Studying IP Level Error Statistics of Streaming Audiovisual Mobile TV Services

Teppo Kurki, Heidi Himmanen and Jarkko Paavola
Department of Information Technology, University of Turku
Turku, FINLAND
e-mail: {teppo.kurki, heidi.himmanen, jarkko.paavola}@utu.fi

***Abstract*—This paper investigates the IP (Internet Protocol) level error statistics and their effects on the application layer for audio and video services transmitted over a DVB-H (Digital Video Broadcasting – Handheld) system. IP is the interface between the transmission system and applications in most Mobile TV systems. Understanding IP level error behavior and its effects to compressed audiovisual material forms an important foundation for the mapping of transmission system performance to subjectively perceived audiovisual quality. In this paper, simulation results and IP packet error statistics for typical parameter sets for audiovisual services transmitted over conventional MobileTV channels are presented.**

***Index Terms*—Mobile TV, digital video, DVB-H, error statistics, QoS**

I. Introduction

Mobile broadcasting services are a strongly growing field in modern telecommunication systems. Several mobile broadcasting standards and systems have emerged in the past years, and DVB-H [1] is one of the most popular among such systems. The DVB-H system is designed to enable two types of services: streaming audiovisual services and file-delivery applications. These service types have very different error tolerances and delay jitter constraints.

Until recently, not much effort has been put into studying the QoS (Quality of Service) criteria in mobile audiovisual services. The need for such criteria is clear since mobile broadcasting systems are typically very complex and include adjustable parameters on multiple layers. Optimizing these parameters to enable the best possible user experience in as many reception conditions as possible is a challenging task.

The DVB-H link layer operations are depicted in Fig. 1. The transmitted application layer data is first encapsulated into IP packets. The IP packets are inserted into an MPE-FEC (Multi-Protocol Encapsulation – Forward Error Coding) frame, and the resulting MPE- and FEC-sections of the MPE-FEC frame are encapsulated into TS (Transport Stream) [2] packets. If FEC is not used, only MPE-sections are transmitted. In both cases, the transmission occurs in bursts, where the transmission of one service is "on air" e.g. 10% of the time. One transmission burst contains one MPE-FEC frame.

Currently the DVB-H implementation guidelines [3] define a simple quality degradation criterion called MFER (MPE-FEC Frame Error Ratio) that is utilized as an acceptability threshold for audiovisual services. The MFER criterion states that up to 5 % of the MPE-FEC frames may be erroneous to still enable acceptable reception quality. In reality the definition is based on an agreement. It is even noted in the implementation guidelines that the MFER criterion may not always correlate with the QoS requirements.

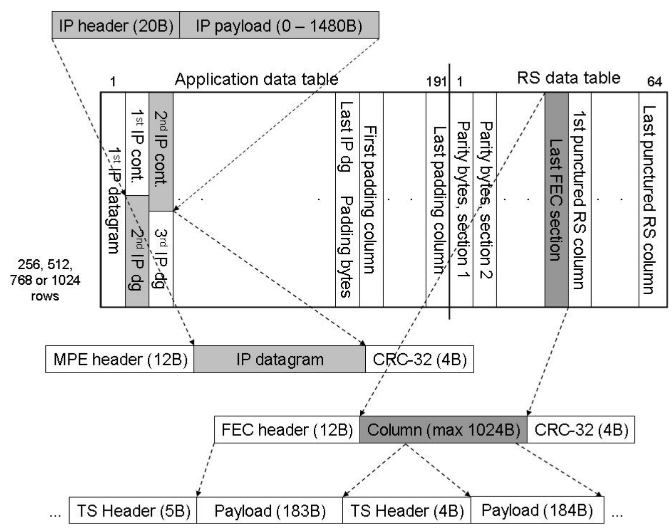


Fig. 1. The DVB-H link layer operations

Recent studies [4], [5] show that the MFER criterion might not reflect the actual user experience with adequate accuracy. Therefore new error criteria should be studied and proposed to form a solid basis for system and network optimization. In this paper the effort is focused on the IP level error behavior, since IP encapsulation is used as an interface not only in DVB-H, but also other current and future systems.

So far the DVB-H system and especially the associated error statistics with various transmission settings and channel models has been studied quite extensively, e.g. in [6]. However, studies on the effects of errors on a streaming audiovisual service have not been carried out before.

The problem has also been approached from an opposite direction by video quality researchers: several methods to approximate the observed video quality from the amount of lost IP packets have been developed in the last 15 years.

However, there still is a lot of work to before these two approaches will meet.

For example, if a system designer wants to know which physical layer, transport layer and application layer parameters they should use to provide the best possible experience to the end user, there is no simple solution. The simple IP PER results obtained with different transmission parameters will only tell how the system would perform in terms of lost IP packets. This knowledge does not necessarily equate to the QoS experienced by the end user, since the audio and especially video encoding parameters bring many more variables to the equation.

Several questions are raised. What is the total amount of audiovisual information lost? Which transmission system bit rate would provide optimal results? How does the transmission burst length affect the audiovisual experience? Does the video bit rate have a relation with the effects of transmission errors? Would shorter more frequent errors be favourable to longer less frequent ones? These are just some of the questions that should be answered before optimizing the parameters used in a mobile TV system. In the following sections we will try to find some answers.

The paper is organized as follows: Next, typical video and audio compression used in DVB-H is presented. In Section III the simulation model is described in details. Simulation results for various parameter combinations are presented in Section IV. Finally, conclusions are drawn, and the need for further research is pointed out.

II. Audio and video compression in DVB-H

The IP data casting specifications of DVB-H allow the use of multiple audio and video compression methods, but the use of high efficiency advanced audio coding version 2 (HE-AAC2) for audio compression and advanced video coding (H.264/AVC) for video compression is recommended.

The basic transmission units for streaming use of both HE-AAC2 and H.264/AVC bit streams are access units and network abstraction layer (NAL) units, respectively. Depending on the size of these units, one or more of them are encapsulated into one transmission packet, which in our case is an IP packet. An access unit of HE-AAC2 contains a coded representation of a frame of audio samples, whereas the contents of a NAL unit can be categorized to video coding layer (VCL) NAL units and non-VCL NAL units. VCL NAL units are typically coded slices of a picture, covering a certain spatial area of the picture, or even the whole picture. Non-VCL NAL units are used to transfer data that is indirectly related to the decoding process of the actual video material.

Primary-coded pictures of H.264/AVC can be categorized to three types: instantaneous decoding refresh (IDR) pictures, other reference pictures, and non-reference pictures. An IDR picture provides a safe restart point for the decoder, since it contains only intra-coded picture information and additionally causes all the previous reference pictures to be no longer used as references for any pictures following the IDR picture. It obviously follows that an IDR picture can be used as a random access point to start the decoding process, and it also provides a resynchronization point for decoding after transmission errors have occurred. In a system with multiple streams the interval between IDR pictures also sets the upper threshold for the channel switching time. When considering DVB-H, each transmission burst should contain at least one IDR picture, which should ideally be located in the beginning of the burst. A H.264/AVC reference picture is stored and maintained as a prediction reference for inter-frame prediction until it is marked as no longer used according to the H.264/AVC reference picture process. A non-reference picture is not used for inter-frame prediction at all and can be removed from the received bit stream without causing error propagation to any other pictures in the stream.

III. Simulation Model

A simulation model has been developed using Matlab to facilitate the studies on the TS, IP and audiovisual level error statistics of the DVB-H system. Physical layer input data for the simulations consist of a TS packet error trace that is obtained from laboratory measurements using a channel simulator and a prototype Nokia receiver. The used channel models are Pedestrian Outdoor (PO), Vehicular Urban (VU) and Motorway Rural (MR) with simulated velocities of 3 km/h, 30 km/h and 100 km/h, respectively. The channel models are presented in [7].

The audiovisual stream data used in the simulations does not consist of real encoded audiovisual sequences, although that is also possible to implement for future studies. To enable statistical comparison of packet errors, it is necessary to fix the packet length. Therefore, typical audio and video IP packet lengths were measured from a functional DVB-H network in Turku, and the obtained values were used to construct data with the same statistical properties to be used in the simulation process. It was found that most of the IP packets containing HE-AAC2 access units are approximately 250 bytes in length. In terms of the video material, the IP packet length tends to be near the upper limit imposed by the system, 1500 bytes. This seems reasonable since the larger IP packet size minimizes the overheads caused by the transport layer. Over 90% of the audio IP packets had a length of 240 – 260 bytes, and over 95% of the video IP packets had a length between 1400 and 1500 bytes. Therefore, in these simulations lengths of 250 bytes and 1450 bytes were used for audio and video IP packets, respectively.

We assume that the multiplex is filled with identical services during one simulation run. First, the TS packet error trace is divided between these services according to the chosen transport level parameters. Next, the data is mapped to each service according to the selected MPE-FEC parameters, and the MPE-FEC encoding and decoding processes are performed. This allows us to find the MFER and IP PER values for each service, as well as the accurate locations of lost audio and video IP packets. Additionally, the following error statistics are calculated for TS, audio IP and video IP packets independently: The average length of an erroneously received burst of packets (AEBL, Average Error Burst Length), the mean amount of correctly received packets

(MTBE, Mean Time Between Errors) and the corresponding variances (VEBL, Variance of Error Burst Lengths and VTBE, Variance of Time Between Errors). These terms are described in more detail in [6]. Once all the values have been obtained, the final results are averaged over all the identical services to make sure that all the conditions of the channel are taken into account. The worst-case values are also recorded.

In addition the TS and IP level results, even more detailed error statistics are calculated from the simulated audiovisual layer. These statistics include the amount of errors per one minute of service, the average length of an error and the total amount of lost audio or video material. In this paper, a loss of audio or video material of any length is considered as one audio or video error.

The physical layer parameters used in the simulations were selected as follows. For each of the three channel models (PO 3 km/h, VU 30 km/h and MR 100 km/h) the simulations were run using two combinations of modulation and physical layer code rate: 16-QAM with 1/2 code rate and QPSK with 2/3 code rate. In each case the OFDM guard interval was set to 1/4. TS level bit rates for the selected parameters are 9.953 Mbps for 16-QAM 1/2 and 6.635 Mbps for QPSK 2/3. Each channel model – settings –pair was simulated at three carrier-to-noise-ratio points: 14, 15 and 16 dB for 16-QAM 1/2, and 12, 13 and 14 dB for QPSK 2/3. These points were chosen so that the obtained error statistics would cover an area from error-free or nearly-error-free viewing to a quality level that most people would rate unacceptable according to [5].

At the link and the application layer, there are far more parameters to choose from. Here, we chose to use two coded video bit rates, 768 and 384 kbps, which correspond to DVB-H device capability classes A and B [8].For the audio, a bit rate of 48 kbps was used throughout the simulations.

To enable fair comparison between the different parameter combinations we decided to simulate all combinations with three different channel switching times, 0.5 seconds, 1 second and 1.5 seconds. The burst length and thus the MPE-FEC frame size needs to be optimized for each case to achieve reasonable power saving. For the MPE-FEC code rate we used 3/4 and 5/6, and the amount of MPE-FEC rows was 256, 512 or 768, depending on the case. At TS level we use full bit rate of the channel for a third of the simulations, and half bit rate for the rest. Using halved bit rate in this case means that two DVB-H service bursts are sent in the channel simultaneously and each of them occupy every second TS packet, so the bit rate is effectively halved. The resulting transmission burst lengths vary from 50 ms to 320 ms. It is assumed that each MPE-FEC frame contains a single IDR picture in the beginning of the frame.

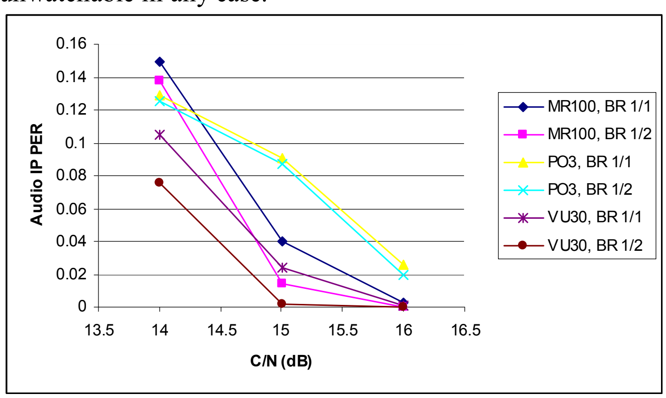
IV. Results

The number of simulated cases in this study is quite high. We use three channel models and three different C/N values for each case. At the physical layer we have two combinations of modulation and physical layer code rate. At TS level we use two different TS level bit rates and several different transmission burst lengths resulting in three different channel switching times. Further, we use two different link layer code rates and two different video bit rates. These combinations give a total of 324 simulations. Therefore, all the results cannot be presented in this paper. Instead, we have concentrated on providing answers to a few questions described in the following subsections.

A. *The effect of transmission burst length*

We can achieve the same audiovisual outcome using two different sets of transport layer parameters that result in shorter or longer transmission bursts. The application layer settings will be exactly the same in both cases. Most importantly, the channel switching time and IDR picture interval are not affected. In this subsection we study which approach would provide the least amount of errors. In practice, this situation is achieved when all the other settings are kept the same, and only the TS bit rate is halved. Thus, the burst length is doubled. Here we will concentrate on the case where video bit rate is 768 kbps, and the MPE-FEC frame has 512 rows and 191 ADT + 64 RS columns. The results for this case had similar trends regardless of the physical layer parameters (16-QAM 1/2 vs. QPSK 2/3) and the MPE-FEC code rate. The presented results are obtained from the simulations with 16-QAM 1/2 using MPE-FEC code rate 3/4. With full TS bit rate the burst length will be 97 ms, and then the TS bit rate is halved, the burst length will double to 194 ms.

Figure 2 shows the IP PER results independently for IP packets containing audio and video. From these results it seems obvious that increasing the burst length results in at least a lower PER for both the audio and video content in all the simulated channels. The video PER in the MR channel with low C/N forms an exception, but it is not very meaningful since at over 0.16 PER the content would be quite unwatchable in any case.



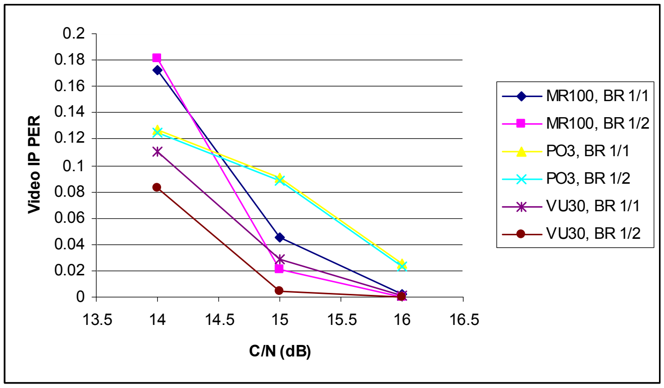
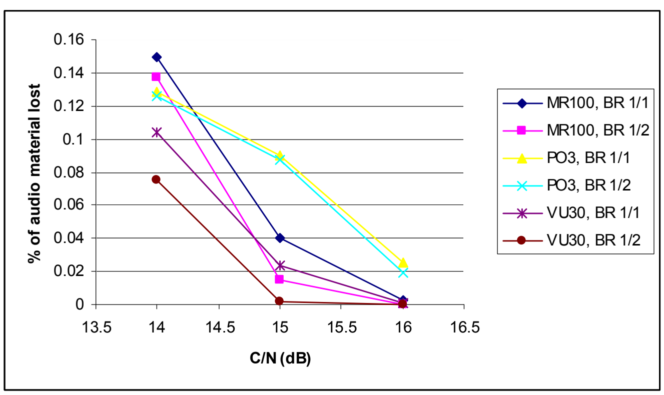


Fig. 2. The effect of transmission burst length, Upper: Audio IP PER versus C/N (dB), Lower: Video IP PER versus C/N (dB). Burst lengths: BR 1/1: 97 ms, BR 1/2: 194 ms

Figure 3 depicts the percentage of lost video and audio material versus C/N values. The values follow the IP PER curves quite closely, but there are also some notable differences, especially in regards to the PO3 channel. Most importantly it can be seen that especially in the C/N = 15 dB point the longer burst time may provide very clear benefits for both audio and video quality. The average lengths of both video and audio errors are quite similar in both cases, there are merely less errors. For the video content this is probably because the IDR picture interval is kept the same, but for audio content the issue is not that clear. The reason might be that the audio errors are overall more random in length, due to the audio encoding not using prediction in the same way as the video encoding does.



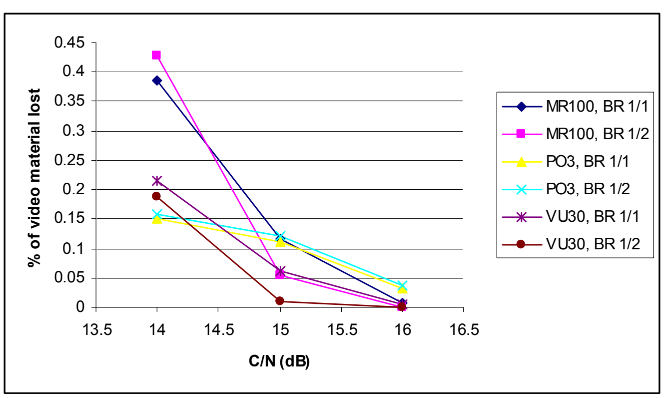
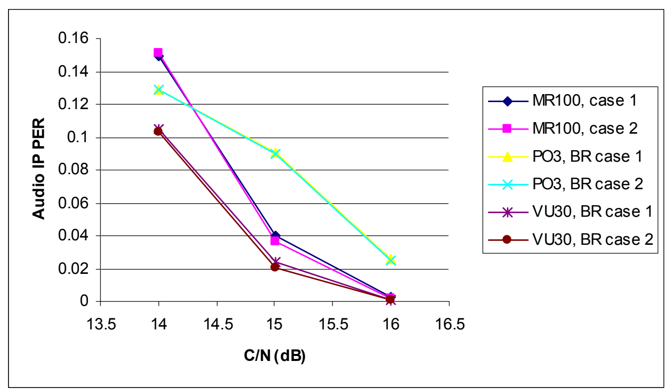


Fig. 3. The effect of transmission burst length, Upper: Percentage of lost audio material versus C/N (dB), Lower: Percentage of lost video material versus C/N (dB)

B. *Effect of channel switching time*

When studying the effect of channel switching time we create two sets of settings that would allow video and audio properties to stay otherwise the same, but the channel switching time should be reduced to half. The used video bit rate is 768 kbps. The radio channel and physical layer parameters are the same as in the previous subsection. For the first case there are 512 MPE-FEC rows, whereas the second case uses only 256 rows. To achieve the same burst lengths we use half of the TS bit rate for the second case. The burst length in both cases is approximately 109 ms. Channel switching times for cases 1 and 2 are 1.0 s and 0.5 s, respectively.

As can be seen from Figure 4, the length of the channel switching time, and thus the interval between IDR pictures, has almost no effect on the IP PER of either audio or video content. Very slight differences can be seen, however, and where these differences exist, the better performance is provided by case 2, which has the shorter channel switching time. Again, the results seem to be the same regardless of modulation and FEC code rates.



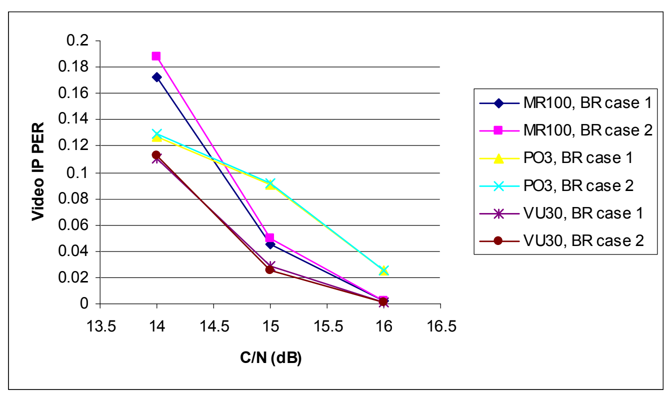


Fig. 4. The effect of channel switching time, Upper: Audio IP PER versus C/N (dB), Lower: Video IP PER versus C/N (dB)

When examining the amount of lost audiovisual material, shown in Figure 5, the results are nearly similar: very small but measurable differences can be found in favor of the shorter channel switching time.

However, the larger differences can be seen in the error distribution, presented in Table 1 for the MR100 channel model. For audio, there is no measurable difference, but for video, the difference is dramatic. It could be said that when you halve the channel switching time, you are effectively halving the length and doubling the amount of video errors. This is where it would come down to subjective video quality measurement to decide which error distribution would be more acceptable to the viewer.

TABLE 1: EFFECTS OF CHANNEL SWITCHING TIMES ON VIDEO ERRORS

	Avg. amount of video errors per minute	Avg. length of video errors	Channel switching time
Case 1	8.1	0.85	1.0 s
Case 2	15.0	0.43	0.5 s

Another issue to consider is the fact that IDR pictures occupy more space than other pictures. Therefore halving their frequency will leave less bits to encode the non-IDR pictures. This might have a measurable effect on the QoS experienced by the viewer. However, this effect is out of the scope of this study.

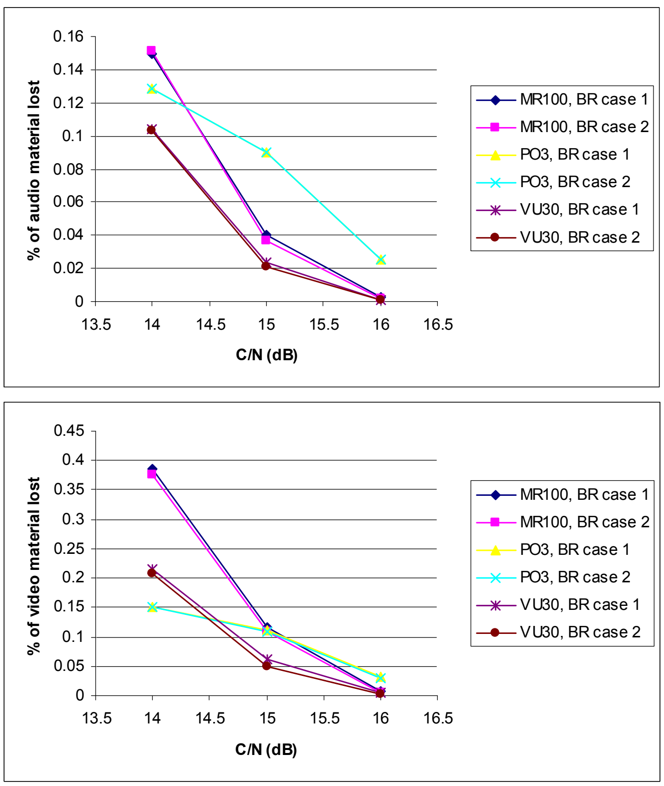


Fig. 5. The effect of channel switching time, Upper: Percentage of lost audio material versus C/N (dB), Lower: Percentage of lost video material versus C/N (dB)

C. *Effect of video bit rate*

In the third case we study the effect of the video bit rate. The only different settings between cases 1 and 2 here, in addition to the video bit rate, are the MPE-FEC row count and the use of bit rate halving. Case 1 represents video bit rate of 768 kbps, and case 2 a video bit rate of 384 kbps. The burst length and channel switching time are kept the same, by utilizing 512 rows and full bit rate in case 1, and 256 rows with halved bit rate in case 2. The burst length in both cases is approximately 109 ms and the channel switching time is 1.0 second.

As seen in Figure 6, the IP PER values for both the audio and video IP packets are very similar to each other. This result is consistent throughout the different modulation schemes and MPE-FEC code rates. It goes to further show that in fact the burst length seems to have a very dominant effect on the audio and video IP PER values.

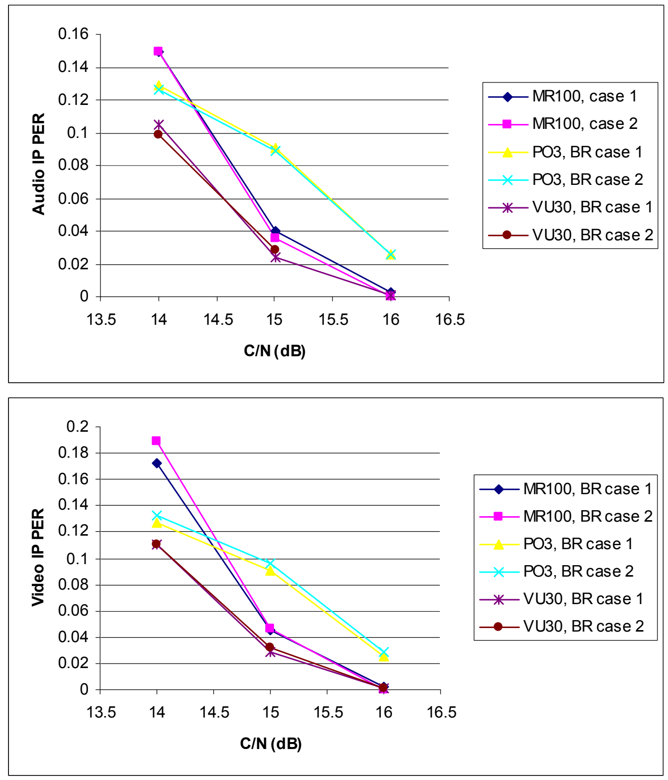


Fig. 6. The effect of video bit rate, Upper: Audio IP PER versus C/N (dB), Lower: Video IP PER versus C/N (dB)

However, we are starting to see some differences when examining the errors from the application level (Figure 7). Logically, there are no huge differences in audio errors, since the audio encoding rate was not changed. However, consistent differences can be observed in the lower part of fig. 7 which shows the percentage of lost video material. It would seem that when the video bit rate is lowered, more video material will be lost. This effect seems to be consistent in all the channel models with all measured C/N values. When examining the error locations and lengths, they seem very similar in principle. The only difference is that with the lower video bit rate of 384 kbps, up to 20 % more video material is lost.

The reason behind this effect is not immediately clear, since the burst length is the same in both cases. However, the effect itself should be real since it will reproduce with different modulation and MPE-FEC code rate combinations.

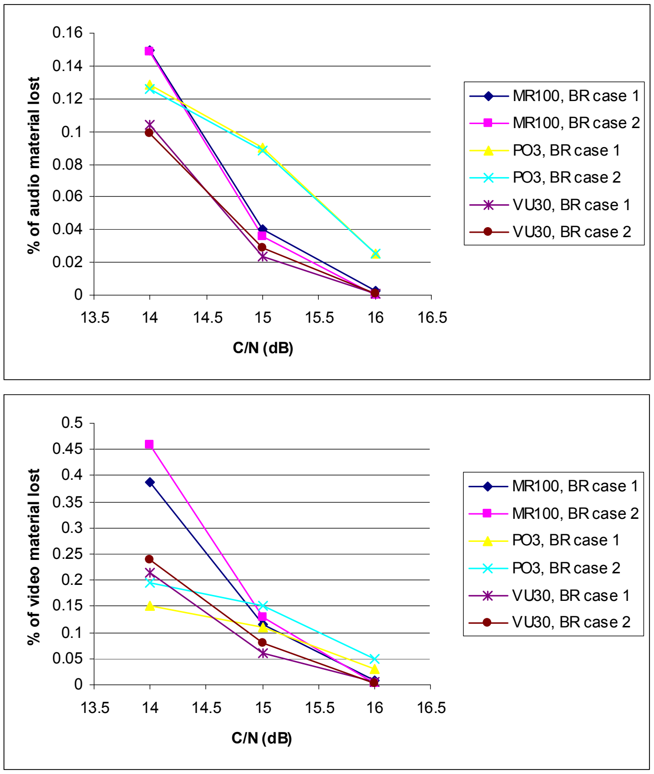


Fig. 7. The effect of channel switching time, Upper: Percentage of lost audio material versus C/N (dB), Lower: Percentage of lost video material versus C/N (dB)

V. Conclusions

Studying IP-level and audiovisual-level error statistics forms an important foundation for a deeper understanding of the error behavior of mobile audiovisual services. In this paper we introduce both a simulation model and the applicable error statistics for the purpose of evaluating different parameter combinations in terms of the resulting audiovisual reception. We also analyze three different scenarios, which could be faced when designing a mobile broadcasting system, and provide possible solutions to them. The conclusion from simulations is to use longer transmission burst length and higher video bit rate where possible. On the other hand, the channel switching time (IDR picture frequency) has only minor effects to the amount of lost audiovisual material. However, shorter IDR picture intervals could be considered to make the video errors shorter and thus possibly less annoying.

The long-term goals of this work include enabling the development of more accurate, combined audiovisual QoS metrics, which bridge transmission error statistics to perceived subjective quality.

References

[1] European Telecommunication Standard ETSI EN 302 304 v1.1.1 (2004-11), *Digital Video Broadcasting (DVB): Transmission System for Handheld Terminals (DVB-H)*, November 2004.

[2] ISO/IEC 13818-1: *Information technology — Generic coding of moving pictures and associated audio information: Systems*, Second Edition, 2000.

[3] European Telecommunication Technical Recommendation ETSI TR 102 377 v.1.2.1 (2005-11), *Digital Video Broadcasting (DVB); DVB-H Implementation Guidelines*, November 2005.

[4] H. Himmanen, M. Hannuksela, T. Kurki and J. Isoaho, ”On Error Criteria for Mobile Broadcasting of Streaming Audiovisual Services”, *EURASIP Journal on Advances in Signal Processing, Special Issue on Wireless Video*, Volume 2008, Article ID 518219, 12 pages, July 2008.

[5] S. Jumisko-Pyykkö, V.K. Malamal Vadakital and J. Korhonen, ”Unacceptability of instantaneous errors in mobile television: from annoying audio to video”, in *Proc. 8th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '06)*, pp. 1-8, Helsinki, Finland, 2006.

[6] H. Himmanen, “Studies on channel models and channel characteristics for mobile broadcasting”, in *Proc. 2008 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, Orlando, United States, pp. 1-9, March 2008.

[7] CELTIC WingTV Project Report, “D11 – WingTV Network Issues”, May 2006

[8] ETSI TS 102 005, “Digital video broadcasting (DVB); specification for the use of video and audio coding in DVB services delivered directly over IP protocols,” ETSI TechnicalSpecification, April 2006.