

Research Article

Comparison of Error Protection Methods for Audio-Video Broadcast over DVB-H

Miska M. Hannuksela,¹ Vinod Kumar Malamal Vadakital,² and Satu Jumisko-Pyykkö³

¹Nokia Research Center, P.O. Box 1000, 33721 Tampere, Finland

²Institute of Signal Processing, Tampere University of Technology, P.O. Box 553, 33101 Tampere, Finland

³Institute of Human-Centered Technology, Tampere University of Technology, P.O. Box 553, 33101 Tampere, Finland

Received 1 September 2006; Revised 21 February 2007; Accepted 16 April 2007

Recommended by Anthony Vetro

The paper discusses methods for robust audio-video broadcast over the digital video broadcasting-handheld (DVB-H) system. DVB-H includes a link-layer forward error correction (FEC) scheme known as multiprotocol encapsulation (MPE) FEC, which provides equal error protection (EEP) to the transmitted media streams. Several approaches for unequal error protection (UEP) have been proposed in the literature, and the applicability of some of them to DVB-H is analyzed in the paper. A link-layer UEP method based on priority segmentation of the media streams is chosen for more detailed analysis. According to the method, audio and the most important coded video pictures are protected by MPE-FEC more robustly compared to the remaining coded pictures. In order to compare EEP and UEP in a DVB-H environment, an error-prone DVB-H channel was simulated, audio-visual clips were sent through it, and a comprehensive subjective quality evaluation was conducted in a controlled laboratory environment. The results of the subjective evaluation revealed that the use of UEP improves the subjective quality of some test clips noticeably when the channel conditions were severe, while in other tested channel conditions and clips, UEP and EEP performed equally well.

Copyright © 2007 Miska M. Hannuksela et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. INTRODUCTION

Mobile television services are expected to gain popularity in the next few years. Digital video broadcasting-handhelds (DVB-H) [1] is among the most used technical solutions for providing low interactivity, mass mobile television services. DVB-H is downward compatible with the DVB-Terrestrial (DVB-T) standard [2], thus enabling it to reuse the same network infrastructure as well as radio frequencies as used by DVB-T. The elementary transmission unit for DVB-H is a 188-byte MPEG-2 transport stream (TS) packet, specified in the MPEG-2 systems specification [3]. In contrast to DVB-T, where usually audio-video elementary streams were directly packetized to MPEG-2 TS packets, DVB-H is primarily designed for carriage of Internet protocol (IP) datagrams. In order to maintain compatibility with DVB-T, IP datagrams are packetized to multi-protocol encapsulation (MPE) sections as specified in [4], which are then carried over MPEG-2 TS packets.

FEC codes transform some number of equal-length k symbols into n symbols, where $n > k$, by adding $(n - k)$ additional symbols, called parity or repair symbols. Ideally,

an FEC code can reconstruct any $(n - k)$ corrupted symbols of the n symbols, when the location of errors is known and $(n - k)/2$ corrupted symbols when the location is not known. This property is called maximum distance separable (MDS) property and most practical FEC systems are bounded by this property. The Reed-Solomon (RS) FEC code [5] is a good example of an FEC code that follows MDS property and is used by DVB-H. Errors in wireless channels typically occur as clusters of bursts rather than isolated errors. Therefore, applications that can endure the longer latency time required for FEC computing are better suited to use the DVB-H transmission.

DVB-H adds additional link-layer features to solve the power constraint and robustness problems associated with handheld mobile terminals. The concept of time-slicing was introduced, reducing the average power consumption of a hand-held mobile terminal by as much as 90–95%. An optional enhancement using Reed-Solomon forward error correction (FEC) codes encapsulated into multiprotocol encapsulated sections (MPE-FEC) was also introduced to provide added error robustness required for hand-held mobile terminals.

Even though DVB-H can convey any IP datagrams, the audio and video codecs for IP-based broadcasting are specified in [6] to facilitate interoperability of DVB-H service providers and receivers. The high efficiency advanced audio coding version 2 (HE AAC v2) [7] is recommended for audio compression, and advanced video coding (H.264/AVC) [8] is recommended for video compression. A number of profiles are specified in H.264/AVC. A profile consists of a subset of the algorithmic features or coding tools of the standard and a set of constraints on those features. A profile is typically targeted for a family of applications sharing similar trade-off between memory, processing, latency, and error resiliency requirements. Decoders conforming to a profile must support all the features of a profile. Five IP integrated receiver-decoder (IRD) capabilities are specified in [6] to facilitate service tailoring for different types of terminals. IP-IRD capabilities for battery-powered devices require the support of H.264/AVC baseline profile with the `constraint_set1_flag` syntax element of H.264/AVC being equal to 1, which is also referred to as the constrained baseline profile.

Unequal error protection (UEP) takes advantage of the fact that different portions of the coded bit stream have different levels of importance to the overall subjective quality of the presentation. UEP aims at providing graceful degradation of subjective quality under harsh transmission conditions and hence the overall quality of all recipients in any transmission conditions is expected to improve in comparison to the quality obtained with equal error protection (EEP). When applied to coded video, UEP requires that video bit streams be partitioned to segments of different priorities according to the segments' impact to subjective quality. Segments are then protected with unequal amount of FEC repair data. The priority partitioning methods can be roughly categorized into data partitioning, region-of-interest prioritization, spatial, quality, and temporal layering.

This paper uses only temporal layering for priority assignment. This is because the goal of the design was to maintain H.264/AVC constrained baseline profile compatibility and using other types of priority partitioning would have required more advanced H.264/AVC profiles support or the scalable extension of H.264/AVC (under development). Temporal layering refers to the encoding of a temporally scalable bit stream. Any bit stream can be partitioned into two temporal layers, one that contains the intra pictures only, and another containing the remaining ones. Many video coding schemes enable nonreference pictures, which are not used for inter prediction of any other picture. Modern video coding standards such as H.264/AVC also enable hierarchical temporal scalability, in which subsequences of coded pictures, including also reference pictures, can be removed from a bit stream without affecting the decoding of the remaining bit stream. It has been shown that temporal scalability improves compression efficiency [9] even with the constrained baseline profile of H.264/AVC, which does not include bi-predictive pictures (also known as B pictures).

In this paper, we analyze which methods for UEP can be applied to DVB-H in a straightforward manner without substantial changes in the system. In addition, we compare the

UEP method that we found the most applicable with the EEP scheme provided by MPE-FEC in different radio conditions.

The rest of this paper is organized as follows. Section 2 reviews the DVB-H protocols and system to an extent that is necessary for understanding of this paper. Section 3 provides an overview of those features of H.264/AVC and its packetization format for real-time transport protocol (RTP) that are essential for the presented UEP method. A brief review of some UEP methods is provided in Section 4 and their applicability to DVB-H is analyzed. Furthermore, one of the reviewed UEP methods is presented in more details in Section 4. The operation of the conventional MPE-FEC-based EEP method and the presented UEP method was simulated in a DVB-H environment and the resulting audio-visual test clips underwent a subjective viewing test. The simulation and test setup is presented in Section 5, and the results are analyzed in Section 6. Finally, Section 7 concludes the paper.

2. OVERVIEW OF DVB-H PROTOCOLS AND SYSTEM

This section introduces the fundamentals of DVB-H and is organized as follows. Section 2.1 presents the protocol stack of DVB-H. The FEC coding of DVB-H is reviewed in Section 2.2. Finally, the method for time-slicing is explained in Section 2.3.

2.1. DVB-H protocol stack

The protocol stack for DVB-H is presented in Figure 1. IP packets are encapsulated to MPE sections for transmission over DVB protocols in the medium access (MAC) sublayer. Each MPE section consists of a header, the IP datagram as a payload, and a 32-byte cyclic redundancy check (CRC) for the verification of payload integrity. The MPE section header contains addressing data among other things. The MPE sections can be logically arranged to application data tables in the logical link control (LLC) sub-layer, over which RS FEC codes are calculated and MPE-FEC sections are formed. The process for MPE-FEC construction is explained in more detail in Section 2.2. The MPE and MPE-FEC sections are mapped onto MPEG-2 TS packets.

2.2. MPE-FEC

MPE-FEC was included in DVB-H to combat long burst errors that cannot be efficiently corrected in the physical layer. MPE-FEC is based on the Reed-Solomon FEC coding. Since Reed-Solomon code is a systematic code, that is, the source data remains unchanged after FEC encoding, MPE-FEC decoding is made optional for DVB-H receivers. MPE-FEC is computed over IP packets and encapsulated into MPE sections. MPE-FEC sections are transmitted such that an MPE-FEC ignorant receiver could just receive the unprotected data while ignoring the protection data that follows.

To compute MPE-FEC, data (IP packets) are filled into an $(N \times 191)$ matrix where each cell of the matrix hosts one byte of information and N denotes the number of rows in the matrix. The standard defines the value of N to be one of

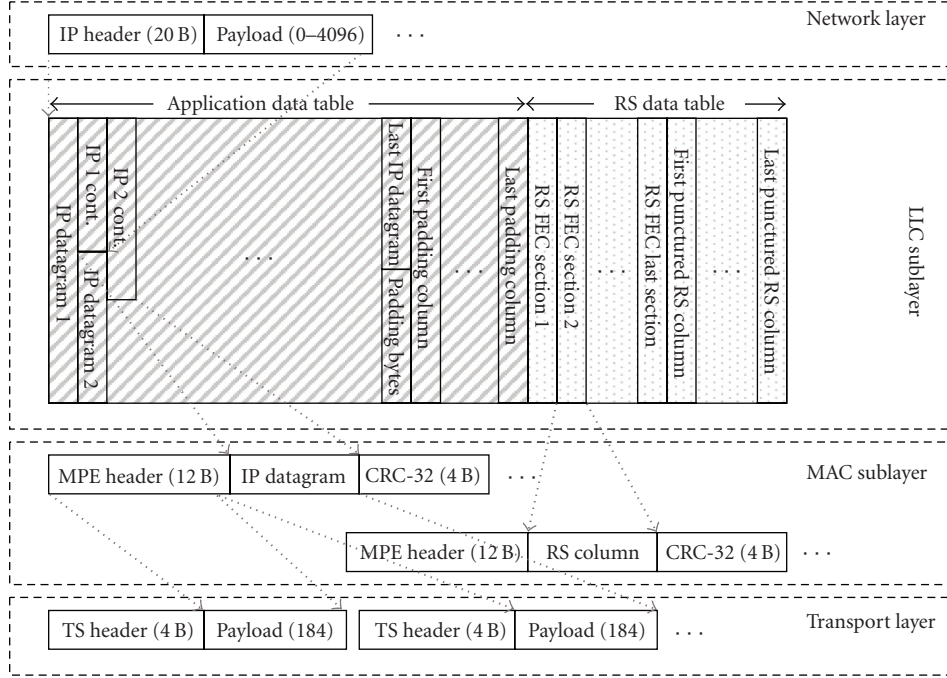


FIGURE 1: A subset of the protocol structure of DVB-H.

256, 512, 768, or 1024. The datagrams are filled into the matrix columnwise. RS codes are computed for each row and concatenated such that the final size of the matrix is of size $(N \times 255)$. The $(N \times 191)$ part of the matrix is called the application data table (ADT) and the adjacent $(N \times 64)$ part of the matrix is called the RS data table (RSDT). For rate control and disallowing of IP packet fragmentation between two MPE-FEC frames in the standard, the ADT need not be completely filled. This unfilled part of the ADT is called padding. To control channel coderate, all 64 columns of RSDT need not be transmitted, that is, the RSDT may be punctured. The structure of an MPE-FEC matrix is shown in Figure 2 and further information on the MPE-FEC matrix construction can be obtained from [4].

2.3. Time slicing

Battery-operated mobile devices have a limited source of power. The power consumed in receiving, decoding, and demodulating a standard full-bandwidth DVB-T signal would use up substantial amount of battery life in a short time. Time slicing of the MPE-FEC frames is used to solve this problem [10]. The data is received in bursts so that the receiver, utilizing control signals, remains inactive when no bursts are to be received. The bursts are sent at a significantly higher bit rate compared to bit rate when conventional bit rate management is used.

Time slicing in DVB-H uses the Delta-T method to signal the relative start of the next burst, that is, the difference between the current time and the start of the next burst. The use of Delta-T method provides flexibility since parameters

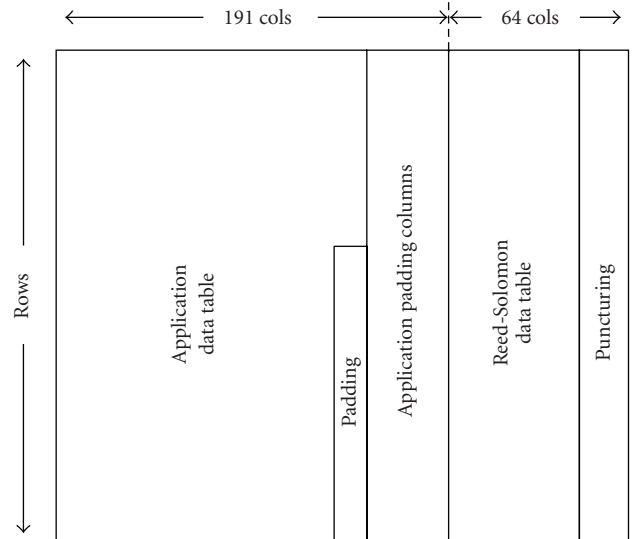


FIGURE 2: The MPE-FEC matrix structure.

such as burst size, burst duration, burst bandwidth, and the offtimes can be freely varied. Figure 3 shows two time-sliced bursts and parameters that define time-sliced bursts.

3. H.264/AVC VIDEO CODING AND RTP ENCAPSULATION

H.264/AVC enables storage of multiple reference pictures for inter prediction and selection of the used reference picture on

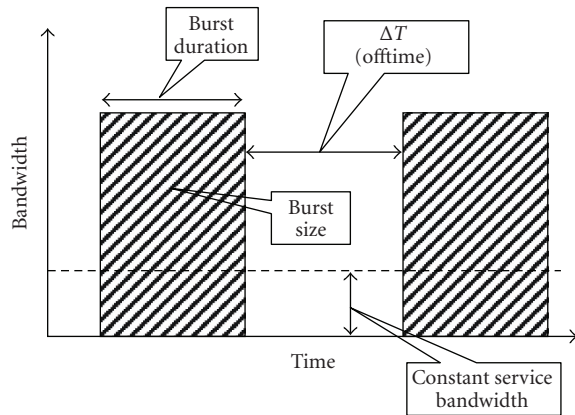


FIGURE 3: Time slicing in DVB-H.

macroblock or macroblock partition basis. In order to maximize compression efficiency, a motion vector is accompanied by a variable-length-coded index to a reference picture list. The reference picture list is initialized according to picture decoding order for inter slices and according to picture output order for bi-predictive slices. Slice headers may contain commands for reference picture list reordering.

Coded pictures of H.264/AVC can be categorized into three types: instantaneous decoding refresh (IDR) pictures, other reference pictures, and nonreference pictures. An IDR picture contains only intra-coded slices and causes marking of all previous reference pictures to be no longer used as references for subsequent pictures. An IDR picture can therefore be used as a random access point for the start of decoding or joining a session. It also provides a resynchronization point for decoding after transmission errors have occurred. A reference picture is stored and maintained as a prediction reference for inter prediction until it is no longer used for reference according to the reference picture marking process of H.264/AVC. A non-reference picture is not used for reference in inter prediction and can therefore be removed from a bit stream without any effect on other pictures.

The elementary unit for the output of an H.264/AVC encoder and the input of an H.264/AVC decoder is a network abstraction layer (NAL) unit. For transport over packet-oriented networks or storage into structured files, NAL units are typically encapsulated into packets or similar structures. NAL units can be categorized into video coding layer (VCL) NAL units, such as coded slices, and non-VCL NAL units, such as sequence and picture parameter sets.

The RTP payload format specification for H.264/AVC [11] includes the syntax and semantics of the RTP payload format, RTP packetization rules for H.264/AVC, informative RTP depacketization guidelines, and multipurpose Internet mail extensions (MIME) definition for use with session description protocol (SDP), including SDP offer-answer model consideration for codec capability exchange. The payload format specification contains three packetization modes: single NAL unit mode, noninterleaved mode, and interleaved mode.

In the single NAL unit packetization mode, one NAL unit is transmitted without any additional payload header in one RTP packet. In the non-interleaved mode, NAL units are transmitted in decoding order and multiple NAL units of one access unit can be encapsulated into the same RTP packet. Encapsulating multiple NAL units into the same RTP packet is especially beneficial when the size of the NAL units is relatively small, which is typically the case for parameter set NAL units, for example. The non-interleaved mode therefore helps to reduce the bit rate overhead caused by protocol headers compared to the transmitting relatively small NAL units with the single NAL unit mode.

The interleaved mode allows transmission of NAL units out of NAL unit decoding order and encapsulating of NAL units from different access units into the same RTP packet. In the interleaved mode, a decoding order number (DON) indicating the decoding order of NAL units is conveyed or derived for each NAL unit. In very low bitrates the interleaved packetization mode allows for encapsulating NAL units from more than one access unit into the same packet, which helps to reduce protocol header overhead. The interleaved mode can also be used for robust packet scheduling for unicast streaming [12, 13]. When interleaved transmission order is used, the decoding order of NAL units must be recovered in the receiver to obtain correct operation of the decoder. The receiver includes a receiver buffer to reorder packets from transmission order to the NAL unit decoding order.

4. UEP METHODS AND THEIR APPLICABILITY TO DVB-H

Priority encoding transmission (PET) [14] established the work towards UEP in packet-oriented systems. The data to be transmitted is partitioned to messages, which are protected one at a time. The messages are then classified to priority segments according to known characteristics of the source signal. For example, a group of pictures (GOP) can be considered as a message, and priority segments can be assigned according to the picture type (I, P, B) [15]. FEC repair data is then generated for each priority segment, and the resulting coded stream is divided into a certain amount of packets, each containing a fixed-length block of data from the resulting coded stream. The amount of FEC repair data is a function of the priority class. The PET scheme results into packets which contain data from each priority segment, and the number of packets required to reconstruct a priority segment can be tuned with the amount of FEC repair data for each priority segment. Horn et al. developed a similar scheme [16] compared to PET and provided details on the practical implementation and application with a spatially scalable video codec.

IETF RFC 2733 [17] specifies an RTP payload format for XOR-based FEC protection. The payload header of FEC packets contains a bit mask identifying the packet payloads over which the bitwise XOR operation is calculated and a few fields for RTP header recovery of the protected packets. One XOR FEC packet enables recovery of one lost source packet. Work is going on to replace IETF RFC 2733 with similar RTP

payload format for XOR-based FEC protection also including the capability of uneven levels of protection (ULP) [18]. The payloads of the protected source packets are split into consecutive byte ranges starting from beginning of the payload. The first byte range starting from the beginning of the packet corresponds to the strongest level of protection and the protection level decreases as a function of byte range order. Hence, the media data in the protected packets should be organized such a way that the data appears in descending order of importance with a payload and a similar number of bytes correspond to similar subjective impact in quality among the protected packets. The number of protected levels in FEC repair packets is selectable and an uneven level of protection is obtained when number of levels protecting a set of source packets is varied. For example, if there are three levels of protection, one FEC packet may protect all three levels, a second one may protect the two first levels, and a third one only the first level.

Both PET and the method proposed by Horn et al. produce packets in an interleaved manner such that they contain data of all priority classes as well as repair data. The packet transmission format therefore requires deinterleaving of payload data even when FEC decoding is not necessary. Furthermore, the packet formats are not compatible with any of the existing standards.

RFC 2733 and ULP operate in application layer and are therefore unable to utilize MPE-FEC efficiently. Both RFC 2733 and ULP are based on XOR, which is known to be clearly inferior to Reed-Solomon FEC when the size of the FEC matrix is relatively large. RFC 2733 and ULP also limit the FEC matrix to a size that may be too small for being efficiently used when applied to DVB-H.

We proposed a UEP scheme first for the 3GPP's multimedia broadcast/multicast service (MBMS) [19] but later specifically tailored for DVB-H [20]. The scheme classifies multimedia data to priority segments and computes an uneven amount of FEC repair data over priority segments similarly to what is done in PET and many subsequent UEP methods. However, in contrast to earlier methods, the packet format remains identical to the case in which EEP is applied. This maintains compatibility with terminals that are not capable of UEP data reception. Furthermore, MPE-FEC is reused instead of introducing any new FEC and packetization scheme at the application layer. Therefore, this method of UEP incurs a small amount of implementation changes compared to the existing DVB-H implementations. In other words this UEP scheme can be considered as a DVB-H-friendly version of PET and the method proposed by Horn et al.

The method proposed in [20] is briefly described next. First, the priority segmentation is performed across all media streams of the same service. In this paper, the audio stream is ranked as high priority, and for video we utilize temporal layering only. It is proposed that H.264/AVC bit streams are encoded in a temporally scalable manner and priority is assigned to temporal level of the pictures. For example, if non-hierarchical temporal scalability is used, that is, one or more non-reference pictures are present between each pair of refer-

ence pictures, the reference pictures can be assigned a higher priority compared to the non-reference pictures.

The multiplexed media datagrams corresponding to certain duration are encapsulated into two or more MPE-FEC matrices according to their priority label. These MPE-FEC matrices are referred to as peer MPE-FEC matrices. The number of peer MPE-FEC matrices in a time-sliced burst is equal to the number of unique priority labels assigned to the datagrams.

To construct the peer MPE-FEC matrices in a time-sliced burst, the datagrams are grouped using their priority labels. The grouping procedure is performed on all the datagrams that go into the time-sliced burst. The grouped datagrams are arranged in ascending order such that the datagrams with the lowest priority come first in the transmission order and the datagrams with the next higher priority comes next and continuing so forth until the datagram group that has the highest priority comes last in the transmission order. Figure 4 illustrates the priority grouping of a service consisting of a temporally scalable video stream and an audio stream. The audio stream and the reference pictures of the video stream are assigned the highest priority, whereas the non-reference pictures are grouped to low-priority MPE-FEC matrices.

The number of RSDT columns for all the MPE-FEC matrices in all the time-sliced bursts in the service should be such that the average service bit rate when using this method will not overshoot the maximum allowed service bit rate. All peer MPE-FEC matrices should be recoverable in normal channel conditions, and in bad channel conditions at least the high priority peer MPE-FEC matrix should be recoverable. Padding and puncturing are used to obtain the desired MPE-FEC code rates.

The estimation of code rates for varying channel error conditions is difficult in DVB-H. Firstly, due to the broadcast nature of the channel some users might be experiencing extremely harsh conditions, while at the same time other users might be having an excellent reception. If a transmitter, sending a service at a single code rate, caters to really harsh channel conditions by using a very low code rate, then there is an inefficient use of bandwidth for users having good reception. On the other hand if the transmitter sends a service at a high code-rate, making efficient use of the bandwidth, the capability of the receivers to receive and decode the service data under bad reception conditions is substantially reduced. Catering to both these groups optimally requires knowledge of the number of users having bad reception versus number of users having good reception. This again is a difficult task because DVB-H by its own does not provide any return channel. However, best practices for adjusting the code rate for sufficient reception quality on average can be derived from network measurement statistics or simulated channel models. For example, in [21] the rate distortion at different error rates for H.264/AVC was evaluated, and the code rate of 3/4 was shown to be most efficient among the tested cases. This code rate was used in the simulations performed in this paper.

In order to obtain identical receiver power consumption compared to conventional data casting over DVB-H, the peer

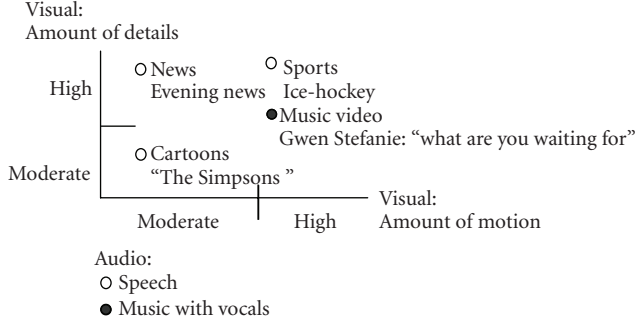


FIGURE 6: Genre of stimuli sequences, contents, and their audio-visual characteristics.

in measuring visual quality in video compression studies, provides consistent results only as long as the video signals being compared are affected by the same type of impairment [22]. Thus, subjective tests were carried out in a controlled laboratory environment to compare EEP provided by MPE-FEC and the UEP method presented in the previous section. Recommendations by International Telecommunication Union (ITU) [23, 24] were modified because no subjective test methodology in literature tuned specifically for this kind of work was found. The audio-visual bit streams presented to the subjective test participants were prepared by simulating a DVB-H channel.

5.1. Participants

45 participants, equally stratified by age group (18–45 years) and gender participated in the quality evaluation experiment. The number of experienced assessors, people engaged in multimedia processing or having extremely positive attitude towards technology [25] was restricted to 20%. All participants were verified to have normal or corrected-to-normal vision and hearing.

5.2. Test material selection and encoding

Four stimuli sequences representing different genre and contents with different audio-visual characteristics were chosen from a set of television broadcast material as described in Figure 6. The duration varied from 61 seconds to 64 seconds, because it was desirable to have semantically complete, meaningful, and understandable sequences for the participants.

The selected test materials were encoded using recommended codecs for the IP data casting service over DVB-H. Advanced audio coding (AAC) was used for audio and H.264/AVC for video encoding. The bit rate, sampling rate, and frame rate were selected according to the results of a previous study [26]. Mono-aural audio, which in [27] is shown to be more preferred than stereo at low bit rates, was coded at a bit rate of 32 kbps with a sampling rate of 16 kHz. Video bit streams were coded at a picture size of 176×144 pixels, a bit rate of 128 kbps, and a frame rate of 12.5 frames per second. Two sets of video sequences were encoded. The first

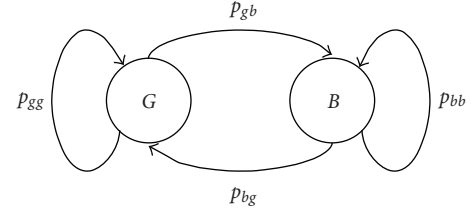


FIGURE 7: Gilbert-Elliot error model.

set of sequences was targeted for the conventional method for audio-video broadcast over DVB-H and therefore contained only reference pictures. The second set of sequences was targeted for the proposed UEP scheme and therefore two non-reference pictures were coded between each pair of reference pictures. In both sets of sequences, at least one IDR frame was coded per DVB-H time slice to reduce the tuning-in delay at the receiver and provide better error resiliency against residual transmission errors. The first set of sequences was conventionally protected with MPE-FEC code rate of 3/4. For the second set of sequences, two MPE-FEC peer matrices were generated as described in Section 4, and the high-priority MPE-FEC peer matrix had a code rate of 3/4 while the low priority MPE-FEC peer matrix was unprotected by MPE-FEC. The time-sliced transmission burst interval for all sequences was set to approximately 1.5 seconds. This choice of code rates for the peer MPE-FEC matrices was chosen based on experimentation. It was found that under such harsh channel conditions as simulated in this paper, the best subjective quality was obtained when all the protection was dedicated to the most important priority while leaving the low-priority data unprotected.

5.3. Channel simulation

Various stochastic models have been proposed for simulation of errors in a wireless channel. Among these, the Gilbert-Elliot (GE) model [28], shown in Figure 7, is popular and widely used because of its simplicity while it still produces a good representation of errors in a wireless channel. The GE model has been confirmed useful for simulating the packet error behavior also in DVB-H [29].

The model consists of two states representing two different channel conditions: the good state G and the bad state B . Each of these states is associated with bit error probabilities: e_g in the good state and e_b in the bad state where $e_g \ll e_b$. The average lengths of the error bursts are determined by the state transition probabilities p_{gb} , p_{bg} between the two states and the bit error probabilities e_g and e_b . In a simplified GE model e_g and e_b are set to zero and one, respectively. The state transition matrix T is then given by the matrix

$$T = \begin{bmatrix} p_{gg} & p_{gb} \\ p_{bg} & p_{bb} \end{bmatrix}. \quad (1)$$

To simulate loss in the DVB-H channel, the results of a field trial carried out in an urban environment with an operable DVB-H system were used as basis. The receiver in the

field trials was located in a car, and the modulation used was 16 QAM. The field test results were used to train a simplified GE model for erroneous time-slices and estimate the state transition matrix.

The field test results were in the form of an MPE-FEC error pattern indicating which MPE-FEC frames contained uncorrectable transmission errors. This error pattern was first used as a training sequence for a simplified GE model resulting into the following state transition matrix:

$$T_{\text{mpe-fec}} = \begin{bmatrix} 0.8478 & 0.1522 \\ 0.4227 & 0.5773 \end{bmatrix}. \quad (2)$$

The state transition matrix was then used to generate an initial MPE-FEC error pattern. Finally, the length of randomly selected error bursts in the initial MPE-FEC error pattern was reduced gradually until error patterns of rates 6.9% and 13.8% were obtained.

MPE-FEC frame error rates (MFER) 6.9% and 13.8% after FEC decoding were chosen into the simulations based on an earlier test [30], in which the boundary of overall acceptability lied between these two rates, that is, the majority of participants considered the audio-visual quality resulting from 6.9% and 13.8% erroneous time-slice rate acceptable and nonacceptable, respectively. It is emphasized that the tested error rates are significantly higher than expected typical error rates for DVB-H services. The aim of the tests was to study the operation of audio-video broadcasting over DVB-H under extreme channel conditions. It is noted that MFER 5% has been conventionally used as an operative quality of restitution (QoR) limit for mobile reception [31].

To generate the error patterns for the transport stream (TS) packets within the uncorrectable MPE-FEC frames, a second simplified GE model was implemented. Based on manual assessment of some TS error patterns, we assumed that the average total number of TS packet errors was 235 and the average error burst length was 95 continuous TS packets. In a simplified GE model the average error rate E is given by $E = (1 - p_{gg})/(2 - p_{gg} - p_{bb})$ and the average burst error lengths B is given by $B = 1/(1 - p_{bb})$. Solving for p_{gg} and p_{bb} a state transition matrix

$$T_{\text{ts}} = \begin{bmatrix} 0.99 & 0.01 \\ 0.01 & 0.99 \end{bmatrix} \quad (3)$$

was obtained, which was used to generate the TS error patterns within an erroneous MPE-FEC frame. The result was a TS error pattern that approximated the results of the actual field test.

The generated TS packet errors were used to corrupt the coded audio-visual sequences. Error correction operation using MPE-FEC was simulated and the resulting residual IP packet error pattern was obtained. The residual IP error pattern reflected the uncorrectable errors in the channel.

5.4. Decoder error concealment

The video decoder used a simple error concealment procedure. When the decoder encountered residual errors in or

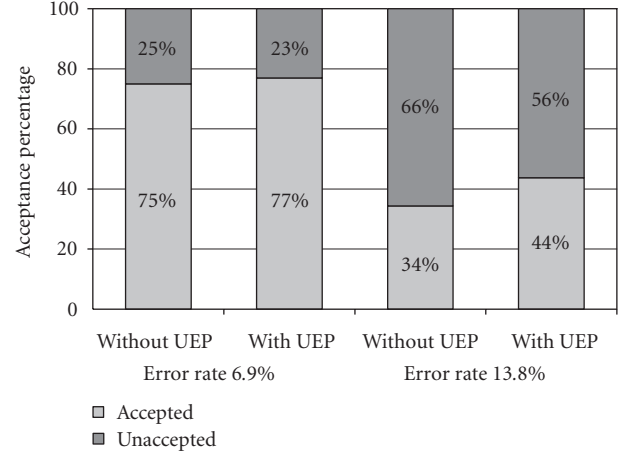


FIGURE 8: Overall acceptability rating of UEP scheme.

losses of reference pictures, it stopped decoding of any subsequent pictures until an IDR picture arrived. During the period when the decoder stopped decoding, it presented the last uncorrupted decoded picture. Subjectively, when this method is used, a transmission error is perceived as discontinuous motion in visual streams. The duration of these discontinuities in visual streams depends on the IDR interval and the placement of the error between two IDR pictures. When the decoder encountered losses of non-reference pictures, the previous correct picture in output order was rendered and decoding continued from the next picture in decoding order. Consequently, if residual errors were present in the peer MPE-FEC matrix for the non-reference pictures but not present in the corresponding peer MPE-FEC matrix for audio and reference pictures, users perceived temporary fluctuations of picture rate, that is, jerky but generally continuous motion.

AAC audio frames are essentially independent of each other and a loss of any one frame of the bit stream does not substantially affect any other frames of an audio channel. When an audio frame was lost, it was replaced with a null frame perceived as discontinuous audio.

5.5. Subjective test procedure

Before the start of the test session, the participants were briefed about the test and their sensorial acuity was measured and they filled the demographic questionnaire. The sensorial tests included in the measurements of visual acuity (20/40), color vision [32, 33], and the aural acuity [34–36].

The subjective test started with a combination of anchoring and training. Participants were shown the extremes of quality range of stimuli to familiarize the participants with the test task, the contents, and the variation in quality they could expect in the actual tests that followed. The tests used retrospective overall evaluation based on the absolute category rating (ACR), also known as single stimulus method, which is typically used in system or performance evaluation [24]. The test sequences were presented one at a time and

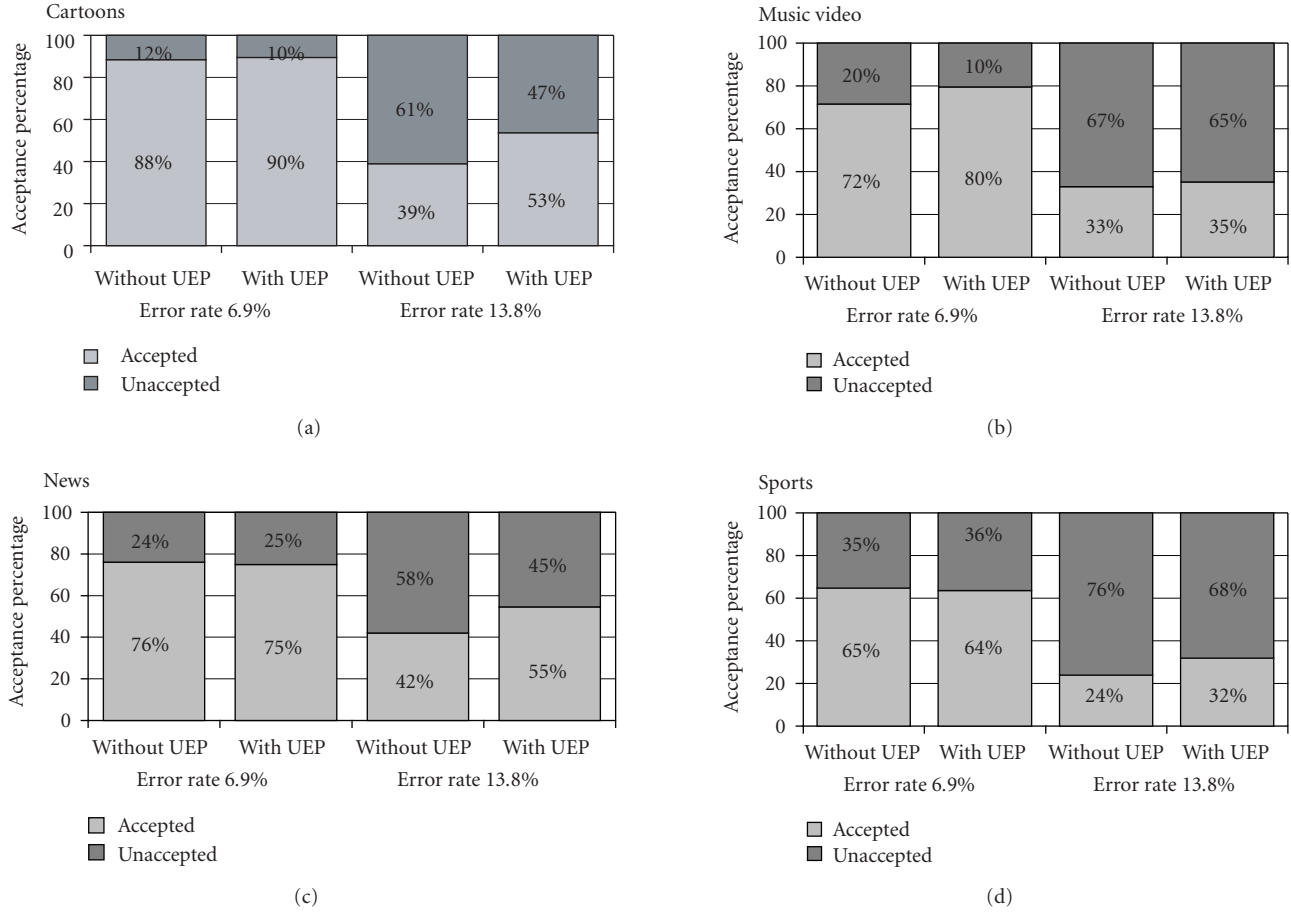


FIGURE 9: Per-sequence acceptability ratings.

they are rated independently after each presentation [24]. The quality ratings were given during a 5-second-long answering time by using a discrete, unlabelled 11-point scale and the acceptance of quality (yes/no choice). The whole test session for a participant consisted of two rounds with two sets of audio-visual clips [A, B] and the starting round was randomized. After the actual test, qualitative data of experiences on the erroneous streams were gathered. One test session lasted about 1.5 hours.

The clips were presented with Nokia 6630 mobile phone, which was enclosed in a stand that left only the screen and buttons of the device visible. The device and the front of the stand were vertically aligned and the viewing distance was set to 44 cm. The headphones delivered in Nokia 6630 sales package were used for audio playback. Audio playback loudness level was adjusted to 75 dB(A) (+ 10 dB(A) for peaks).

5.6. Data analysis methods

For data analysis, two different nonparametric methods were used. Overall quality ratings were analyzed with Wilcoxon matched pair signed rank test which was used to measure the differences between two related and ordinal data sets because

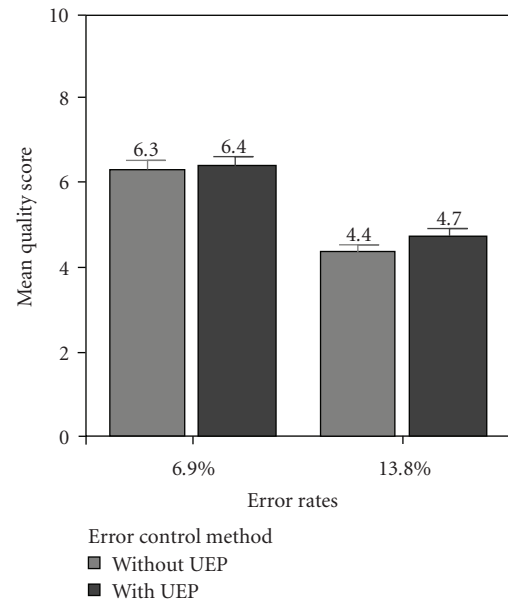


FIGURE 10: Overall mean satisfaction ratings for UEP scheme. The error bars show 95% CI of mean.

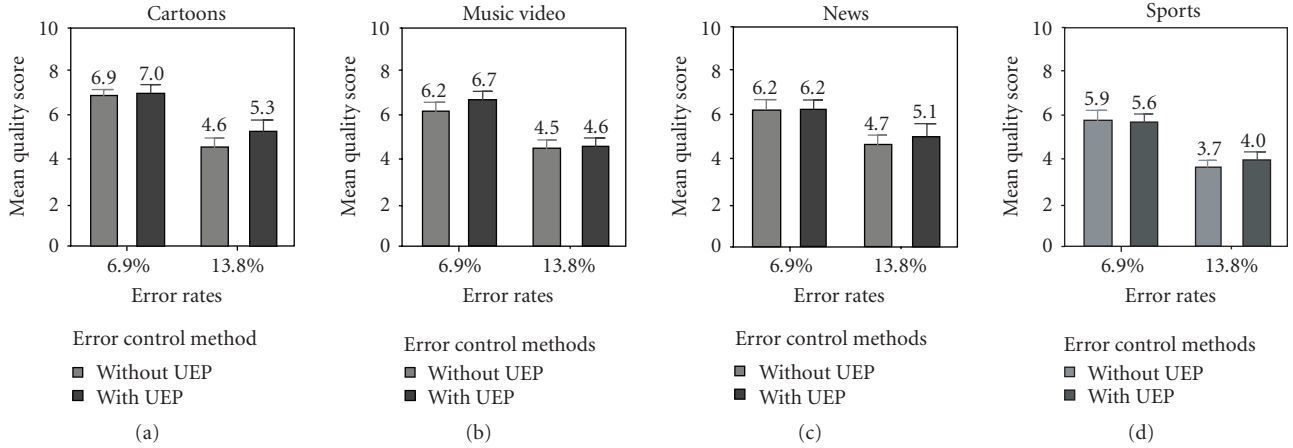


FIGURE 11: Per-sequence satisfaction ratings for UEP scheme. The error bars show 95% CI of mean.

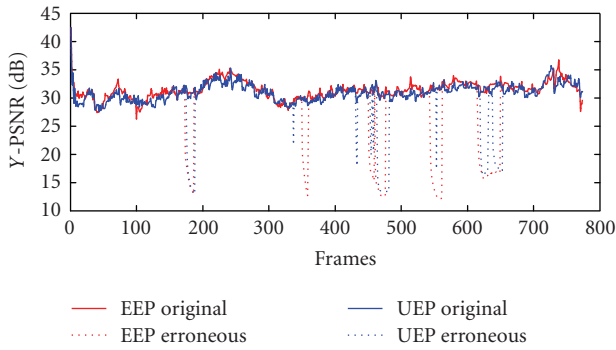


FIGURE 12: Per-frame PSNR for sports sequence at 13.8% MFER.

the preassumption of parametric methods (normality) was not filled [37]. For the nominal acceptance evaluations McNemar's test was applied to test the differences between two categories in the related data [37]. The significance level of $P < .05$ was adopted in this study.

6. RESULTS

Figure 8 shows the cumulative acceptability statistics and Figure 10 shows mean satisfaction scores for all audio-visual sequences at the two simulated error rates. When the residual time slice error rate was 6.9%, the proposed UEP method did not have a significant impact on overall acceptance or satisfaction rating compared to the conventional method (McNemar $P > .05$, Wilcoxon $Z = -0.71$, $P > .05$). A majority of participants rated sequences of both error control methods as acceptable. When the residual time slice error rate was 13.8%, the proposed UEP method outperformed the conventional method significantly (McNemar $P < .001$, Wilcoxon $Z = -4.1$, $P < .001$), which can also be seen in the number of acceptable clips in Figure 8. However, on average, the sequences of both the proposed UEP method and the conventional method remained unacceptable.

Figures 9 and 11 show the acceptability and mean satisfaction statistics for each of the four audio-visual sequences at 6.9% and 13.8% residual MPE-FEC time slice error rates, respectively. At the error rate of 6.9%, the improvement provided by the proposed UEP method was not significant in any sequences (McNemar, Wilcoxon $P > .05$). However, at the error rate of 13.8% the proposed UEP scheme outperformed the conventional scheme significantly in animation (McNemar $P < .01$, Wilcoxon $Z = -3.7$, $P < .001$), news (Wilcoxon $Z = -2.0$, $P < .05$), and sports (McNemar $P < .05$). Moreover, a majority of participants rated the animation and news sequences of the proposed UEP scheme as acceptable under residual time slice error rate of 13.8%, whereas the corresponding conventionally coded and transmitted sequences were rated as unacceptable by a majority of participants. In other words, the threshold for a transmission error rate yielding an unacceptable audio-visual quality was increased due to the proposed UEP scheme.

Figure 12 shows the per-frame PSNR behavior for the sports sequence at 13.8% MFER for both EEP and UEP. It clearly illustrates how some burst errors in the EEP case can be transformed into isolated single picture errors in the UEP case.

7. CONCLUSIONS

The paper reviewed some methods for unequal error protection (UEP) and analyzed their applicability to DVB-H. A method based on priority segmentation of the media streams of a service was chosen for more detailed analysis. The presented UEP method was compared to equal error protection (EEP) provided by the link layer forward error correction scheme (MPE-FEC) of DVB-H. Several audio-visual streams were processed through a DVB-H channel model for the comparison, and the resulting streams were presented in a comprehensive subjective quality evaluation conducted in a controlled laboratory environment. Two MPE-FEC error rates (MFER) were selected for the evaluation, 6.9% and

13.8%, which resulted into acceptable and unacceptable average quality, respectively, according to a previous study. The results of the evaluation revealed that, at MFER of 6.9%, the presented UEP scheme was at least as good as the EEP case obtained by conventional use of MPE-FEC. However, at MFER of 13.8%, the use of the proposed UEP method improved the subjective acceptability of the tested multimedia sequences on average, as the share of participants rating the sequences acceptable was 10 percent units higher in the UEP case compared to the EEP case.

ACKNOWLEDGMENTS

This study was funded by Radio- ja Televisiotekniikan Tutkimus Oy (RTT). RTT is a nonprofit organization that contributes to the research and development of new radio and television technologies in Finland. Satu Jumisko-Pyykkö's work is supported by the Graduate School in User-Centered Information Technology.

REFERENCES

- [1] European Telecommunications Standards Institute (ETSI), "Digital video broadcasting (DVB); transmission system for handheld terminals (DVB-H)," European Standard EN 302 304, version 1.1.1, November 2004.
- [2] European Telecommunications Standards Institute (ETSI), "Digital Video Broadcasting (DVB): framing structure, channel coding and modulation for digital terrestrial television," ETSI standard, EN 300 744, 2001.
- [3] ISO/IEC 13818-1, "Information technology—generic coding of moving pictures and associated audio information: systems," November 1994.
- [4] European Telecommunications Standards Institute (ETSI), "Digital video broadcasting (DVB); DVB specification for data broadcasting," European Standard EN 301 192, version 1.4.1, November 2004.
- [5] I. S. Reed and G. Solomon, "Polynomial codes over certain finite fields," *SIAM Journal of Applied Mathematics*, vol. 8, no. 2, pp. 300–304, 1960.
- [6] European Telecommunications Standards Institute (ETSI), "Digital Video Broadcasting (DVB); specification for the use of video and audio coding in DVB services delivered directly over IP protocols," European standard TS 102 005 V1.2.1, November 2005.
- [7] International Organization for Standardization (ISO)/ International Engineering Consortium (IEC), "Information technology—generic coding of moving picture and associated audio information—part 3: audio," International Standard 14496-3 (2001), including "Bandwidth Extension," International Standard 14496-3 AMD-1 (2001) and "Parametric Coding for High Quality Audio," International Standard 14496-3 AMD-2 (2004).
- [8] Telecommunication Standardization Sector of International Telecommunication Union (ITU-T), "Advanced video coding for generic audiovisual services," Recommendation H.264, March 2005.
- [9] D. Tian, M. M. Hannuksela, and M. Gabbouj, "Sub-sequence video coding for improved temporal scalability," in *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS '05)*, vol. 6, pp. 6074–6077, Kobe, Japan, May 2005.
- [10] J. Aaltonen, H. Pekonen, T. Auranen, K. Laiho, and P. Talmola, "Power saving considerations in mobile datacasting terminals," in *Proceedings of IEEE International Symposium on Consumer Electronics*, pp. F43–F48, Erfurt, Germany, September 2002.
- [11] S. Wenger, M. M. Hannuksela, T. Stockhammer, M. Westerlund, and D. Singer, "RTP Payload Format for H.264 Video," IETF RFC 3984, February 2005.
- [12] T. Schierl, M. Kampmann, and T. Wiegand, "H.264/AVC interleaving for 3G wireless video streaming," in *Proceedings of IEEE International Conference on Multimedia and Expo (ICME '05)*, pp. 868–871, Amsterdam, The Netherlands, July 2005.
- [13] T. Schierl, T. Wiegand, and M. Kampmann, "3GPP compliant adaptive wireless video streaming using H.264/AVC," in *Proceedings of IEEE International Conference on Image Processing (ICIP '05)*, vol. 3, pp. 696–699, Genova, Italy, September 2005.
- [14] A. Albanese, J. Blömer, J. Edmonds, M. Luby, and M. Sudan, "Priority encoding transmission," *IEEE Transactions on Information Theory*, vol. 42, no. 6, part 1, pp. 1737–1744, 1996.
- [15] C. Leicher, "Hierarchical encoding of MPEG sequences using priority encoding transmission (PET)," Tech. Rep. TR-94-058, International Computer Science Institute, Berkeley, Calif, USA, November 1994.
- [16] U. Horn, K. Stuhlmüller, M. Link, and B. Girod, "Robust internet video transmission based on scalable coding and unequal error protection," *Signal Processing: Image Communication*, vol. 15, no. 1-2, pp. 77–94, 1999.
- [17] J. Rosenberg and H. Schulzrinne, "An RTP payload format for generic forward error correction," Internet Engineering Task Force Request for Comments 2733, December 1999.
- [18] A. H. Li, "RTP payload format for generic forward error correction," Internet Engineering Task Force Internet Draft draft-ietf-avt-ulp-18.txt, June 2006.
- [19] D. Tian, V. K. Malamal Vadakital, M. M. Hannuksela, S. Wenger, and M. Gabbouj, "Improved H.264/AVC video broadcast /multicast," in *Visual Communications and Image Processing 2005*, vol. 5960 of *Proceedings of SPIE*, pp. 71–82, Beijing, China, July 2005.
- [20] V. K. Malamal Vadakital, M. M. Hannuksela, M. Rezaei, and M. Gabbouj, "Method for unequal error protection in DVB-H for mobile television," in *Proceedings of the 17th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '06)*, pp. 1–5, Helsinki, Finland, September 2006.
- [21] V. K. Malamal Vadakital, M. M. Hannuksela, H. Pekkonen, and M. Gabbouj, "On datacasting of H.264/AVC over DVB-H," in *Proceedings of the 7th IEEE Workshop on Multimedia Signal Processing*, pp. 1–4, Shanghai, China, October 2005.
- [22] B. Girod, "Psychovisual aspects of image communication," *Signal Processing*, vol. 28, no. 3, pp. 239–251, 1992.
- [23] International Telecommunications Union—Radiocommunication sector, "Methodology for the subjective assessment of the quality of television pictures," ITU-R BT.500-11, 2002.
- [24] International Telecommunications Union—Telecommunication sector, "Subjective video quality assessment methods for multimedia applications," ITU-T P.910, 1999.
- [25] E. M. Rogers, *Diffusion of Innovation*, Free Press, New York, NY, USA, 5th edition, 2003.
- [26] S. Jumisko-Pyykkö and J. Häkkinen, "Evaluation of subjective video quality on mobile devices," in *Proceedings of the 13th ACM International Conference on Multimedia*, pp. 535–538, Singapore, November 2005.

- [27] S. Winkler and C. Faller, "Maximizing audiovisual quality at low bitrates," in *Proceedings of the 1st International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM '05)*, Scottsdale, Ariz, USA, January 2005.
- [28] E. N. Gilbert, "Capacity of a burst-noise channel," *Bell Systems Technical Journal*, vol. 39, pp. 1253–1265, 1960.
- [29] J. Poikonen and J. Paavola, "Comparison of finite-state models for simulating the DVB-H link layer performance," in *Proceedings of the 2nd International Symposium on Wireless Communications Systems Conference (ISWCS '05)*, pp. 153–157, Siena, Italy, September 2005.
- [30] S. Jumisko-Pyykkö, V. K. Malamal Vadakital, M. Liinasuo, and M. M. Hannuksela, "Acceptance of audiovisual quality in erroneous television sequences over a DVB-H channel," in *Proceedings of the 2nd International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM '06)*, pp. 1–5, Scottsdale, Ariz, USA, January 2006.
- [31] G. Faria, J. A. Henriksson, E. Stare, and P. Talmola, "DVB-H: digital broadcast services to handheld devices," *Proceedings of the IEEE*, vol. 94, no. 1, pp. 194–209, 2006.
- [32] Ishihara's tests for color deficiency 24 Plates edition. Tokyo, Kanehara trading INC, 2005.
- [33] International Standardization Organization, "Visual acuity testing," ISO 8596, 1994.
- [34] International Standardization Organization, "Audiometric test methods," ISO 8523-1- 2, 1989.
- [35] International Standardization Organization, "ISO Standards Handbook 35," Acoustics. 1st edition, p. 386. Switzerland, 1990.
- [36] International Standardization Organization, "Statistical distribution of hearing threshold," ISO 7029 SFS-EN Acoustics, 2000.
- [37] H. Coolican, *Research Methods and Statistics in Psychology*, J. W. Arrowsmith, London, UK, 4th edition, 2004.

Miska M. Hannuksela is a Research Leader in Nokia Research Center, Tampere, Finland. He has more than 10 years of experience in video compression and multimedia communication systems. He has been an active delegate in international standardization organizations, such as the Joint Video Team, the Digital Video Broadcasting Project, and the 3rd Generation Partnership Project. His research interests include scalable and error-resilient video coding, real-time multimedia broadcast systems, and human perception of audiovisual quality. He holds more than 15 international patents and has authored several tens of academic papers.



Vinod Kumar Malamal Vadakital received his B.Tech. degree in computer science and engineering from Bangalore University, Bangalore, India, and an M.S. degree in information technology from Tampere University of Technology, Tampere, Finland, in 1998 and 2005, respectively. From 1999 to 2001, he worked as a Project Assistant at the Indian Institute of Science, Bangalore, India. From 2001 to 2003 he was a Research Engineer at Fraunhofer Institute of Integrated Circuits (IIS-B), Erlangen, Germany. From 2003 to 2005, he worked as a Research Assistant at Tampere University of Technology. Currently he is



a researcher at the Tampere University of Technology and he is working towards his doctoral degree. His research interests are in the areas of video coding algorithms, video quality analysis, and mobile multimedia communications.

Satu Jumisko-Pyykkö received the M.S. degree in software engineering in 2005 from Tampere University of Technology. She has broad studies in multimedia, human-computer interaction, computer-aided learning, and psychology from the University of Helsinki. She is currently a Ph.D. student in the Graduate School in User-Centered Information Technology and is working as a researcher in the Institute of Human-Centered Technology at Tampere University of Technology. Her research interests are focused on human-centered approach to multimedia quality and development of research methods for understanding and measuring experienced multimodal quality.

