Audio-Video databases for H.264-bitstream-based quality assessment of IPTV services

Marie-Neige Garcia*, Peter List†, Bernhard Feiten†, Ulf Wüstenhagen† and Alexander Raake‡
*Telekom Innovation Laboratories, Assessment of IP-based Applications, Technische Universität Berlin, Germany
†Telekom Innovation Laboratories, Deutsche Telekom AG, Berlin, Germany
‡ Technische Universität Ilmenau, Audiovisual Technology Group, Ilmenau, Germany

***Abstract*—The paper introduces one audio, two video and three audiovisual databases that were produced for training and validating the standardized models ITU-T P.1201.2 and P.1202.2. Both models are bitstream-based no-reference models targeting IPTV services and high-resolution video sequences (Standard and High Definition). They cover compression and packet-loss impairments. Each database contains about 240 sequences. For video, the H.264 codec was used, while the audio was encoded with four different codecs (AAC-LC, HE-AAC, MP2, AC-3). Both uniform (random) and bursty losses were addressed for different types of packet-loss concealment. Databases include per-subject quality scores as well as per-frame (video and audio) and per sequence bitstream- and packet loss-related information.**

I. Introduction

Bitstream-based no-reference quality models have shown to be useful to network providers for service monitoring of video streaming applications. By using bitstream-based information, these models may be sensitive to encoding settings. In general, the scope of quality models is limited to that of the databases they have been developed from. To extend the applicability of bitstream-based models, the number of databases used for developing and validating the models should be increased, for instance by regrouping publicly available databases as supported by the Joint Effort Group (JEG) of the Video Quality Experts Group (VQEG).

We want to join this effort by sharing one audio, two video and three audiovisual databases. These databases were created in the course of two ITU-T competitions on bitstream-based quality models for higher resolution applications such as IPTV. The competitions resulted in the ITU-T P1201.2 [1] and ITU-T P.1202.2 standards.

With the H.264 codec and High Definition resolutions, these databases address more "traditional" video configurations than recently created databases covering for instance HEVC/H.265 and Ultra HD video. In our view they are still relevant for research, since: A) The video databases can be used in combination with already publicly available databases with similar or complementary scopes [2]–[5] for extending the scope of existing bitstream-based models. B) The contents of the audio, video and audiovisual databases are generated using the same original source sequences, therefore enabling better development of full audiovisual models. To our knowledge, such combination of databases has never been made publicly available so far. C) The shared audio database compares four different codecs for compression and network impairments, which is, to our knowledge, unique for a publicly available database. D) When combined with databases addressing other codecs, the shared databases are useful for comparing bitstream-based parameters between different codecs.

The databases creation process is described in Section II and details about the subjective test experiments are provided in Section III.

II. Database creation

A. Test material

Eight $10\ s$ duration contents (*SRCs*) were used per test and applied on 30 test conditions (*HRCs*, resp. 34 for audio) resulting in 240 sequences per test (resp. 272 for audio). Audio and video SRCs are representative of typical TV or movie contents. Special care was taken, mainly by visual inspection of the video sequences, that the contents cover a wide range of spatio-temporal (ST) complexities. This large range of ST complexity values is also captured by the "Spatial (SI) and Temporal (TI) perceptual Information" indicators defined in ITU-T Rec. P.910 (e.g., $TI : [3, 90]$ and $SI : [40, 90]$ for one of the video-only databases).

As can be observed in Table I, databases address High Definition (HD720: 1280x720, HD1080: 1920x1080) video resolutions and both compression and network impairments typical of unreliable transport (involving typical combinations of UDP, RTP, MPEG2-TS).

The video stream was H.264 encoded with constant bitrates (CBR) with the x264 encoder revision r1867 (http://www.videolan.org/developers/x264.html). As shown in Table I, high profile was used as encoder setting for the HD1080 format and main profile for HD720. Different Group of Picture (GOP) structures were used, referred to as $MxNy$ in Table I, where $x \in \{3, 4\}$ corresponds to the number of B-frames between I- or P-frames and $y \in \{25, 50\}$ indicates the number of frames between two I-frames.

In the case of audio, the audio stream was encoded into MPEG1-LII ($MP2$) or $AC3$ using ffmpeg (https://ffmpeg.org/). For encoding the audio into the Advanced Audio Coding Low Complexity ($AAC - LC$) and High Efficiency AAC version 2 ($HE - AACv2$), the Nero encoder v1.5.4.0 (http://www.nero.com/) was used.

The encoded video and audio streams were packetized with the Sirannon v0.6.8 software (http://sirannon.atlantis.ugent.be) into MPEG2-TS and RTP packets. The resulting packet capture (PCAP) files were impaired with packet loss using a software from Telchemy (http://vqegstl.ugent.be/?q=node/27). The range of applied packet loss percentage values are shown in Table I. The loss pattern followed a four-state Markov-Model

Parameters	Audio-only ($tr16$)	Video-only ($tr10$, $vl14$)	Audiovisual ($tr13$, $vl23$, $vl24$)
Video Format	n.a	720p50 ($tr10$ and $vl14$)	720p50 ($tr13$), 1080p25 ($vl23$), 1080i25 ($vl24$)
Video Codec	n.a.	H.264 main profile	H.264 (HD1080: high profile, HD720: main profile)
Video CBR	n.a.	0.5 Mbps to 30 Mbps	1 Mbps to 30 Mbps
GOP structure	n.a.	M3N25, M3N50, M4N50	M3N25, M3N50, M4N25, M4N50
Audio Format	44.1 & 48 kHz, 16 bit, mono & stereo	n.a.	44.1 & 48 kHz, 16 bit, mono & stereo
Audio Codec	AAC-LC, MP2, HEAACv2, AC3	n.a.	AAC-LC, MP2
Audio CBR	AAC-LC: 48 to 576 kbps	n.a.	AAC-LC: 48 to 576 kbps
	MP2: 64 to 384 kbps	n.a.	MP2: 64 to 384 kbps
	HEAAC v2: 16 to 96 kbps	n.a.	n.a.
	AC3: 96 to 256 kbps	n.a.	n.a.
Ppl	0 % and [0.5, 5] %	0 % and [0.06, 0.5] % (freezing)	0 % and [0.01, 1.03] % (freezing)
		0 % & [0.125, 1] % (1 slice/frame)	0 % and [0.02, 1.2] % (slicing: 1 slice/frame)
		n.a	0 % and [0.04, 1.4] % (slicing: 1/MBrow)
PLD		random & bursty (four-state-Markov Model)	
PLC	codec built-in	freezing & slicing	freezing & slicing

TABLE I. OVERVIEW OF TEST CONDITIONS FOR AUDIO-ONLY, VIDEO-ONLY AND AUDIOVISUAL TESTS.

so that both random loss and different strength of bursty losses with gaps (no loss) in the burst period were achieved.

The impaired PCAP files were decoded with a proprietary H.264 decoder from T-Labs. For audio, the audio elementary stream (ES) was first extracted from the impaired PCAP file using the same T-Labs decoder. Then, the audio ES was decoded with ffmpeg in the case of MP2 and AC3. For AAC and HE-AACv2, the audio ES was first packetized into an MP4 container with MP4Box (MP4Box-0.4.6-rev2735) and then decoded with the Nero AAC decoder.

Audio concealment was codec-dependent. For video, both "freezing" and "slicing" were used for error handling. For "freezing", the frames directly affected by loss or through temporal propagation are discarded and replaced by the last unimpaired reference frame till the next unimpaired I-frame. If slicing is applied, its impact depends on the number of slices per frame (1 slice per frame, $1/frame$, or 1 slice per MacroBlock row, $1/MBrow$). In that case, the packet loss concealment (PLC) operated Macroblock (MB) by MB. Spatial intra-frame concealment was used when loss occured in the first picture of a scene, column-wise, left-to-right. For each lost MB the nearest correctly decoded MB in the same column was copied. Temporal inter-frame concealment is applied when a P-, B- or I-frame that is not the start of a scene is affected. Concealment is applied column-wise left-to-right, by copying MBs from co-located positions in the temporally nearest previous unaffected reference frame.

B. Database structure

Each database is composed of subjective test results, details on test conditions, side information and XML files. Due to content rights, video and audio signals are not available. Test results are provided in the form of a spreadsheet with individual scores per subject and the mean opinion score for each sequence. Standard deviations and confidence intervals are also given. There is one file with side information and one XML file per sequence. Side information in text form includes audio and video codecs and profiles, transport format, video resolution, packet loss handling or the number of slices per frame. XML files include frame-level information such as frame type, frame size, number of lost or found RTP and TS packets in the frame. Bitstream parameters such as macroblock types, averaged quantization parameters and motion vector sizes are also provided per frame and summary statistics can be found at the end of each XML file. A full description of the parameters is provided with the databases.

III. SUBJECTIVE TESTS

Listening and viewing conditions were compliant to ITU-T Rec. P.800, Rec. ITU-R BT-500-11 and ITU-T P.910. Professional high-performance systems were used for audio (headphones) and video presentation. Subjective scores were collected using the Absolute Category Rating (ACR) method as specified in ITU-T P.910, with the 5-point scale recommended in ITU-R BT-500-11. Up to 30 subjects participated in each test, each subject in only one test. They were screened for visual acuity and color blindness using the Wenzel plates and Ishihara test. Ratings of a subject were rejected if the Pearson Correlation Coefficient between the ratings of this subject and the ratings averaged over all subjects was smaller than 0.73 for that database. This rejection procedure resulted in a minimum of 24 "valid" participants per test.

IV. CONCLUSION

One audio, two video and three audiovisual databases are made publicly available. They are useful for developing and validating bitstream-based audio, video and audiovisual quality models for high resolution sequences and contain both coding and packet loss impairments. The databases include subjective test results as well as per-frame (video, audio) and per-sequence bitstream- and packet-loss parameters. They are included in the Qualinet databases as "TLABS P1201 IPTVH264HD" and can be found with detailed information on the used sources and test conditions on ftp://ftp.ivc.polytech.univ-nantes.fr/VQEG/JEG/HYBRID/database_TLABS_P1201_IPTVH264HD_Audiovisual.

REFERENCES

[1] M.-N. Garcia, P. List, S. Argyropoulos, D. Lindegren, M. Pettersson, B. Feiten, J. Gustafsson, and A. Raake, "Parametric model for audiovisual quality assessment in IPTV: ITU-T Rec. P.1201.2," in *Proc. of MMSP*, 2013.

[2] N. Staelens, G. V. Wallendael, R. V. de Walle, F. D. Turck, and P. Demeester, "High Definition H.264/AVC Subjective Video Database for Evaluating the Influence of Slice Losses on Quality Perception," in *Proc. of QoMEX*, 2013.

[3] M. Barkowsky, N. Staelens, L. Janowski, Y. Koudota, M. Leszczuk, M. Urvoy, P. Hummelbrunner, I. Sedano, and K. Brunnström, "Subjective experiment dataset for joint development of hybrid video quality measurement algorithms," in *Proc. of QoEMCS*, 2012.

[4] M. Barkowsky, M. Pinson, R. Pépion, and P. Callet, "Analysis of Freely Available Dataset for HDTV including Coding and Transmission Distortions," in *Proc. of VPQM*, 2010.

[5] T. Maki, D. Kukolj, D. Dordevic, and M. Varela, "A reduced-reference parametric model for audiovisual quality of IPTV services," in *Proc. of QoMEX*, 2013.