International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

J.343

Amendment 1
(05/2018)

SERIES J: CABLE NETWORKS AND TRANSMISSION OF TELEVISION, SOUND PROGRAMME AND OTHER MULTIMEDIA SIGNALS

Measurement of the quality of service - Part 3

Hybrid perceptual bitstream models for objective video quality measurements

Amendment 1

Recommendation ITU-T J.343 (2014) – Amendment 1

ITU-T

International Telecommunication Union

Recommendation ITU-T J.343

Hybrid perceptual bitstream models for objective video quality measurements

Amendment 1

Summary

Recommendation ITU-T J.343 specifies objective video quality measurement methods which use bitstream data in addition to processed video sequences. From bitstream data, the models can obtain additional information on the codec type, bit rate, frame rate, some transmission errors and spatial/temporal shifts. Consequently, such models may provide improved performance compared to objective video quality models, which use only processed video sequences.

Amendment 1 adds a new Appendix I containing information about test vectors for the ITU-T J.343 family of Recommendations.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T J.343	2014-11-29	9	11.1002/1000/12315
1.1	ITU-T J.343 (2014) Amd. 1	2018-05-10	12	11.1002/1000/13619

Keywords

Bitstream data, objective methods, test vectors, video.

* To access the Recommendation, type the URL http://handle.itu.int/ in the address field of your web browser, followed by the Recommendation's unique ID. For example, http://handle.itu.int/11.1002/1000/11830-en.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure, e.g., interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database at http://www.itu.int/ITU-T/ipr/.

Table of Contents

Introduction

Generally video quality estimation models are, depending upon the required input signals, categorized as no reference (NR), reduced reference (RR) and full reference (FR) models. NR models are provided with the processed video sequences only. RR models require that features extracted from the reference video sequences and the processed video sequences are provided. For FR models, the unimpaired reference and the processed video sequences must be provided.

In addition, the models described in this Recommendation need access to the received bitstream data from which the model can obtain information on transmission errors (e.g., delay, packet loss), codec (e.g., type, bit-rates, frame rates, codec parameters), etc.

Consequently, the models described here are categorized as hybrid models (i.e., Hybrid-NR, Hybrid-RR and Hybrid-FR).

Recommendation ITU-T J.343

Hybrid perceptual bitstream models for objective video quality measurements

Amendment 1

Editorial note: This is a complete-text publication. Modifications introduced by this amendment are shown in revision marks relative to Recommendation ITU-T J.343 (2014).

1 Scope

This Recommendation describes recommended objective models for non-intrusive monitoring of the video quality of IP-based video services based on the decoded video frames and packet-header. Some types of models also utilize the reference video or bitstream information. This Recommendation addresses six application areas:

- [ITU-T J.343.1] specifies Hybrid-NRe models
- [ITU-T J.343.2] specifies Hybrid-NR models
- [ITU-T J.343.3] specifies Hybrid-RRe models
- [ITU-T J.343.4] specifies Hybrid-RR models
- [ITU-T J.343.5] specifies Hybrid-FRe models
- [ITU-T J.343.6] specifies Hybrid-FR models

[ITU-T J.343.1] includes two models, [ITU-T J.343.2] includes one model, [ITU-T J.343.3] includes one model that operates at multiple side channel bandwidths to transmit the reduced reference information, [ITU-T J.343.4] includes one model that operates at multiple side channel bandwidths to transmit the reduced reference information, [ITU-T J.343.5] includes two models and [ITU-T J.343.6] includes two models.

All of these models predict video quality in terms of mean opinion score (MOS), for example on a five-level absolute category rating (ACR) scale (see [ITU-T P.910] or [ITU-T P.913]).

1.1 Applications

This Recommendation describes models that estimate perceptual video quality. The applications for the estimation models described in this Recommendation include, but are not limited to:

– real-time, in-service quality monitoring at the source;

– remote destination quality monitoring;

– quality measurement of transmission systems that utilize video compression and decompression techniques, including concatenations of such techniques.

More information about applications can be found in the individual Recommendations that address these six application areas.

2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

[ITU-T J.340] Recommendation ITU-T J.340 (2010), *Reference algorithm for computing peak signal to noise ratio of a processed video sequence with compensation for constant spatial shifts, constant temporal shift, and constant luminance gain and offset*.

[ITU-T J.343.1] Recommendation ITU-T J.343.1 (2014), *Hybrid-NRe objective perceptual video quality measurement for HDTV and multimedia IP-based video services in the presence of encrypted bitstream data*.

[ITU-T J.343.2] Recommendation ITU-T J.343.2 (2014), *Hybrid-NR objective perceptual video quality measurement for HDTV and multimedia IP-based video services in the presence of non-encrypted bitstream data*.

[ITU-T J.343.3] Recommendation ITU-T J.343.3 (2014), *Hybrid-RRe objective perceptual video quality measurement for HDTV and multimedia IP-based video services in the presence of a reduced reference signal and encrypted bitstream data*.

[ITU-T J.343.4] Recommendation ITU-T J.343.4 (2014), *Hybrid-RR objective perceptual video quality measurement for HDTV and multimedia IP-based video services in the presence of a reduced reference signal and non-encrypted bitstream data*.

[ITU-T J.343.5] Recommendation ITU-T J.343.5 (2014), *Hybrid-FRe objective perceptual video quality measurement for HDTV and multimedia IP-based video services in the presence of a full reference signal and encrypted bitstream data*.

[ITU-T J.343.6] Recommendation ITU-T J.343.6 (2014), *Hybrid-FR objective perceptual video quality measurement for HDTV and multimedia IP-based video services in the presence of a full reference signal and non-encrypted bitstream data*.

[ITU-T P.910] Recommendation ITU-T P.910 (2008), *Subjective video quality assessment methods for multimedia applications*.

[ITU-T P.913] Recommendation ITU-T P.913 (2014), *Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment*.

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following terms defines elsewhere:

3.1.1 processed [ITU-T P.913]: The reference stimuli presented through a system under test.

3.1.2 processed video sequence [ITU-T P.913]: The processed video sequence (PVS) is the impaired version of a video sequence.

3.1.3 reference [ITU-T P.913]: The original version of each source stimulus. This is the highest quality version available of the audio sample, video clip or audiovisual sequence.

3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

3.2.1 hybrid full reference model: An objective video quality model that predicts subjective quality using the reference video, the decoded video frames, packet headers, and the video payload. Such models cannot analyse encrypted video.

3.2.2 hybrid full reference encrypted model: An objective video quality model that predicts subjective quality using the reference video, the decoded video frames, and packet headers. Such models are suitable for use with encrypted video.

3.2.3 hybrid no reference model: An objective video quality model that predicts subjective quality using the decoded video frames, packet headers, and video payload. Such models can be deployed in-service but cannot analyse encrypted video.

3.2.4 hybrid no reference encrypted model: An objective video quality model that predicts subjective quality using the decoded video frames and packet headers. Such models can be deployed in-service and are suitable for use with encrypted video.

3.2.5 hybrid reduced reference model: An objective video quality model that predicts subjective quality using the decoded video frames, packet headers, video payload and features extracted from the reference video. Such models can be deployed in-service but cannot analyse encrypted video.

3.2.6 hybrid reduced reference encrypted model: An objective video quality model that predicts subjective quality using the decoded video frames, packet headers, and features extracted from the reference video. These models can be deployed in-service and are suitable for use with encrypted video.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

ACR	Absolute Category Rating
CODEC	Coder-Decoder
ES	Elementary bitstream
FR	Full Reference
Hybrid-FR	Hybrid Full Reference
Hybrid-FRe	Hybrid Full Reference encrypted
Hybrid-NR	Hybrid No Reference
Hybrid-NRe	Hybrid No Reference encrypted
Hybrid-RR	Hybrid Reduced Reference
Hybrid-RRe	Hybrid Reduced Reference encrypted
MOS	Mean Opinion Score
MPEG	Moving Picture Experts Group
NR	No (or zero) Reference
PES	Packetized Elementary bitstream
PSNR	Peak Signal to Noise Ratio
PVS	Processed Video Sequence
RMSE	Root-Mean Square Error
RR	Reduced Reference
SRC	Source Reference Channel or Circuit
VQEG	Video Quality Experts Group

5 Conventions

None.

6 Description of hybrid perceptual bitstream model types

This Recommendation specifies objective video quality measurement methods which use both processed video sequences and bitstream data. The bitstream data may be provided in the forms of elementary bitstream (ES), packetized elementary bitstream (PES) or packet video (Figure 1). Table 1 shows required inputs for each model.

Table 1 – Required inputs

Model type	Model name	Required inputs
Hybrid NRe	RST-V model YHyNRe	Processed video sequence (PVS) Encrypted bitstream
Hybrid NR	YHyNR	PVS Non-encrypted bitstream
Hybrid RRe	YHyRRe	PVS Features extracted from source reference channel (SRC) Encrypted bitstream
Hybrid RR	YHyRR	PVS Features extracted from SRC Non-encrypted bitstream
Hybrid FRe	PEVQ-S (e) YHyFRe	PVS SRC Encrypted bitstream
Hybrid FR	PEVQ-S YHyFR	PVS SRC Non-encrypted bitstream

Hybrid-NR and Hybrid-NRe models use only PVS and bitstream data, as shown in Figure 1 and Figure 2. Where Hybrid-NR models have access to all of this data, Hybrid-NRe models do not have access to the video payload. Therefore, these models can be used with encrypted bitstreams.

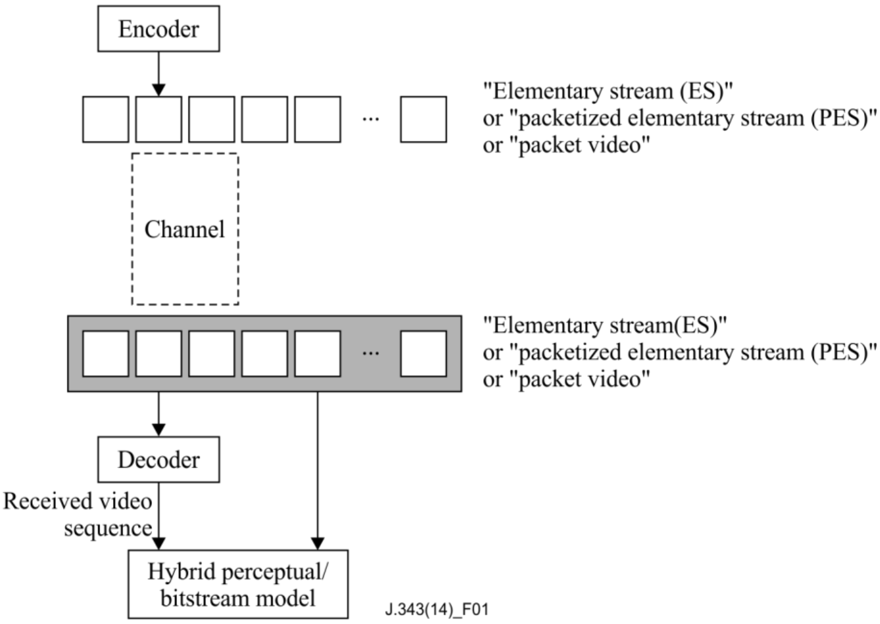
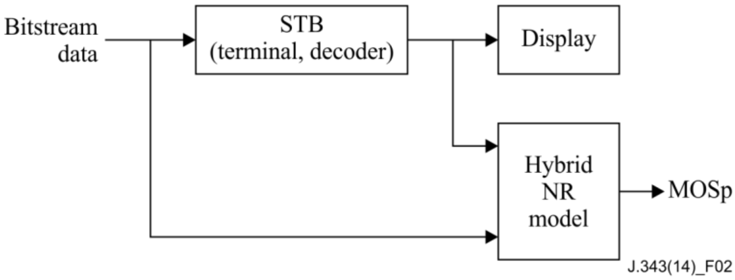


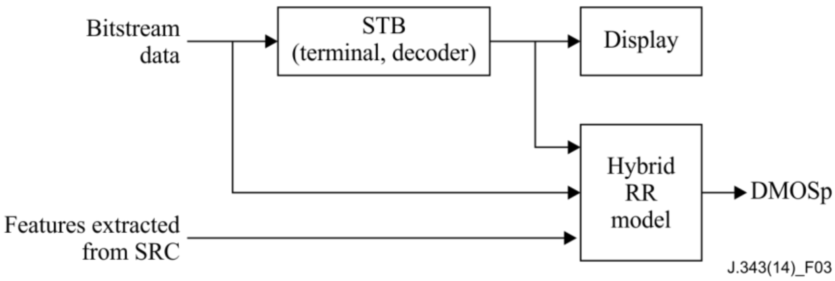
Figure 1 – Block-diagram depicts the core concept of hybrid perceptual bitstream models



MOSp: predicted MOS by the model

Figure 2 – Block-diagram of the Hybrid-NR model

In addition to the data available to a Hybrid-NR model, Hybrid-RR and Hybrid-RRe models also use features extracted from source video sequences. Figure 3 shows a Hybrid-RR model. In addition to the bitstream data, the Hybrid-RR model uses the features extracted from the SRC. Where Hybrid-RR models have access to all of this data, Hybrid-RRe models do not have access to the video payload. Therefore, these models can be used with encrypted bitstreams.



DMOSp: predicted DMOS by the model

Figure 3 – Block-diagram depicts the Hybrid-RR model

In addition to the data available to a Hybrid-NR model, the Hybrid-FR and Hybrid-FRe models also use reference video sequences. Figure 4 shows a Hybrid-FR model. The Hybrid-FR and Hybrid-FRe models needs the SRC. Where Hybrid-FR models have access to all of this data, Hybrid-FRe models do not have access to the video payload. Therefore, these models can be used with encrypted bitstreams.

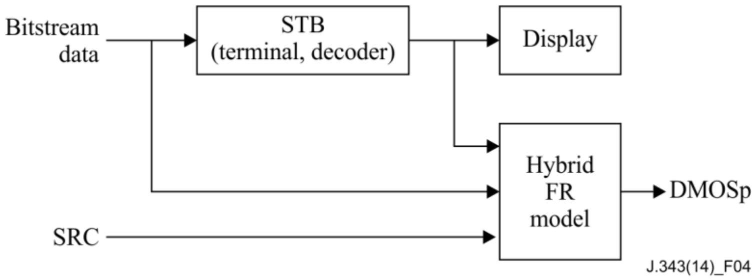


Figure 4 – Block-diagram depicts the Hybrid-FR model

Annex A

Summary of VQEG validation of hybrid models

(This annex forms an integral part of this Recommendation.)

NOTE – The text in this annex is taken directly from Section 1 (Summary) of [b-VQEG Hybrid] and differs only as follows:

- Models not present in this series of Recommendations have been omitted, with one exception: peak signal to noise ratio (PSNR) from [ITU-T J.340] is included as a reference point for model accuracy.
- The RST-V model is referred to as "TVM-Hybrid Encrypted" within [b-VQEG Hybrid].
- The PEVQ-S (e) model is referred to as "PEVQ-S" within [b-VQEG Hybrid].
- Table 1 from [b-VQEG Hybrid] is re-named Table A.1, Table 2 from [b-VQEG Hybrid] is renamed Table A.2, etc. to align with numbering in this annex.
- Grammatical edits were made (e.g., references to [b-VQEG Hybrid], clarifications within footnotes).

This annex presents results from the hybrid project of the Video Quality Experts Group (VQEG). The goal was to evaluate the performance of hybrid perceptual bitstream models predicting the perceived video quality based on input consisting of video frames and bitstream information. The Hybrid Test addressed no-reference, reduced reference and full reference hybrid models as well as one no-reference non-hybrid model. This Hybrid Test addresses the following video formats: 1080p at 25 and 29.97 frames-per-second, 1080i at 50 and 59.94 fields-per second, VGA at 25 and 30 frames-per-second, and WVGA at 25 and 30 frames-per-second. The Hybrid Test addresses videos encoded using H.264 and transmitted over RTP/UDP for VGA/WVGA formats, and transmitted over TS/RTP/UDP for HD.

A total of eleven testing laboratories coordinated to perform subjective testing. Ten subjective experiments provided data against which model validation was performed. The experiments were divided as follows: five HD experiments, three VGA experiments, and two WVGA experiments. One of the WVGA experiments was assessed twice, once with rebuffering conditions included and once with rebuffering conditions excluded. The impairments examined were restricted to coding artifacts, packet loss, tandem coding, rebuffering, scaling of the resolution, frame rate reduction, error concealment, slicing, freezing, as well as live and simulated transmission errors.

Two common sets of video sequences were created: one for the HD experiments and another for the VGA/WVGA experiments. These common sets were inserted into each experiment, to anchor the video experiments to one another and assist in comparisons between the subjective experiments. These common sequences were used to map the experiments onto a single scale (called "HD merge", "VGA merge" in this report).

Models were submitted by four proponents: Deutsche Telekom AG, OPTICOM GmbH, SwissQual AG, and Yonsei University. Different models were submitted for different encryption levels (encrypted and non-encrypted) and reduced reference side channel bitrates (56, 128 and 256 kbit/s). Six hybrid no-reference models, two hybrid reduced reference models, four hybrid full reference models and one no-reference model were submitted. No models were withdrawn. Thirteen models are presented in this final report.

The hybrid data may not be used as evidence to standardize any other objective video quality model that was not tested within this phase. This comparison would not be fair, because another model could have been trained on the hybrid data.

A.1 Subjective datasets

- HybridVGA1: This test focuses on live video recording of video streams transmitted over a commercially operated 3G mobile network or transmitted over LAN with simulated network impairments. This dataset has 15 different source videos of VGA resolution at 30 fps. Video sequences contain coding impairments and transmission impairments with packet loss, packet delay, and rebuffering due to limited throughput.

- HybridVGA2: This VGA experiment included x264 encoding (from 128 kbit/s to 1200 kbit/s) with simulated transmission errors (burst/random, from 0.5%~1.5%). Down-sampling (to QVGA) followed by up-sampling and error concealments (slicing/freezing) were applied to some sequences. The frame rate was set to 30 fps except for some sequences (10 fps).
- HybridVGA3: This database is targeting transmission errors without player rebuffering effects. The database includes simulated transmission distortions, as well as transmission over a commercially operated IP Network. Transcoding, scaling, and error concealment were applied to some sequences as pre- and post-processing. The resolution is VGA (640 × 480 pixels) and the frame rate is 25 frames per second.
- HybridWVGA1: This WVGA experiment includes x264 encoding (from 128 kbit/s to 1200 kbit/s) with simulated transmission errors (burst/random, from 0.5%~1.5%). Down-sampling (to QWVGA) followed by up-sampling and error concealments (slicing/freezing) were applied to some sequences. The frame rate was set to 30 fps except for some sequences (15 fps).
- HybridWVGA2: This WVGA experiment focuses on simulated rebuffering. Videos were streamed over a local loopback, and changing buffer sizes resulted in packet delay and rebuffering. In addition, the test set contains videos with coding only distortions, and down-sampling before transmission followed by up-sampling at the video decoder. A total of 8 source videos were paired with 11 HRCs, resulting in a total of 88 PVSs each at 25 fps.
- HybridWVGA2 no rebuf.: Contains the HRCs from dataset Hybrid WVGA2 that do not contain rebuffering (see section 3.1 of [b-VQEG Hybrid]).
- VGA merge: Datasets VGA1, VGA2, VGA3, WVGA1 and WVGA2 are combined into a single dataset, which provides an estimate of the model's overall VGA/WVGA performance. The algorithm used to combine datasets has some limitations (see section 4.2.3 of [b-VQEG Hybrid] for details).
- VGA merge no rebuf: Does the same but eliminates rebuffering from dataset HybridWVGA2 (see section 3.1 of [b-VQEG Hybrid]).
- Hybrid HD1: This 1080i 60fps experiment contains x264 encoding / simulated loss (uniform-bursty distributions, low/medium/high packet loss rates) / VLC and T-Labs decoder. Many sequences contained network impairments, which resulted in a cluster of low quality data points (from 1 to 2.5).
- Hybrid HD2: This 1080i 50 fps experiment presents typical H264 over UDP streaming scenarios at bit rates from 2 Mbit/s to 15 Mbit/s, with transcoding from lower bit rate to higher bit rate, packet losses (from 5-10 packets up to 0.125%), a relatively short GOP structure (12 or 15 frames in a single GOP) and short IP packets (242 bytes long).
- Hybrid HD3: This 1080p 30 fps experiment includes x264-encoded sequences with coding distortions and simulated network errors (uniform and bursty loss), targeting H.264 over UDP streaming scenarios with low (1.5 Mbit/s) to high (8 Mbit/s) bitrates and low (0.125%) to high (0.5%) packet loss ratios. Packet loss was concealed, resulting in slicing and freezing artifacts.
- Hybrid HD4: This 1080p25 database consists of sequences containing encoding-only artifacts or degradations caused by packet losses during video streaming over UDP. Furthermore, some more advanced features of H.264 video encoding are used such as Intra-refresh, open GOP structures, and hierarchical B-pictures.

- Hybrid HD5: This database contains 10 different source video sequences (1080i, 60 fps). This experiment includes x264 encoding (from 2 Mbit/s to 14 Mbit/s) with simulated transmission errors (burst/random, from 0.1%~1.3%). Down-sampling (by a factor 1/2, 1/3) followed by up-sampling and error concealments (slicing/freezing) were applied to some sequences.
- HD merge: Datasets HD1, HD2, HD3, HD4 and HD5 are combined into a single dataset, which provides an estimate of the model's overall HD performance. The algorithm used to combine datasets has some limitations (see section 4.2.3 of [b-VQEG Hybrid] for details).

A.2 Model performance summary

The models were evaluated using three statistics that provide insights into model performance: Root-mean square error (RMSE), Pearson Correlation and Epsilon Independent RMSE. Each model was fitted to each subjective experiment using a 3rd order monotonic polynomial function. RMSE is considered the primary metric for analysis in this report. Thus, RMSE is used to determine whether a model is in the group of top performing models for one video format/resolution (i.e., a group of models that include the top performing model and models that are statistically equivalent to the top performing model).

Table A.1 through Table A.6 provide RMSE and Pearson Correlation for each type of model. Note that:

- Mean opinion score (MOS) was calculated using a 5-level absolute category rating (ACR) scale, and thus spans the range [1, 5]. Hybrid-NR and NR models are analysed using MOS.
- DMOS was calculated on the same [1, 5] scale using the ACR-HR method. Hybrid-FR and Hybrid-RR models are analysed using DMOS.
- PSNR is computed according to [ITU-T J.340], for comparison purposes.
- Within one table, all RMSE values can be directly compared.
- On the top half of each table (RMSE), the yellow highlights indicate that this model is statistically equivalent to the top performing model on this particular dataset. This statistical equivalence is computed using RMSE.
- On the bottom half of each table (Pearson correlation), the light blue highlights indicate that this model is equivalent to or better than PSNR for this particular dataset. This comparison is made for all models that can be deployed in-service (Hybrid-NR, Hybrid-RR and NR). This statistical equivalence is computed using RMSE.
- On the bottom half of each table (Pearson correlation), the light green highlights indicate that this model is better than PSNR for this particular dataset. This comparison is made for all models that cannot be deployed in-service (Hybrid-FR). This statistical equivalence is computed using RMSE.
- "Mean of VGA" computes the averages for that model over all five VGA and WVGA datasets (i.e., HybridVGA1, HybridVGA2, HybridVGA3, HybridWVGA1, and HybridWVGA2).
- "Mean of VGA no rebuf" does the same but eliminates rebuffering from dataset HybridWVGA2, i.e., computes the average of HybridVGA1, HybridVGA2, HybridVGA3, HybridWVGA1, and HybridWVGA2 no rebuf, (see section 3.1 of [b-VQEG Hybrid]).
- "Mean of HD" computes the average for that model over all five HD datasets (i.e., HybridHD1, HybridHD2, HybridHD3, HybridHD4 and HybridHD5).

Comparisons between different types of models are presented in the body of [b-VQEG Hybrid].

Table A.1 – Hybrid-NR non-encrypted model performance summary, using MOS[1]

Statistic	Dataset	YHyNR	PSNR
RMSE	HybridVGA1	0.70	0.68
	HybridVGA2	0.49	0.62
	HybridVGA3	0.52	0.59
	HybridWVGA1	0.59	0.62
	HybridWVGA2	0.49	0.65
	HybridWVGA2 no rebuf	0.42	0.65
	VGA merge/no rebuf	0.59 / 0.59	0.69 / 0.70
	Mean of VGA/no rebuf	0.56 / 0.55	0.63 / 0.63
RMSE	HybridHD1	0.43	0.47
	HybridHD2	0.54	0.65
	HybridHD3	0.47	0.63
	HybridHD4	0.70	0.59
	HybridHD5	0.50	0.71
	HD merge	0.56	0.63
	Mean of HD	0.52	0.61
Pearson correlation	HybridVGA1	0.69	0.72
	HybridVGA2	0.88	0.80
	HybridVGA3	0.79	0.72
	HybridWVGA1	0.83	0.81
	HybridWVGA2	0.87	0.76
	HybridWVGA2 no rebuf	0.93	0.82
	VGA merge/no rebuf	0.78 / 0.79	0.68 / 0.69
	Mean of VGA/ no rebuf	0.81 / 0.82	0.76 / 0.79
Pearson correlation	HybridHD1	0.89	0.87
	HybridHD2	0.84	0.76
	HybridHD3	0.88	0.78
	HybridHD4	0.80	0.86
	HybridHD5	0.82	0.58
	HD merge	0.84	0.79
	Mean of HD	0.85	0.77

[1] Yellow highlight (top) indicates model is either the top performing model for this dataset or has equivalent performance. If no model has a yellow highlight, then [b-VQEG Hybrid] contains a model that had superior performance for that dataset. Blue highlight (bottom) indicates model performs equivalently to or better than PSNR. These statistical comparisons are computed using RMSE for both the top (yellow) and bottom (blue) part. Highlights in the "VGA merge/- no rebuff" rows mark the performance of "VGA merge".

Table A.2 – Hybrid-NR encrypted model performance summary, using MOS[2]

Statistic	Dataset(s)	RST-V	YHyNRe	PSNR
RMSE	HybridVGA1	0.51	0.70	0.68
	HybridVGA2	0.59	0.57	0.62
	HybridVGA3	0.54	0.60	0.59
	HybridWVGA1	0.71	0.67	0.62
	HybridWVGA2	0.62	0.56	0.65
	HybridWVGA2 no rebuf	0.41	0.47	0.65
	VGA merge/no rebuf	0.61 / 0.61	0.66 / 0.66	0.69 / 0.70
	Mean of VGA/no rebuf	0.59 / 0.55	0.62 / 0.60	0.63 / 0.63
RMSE	HybridHD1	0.48	0.34	0.47
	HybridHD2	0.70	0.72	0.65
	HybridHD3	0.74	0.70	0.63
	HybridHD4	0.81	0.74	0.59
	HybridHD5	0.49	0.55	0.71
	HD merge	0.69	0.63	0.63
	Mean of HD	0.64	0.61	0.61
Pearson correlation	HybridVGA1	0.85	0.69	0.72
	HybridVGA2	0.82	0.83	0.80
	HybridVGA3	0.77	0.71	0.72
	HybridWVGA1	0.74	0.77	0.81
	HybridWVGA2	0.78	0.83	0.76
	HybridWVGA2 no rebuf	0.93	0.91	0.82
	VGA merge/no rebuf	0.76 / 0.78	0.72 / 0.73	0.68 / 0.69
	Mean of VGA/no rebuf	0.79 / 0.82	0.77 / 0.78	0.76 / 0.77
Pearson correlation	HybridHD1	0.86	0.93	0.87
	HybridHD2	0.72	0.70	0.76
	HybridHD3	0.67	0.71	0.78
	HybridHD4	0.72	0.77	0.86
	HybridHD5	0.83	0.78	0.58
	HD merge	0.74	0.79	0.79
	Mean of HD	0.77	0.78	0.77

[2] Yellow highlight (top) indicates model is either the top performing model for this dataset or has equivalent performance. If no model has a yellow highlight, then [b-VQEG Hybrid] contains a model that had superior performance for that dataset. Blue highlight (bottom) indicates model performs equivalently to or better than PSNR. These statistical comparisons are computed using RMSE for both the top (yellow) and bottom (blue) part. Highlights in the "VGA merge/- no rebuff" rows mark the performance of "VGA merge".

Table A.3 – Hybrid-RR non-encrypted model performance summary, using DMOS[3]

Statistic	Dataset(s)	YHyRR at 56 kbit/s	YHyRR at 128 kbit/s	YHyRR at 256 kbit/s	PSNR
RMSE	HybridVGA1	0.79	0.79	–	0.66
	HybridVGA2	0.49	0.49	–	0.63
	HybridVGA3	0.41	0.41	–	0.56
	HybridWVGA1	0.50	0.50	–	0.59
	HybridWVGA2	0.39	0.39	–	0.60
	HybridWVGA2 no rebuf	0.31	0.30	–	0.59
	VGA merge/no rebuf	0.57 / 0.58	0.57 / 0.58	–	0.66 / 0.66
	Mean of VGA/no rebuf	0.52 / 50	0.51 / 0.50	–	0.61 / 0.61
RMSE	HybridHD1	0.41	0.41	0.41	0.42
	HybridHD2	0.67	0.66	0.66	0.59
	HybridHD3	0.52	0.52	0.52	0.60
	HybridHD4	0.56	0.56	0.56	0.60
	HybridHD5	0.46	0.46	0.46	0.72
	HD merge	0.55	0.55	0.55	0.61
	Mean of HD	0.52	0.52	0.52	0.59
Pearson correlation	HybridVGA1	0.63	0.63	–	0.77
	HybridVGA2	0.89	0.89	–	0.81
	HybridVGA3	0.88	0.88	–	0.75
	HybridWVGA1	0.88	0.88	–	0.83
	HybridWVGA2	0.92	0.92	–	0.79
	HybridWVGA2 no rebuf	0.96	0.96	–	0.86
	VGA merge/no rebuf	0.80 / 0.80	0.80 / 0.81	–	0.72 / 0.73
	Mean of VGA/no rebuf	0.84 / 0.85	0.84 / 0.85	–	0.79 / 0.80
Pearson correlation	HybridHD1	0.91	0.91	0.91	0.91
	HybridHD2	0.78	0.79	0.79	0.84

3 Yellow highlight (top) indicates model is either the top performing model for this dataset or has equivalent performance. Blue highlight (bottom) indicates model performs equivalently to or better than PSNR. These statistical comparisons are computed using RMSE for both the top (yellow) and bottom (blue) part. Highlights in the "VGA merge/- no rebuff" rows mark the performance of "VGA merge".

Table A.3 – Hybrid-RR non-encrypted model performance summary, using DMOS[3]

Statistic	Dataset(s)	YHyRR at 56 kbit/s	YHyRR at 128 kbit/s	YHyRR at 256 kbit/s	PSNR
	HybridHD3	0.86	0.86	0.86	0.80
	HybridHD4	0.88	0.87	0.88	0.86
	HybridHD5	0.86	0.86	0.86	0.60
	HD merge	0.86	0.86	0.86	0.83
	Mean of HD	0.86	0.86	0.86	0.80

Table A.4 – Hybrid-RR encrypted model performance summary, using DMOS[4]

Statistic	Dataset(s)	YHyRRe at 56 kbit/s	YHyRRe at 128 kbit/s	YHyRRe at 256 kbit/s	PSNR
RMSE	HybridVGA1	0.79	0.78	–	0.66
	HybridVGA2	0.49	0.49	–	0.63
	HybridVGA3	0.44	0.44	–	0.56
	HybridWVGA1	0.49	0.49	–	0.59
	HybridWVGA2	0.42	0.41	–	0.60
	HybridWVGA2 no rebuf	0.30	0.30	–	0.59
	VGA merge/no rebuf	0.58 / 0.59	0.58 / 0.58	–	0.66 / 0.66
	Mean of VGA/no rebuf	0.53 / 0.50	0.52 / 0.50	–	0.61 / 0.61
RMSE	HybridHD1	0.38	0.38	0.38	0.42
	HybridHD2	0.74	0.73	0.73	0.59
	HybridHD3	0.64	0.64	0.64	0.60
	HybridHD4	0.60	0.60	0.59	0.60
	HybridHD5	0.39	0.39	0.38	0.72
	HD merge	0.58	0.58	0.58	0.61
	Mean of HD	0.55	0.55	0.55	0.59
Pearson correlation	HybridVGA1	0.64	0.64	–	0.77
	HybridVGA2	0.89	0.89	–	0.81

[4] Yellow highlight (top) indicates model is either the top performing model for this dataset or has equivalent performance. Blue highlight (bottom) indicates model performs equivalently to or better than PSNR. These statistical comparisons are computed using RMSE for both the top (yellow) and bottom (blue) part. Highlights in the "VGA merge/- no rebuff" rows mark the performance of "VGA merge".

Table A.4 – Hybrid-RR encrypted model performance summary, using DMOS[4]

Statistic	Dataset(s)	YHyRRe at 56 kbit/s	YHyRRe at 128 kbit/s	YHyRRe at 256 kbit/s	PSNR
Pearson correlation	HybridVGA3	0.86	0.86	–	0.75
	HybridWVGA1	0.89	0.89	–	0.83
	HybridWVGA2	0.91	0.91	–	0.79
	HybridWVGA2 no rebuf	0.96	0.96	–	0.86
	VGA merge/no rebuf	0.79 / 0.80	0.79 / 0.80	–	0.72 / 0.73
	Mean of VGA/no rebuf	0.84 / 0.85	0.84 / 0.85	–	0.79 / 0.80
	HybridHD1	0.92	0.92	0.92	0.91
	HybridHD2	0.72	0.73	0.73	0.84
	HybridHD3	0.78	0.78	0.78	0.80
	HybridHD4	0.86	0.86	0.86	0.86
	HybridHD5	0.91	0.91	0.91	0.60
	HD merge	0.85	0.85	0.85	0.83
	Mean of HD	0.84	0.84	0.84	0.80

Table A.5 – Hybrid-FR non-encrypted model performance summary, using DMOS[5]

Statistic	Dataset(s)	PEVQ-S	Yonsei-hFR	PSNR
RMSE	HybridVGA1	0.65	0.79	0.66
	HybridVGA2	0.51	0.49	0.63
	HybridVGA3	0.52	0.41	0.56
	HybridWVGA1	0.54	0.50	0.59
	HybridWVGA2	0.53	0.39	0.60
	HybridWVGA2 no rebuf	0.51	0.31	0.59
	VGA merge/no rebuf	0.57 / 0.55	0.57 / 0.58	0.66 / 0.66
	Mean of VGA/no rebuf	0.55 / 0.55	0.52 / 0.50	0.61 / 0.61
RMSE	HybridHD1	0.34	0.41	0.42
	HybridHD2	0.51	0.66	0.59
	HybridHD3	0.41	0.52	0.60
	HybridHD4	0.64	0.57	0.60
	HybridHD5	0.50	0.46	0.72
	HD merge	0.51	0.55	0.61
	Mean of HD	0.48	0.52	0.59
Pearson correlation	HybridVGA1	0.77	0.63	0.77
	HybridVGA2	0.88	0.89	0.81
	HybridVGA3	0.79	0.88	0.75
	HybridWVGA1	0.86	0.88	0.83
	HybridWVGA2	0.84	0.92	0.79
	HybridWVGA2 no rebuf	0.89	0.96	0.86
	VGA merge/no rebuf	0.81 / 0.83	0.80 / 0.81	0.72 / 0.73
	Mean of VGA/no rebuf	0.83 / 0.84	0.84 / 0.85	0.79 / 0.80
Pearson correlation	HybridHD1	0.94	0.91	0.90
	HybridHD2	0.88	0.79	0.84
	HybridHD3	0.91	0.86	0.80
	HybridHD4	0.83	0.87	0.86
	HybridHD5	0.84	0.86	0.60
	HD merge	0.88	0.86	0.83
	Mean of HD	0.88	0.86	0.80

5 Yellow highlight (top) indicates model is either the top performing model for this dataset or has equivalent performance. Green highlight (bottom) indicates model performs better than PSNR. These statistical comparisons are computed using RMSE for both the top (yellow) and bottom (green) part. Highlights in the "VGA merge/- no rebuff" rows mark the performance of "VGA merge".

Table A.6 – Hybrid-FR encrypted model performance summary, using DMOS[6]

Statistic	Dataset(s)	PEVQ-S (e) for (pes+rtp)	PEVQ-S (e) for (ts+rtp)	YHyFRe	PSNR
RMSE	HybridVGA1	0.65	0.65	0.78	0.66
	HybridVGA2	0.51	0.51	0.49	0.63
	HybridVGA3	0.52	0.52	0.44	0.56
	HybridWVGA1	0.54	0.54	0.49	0.59
	HybridWVGA2	0.53	0.53	0.41	0.60
	HybridWVGA2 no rebuf	0.51	0.51	0.30	0.59
	VGA merge/no rebuf	0.57 / 0.55	0.57 / 0.55	0.58 / 0.59	0.66 / 0.66
	Mean of VGA/no rebuf	0.55 / 0.55	0.55 / 0.55	0.52 / 0.50	0.61 / 0.61
RMSE	HybridHD1	0.34	0.34	0.38	0.42
	HybridHD2	0.51	0.50	0.73	0.59
	HybridHD3	0.41	0.41	0.64	0.60
	HybridHD4	0.64	0.64	0.60	0.60
	HybridHD5	0.50	0.50	0.38	0.72
	HD merge	0.51	0.51	0.58	0.61
	Mean of HD	0.48	0.48	0.55	0.59
Pearson correlation	HybridVGA1	0.77	0.77	0.64	0.77
	HybridVGA2	0.88	0.88	0.89	0.81
	HybridVGA3	0.79	0.79	0.86	0.75
	HybridWVGA1	0.86	0.86	0.89	0.83
	HybridWVGA2	0.84	0.84	0.91	0.79
	HybridWVGA2 no rebuf	0.89	0.89	0.96	0.86
	VGA merge/no rebuf	0.81 / 0.83	0.81 / 0.83	0.79 / 0.80	0.72 / 0.73
	Mean of VGA/no rebuf	0.83 / 0.84	0.83 / 0.84	0.84 / 0.85	0.79 / 0.80
Pearson correlation	HybridHD1	0.94	0.94	0.92	0.91
	HybridHD2	0.88	0.88	0.73	0.84

[6] Yellow highlight (top) indicates model is either the top performing model for this dataset or has equivalent performance. Green highlight (bottom) indicates model performs better than PSNR. These statistical comparisons are computed using RMSE for both the top (yellow) and bottom (green) part. Highlights in the "VGA merge/- no rebuff" rows mark the performance of "VGA merge". The PEVQ-S (e) model is analysed in two different ways: for handling PES payload encryption only (pes+rtp) and for handling TS payload encryption (ts+rtp).

Table A.6 – Hybrid-FR encrypted model performance summary, using DMOS[6]

Statistic	Dataset(s)	PEVQ-S (e) for (pes+rtp)	PEVQ-S (e) for (ts+rtp)	YHyFRe	PSNR
	HybridHD3	0.91	0.91	0.78	0.80
	HybridHD4	0.83	0.83	0.86	0.86
	HybridHD5	0.84	0.84	0.91	0.60
	HD merge	0.88	0.88	0.85	0.83
	Mean of HD	0.88	0.88	0.84	0.80

Appendix I

Test vectors for the ITU-T J.343 family of standards

(This appendix does not form an integral part of this Recommendation.)

The Consumer Digital Video Library (CDVL) kindly offered some of the video clips used in the development of this Recommendation. These video clips may be used freely for research and development purposes, and are available at [b-CDVL1].

Some test vectors for this Recommendation are available at [b-CDVL2], courtesy of CDVL.

The test vector ZIP-file contains the test vectors for all models in the ITU-T J.343 family of Recommendations.

Each sample is composed of a reference and a processed video sequence. The processed video sequence is available in compressed format (stored in a pcap file) and a decoded format (stored in an avi file).

In addition, the reference and processed uncompressed videos are available in two versions, with pre-/post-roll cut and in full length (extended).

The target values (what the models need to predict for the test vectors) are provided in the tables below:

*sample_name * **model**	**RST-V**	**PEVQ-S**	**PEVQ-S(e)**	**YHyFR**	**YHyFRe**	**YHyNR**	**YHyNRe**	**YHyRR_56k**	**YHyRRe_56k**	**YHyRR_128k**	**YHyRRe_128k**	**YHyRR_256k**	**YHyRRe_256k**
sampleHD1	2.433	4.299	4.299	3.58	3.37	3.418	3.397	3.572	3.362	3.585	3.375	3.578	3.368
sampleHD2	4.225	4.886	4.886	3.881	3.889	4.075	4.09	3.877	3.884	3.879	3.886	3.882	3.89
sampleHD3	3.76	2.802	2.802	3.356	3.483	2.982	3.236	3.356	3.483	3.356	3.483	3.356	3.483
sampleHD4	2.487	2.361	2.361	2.667	2.546	2.474	2.889	2.741	2.621	2.708	2.588	2.691	2.57

*sample_name * **model**	**RST-V**	**PEVQ-S**	**PEVQ-S(e)**	**YHyFR**	**YHyFRe**	**YHyNR**	**YHyNRe**	**YHyRR_56k**	**YHyRRe_56k**	**YHyRR_128k**	**YHyRRe_128k**
sampleVGA1	3.484	3.757	3.757	1.359	1.148	2.071	1.649	1.282	1.071	1.359	1.148
sampleVGA2	4.597	4.677	4.677	2.27	2.043	2.039	1.585	2.27	2.043	2.27	2.043
sampleVGA3	4.239	4.67	4.67	2.113	1.835	1.938	1.383	2.089	1.812	2.095	1.817
sampleVGA4	3.554	4.172	4.172	2.408	2.182	2.04	1.588	2.368	2.142	2.349	2.122
sampleWVGA1	4.605	4.35	4.35	3.85	3.882	3.424	3.488	3.857	3.889	3.854	3.886
sampleWVGA2	2.654	2.401	2.401	2.66	2.593	2.06	1.925	2.657	2.589	2.658	2.59
sampleWVGA3	2.593	1.248	1.248	2.24	2.333	1.954	2.141	2.239	2.332	2.241	2.334
sampleWVGA4	2.264	2.618	2.618	2.837	2.731	2.278	2.066	2.858	2.752	2.845	2.739

Bibliography

[b-CDVL1] CDVL website,
https://www.cdvl.org/

[b-CDVL2] J.343 test vectors,
https://www.cdvl.org/find-videos/details.php?id=2956 (registration required)

[b-VQEG Hybrid] Video Quality Experts Group (2014), *Hybrid Perceptual/Bitstream Validation Test Final Report*.

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series D	Tariff and accounting principles and international telecommunication/ICT economic and policy issues
Series E	Overall network operation, telephone service, service operation and human factors
Series F	Non-telephone telecommunication services
Series G	Transmission systems and media, digital systems and networks
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	**Cable networks and transmission of television, sound programme and other multimedia signals**
Series K	Protection against interference
Series L	Environment and ICTs, climate change, e-waste, energy efficiency; construction, installation and protection of cables and other elements of outside plant
Series M	Telecommunication management, including TMN and network maintenance
Series N	Maintenance: international sound-programme and television-transmission circuits
Series O	Specifications of measuring equipment
Series P	Telephone transmission quality, telephone installations, local line networks
Series Q	Switching and signalling, and associated measurements and tests
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks, open system communications and security
Series Y	Global information infrastructure, Internet protocol aspects, next-generation networks, Internet of Things and smart cities
Series Z	Languages and general software aspects for telecommunication systems

Printed in Switzerland
Geneva, 2018