3.3. QoS and QoE issues

Quality of service is a term that has been mostly used in network applications to describe data integrity during transmission, including timely arrival demands. As already mentioned, in video streaming/broadcasting application we may distinguish two sequential stages, affecting the overall QoS: the application QoS ($\mathrm{QoS_A}$) and the broadcasting-network QoS ($\mathrm{QoS_B}$) [47–49]. This 2-stage model has been broadly applied in related application [47–49], and it was also adopted during the experimental phase of the current work for the broadcasting simulation setups and their statistical analysis (refer to Figure 2). Specifically, in the application level we may consider the influence of the bitrate [Mbps], input format [1080i/720p], encoding type [CBR/VBR], content motion activity [motionActivity], deinterlacing, and error concealment strategy for $\mathrm{QoS_A}$. As far as the network broadcasting level is concerned, the $\mathrm{QoS_B}$ is influenced by the network direct packet losses, the jitter, which produces indirect/secondary packet losses according to the routing/streaming/buffering/strategy (buffer size, UDP/RTP). In general, QoS metrics (Qsm) may be configured using the already presented FR metrics PSNR/MSE (use of a single FR-metric or a combination of them) to evaluate the video quality at each stage. Hence, given the preceded analysis we may form simple expressions to model Qsm as function of the application/broadcasting characteristics:

$$\begin{aligned}
\mathrm{Qsm_{e\text{-}e}} &= \operatorname*{avr}_n[\mathrm{Qsm_{e\text{-}e}}(n)] \\
&= \operatorname*{avr}_n[f(\mathrm{Qsm_A}, \mathrm{Qsm_B})], \\
\mathrm{Qsm_A} &= f_A(\text{encoding, format, content, decoding}) \\
&= f_A\begin{pmatrix}\text{bit-rate, 720p/1080i, CBR/VBR,}\\ \text{GoP-length, motionActivity,}\\ \text{de-interlacing, error concealement}\end{pmatrix}, \\
\mathrm{Qsm_B} &= f_B(\text{packet losses, jitter, routing/streaming/} \\
&\qquad \text{buffering/error correction}),
\end{aligned} \tag{3}$$

where $\mathrm{Qsm_{e\text{-}e}}$, $\mathrm{Qsm_A}$, and $\mathrm{Qsm_B}$ are the involved end-to-end, application, and broadcasting QoS metrics, respectively, n is the number of video sequence broadcasted frames, and f, f_A, f_B are, in general, complicated functions aiming to relate the input independent variables ($\mathrm{Qm_A}$-vars, $\mathrm{Qm_B}$-vars) with the QoS estimates.

As already stated, video stream, packet losses, and PSNR do not reflect linearly on the "video quality" during reproduction/projection of the AV broadcasted content, since lost pixels might or not be visible, and in general have different impact on the perceived QoE [29, 32, 35, 50]. In order to be able to account for the gained experience and model the QoE, perceptual criteria related to the HVS should be considered. In the above context, the simplest rule to express QoE metrics (Qem) as function of the input parameters ($\mathrm{Qm_A}$-vars, $\mathrm{Qm_B}$-vars) is to pass all the $\mathrm{Qsm_{e\text{-}e}}$, $\mathrm{Qsm_A}$, $\mathrm{Qsm_B}$ metrics through a filter that emulates the HVS behavior to obtain the desired $\mathrm{Qem_{e\text{-}e}}$, $\mathrm{Qem_A}$, and $\mathrm{Qem_B}$. A different approach would be to deploy perceptually adapted metrics, such as the SSIM, or even subjective MOS results. Thus, we may replace Qsm and Qem with the generic expression Qm

to parameterize QoS/QoE as function of the corresponding application and network-oriented QoS/QoE estimates

$$\begin{aligned} \mathrm{Qm}_{\text{e-e}} &= f(\mathrm{Qm}_{\mathrm{A}}, \mathrm{Qm}_{\mathrm{B}}) \\ &= f[f_{\mathrm{A}}(\mathrm{Qm}_{\mathrm{A}}\text{-vars}), f_{\mathrm{B}}(\mathrm{Qm}_{\mathrm{B}}\text{-vars})], \end{aligned} \tag{4a}$$

$$\begin{aligned} \mathrm{Qem}_{\text{e-e}} &= f(\mathrm{Qem}_{\mathrm{A}}, \mathrm{Qem}_{\mathrm{B}}) \\ &= g(\mathrm{Qsm}_{\mathrm{A}}, \mathrm{Qsm}_{\mathrm{B}}), \end{aligned} \tag{4b}$$

where f, f_{A}, f_{B}, g are again nontrivial parametric functions controlled by the Qm_{A}-vars/Qm_{B}-vars independent variables (inputs) previously discussed. According to (4b), the wanted in the current approach is to model the QoE outcomes as a function of both ($\mathrm{QoS}_{\mathrm{A}}$, $\mathrm{QoS}_{\mathrm{B}}$) and ($\mathrm{QoE}_{\mathrm{A}}$, $\mathrm{QoE}_{\mathrm{B}}$) pairs, expressed by means of PSNR and SSIM, respectively.

Let us take a closer inspection to (3), (4a), and (4b), trying to predict the Qm changes with respect to single-input variations. We will assume that QoS and QoE are correlated with increasing monotony, so that a Qsm raise (improved QoS) would reflect to a relative Qem increase (improved QoE), and vice versa. Hence, we may form simple rules to describe the influence of bitrate, format (720p/1080i), and content motion activity to the $\mathrm{QoS}_{\mathrm{A}}$/$\mathrm{QoE}_{\mathrm{A}}$, and the influence of jitter to the $\mathrm{QoS}_{\mathrm{B}}$/$\mathrm{QoE}_{\mathrm{B}}$. It is obvious that as the bitrate increases, Qm_{A} get also higher due to the fact that better image quality is obtained with fewer compression artifacts. Another issue connected with the bitrate is the content itself. For instance, the AV streams feature high video motion activity, this makes the encoding process more demanding and complicated, so that video degradation worsens in order to be able to attain the desired bandwidth. On the other hand, increasing the bitrate may cause Qm_{B} metrics to decrease, since the effect of packet losses and jitter to the increasing network-traffic demands is rising. Finally, it is clear that as the jitter increases the Qm_{B} metrics go worse, since the indirect packet loss rates get higher. The affection of the remaining parameter is not considered here as less important for the specific demands of the current application. For instance, CBR/VBR variations were not tested, on the basis that CBR is more robust and reliable to be deployed in broadband networks, while deinterlacing and error concealment options were excluded from the current study for the sake of simplicity (to avoid confusion by using too many parameters). Similarly, since direct packet losses are not quite common in broadband networks, we considered only the influence of jitter. As far as routing strategies are concerned, it is obvious that the use of RTP is superior over UDP (related experiments validated this statement), so that the use of RTP was decided as fixed option. The above remarks could be formalized and expressed by the following equations:

$$\begin{aligned} &\frac{\partial(\mathrm{Qm}_{\mathrm{A}})}{\partial(\text{bit-rate})} > 0, \qquad \frac{\partial(\mathrm{Qm}_{\mathrm{A}})}{\partial(\text{motionActivity})} < 0, \\ &\frac{\partial(\mathrm{Qm}_{\mathrm{B}})}{\partial(\text{bit-rate})} < 0, \qquad \frac{\partial(\mathrm{Qm}_{\mathrm{B}})}{\partial(\text{jitter})} < 0, \end{aligned} \tag{5}$$

where the partial derivative $\partial(\cdot)$ expresses the influence of each independent variable, considering that all the other input variables (Qm_{A}-vars, Qm_{B}-vars) remain unchanged. We may observe that some of the above changes have complete different impact to the partial system responses (Qm_{A}, Qm_{B}), so that none could easily say which of the above parameters will prevail to the determination of the end-to-end metrics $\mathrm{Qm}_{\text{e-e}}$ in a nontrivial/realistic broadcasting-configuration scenario, like the one dealt with in the current work.

Related to the motion activity characteristics is also the use of interlaced (1080i) or progressive (720p) scanning. For instance, it would be preferable for a high motion video sequence to be encoded to progressive format (i.e., 720p) in order to avoid filtering out motion details due to interlaced scanning (in case of 1080i). Although these motion artifacts are quite annoying and they are easily perceived by the subjects/spectators, both the FR-metrics PSNR and SSIM are unable to measure this limitation due to the fact that it also inherently exists in the original, source material which is used as reference. Hence, once again the influence of a single parameter (interlaced/progressive) provides controversial effects to the overall system behavior. It is important to mention that the use of 720p and 1080i was not intended for direct comparisons between the two formats, using the previously mentioned FR-metrics, but they were proposed as alternative source-content choices to confirm that they both follow certain rules and pose similar Qm dependencies from the remaining independent variables (Qm_{A}-vars, Qm_{B}-vars).

Based on the above analysis, evaluation procedures were contacted using both simulation setups and real video-transmission according to the layout drawn in Figure 2.

4. EXPERIMENTAL RESULTS AND DISCUSSION

Following the design and the implementation of the system as well as the related methodology, several applications involving the organization of transmissions and receptions were conducted. The experimental transmissions among others included four actual real-time transmissions from the Thessaloniki Concert Hall (three from the foyer and one from the main hall). Receiving projection venues was set up in four cities in Greece (Athens, Thessaloniki, Patras, and Heraklion) and one in EU (Dublin, Ireland) involving a total of seven virtual halls varying from very small to medium size. The decoding devices we encouraged the organizers to set up and finally used were PC-based decoders using the VLC Media Player, whereas the case of hardware decoding was only tested by our team. The acceptance of the audience was quite encouraging for this kind of activities and was expressed by the increasing number of spectators and the desire to continue to provide the service.

We also tested the proposed methodology concerning the virtual hall arrangement via the organization of two public projections of the Global Concert Series, considering also the organizational aspects of such an event. Several promotional acts were taken (TV promotional videos, posters, invitations, etc.) in order to stimulate and measure the public interest on the subject. After the event, the spectators were asked to fill in a questionnaire in order to measure certain experience

factors. The similar results to the ones received from our own transmissions are quite promising. A more exhaustive analysis of the parameter selection and the tests conducted is presented below.

4.1. Quantitative analysis by means of metric-based evaluation

Before any NR qualitative evaluation, FR methods were also enabled aiming at evaluating the video degradation issues by means of metric-extracted objective quantities. This evaluation procedure was carried out only for the video content due to the fact that audio coding/multiplexing was based on a well-tested technology (AC-3) that has been successfully implemented [45] during the last decades. However, the evaluation of the new recording layout is worthwhile, and this is why audio-related subjective tests were included in the qualitative evaluation procedure during content reproduction at the remote site(s) (this issue is further analyzed in the next paragraph). In addition, more thoroughly subjective evaluation, in combination with quantitative analysis and the adoption/definition of appropriate audio-metrics, is currently scheduled.

As far as it concerns video evaluation, uncompressed HD video content [33, 34, 51] was selected as reference material in order to be compared with the received/decompressed video at the simulated remote site. In general, the evaluation procedure had to be carried according to the following variables: (a) content type, (b) compression parameters, (c) streaming parameters, and (d) routing/QoS settings. The first category divides content into two major categories according to the original HDTV format (720p,1080i). Five subcategories are formed for each format type, based on the involved motion activity of the content, which implies pace of action [33, 34]. The involved "Sverige Television AB" (SVT) reference video set has been used in the past for similar evaluation in HDTV-related application [33, 34] and this is an additional reason for its selection that enriches its suitability for the demands of the current application. The 10 different *content type* reference videos were edited separately for each format and two different video clips were produced, one 720p sequence and one 1080i sequence. Each content type was used three times in each sequence and between the different content types a grey mate of 2 seconds was added. The five different content types were ranked by the motion activity that they contain by us. They were graded in a scale from 1 to5, with 1 being the one containing the smallest motion activity.

Besides content type, the compression parameters provide an additional variable that determines the encoding bitrate. Given the use of MPEG-2 compression, three different bitrates were involved during simulations (high = 18.1 Mbps, moderate = 17 Mbps, and low = 15 Mbps), these values were selected as recommended and used as reference from IPTV Focus Group of ITU [23, 25–29]. Other parameters of compression like CBR versus VBR, type, and length of GOP as already stated were not further examined [23]. However, the unavoidable network layer issues were put under investigation in order to balance the factors of video stream bitrate/protocol versus quality in jitter conditions. By definition, jitter is the variation presented in trip time from a transmitter to a receiver leading to the deterioration of the stream quality especially in the case of synchronization sensitive services such as the multimedia applications. It is also highly dependent of the network topology of a packet switching infrastructure. Since in the Internet framework the network complexity and therefore the jitter involved is increased as the geographic distances of the venues grow, it is crucial to investigate this factor in the current context.

As for the streaming parameters, the protocol selection (UDP versus RTP) was the only variable involved given that the MPEG2-TS formation was used. The superiority of RTP is rather obvious that the use of it is preferable, whenever this is possible. Nevertheless, the presence of extra buffering memory that RTP implies is a cost proportional to the transmitted stream bitrate that cannot be unnoticed. Especially in cases of streams of high bitrates like the ones we transmit, the extra cost of using RTP is quite significant and so it is supposed to be the second, more costly choice, after UDP. Since RTP actually adds a predefined immunity of certain milliseconds of jitter, according to the buffer size used, the relation of the jitter effects between RTP and UDP can be expressed by the following equations:

$$\mathrm{Qsm_B}(\mathrm{jitter_{ms}})\big|_{\mathrm{RTP}} = \begin{cases} \mathrm{Qsm_B}(\mathrm{jitter_{ms}} - \mathrm{buffer_{ms}})\big|_{\mathrm{UDP}}, \\ \qquad \mathrm{jitter_{ms}} > \mathrm{buffer_{ms}}, \\ \max\{\mathrm{Qsm_B}\}, \\ \qquad \mathrm{jitter_{ms}} \le \mathrm{buffer_{ms}}, \end{cases}$$

$$\mathrm{buffer_{ms}} \approx \frac{8000 \cdot \mathrm{buffer_{MB}}}{\mathrm{bitrate_{MBps}}} \approx \frac{8000 \cdot \mathrm{buffer_{MB}} \cdot \mathrm{compression\_ratio}}{H \cdot V \cdot \mathrm{bpp} \cdot \mathrm{fps}}, \tag{6}$$

where H, V are the horizontal and vertical video resolution, respectively, bpp stands for bits per pixel, fps for frames per second, and *compression_ratio* is the ratio of the original versus the compressed stream size. Due to the above, the case of UDP, which may be used for general conclusions, was examined.

The network performance was simulated by the NETEM (NETwork EMulator) module [52] which can be totally parameterized. Network latency was set at constant typical value of 50 milliseconds as it does not affect the one way transmission quality except the addition of a constant delay. The value of jitter (latency variation) was used as a control variable using values of $0, 0.05, 0.09, 0.11$, and 0.12 milliseconds. As a result, 30 different simulations (2 *resolution formats* $\times$ 3 *bitrates* $\times$ 5 *jitter*) had to be implemented, in order for all the possible combinations to be recorded. For this purpose, fifteen different hypothetical reference circuits (HRCs) were used which are presented in Table 1.

One high performance windows-based PC, equipped with an HD-SDI video card and a high data transfer rate striped disk array was used as player for playing the reference

Table 1: The HRC profiles that were tested during quantitative evaluation.

HRC	Bitrate	Jitter
1	15 Mbps	0
2	17 Mbps	0
3	18.1 Mbps	0
4	15 Mbps	0.05
5	17 Mbps	0.05
6	18.1 Mbps	0.05
7	15 Mbps	0.09
8	17 Mbps	0.09
9	18.1 Mbps	0.09
10	15 Mbps	0.11
11	17 Mbps	0.11
12	18.1 Mbps	0.11
13	15 Mbps	0.12
14	17 Mbps	0.12
15	18.1 Mbps	0.12

video clips. A Linux-based PC with NETEM module installed was used as the network simulator and finally the capturing of projected video was done by another PC similar to the first. The encoding and decoding were done by the standard Tandberg encoder-decoder system used in all transmissions. For compatibility reasons, the set of reference video contents that was used was converted and edited by taking special care about not having any quality degradations during the whole pretransmission process. The format that was used was uncompressed YUV 4:2:2 8-bit in AVI file container. The coded and transmitted video signals had to be edited and converted to the same format in order for the comparisons to be done. Editing of the captured video was mandatory since both recording softwares that we tried were unable to synchronize at once with the playback system through network connection.

The comparison and evaluation of transmitted video signals were done with Semaca's software VQLab [53] which can extract the PSNR and SSIM metrics of each video signal compared to a reference video. Both metrics were extracted once for the video degradation ($Qm_{e\text{-}e}$ and Qm_A) caused by each HRC end-to-end, using as references the original played content videos and once for video degradation (Qm_B) caused just by the network using as reference the video produced right after the codec system. The latter is the video signal coded and transmitted by an ideal network (jitter and packet loss are zero) and so it coincides with the video produced from systems HRC 1, 2, and 3 for different coding bitrates.

The modeling of $Qm_{e\text{-}e}$ as a function of Qm_A and Qm_B is our intention as we have already stated in Section 3.3. By the extracted values of PSNR and SSIM metrics of HRCs 1, 2, and 3 (Jitter = 0) Qm_A can be modeled as a function of bitrate. In Figure 3 the graphs of experimental data are presented where the logarithmic trend of them can be seen. A logarithmic function is also used in [49] for standard definition video SSIM modelling and so it may be assumed that the relation of quality metrics and bitrate can be described by the following equation in the case of high definition too:

$$y = a_1 \cdot \ln(x) + a_2. \tag{7}$$

By using our experimental data and the Levenberg-Marquardt algorithm for nonlinear curve fitting in LabVIEW 7.1, we calculated the coefficients of this function for each different content type and for both metrics. The two equations for each metric are as follows:

$$\begin{aligned} \mathrm{SSIM_A} &= a_{s1} \cdot \ln(\mathrm{bitrate}) + a_{s2}, \\ \mathrm{PSNR_A} &= a_{p1} \cdot \ln(\mathrm{bitrate}) + a_{p2}. \end{aligned} \tag{8}$$

The model works fine for all content types as it can be seen from the mean errors and the standard deviations of it, which are presented in Table 2. All the coefficients for each content type can be viewed in Table 3.

Following the same strategy, Qm_B can be modeled from the measured PSNR and SSIM values of HRCs 6, 9, 12, and 15 as a function of jitter using as reference the measurements of HRC 3. Based on the observed logarithmic decay of the readings as can be viewed in Figure 4, the following exponential equation was initially used:

$$y = b_0 + b_1 \cdot e^{b_2 \cdot x}. \tag{9}$$

By applying our experimental data to the previously mentioned method, we calculated the coefficients of this function for each different content type and for both metrics. The two equations for each metric are presented as follows:

$$\begin{aligned} \mathrm{SSIM_{B_1}} &= b_{s0} + b_{s1} \cdot e^{b_{s2} \cdot (\mathrm{jitter})}, \\ \mathrm{PSNR_{B_1}} &= b_{p0} + b_{p1} \cdot e^{b_{p2} \cdot (\mathrm{jitter})}. \end{aligned} \tag{10}$$

The evaluation of the above model, through the examination of the mean errors and the standard deviations, showed undesirable behavior for the case of PSNR, which provided concrete evidence for the model-data incompatibility. However, this model proved to be acceptable for the case of SSIM for both cases of 720p and 1080i. These remarks are evident in Table 4. All the coefficients for each content type can be viewed in Table 5.

To overcome the instability of the above model in the case of PSNR, a linear mixture of exponential models was tested, which was expressed by the following equation:

$$y = \sum_{k=0}^{K} b_k \cdot \exp(k \cdot x), \quad K = 3. \tag{11}$$

Thus, the resulting equations were formed to the following and the fitting was based on the least squares algorithm using the LabVIEW 7.1 environment, for each different content type and for both metrics:

$$\begin{aligned} \mathrm{SSIM_{B_2}} &= \sum_{k=0}^{K} b_{sk} \cdot \exp\left(k \cdot (\mathrm{jitter})\right), \quad K = 3, \\ \mathrm{PSNR_{B_2}} &= \sum_{k=0}^{K} b_{pk} \cdot \exp\left(k \cdot (\mathrm{jitter})\right), \quad K = 3. \end{aligned} \tag{12}$$

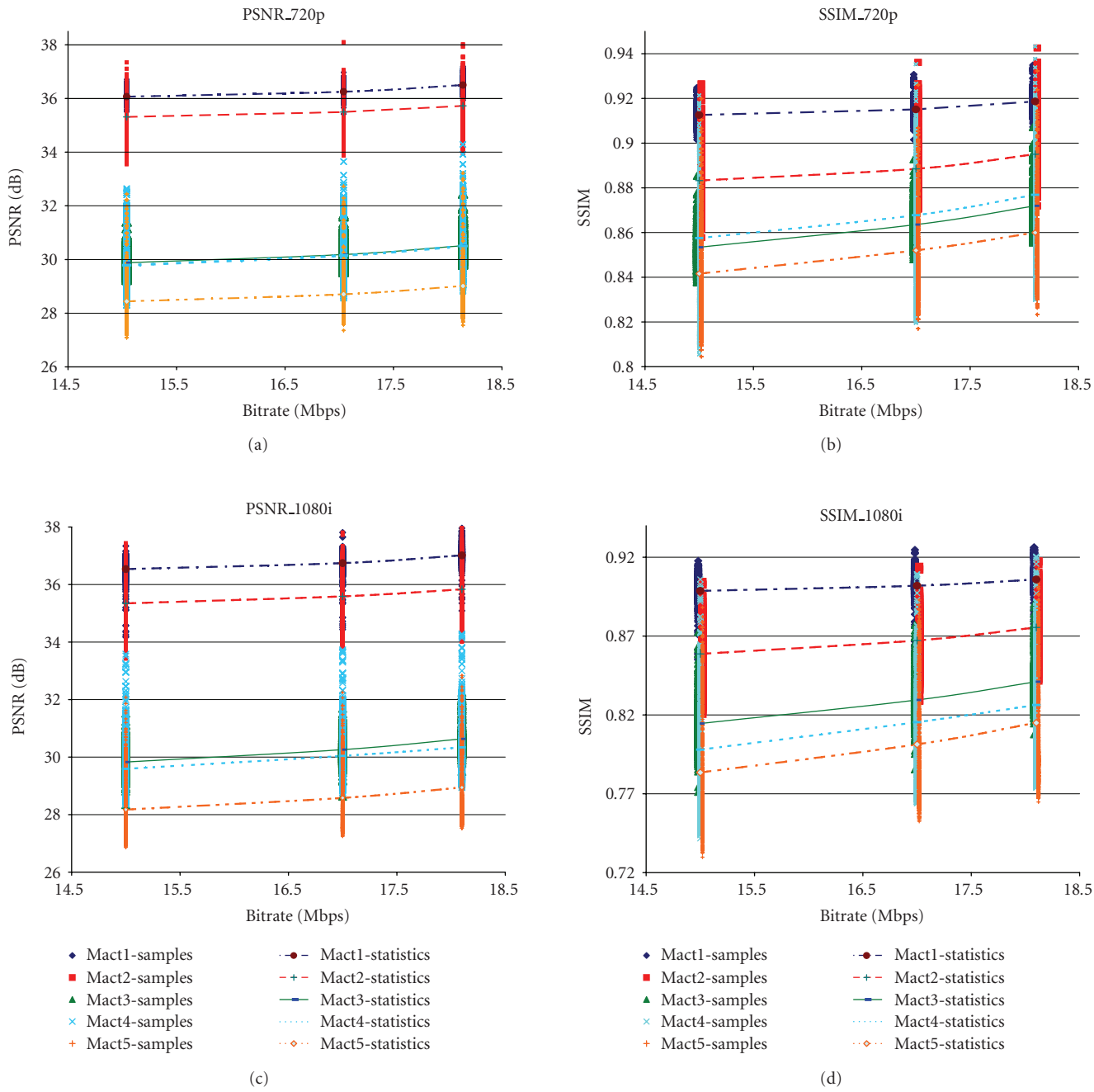

Figure 3: Graphs of the samples and mean curves of experimental data for the model Qm_A.

Table 2: Mean errors and standard deviations of model for Qm_A.

	720p				1080i			
	SSIM		PSNR		SSIM		PSNR	
Motion activity	Mean error	Error STD	Mean error	Error STD	Mean error	Error STD	Mean error	Error STD
1	4.81265E-9	0.00413522	2.18991E-11	0.228754	4.73121E-9	0.00648127	5.64537E-9	0.32641
2	4.47919E-9	0.0111251	4.70308E-9	0.763141	4.13375E-9	0.0135428	8.15997E-9	0.967442
3	4.02524E-9	0.00762155	1.12177E-11	0.378139	3.49324E-9	0.0121462	1.80827E-8	0.373196
4	3.95311E-9	0.0280775	1.19821E-11	1.07011	3.34799E-9	0.0333977	1.96567E-8	1.00888
5	3.99949E-9	0.0214723	1.22118E-11	0.927963	3.16903E-9	0.0377795	1.54833E-8	1.10852

Table 3: Coefficients for the model of Qm_A.

Qm_A_Model								
	720p				1080i			
	SSIM		PSNR		SSIM		PSNR	
Motion activity	α_{s1}	α_{s2}	α_{p1}	α_{p2}	α_{s1}	α_{s2}	α_{p1}	α_{p2}
1	0.0305431	0.829554	2.18079	30.1913	0.0366121	0.799194	2.41602	29.9708
2	0.0597733	0.720824	2.09925	29.6575	0.0861779	0.624756	2.51942	28.5038
3	0.0959924	0.593017	3.24086	21.1299	0.137778	0.440884	4.16464	18.5274
4	0.100119	0.58582	3.68974	19.822	0.149051	0.394054	3.9214	18.9615
5	0.0955935	0.582348	2.91985	20.5602	0.164042	0.338607	3.98058	17.3714

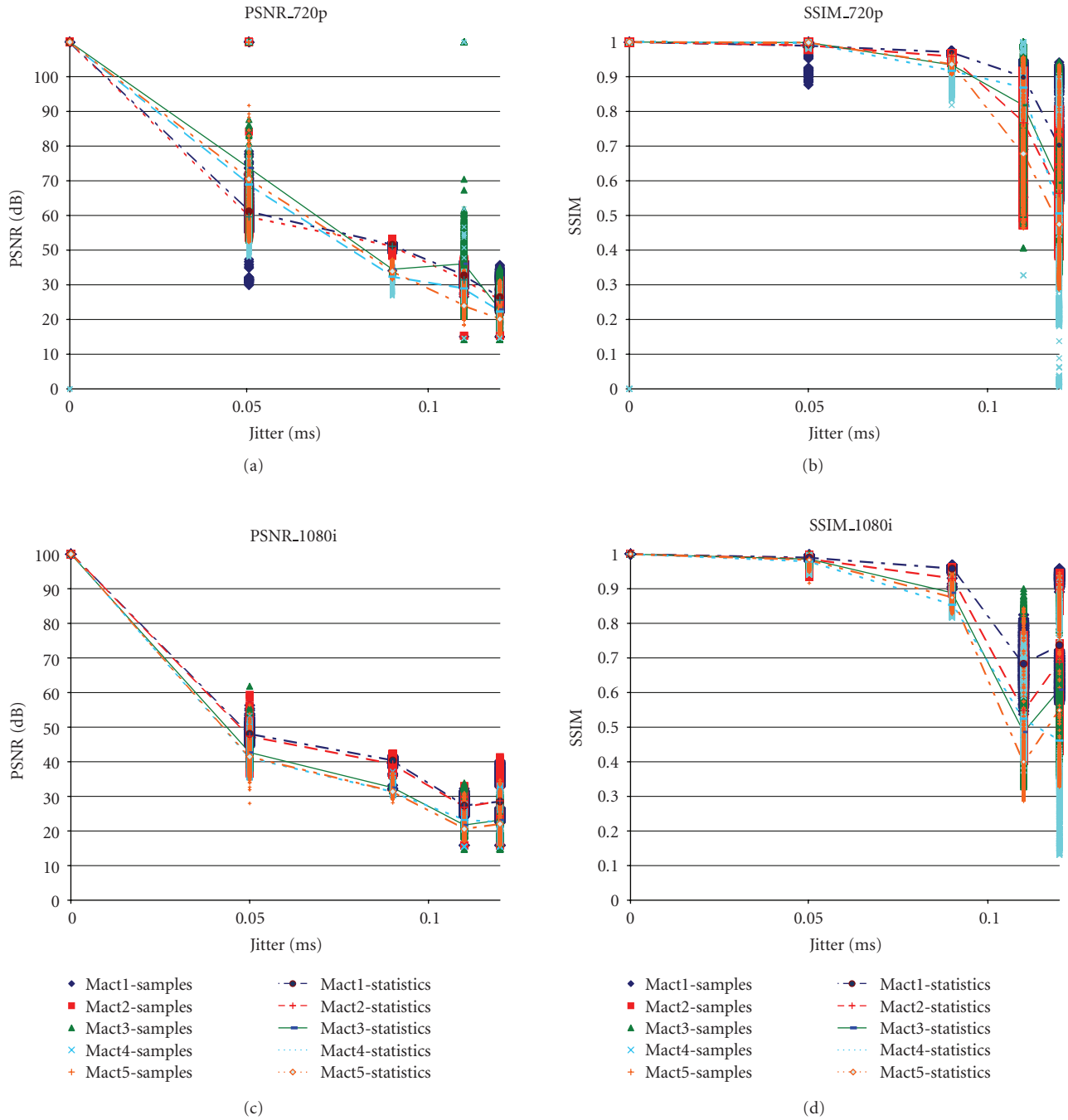

Figure 4: Graphs of the samples and mean curves of experimental data for the model Qm_{B2}.

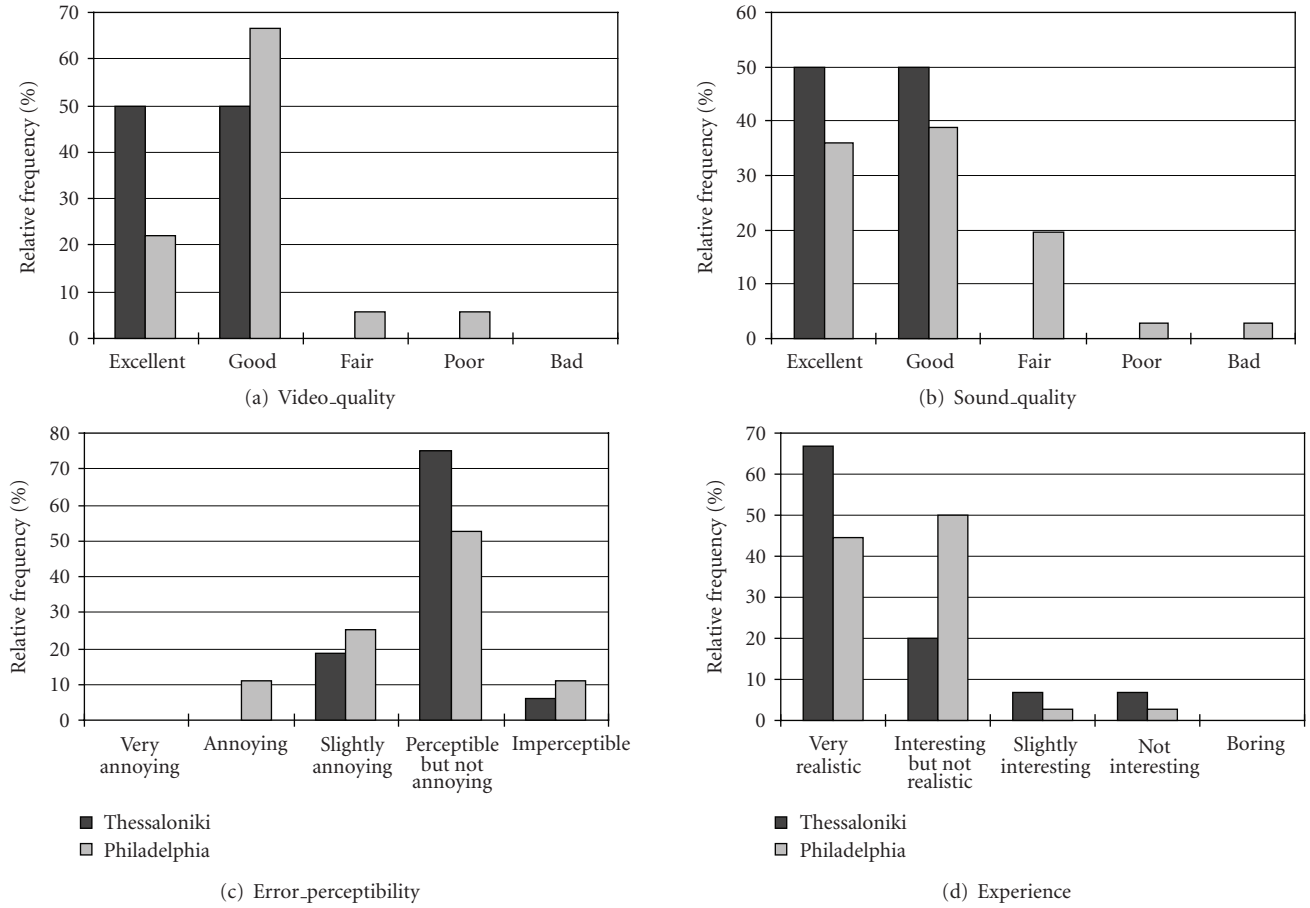

Figure 5: Qualitative evaluation results: (a) video quality (b) sound quality (c) perceptibility of errors and impairments (d) total experience.

Table 4: Mean errors and standard deviations of model for Qm_{B1}.

	720p		1080i	
	SSIM		SSIM	
Motion activity	Mean error	Error STD	Mean error	Error STD
1	−8.48932E-13	0.0940386	−1.06714E-12	0.106202
2	−3.39184E-13	0.128918	−2.24342E-11	0.140298
3	−5.36862E-13	0.148489	−1.01207E-11	0.157717
4	−1.4043E-11	0.174185	−4.34895E-12	0.170075
5	−1.26526E-12	0.157883	−2.00221E-11	0.180327

The above model presented acceptable behavior for the SSIM case as well as for the PSNR case of 1080i, but failed to adequately model the PSNR case of 720p, the fact that is evident from the mean errors and the standard deviations of the second Qm_B model fitting in Table 6. All the coefficients for each content type can be viewed in Tables 7 and 8.

Finally, as far as the end-to-end model is concerned, a function combining the results of subsystems A and B as defined in (4a) and (4b) is needed. From (2), it can be proven that the values of SSIM range from (0-1) according to the similarity of the original and the processed video frame. Moreover, the subsystems A and B are connected in cascade resulting in additive deterioration of the image quality related to the behavior of each. Based on the above remarks, the proposed model of the end-to-end system for the SSIM metric was defined by the following function:

$$\mathrm{SSIM}_{\text{e-e}} = \mathrm{SSIM}_{\mathrm{A}} \cdot \mathrm{SSIM}_{\mathrm{B}}. \tag{13}$$

On the other hand, the PSNR metric is a logarithmic measure expressed in dB as presented in (1). Therefore, the final result of the end-to-end metric is in fact the superimposition of the partial subsystems metrics, in which the minimum value is defined from the minimum partial value. This is similar to the superimposition applied to the uncorrelated sound sources for the calculation of the equivalent sound pressure level [54] as follows:

$$L_{\mathrm{EQ}} = 10 \cdot \log_{10} \sum_{k=1}^{K} 10^{L_k/10}. \tag{14}$$

In our case, the calculation formula based on (14) is transformed to the following:

$$\mathrm{PSNR}_{\text{e-e}} = -10 \cdot \log_{10}\left(10^{-\mathrm{PSNR}_{\mathrm{A}}/10} + 10^{-\mathrm{PSNR}_{\mathrm{B}}/10}\right). \tag{15}$$

The functions (13) and (15) were applied to the observations of all the HRCs defined in Table 1 for 720p, 1080i and the combination of the two and the results are summarized in Table 9.

Table 5: Coefficients for the first model of Qm_B.

Qm_B_Model1 (SSIM)						
	720p			1080i		
Motion activity	b_{s0}	b_{s1}	b_{s2}	b_{s0}	b_{s1}	b_{s2}
1	1.04583	623.413	−0.00289463	1.05583	741.162	−0.00332359
2	1.07585	985.212	−0.00301709	1.06958	799.006	−0.00404421
3	1.06874	964.321	−0.00280015	1.08267	968.93	−0.00403208
4	1.07646	919.701	−0.00321425	1.0966	1365.86	−0.00329577
5	1.09715	1303.84	−0.00290083	1.096	1076.83	−0.00420267

Table 6: Mean errors and standard deviations of model for Qm_{B2}.

	720p				1080i			
	SSIM		PSNR		SSIM		PSNR	
Motion activity	Mean error	Error STD	Mean error	Error STD	Mean error	Error STD	Mean error	Error STD
1	9.12609E-11	0.0541336	−1.63586E-9	5.32856	1.87671E-10	0.068829	7.62564E-9	2.8576
2	1.7857E-11	0.0724123	−3.20952E-9	3.55376	1.0412E-10	0.0862927	3.33925E-9	3.08074
3	9.23496E-11	0.113849	1.8E-8	14.0633	3.42642E-11	0.105127	3.10354E-9	3.00685
4	3.83357E-11	0.117794	−1.57997E-9	9.95972	6.21223E-11	0.128392	2.04578E-9	2.79959
5	−2.00853E-11	0.0963483	6.46447E-11	9.18432	2.01423E-10	0.115457	1.85578E-9	3.13145

4.2. Qualitative evaluation

After examining the whole set of possible parameters of the system that can define its performance and taking into account all the restrictions, a core system configuration was chosen which was to be under minor changes. This configuration was based in the conclusions of other relevant research, as well as in economical and practical reasons. The implementation of technologies that have already been tested and evaluated separately or respectively to other use was decisive in proceeding to real-world experiments. So the chance to evaluate system's performance by transmitting or receiving real-world events was given.

As it has already been stated, three transmissions and two receptions have taken place by us. From the first three transmissions, useful conclusions have been conducted, respectively, to the potentials of the system. Alternative formats, recording techniques and equipment were tested and the results were promising about the feasibility of the system and there have been also interest by the potential audience wherever projection has taken place in big auditorium (University of Patras). The audience accepted the projections of events with enthusiasm and the feeling of high-definition video and surround sound was noticeable and positive evaluated by everyone. The use of just one camera was evaluated positive too and in any case not monotonic, in spite of the fact that the size of the screen was smaller than the recommended.

The reception of two transmissions from the Global Concerts Series by the Philadelphia Orchestra gave us the chance to evaluate systems performance one more time and compare a different approach in a fully controlled projection site and by a survey that was conducted in a larger group of people. Although the use of just one camera was not the case in these transmissions, the rest of the transmission system was almost identical to the one we used so the survey's results can be useful for the evaluation.

The network over which the transmission has taken place provided a high bandwidth of 100 Mbps and a very high QoS. Thus, the encoder's bitrate was 18 Mbps, the transmitting format 1080i25 with MP@HL(4:2:0) profile and the packet formation protocol UDP. The other parameters for reasons that have already been mentioned were constant bitrate (CBR), IBBP format for GOP, and a relatively low length of 12 frames. The encoder that was used was Tandberg E5280 with HD expansion module, as decoder was used mainly Tandberg T1228 decoder and secondary VLC software installed in PC and a DLP-technology projector was used.

With this configuration in the cases of one transmission and one reception, 56 people participated in the survey and completed questionnaires. The subjects were not experts and were randomly selected. They were given the questionnaires before the projection or in the break and supplement it at the end. The instructions given were to read it in advance in order to ensure the worst case scenario. This is because the subjects after reading the questions concentrated and paid attention even for minor quality degradation. The subjects were asked among others to characterize the total quality of video and audio separately using a five-degree quality scale and also to evaluate the quality degradations. The scale that was used is similar to the one proposed by ITU for single stimulus continuous quality evaluation (SSCQE) [32–34, 55]. Moreover, they were asked about the total experience of watching an event by this way and had to answer how much realistic they have found the projection. Another question was, if they would like to watch another event, what was their motive to watch it and what improvements they would suggest? From the validation tests, three subjects were rejected for inconsistent answers or incomplete questions.

Table 7: Coefficients for the second model of Qm_B.

Qm_B_Model 2-720p								
	SSIM				PSNR			
Motion activity	b_{s0}	b_{s1}	b_{s2}	b_{s3}	b_{p0}	b_{sp}	b_{p2}	b_{p3}
1	−46725.1	176095	−248735	156070	736328	−2.65946E+6	3.59846E+6	−2.16142E+6
2	−10470.1	40397.6	−58424.9	37541	1.36928E+6	−5.02405E+6	6.90881E+6	−4.21977E+6
3	−46514.1	175285	−247582	155348	−7.54831E+6	2.8292E+7	−3.97343E+7	2.47832E+7
4	−120208	451941	−636803	398563	−3.19986E+6	1.19944E+7	−1.68435E+7	1.0503E+7
5	8717.2	−31553.1	42673.6	−25544.3	−26462	95574.4	−125026	70375.6

Table 8: Coefficients for the second model of Qm_B.

Qm_B_Model 2-1080i								
	SSIM				PSNR			
Motion activity	b_{s0}	b_{s1}	b_{s2}	b_{s3}	b_{p0}	b_{sp}	b_{p2}	b_{p3}
1	109072	−408678	573804	−357801	5.07703E+6	−1.89401E+7	2.64827E+7	−1.64484E+7
2	172349	−646148	907766	−566398	5.06641E+6	−1.89048E+7	2.64395E+7	−1.64257E+7
3	169289	−634746	891844	−556520	4.23374E+6	−1.57846E+7	2.2059E+7	−1.36947E+7
4	76416.1	−286047	401204	−249889	2.66162E+6	−9.88401E+6	1.37599E+7	−8.51072E+6
5	201941	−757169	1.06384E+6	−663842	4.28908E+6	−1.59895E+7	2.23434E+7	−1.38702E+7

Table 9: Mean errors and standard deviations of model for $Qm_{e\text{-}e}$.

End-to-end quality metric				
	SSIM		PSNR	
Test set	Error mean	Error STD	Error mean	Error STD
ALL	0.02	0.03	0.45	0.73
720p	0.02	0.02	0.37	0.62
1080i	0.03	0.04	0.59	0.87

The results presented in Figure 5 show that in both cases, the transmission from Thessaloniki and from Philadelphia, the subjects evaluated the quality of video as "very good" or "excellent," more than 80% and the quality of sound more than 70%. Most of the subjects, more than 60%, evaluated the degradations and impairments of the video and audio as "perceptible, but not annoying" or "imperceptible." Lastly, the total experience of the event was evaluated as "very realistic" or "interesting" by more than 70%.

It is obvious that the results are influenced by the "hollow effect." This can be explained by the fact that the audience was for the first time watching a transmission like this. However, the percentages of positive answers were very high to be caused just by that effect. Another interesting finding was that comparing the answers of the subjects answering in the first survey (transmission) the frequency of answering "Realistic" in the evaluation of the total experience was higher than the relative one in the second survey (reception). The small size of the samples in both surveys results in a big confidence interval for the percentages of answers "realistic" so we can assume that there is a trend by the audience to evaluate the use of just one camera more realistic. This can be claimed once all the other parameters were identical in both cases and the relative percentages have a quite big difference in favor of the first transmission and consequently of the one camera use. In order to prove the validity of this statement, a more elaborate survey is necessary.

Figure 6 represents the spatial distribution of subjects' answers in the case of the second survey. The respective results of the first survey cannot be evaluated because of the size of the sample. A1 is located in front of the left part of the projection screen facing the audience. It may be observed the distribution of subjects, respectively, and the evaluation of several parameters. It is obvious from Figure 6(c) that the subjects sited near the minimum viewing distance evaluated the video quality better than the ones in front or behind of them. This figure in combination with Figure 6(e) proves that sitting closer than the minimum viewing distance, the viewing experience decreases and the video errors and impairments are more perceptible. Figure 6(d) reveals that because of the poor acoustic performance of the remote auditorium, the listening experience was better close to the loudspeakers which were besides the screen and at the back of side walls.

4.3. Conclusion and further work

The purpose of the current work was to implement and evaluate HDTV over IP technologies in real-world multimedia-broadcasting applications, for the live transmission of cultural events via broadband networks (such as Internet2 and Geant backbones). Various evaluation procedures were conducted in combination with network simulations and different configuration setups, before the finally selected architecture was decided and deployed. The soundness of the current work stems from the fact that similar experimentation procedures have not been considered for specific "enhanced reality" applications, such as e-learning, teleworking, telecollaboration, and others. The proposed system

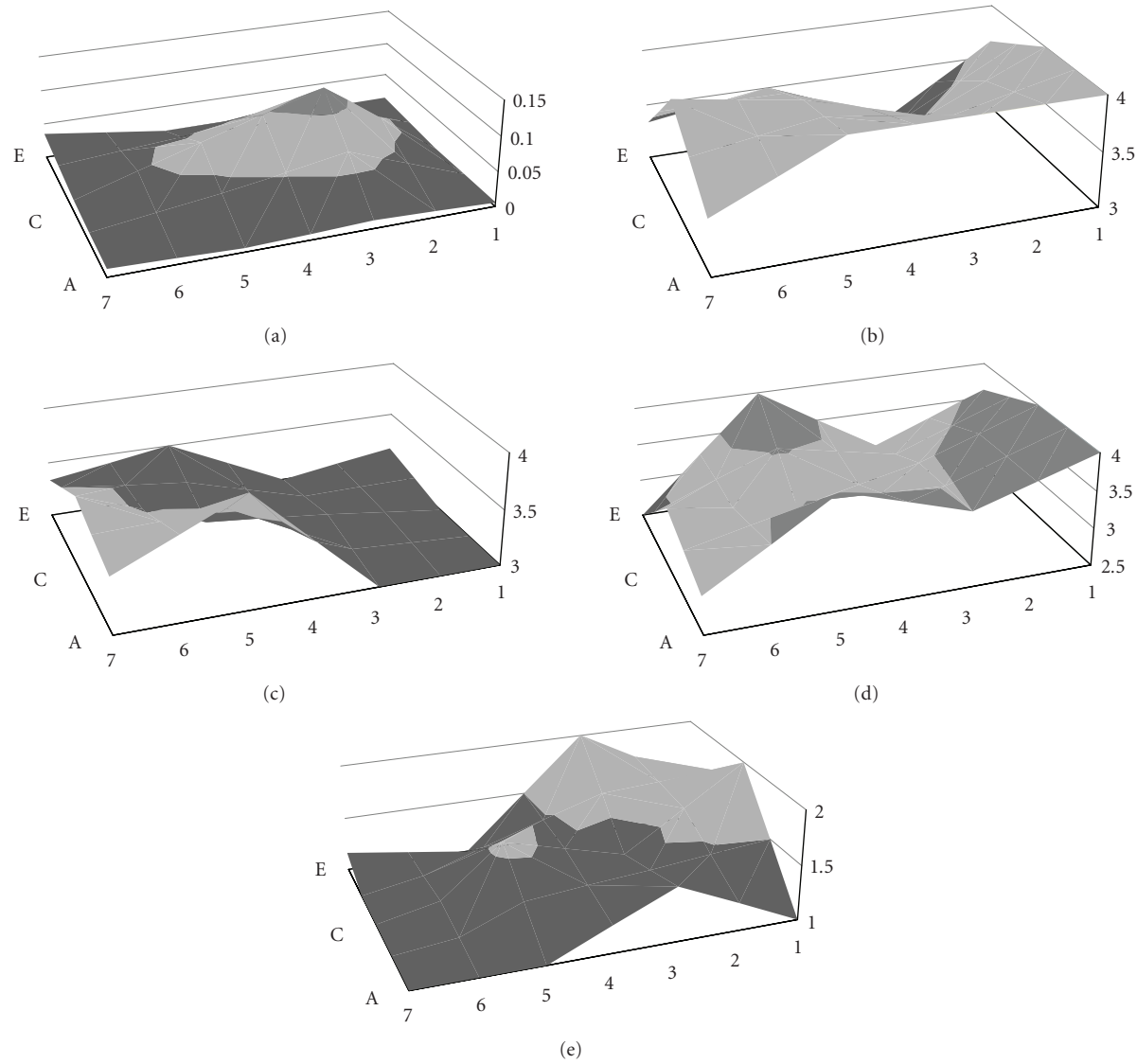

Figure 6: Subjective evaluation (5-degree level: from 0—"poor" to 4—"excellent") of the result throughout the projection venue (A1 is located in front of the left part of the projection screen facing the audience): (a) samples distribution, (b) total experience, (c) video quality, (d) audio quality, and (e) perception of specific video errors.

was enthusiastically accepted by the audience, proving to be feasible and reliable. The adopted methodology to use a single-still camera in combination with large projection screen and theater-adapted surround sound seems to provide increased realism. Given the above scenario combined with the optimization tests that we conducted, resulted to the specification of minimal requirements (bitrate, jitter) for such a task. Specifically, CBR encoding of 18 Mbps UDP in jitter conditions of 0.10 millisecond substantiated to be a minimal choice for high-quality transmission, although further experimentation could lead to more optimal utilization of network resources and increased tolerance to QoS variations. For instance, the use of additional combinations of VBR, ProMPEG FEC for today's MPEG-2, different compression formats (MPEG-4, WMV, etc.), various types and lengths of GOP, and lower/higher bitrates are under consideration. The potential of full interaction of the system is another issue.

Furthermore, a more elaborate investigation of transmission modeling was made based on simulations of hypothetical reference circuits (HRCs). The creation of a conditional model setup showed the feasibility of end-to-end performance estimation from the distinct properties of two subsystems regarding the encoding process (logarithmic curve fitting) and the transmission process (exponential mixture curve fitting), respectively. The model presented acceptable performance for all the cases except from the case of network subsystem PSNR modelling for the case of 720p which calls for further investigation. Also, a qualitative evaluation of the applied system is presented, proving the assumptions made during the design process mostly regarding the physical aspects of the project. Evolutions of

the present model could include the incorporation of such subjective tests and perceptually adapted metrics into the performance definition QoS/QoE of a system, the extension of the properties for the subsystem parameterization as well as audio performance estimation.

In any case, we may conclude that the impact of broadband networks in digital multimedia broadcasting, like the one described, will bring a new era to the cultural and educational world prospects.

ACKNOWLEDGMENTS

The authors would like to acknowledge the valuable collaboration of the Philadelphia Orchestra-Global Concert Series project team as well as the contribution of Sound Engineer Ph.D. candidate K. Kontos during the development phases of the project.

REFERENCES

[1] O. Friedrich, A. Al-Hezmi, S. Arbanowski, and T. Magedanz, "Evolution of next generation networks towards an integrated platform for IMS-based IPTV services," in *Proceedings of the Internalional Symposium on Applications and the Internet Workshops (SAINT '07)*, p. 10, Hiroshima, Japan, January 2007.

[2] O. Friedrich, A. Al-Hezmi, S. Arbanowski, and T. Magedanz, "Next generation IPTV services for an extended IMS architecture," in *Proceedings of the 8th International Symposium on Autonomous Decentralized Systems (ISADS '07)*, pp. 429–436, Sedona, Ariz, USA, March 2007.

[3] A. Al-Hezmi, O. Friedrich, S. Arbanowski, and T. Magdanz, "Requirements for an IMS-based quadruple play service architecture," *IEEE Network*, vol. 21, no. 2, pp. 28–33, 2007.

[4] A. Al-Hezmi, Y. Rebahi, T. Magedanz, and S. Arbanowski, "Towards an interactive IPTV for mobile subscribers," in *Proceedings of International Conference on Digital Telecommunications (ICDT '06)*, p. 45, Cote d'Azur, France, August-September 2006.

[5] H. J. Kang, Y. H. Jeong, and S. G. Choi, "A method of forking stream using IP multicast in HDTV media gateway," in *Proceedings of the 7th International Conference on Advanced Communication Technology (ICACT '05)*, vol. 1, pp. 393–396, Phoenix Park, South Korea, February 2005.

[6] T. Kuge, "Development of JPEG2000 HDTV program production system," in *Proceedings of the IEEE International Conference on Image Processing*, pp. 505–508, Atlanta, Ga, USA, October 2006.

[7] J. Lee and K. Chon, "Compressed high definition television (HDTV) over IPv6," in *Proceedings of the Internalional Symposium on Applications and the Internet Workshops (SAINT '06)*, pp. 22–25, Phoenix, Ariz, USA, January 2006.

[8] K. Jack, *Video Demystified: A Handbook for the Digital Engineer*, Elsevier, Oxford, UK, 4th edition, 2005.

[9] G. Steinke, "High-definition surround sound with accompanying HD picture," in *Proceedings of International Tonmeister Symposium*, Bavaria, Germany, October-November 2005.

[10] M. Sugawara, K. Mitani, M. Kanazawa, F. Okano, and Y. Nishida, "Future prospects of HDTV—technical trends toward 1080p," in *Proceedings of the the International Broadcasting Conference (IBC '05)*, Compiegne, France, September 2005.

[11] J. C. Whitaker, "Search of the new viewing experience," in *Creating Digital Content*, J. Rice and B. McKernan, Eds., pp. 351–366, McGraw-Hill, New York, NY, USA, 2002.

[12] M. Emoto, K. Masaoka, M. Sugawara, and F. Okano, "Viewing angle effects from wide field video projection images on the human equilibrium," *Displays*, vol. 26, no. 1, pp. 9–14, 2005.

[13] H. Kimio, N. Toshiyuki, H. Koichiro, and O. Kazuho, "Advanced multichannel audio systems with superior impression of presence and reality," in *Proceedings of the 116th AES Convention*, Berlin, Germany, May 2004, paper number: 6053.

[14] H. Kimio, H. Koichiro, N. Toshiyuki, and O. Reiko, "Effectiveness of height information for reproducing the presence and reality in multichannel audio system," in *Proceedings of the 120th AES Convention*, Paris, France, May 2006, paper number: 6679.

[15] H. Kimio, H. Koichiro, and O. Reiko, "The 22.2 multichannel sound system and its application," in *Proceedings of the 118th AES Convention*, Barcelona, Spain, May 2005, paper number: 6406.

[16] H. Kimio, N. Yasushige, N. Toshiyuki, and O. Reiko, "Wide listening area with exceptional spatial sound quality of a 22.2 multichannel sound system," in *Proceedings of the 122th AES Convention*, Vienna, Austria, May 2007, paper number: 7073.

[17] M. Emoto, K. Masaoka, M. Sugawara, and F. Okano, "Quantitative evaluation of sensation of presence in viewing the "Super Hi-Vision" 4000-scanning-line wide-field video system," in *Proceedings of Presence*, pp. 62–66, Valencia, Spain, October 2004.

[18] F. Okano, M. Kanazawa, K. Mitani, et al., "Ultrahigh-definition television system with 4000 scanning lines," in *Proceedings of NAB BEC*, Guobin (Jacky) Shen and J. Cai, Eds., pp. 437–440, Las Vegas, Nev, USA, April 2004.

[19] G. (Jacky) Shen and J. Cai, Eds., "Multimedia networking," *Advances in Multimedia*, vol. 2007, Article ID 97262, 1 page pages, 2007.

[20] Digital Cinema Initiatives LLC, http://www.dcimovies.com/.

[21] Internet2, http://www.internet2.org/.

[22] GEANT, http://www.geant.net/.

[23] Laboratory of Electroacoustics and TV Systems, "HDTV over IP: live broadcasting of cultural events," project report, 2007, http://avlab.ee.auth.gr/hdtv/index_en.html.

[24] The Philadelphia Orchestra, "Global Concert Series: broadcasting live concerts to your location with interactive content throughout Internet2," http://www.philorch.org/internet2_1.html.

[25] B. Pesquet-Popescu, A. Dumitras, and B. Macq, Eds., "Video analysis and coding for robust transmission," *EURASIP Journal on Applied Signal Processing*, vol. 2006, Article ID 53981, 3 pages, 2006.

[26] P. Seeling and M. Reisslein, "Evaluating multimedia networking mechanisms using video traces," *IEEE Potentials*, vol. 24, no. 4, pp. 21–25, 2005.

[27] I. E. G. Richardson, *H.264 and MPEG-4 Video Compression: Video Coding for Next-Generation Multimedia*, John Wiley & Sons, Chichester, UK, 2003.

[28] J.-F. (Kevin) Yang, H.-M. Hang, E. Steinbach, and M.-T. Sun, Eds., "Advanced video technologies and applications for H.264/AVC and beyond," *EURASIP Journal on Applied Signal Processing*, vol. 2006, Article ID 27579, 3 pages, 2006.

[29] S. Kanumuri, P. Cosman, and A. R. Reibman, "A generalized linear model for MPEG-2 packet-loss visibility," in *Proceedings of the 14th International Packet Video Workshop (PV '04)*, Irvine, Calif, USA, December 2004.

[30] O. A. Lotfallah, M. Reisslein, and S. Panchanathan, "A framework for advanced video traces: evaluating visual quality for video transmission over lossy networks," *EURASIP Journal on Applied Signal Processing*, vol. 2006, Article ID 42083, 21 pages, 2006.

[31] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[32] S. Winkler, *Digital Video Quality: Vision Models and Metrics*, John Wiley & Sons, Chichester, UK, 2005.

[33] H. Hoffmann, T. Itagaki, D. Wood, and A. Bock, "Studies on the bit rate requirements for a HDTV format with 1920 × 1080 pixel resolution, progressive scanning at 50 Hz frame rate targeting large flat panel displays," *IEEE Transactions on Broadcasting*, vol. 52, no. 4, pp. 420–434, 2006.

[34] H. Hoffmann, T. Itagaki, D. Wood, T. Hinz, and T. Wiegand, "A novel method for subjective picture quality assessment and further studies of HDTV formats," *IEEE Transactions on Broadcasting*, vol. 54, no. 1, pp. 1–13, 2008.

[35] ITU-T Tutorial, "Objective perceptual assessment of video quality: full reference television," International Telecommunication Union, Geneva, Switzerland, 2004.

[36] A. Akio, H. Kimio, I. Atsushi, et al., "Production and live transmission of 22.2 multichannel sound with ultra-high definition TV," in *Proceedings of the 122th AES Convention*, Vienna, Austria, May 2007, paper number: 7137.

[37] C. A. Dimoulas, K. A. Avdelidis, G. M. Kalliris, and G. V. Papanikolaou, "Joint wavelet video denoising and motion activity detection in multimodal human activity analysis: application to video-assisted bioacoustic/psychophysiological monitoring," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, Article ID 792028, 19 pages, 2008.

[38] K. Hamasaki, S. Komiyama, K. Hiyama, and H. Okubo, "5.1 and 22.2 multichannel sound productions using an integrated surround sound panning system," in *Proceedings of NAB BEC*, pp. 382–387, Las Vegas, Nev, USA, April 2005.

[39] N. H. Kyokai, "Reports on surround sound for radio dramas," in *ABU Technical Committee Annual Meeting*, Beijing, China, November 2006.

[40] M. A. Gerzon, "Ambisonics in multichannel broadcasting and video," *Journal of the Audio Engineering Society*, vol. 33, no. 11, pp. 859–871, 1985.

[41] SoundField, "SoundField Technology," 2008, http://www.soundfield.com/.

[42] K. Avdelidis, C. Dimoulas, G. Kalliris, and G. Papanikolaou, "Sound source localization and B-format enhancement using sound field microphone sets," in *Proceedings of the 122nd AES Convention*, Vienna, Austria, May 2007, paper Number: 7091.

[43] P. Salembier, "Overview of the MPEG-7 standard and of future challenges for visual information analysis," *EURASIP Journal on Applied Signal Processing*, vol. 2002, no. 4, pp. 343–353, 2002.

[44] C. Vegiris, K. Avdelidis, and G. Papanikolaou, "Surround sound systems implementation for the audience sense experience enhancement. An application study," in *Proceedings of the 4th National Conference of Acoustics*, Xanthi, Greece, 2008.

[45] ETSI, "Specification for the use of Video and Audio Coding in Broadcasting Applications based on the MPEG-2 Transport Stream," ETSI *TR 101 154*, DVB Document A001 Rev., 7 February 2007.

[46] ITU-T, IPTV Focus Group Proceedings, 2008.

[47] M. Mbise and J. Woods, "Multimedia interaction and a quality of experience metric based on application and network level considerations," in *Proceedings of the 13th International Packet Video Workshop (PVW '03)*, pp. 28–29, Nantes France, April 2003.

[48] M. Siller and J. Woods, "Improving quality of experience for multimedia services by QoS arbitration on a QoE framework," in *Proceedings of the 13th International Packet Video Workshop (PVW '03)*, Nantes, France, April 2003.

[49] H. Koumaras, A. Kourtis, C.-H. Lin, and C.-K. Shieh, "A theoretical framework for end-to-end video quality prediction of MPEG-based sequences," in *Proceedings of the 3rd International Conference on Networking and Services (ICNS '07)*, p. 62, Athens, Greece, June 2007.

[50] ITU-T Recommendation P.10./G.100 Appendix I, "*Definition of Quality of Experience (QoE)*," January 2007.

[51] L. Haglund, "The SVT high definition multi format test set," *Swedish Television Stockholm*, 2006, ftp://vqeg.its.bldrdoc.gov/.

[52] The Linux Foundation, "NETEmml: network emulator," 2008, http://www.linux-foundation.org/en/Net:Netem.

[53] Semaca Ltd., http://www.semaca.co.uk/content.php?produs=2.

[54] C. Dimoulas, G. Kalliris, G. Papanikolaou, and A. Kalampakas, "Long-term signal detection, segmentation and summarization using wavelets and fractal dimension: a bioacoustics application in gastrointestinal-motility monitoring," *Computers in Biology and Medicine*, vol. 37, no. 4, pp. 438–462, 2007.

[55] Recommendation ITU-R BT.500-11 (Question ITU-R 211/11), "Methodology for the subjective assessment of the quality of television pictures," June 2006.