FINAL REPORT FROM THE VIDEO QUALITY EXPERTS GROUP ON THE VALIDATION OF OBJECTIVE MODELS OF MULTIMEDIA QUALITY ASSESSMENT, PHASE I ©2008 VQEG

Version 1.2 March 28, 2008

Version 1.2.2 April 1, 2008

Version 1.3.1 April 7, 2008

Version 1.4.1 April 15, 2008

Version 1.4.2 April 17, 2008

Version 1.5.1 April 21, 2008

Version 1.5.2 April 21, 2008

Version 1.5.3 April 23, 2008

Version 2.0 May 1, 2008 – Draft Submitted to ITU

Version 2.1 May, 2008

Version 2.2 July, 2008

Version 2.3 July, 2008

Version 2.4 August, 2008

Version 2.5 September 12, 2008

Version 2.6 September 12, 2008

VQEG_MM_Report_Final_v2.6.doc

Copyright Information

Draft VQEG Final Report of MM Phase I Validation Test ©2008 VQEG

http://www.vqeg.org

For more information contact:

Arthur Webster	webster@its.bldrdoc.gov	Co-Chair VQEG
Filippo Speranza	Filippo.Speranza@crc.ca	Co-Chair VQEG

Regarding the use of VQEG's Multimedia Phase I data:

Subjective data is available to the research community [Note: The subjective data will not be released outside the participants of VQEG's MM Phase I validation test until 1 year from September 12, 2008]. Some video sequences are owned by companies and permission must be obtained from them. See the VQEG Multimedia Phase I Final Report for the source of various test sequences.

VQEG validation subjective test data is placed in the public domain. Video sequences are available for further experiments with restrictions required by the copyright holder. Some video sequences have been approved for use in research experiments. Most may not be displayed in any public manner or for any commercial purpose. Some video sequences (such as 'Mobile and Calendar') will have less or no restrictions. VQEG objective validation test data may only be used with the proponent's approval. Results of future experiments conducted using the VQEG video sequences and subjective data may be reported and used for research and commercial purposes, however the VQEG final report should be referenced in any published material.

Acknowledgments

This report is the product of efforts made by many people over the past two years. It will be impossible to acknowledge all of them here but the efforts made by individuals listed below at dozens of laboratories worldwide contributed to the report.

Editing Committee:

Greg Cermak, Verizon (USA)

Kjell Brunnström, Acreo AB (Sweden)

David Hands, BT (UK)

Margaret Pinson, NTIA (USA)

Filippo Speranza, CRC (Canada)

Arthur Webster, NTIA (USA)

List of Contributors:

Ron Renaud, CRC (Canada)

Vittorio Baroncini, FUB (Italy)

Chulhee Lee, Yonsei University (Korea)

Stephen Wolf, NTIA/ITS (USA)

Quan Huynh-Thu, Psytechnics (UK)

Christian Schmidmer, OPTICOM (Germany)

Marcus Barkowsky, OPTICOM (Germany)

Roland Bitto, OPTICOM (Germany)

Alex Bourret, BT (France)

Jörgen Gustafsson, Ericsson (Sweden)

Patrick Le Callet, University of Nantes (France)

Ricardo Pastrana, Orange-FT (France)

Stefan Winkler, Symmetricom (USA)

Yves Dhondt, Ghent University - IBBT (Belgium)

Nicolas Staelens, Ghent University - IBBT (Belgium)

Phil Corriveau, INTEL (USA)

Jens Berger, SwissQual (Switzerland)

Romuald Pépion, IRCCyN (France)

Jun Okamoto, NTT (Japan)

Keishiro Watanable, NTT (Japan)

Akira Takahashi, NTT (Japan)
Osamu Sugimoto, KDDI (Japan)
Toru Yamada, NEC (Japan)
Yuukou Horita, University of Toyama (Japan)
Yoshikazu Kawayoke, University of Toyama (Japan)
Leigh Thorpe, Nortel (Canada)
Tim Rahrer, Nortel (Canada)
Irina Cotanis, Ericsson (Sweden)
Carolyn Ford, NTIA (USA)
Bruce Adams, Telchemy (USA)
Kevin Ferguson, Tektronix (USA)
Pero Juric, SwissQual (Switzerland)
Eugen Rodel, SwissQual (Switzerland)
Rene Widmer, SwissQual (Switzerland)
Jean-Louis Blin, Orange-FT (France)
Marie-Neige Garcia, Deutsche Telekom AG (Germany)
Alexander Raake, Deutsche Telekom AG (Germany)

Table of Contents

EXECUTIVE SUMMARY 8

1 INTRODUCTION 16

2 LIST OF DEFINITIONS 20

3 LIST OF ACRONYMS 22

4 TEST LABORATORIES 24

4.1 Independent Laboratory Group (ILG) 24
4.2 Proponent Laboratories 24
4.3 Other Laboratories 24

5 DESIGN OVERVIEW: SUBJECTIVE EVALUATION PROCEDURE 25

5.1 Subjective Test Method: ACR Method with Hidden Reference 25
5.2 Viewing distance 26
5.3 Display Specification and Set-up 26
5.4 Subjective Test Control Software 27
5.5 Subjects 28
5.6 Viewing Conditions 29
5.7 Experiment design 29
5.8 Randomization 29
5.9 Data Collection 29

6 LIMITATIONS ON SOURCE SCENES, HRCS & CALIBRATION 31

6.1 Source Video Processing Overview 31
6.2 Source Video Selection Criteria 31
6.3 Hypothetical Reference Circuit (HRC) Limitations 33
6.4 Processed Video Sequence Calibration: Limitations and Validation 37

7 MODEL EVALUATION CRITERIA 39

7.1 Evaluation Procedure 39
7.2 PSNR 39
7.3 Data Processing 40
7.4 Evaluation Metrics 41
7.5 Statistical Significance of the Results 45

8 COMMON VIDEO CLIP ANALYSIS AND INTERPRETATION 48

9 OFFICIAL ILG DATA ANALYSIS 50

9.1 VGA Primary Analysis 51

9.2 CIF Primary Data Analysis 59
9.3 QCIF Primary Data Analysis 67
10 SECONDARY DATA ANALYSIS 75
10.1 Explanation and Warnings 75
10.2 Official ILG Secondary Data Analysis 77
11 CONCLUSIONS 86
12 REFERENCES 86
Appendix I Model Descriptions 87
Appendix I.1 Proponent A, NTT 87
Appendix I.2 Proponent B, OPTICOM 87
Appendix I.3 Proponent C, Psytechnics 88
Appendix I.4 Proponent D, Yonsei University 89
Appendix I.5 Proponent E, SwissQual 89
Appendix II Subjective Testing Facilities 90
Appendix II.1 KDDI 90
Appendix II.2 NTT 91
Appendix II.3 OPTICOM 95
Appendix II.4 Psytechnics 96
Appendix II.5 SwissQual 97
Appendix II.6 Symmetricom 99
Appendix II.7 Yonsei University 100
Appendix II.8 Orange France Telecom 101
Appendix II.9 IRCCyN 102
Appendix II.10 Verizon 106
Appendix II.11 CRC-Nortel 109
Appendix II.12 Acreo 113
Appendix II.13 FUB 115
Appendix III SRC Associated with Each Individual Experiment 117
Appendix III.1 Scene Descriptions and Classifications 117
Appendix III.2 SRC in Each Common Set 124
Appendix III.3 SRC in Each Experiment's Scene Pool 124
Appendix III.4 Mapping of Scene Pools to Subjective Experiment 124
Appendix IV HRCs Associated with Each Individual Experiment 124
Appendix V Plots 124
Appendix V.1 VGA Plots 124

Appendix V.2 CIF Plots _____ 124
Appendix V.3 QCIF Plots _____ 124
Appendix VI Proponent Comments _____ 124
Appendix VI.1 NTT _____ 124
Appendix VI.2 OPTICOM _____ 124
Appendix VI.3 Psytechnics _____ 124
Appendix VI.4 SwissQual _____ 124
Appendix VI.5 Yonsei University _____ 124

EXECUTIVE SUMMARY

FINAL REPORT FROM THE VIDEO QUALITY EXPERTS GROUP ON THE VALIDATION OF OBJECTIVE MODELS OF MULTIMEDIA QUALITY ASSESSMENT, PHASE I

This document presents results from the Video Quality Experts Group (VQEG) Multimedia validation testing of objective video quality models for mobile/PDA and broadband internet communications services. This document provides input to the relevant standardization bodies responsible for producing international Recommendations.

The Multimedia Test contains two parallel evaluations of test video material. One evaluation is by panels of human observers (i.e., subjective testing). The other is by objective computational models of video quality (i.e., proponent models). The objective models are meant to predict the subjective judgments. Each subjective test will be referred to as an "experiment" throughout this document.

This Multimedia (MM) Test addresses three video resolutions (VGA, CIF, and QCIF) and three types of models: full reference (FR), reduced reference (RR), and no reference (NR). FR models have full access to the source video; RR models have limited bandwidth access to the source video; and NR models do not have access to the source video. RR models can be used in certain applications which cannot be addressed by FR models, such as in-service monitoring in networks. NR models can be used in certain applications which cannot be addressed by FR or RR approaches. Typically, no-reference models are applied in situations where the user doesn't have access to the source. Proponents were given the option of submitting different models for each video resolution and model type.

Forty-one subjective experiments provided data against which model validation was performed. The experiments were divided between the three video resolutions and two frame rates (25 fps and 30 fps). A common set of carefully chosen video sequences were inserted identically into each experiment at a given resolution, to anchor the video experiments to one another and assist in comparisons between the subjective experiments. The subjective experiments included processed video sequences with a wide range of quality, and both compression and transmission errors were present in the test conditions. These forty-one subjective experiments included 346 source video sequences and 5320 processed video sequences. These video clips were evaluated by 984 viewers.

A total of 13 organizations performed subjective testing for Multimedia. Of these organizations, 5 were model proponents (NTT, OPTICOM, Psytechnics, SwissQual, and Yonsei University) and the remainder were independent testing laboratories (Acreo, CRC, IRCCyN, France Telecom, FUB, Nortel, NTIA, and Verizon), or laboratories that helped by running processed video sequences (PVS) and subjective experiments (KDDI and Symmetricom). Objective models were submitted prior to scene selection, PVS generation, and subjective testing, to ensure none of the models could be trained on the test material. 31 models were submitted, 6 were withdrawn, and 25 are presented in this report. A model is considered in this context to be a model type (i.e., FR or RR or NR) for a specified resolution (i.e., VGA or CIF or QCIF).

Results for models submitted by the following five proponent organizations are included in this Multimedia Final Report:

- NTT (Japan)

- OPTICOM (Germany)
- Psytechnics (UK)
- SwissQual (Switzerland)
- Yonsei University (Korea)

The intention of VQEG is that the MM 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 MM data.

MODEL PERFORMANCE EVALUATION TECHNIQUES

The models were evaluated using three statistics that provide insights into model performance: Pearson Correlation, Root-Mean Squared Error (RMSE) and Outlier Ratios. These statistics compare the objective model's predictions with the subjective quality as judged by a panel of human observers. Each model was fitted to each subjective experiment, by optimizing Pearson Correlation with subjective data first, and minimizing RMSE second.

Each of these statistics (Pearson Correlation, RMSE, and Outlier Ratios) can be 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). Note that a model that is not in the top performing group and is statistically worse than the top performing model may still be statistically equivalent to one or more of the models that are in the top performing group. Statistical significances are computed for each metric separately, and therefore the models' ranking per video resolution is accomplished per each statistical metric.

When examining the total number of times a model is statistically equivalent to the top performing model for each resolution, comparisons between models should be performed carefully. Determining which differences in totals are statistically significant requires additional analysis not available in this document. As a general guideline, small differences in these totals do not indicate an overall difference in performance. This refers to the tables below.

Primary analysis considers each video sequence separately. Secondary analysis averages over all video sequences associated with each video system (or condition), and thus reflects how well the model tracks the average Hypothetical Reference Circuit (HRC) performance. The common set of video sequences are included in primary analysis but eliminated from secondary analysis. The following sections of the executive summary report on model performance across model type and resolution. The reader should be aware that performance is reported according to primary evaluation metrics and secondary evaluation metrics. Secondary analysis is presented to supplement the primary analysis. The primary analysis is the most important determinant of a model's performance.

PSNR was computed as a reference measure, and compared to all models. PSNR was computed using an exhaustive search for calibration and one constant delay for each video sequence. Models were required to perform their own calibration, where needed. While PSNR serves as a reference measure, it is not necessarily the most useful benchmark for recommendation of models.

FR MODEL PERFORMANCE

FR model results from NTT, OPTICOM, Psytechnics, and Yonsei for all three resolutions (VGA, CIF and QCIF) are included in this report.

Primary Analysis of FR Models

The average correlations of the primary analysis for the FR VGA models ranged from 0.79 to 0.83, and PSNR was 0.71. Individual model correlations for some experiments were as high as 0.94. The average RMSE for the FR VGA models ranged from 0.57 to 0.62, and PSNR was 0.71. The average outlier ratio for the FR VGA models ranged from 0.50 to 0.54, and PSNR was 0.62. All proposed models performed statistically better than PSNR for at least 8 of the 13 experiments. Based on each metric, each FR VGA model was in the group of top performing models the following number of times:

VGA

Statistic	Psy_FR	Opt_FR	Yon_FR	NTT_FR	PSNR
Correlation	11	10	10	8	3
RMSE	10	8	6	4	0
Outlier Ratio	12	11	8	9	4

The average correlations of the primary analysis for the FR CIF models ranged from 0.78 to 0.84, and PSNR was 0.66. Individual model correlations for some experiments were as high as 0.92. The average RMSE for the FR CIF models ranged from 0.53 to 0.60, and PSNR was 0.72. The average outlier ratio for the FR CIF models ranged from 0.51 to 0.54, and PSNR was 0.63. All proposed models performed statistically better than PSNR for at least 10 of the 14 experiments. Based on each metric, each FR CIF model was in the group of top performing models the following number of times:

CIF

Statistic	Psy_FR	Opt_FR	Yon_FR	NTT_FR	PSNR
Correlation	14	13	10	8	0
RMSE	13	10	9	6	0
Outlier Ratio	12	13	11	10	1

The average correlations of the primary analysis for the FR QCIF models ranged from 0.76 to 0.84, and PSNR was 0.66. Individual model correlations for some experiments were as high as 0.94. The average RMSE for the FR QCIF models ranged from 0.52 to 0.62, and PSNR was 0.72. The average outlier ratio for the FR QCIF models ranged from 0.46 to 0.52, and PSNR was 0.60. All proposed models performed statistically better than PSNR for at least 8 of the 14 experiments. Based on each metric, each FR QCIF model was in the group of top performing models the following number of times:

QCIF

Statistic	Psy_FR	Opt_FR	Yon_FR	NTT_FR	PSNR
Correlation	12	11	4	9	1
RMSE	11	10	2	7	1
Outlier Ratio	12	11	8	10	4

The gaps in performance between all of the models for individual experiments are very small. The models from Psytechnics and OPTICOM tend to perform slightly better than the NTT and Yonsei models in some resolutions; however for some experiments this difference is not statistically significant. The Psytechnics and OPTICOM models usually produce statistically equivalent results. For QCIF the model from NTT is often statistically equivalent to the models of Psytechnics and OPTICOM. For VGA, the Yonsei model is typically statistically equivalent to the Psytechnics and OPTICOM models.

Secondary Analysis of FR

The secondary analysis shows in principle a similar picture. The correlation coefficients generally increase. For VGA the FR models from OPTICOM and Psytechnics tend to perform a bit better than the two other ones. However, all tested models show disadvantages for individual experiments. For CIF the performance of all FR models is very similar. For QCIF, the performance of all FR models is very similar. The NTT model shows no disadvantages for any experiment (all correlation coefficients above 0.90).

FR Model Conclusions

- VQEG believes that some FR models perform well enough to be included in normative sections of Recommendations.
- The scope of these Recommendations should be written carefully to ensure that the use of the models is defined appropriately.
- If the scope of these Recommendations includes video system comparisons (e.g., comparing two codecs), then the Recommendation should include instructions indicating how to perform an accurate comparison.
- None of the evaluated models reached the accuracy of the normative subjective testing.
- All of the FR models performed statistically better than PSNR.
- The secondary analysis requires averaging over a well defined set of sequences while the tested system including all processing steps for the video sequences must remain exactly the same for all clips. Averaging over arbitrary sequences will lead to much worse results.

It should be noted that in case of new coding and transmission technologies, which were not included in this evaluation, the objective models can produce erroneous results. Here a subjective evaluation is required.

RR MODEL PERFORMANCE

RR models were submitted by Yonsei for the following resolutions and bit-rates: VGA at 128 kbits/s, 64 kbits/s and 10 kbits/s; CIF at 64 kbits/s and 10 kbits/s; and QCIF at 10 kbits/s and 1 kbits/s. When comparing these RR models to PSNR, it must be noted that PSNR is an FR model (i.e., PSNR needs full access to the source video).

Primary Analysis of RR Models

The average correlations of the primary analysis for the RR VGA models were all 0.80, and PSNR was 0.71. Individual model correlations for some experiments were as high as 0.93. The average RMSE for the RR VGA models were all 0.60, and PSNR was 0.71. The average outlier ratio for the RR VGA models ranged from 0.55 to 0.56, and PSNR was 0.62. All proposed models performed statistically better than PSNR for 7 of the 13 experiments. Based on each metric, each

RR VGA model was in the group of top performing models the following number of times:

VGA

Statistic	Yon_RR10k	YonRR64k	YonRR128k	PSNR
Correlation	13	13	13	7
RMSE	13	13	13	6
Outlier Ratio	13	13	13	10

The average correlations of the primary analysis for the RR CIF models were 0.78, and PSNR was 0.66. Individual model correlations for some experiments were as high as 0.90. The average RMSE for the RR CIF models were all 0.59, and PSNR was 0.72. The average outlier ratio for the RR CIF models were 0.51 and 0.52, and PSNR was 0.63. All proposed models performed statistically better than PSNR for 10 of the 14 experiments. Based on each metric, each RR CIF model was in the group of top performing models the following number of times:

CIF

Statistic	Yon_RR10k	YonRR64k	PSNR
Correlation	14	14	5
RMSE	14	14	4
Outlier Ratio	14	14	5

The average correlations of the primary analysis for the RR QCIF models were 0.77 and 0.79, and PSNR was 0.66. Individual model correlations for some experiments were as high as 0.89. The average RMSE for the RR QCIF models were 0.58 and 0.60, and PSNR was 0.72. The average outlier ratio for the RR QCIF models were 0.49 and 0.51, and PSNR was 0.60. All proposed models performed statistically better than PSNR for at least 9 of the 14 experiments. Based on each metric, each RR QCIF model was in the group of top performing models the following number of times:

QCIF

Statistic	Yon_RR1k	YonRR10k	PSNR
Correlation	14	14	5
RMSE	14	14	4
Outlier Ratio	12	13	4

Secondary Analysis of RR Models

The secondary analysis shows in principle a similar picture. The VGA RR models all tend to perform similarly. The CIF RR models all tend to perform similarly. For QCIF, Yonsei's 10k RR model slightly outperforms Yonsei's 1k RR model. The average correlation coefficients increase to 0.87 for VGA, 0.85 for CIF, and 0.91 for Yonsei's 10k model.

RR Model Conclusions

- VQEG believes that some of the RR models may be considered for standardization making sure that the scopes of these Recommendations are written carefully to ensure that the use of the models is defined appropriately.
- If the scope of these Recommendations includes video system comparisons (e.g., comparing two codecs), then the Recommendation should include instructions indicating

how to perform an accurate comparison.

- None of the evaluated models reached the accuracy of the normative subjective testing.
- All of the RR models performed statistically better than PSNR. It must be noted that PSNR is a FR model requiring full access to the source video.
- The secondary analysis requires averaging over a well defined set of sequences while the tested system including all processing steps for the video sequences must remain exactly the same for all clips. Averaging over arbitrary sequences will lead to much worse results.

It should be noted that in case of new coding and transmission technologies, which were not included in this evaluation, the objective models can produce erroneous results. Here a subjective evaluation is required.

NR MODEL PERFORMANCE

NR models were submitted by Psytechnics and Swissqual for all resolutions (VGA, CIF and QCIF). When comparing these NR models to PSNR, it must be noted that PSNR is an FR model (i.e., PSNR needs full access to the source video).

Primary Analysis of NR Models

The average correlations of the primary analysis for the NR VGA models were 0.44 and 0.57, and PSNR was 0.79. The average RMSE for the NR VGA models were 0.87 and 0.97, and PSNR was 0.65. The average outlier ratio for the NR VGA models were 0.78 and 0.80, and PSNR was 0.62. None of the proposed models performed better than PSNR. Based on each metric, each NR VGA model was in the group of top performing models the following number of times:

VGA

Statistic	Psy_NR	Swi_NR	PSNR
Correlation	1	1	13
RMSE	1	0	13
Outlier Ratio	13	12	*

* Note: statistical significance testing for NR models using Outlier Ratio did not include PSNR.

The average correlations of the primary analysis for the NR CIF models were 0.58 and 0.55, and PSNR was 0.76. The average RMSE for the NR CIF models were 0.82 and 0.85, and PSNR was 0.66. The average outlier ratio for the NR CIF models were 0.73 and 0.74, and PSNR was 0.65. None of the proposed models performed better than PSNR. Based on each metric, each NR CIF model was in the group of top performing models the following number of times:

CIF

Statistic	Psy_NR	Swi_NR	PSNR
Correlation	4	3	14
RMSE	3	3	14
Outlier Ratio	4	3	14

The average correlations of the primary analysis for the NR QCIF models were 0.70 and 0.64, and PSNR was 0.75. The average RMSE for the NR QCIF models were 0.74 and 0.80, and PSNR was 0.69. The average outlier ratio for the NR QCIF models were 0.68 and 0.71, and PSNR was 0.63. Each of the proposed models performed better than PSNR for at most 1 of the 14

experiments. Based on each metric, each NR QCIF model was in the group of top performing models the following number of times:

QCIF

Statistic	Psy_NR	Swi_NR	PSNR
Correlation	10	5	13
RMSE	10	5	13
Outlier Ratio	14	12	*

* Note: statistical significance testing for NR models using Outlier Ratio did not include PSNR.

Secondary Analysis of NR Models

In general, NR models show a content dependency. NR models use visual pattern matching to identify distortions caused by compressing and transmission. The problem is that the source video content (undistorted) occasionally looks like a compression or transmission artifact to the NR model. The secondary analysis addresses this issue by averaging over video clips with different contents. This decreases the content dependency of the NR models.

The secondary analysis shows improved performance for the NR models. The average correlations of the secondary analysis for the NR VGA models were 0.70 for Psytechnics' model, 0.79 for SwissQual's model, and 0.80 for PSNR. The average correlations of the secondary analysis for the NR CIF models were 0.82 for Psytechnics' model, 0.80 for SwissQual's model, and 0.74 for PSNR. The average correlations of the secondary analysis for the NR QCIF models were 0.91 for Psytechnics' model, 0.86 for SwissQual's model, and 0.81 for PSNR.

NR Model Conclusions

- The VGA and CIF NR models did not perform well enough to be considered in normative portions of Recommendations.
- VQEG believes that the QCIF NR models may be considered for standardization making sure that the scopes of these Recommendations are written carefully to ensure that the use of the models is defined appropriately.
- The scope of these Recommendations should be limited to quality monitoring. Use of QCIF NR models for video system comparisons is not recommended.
- The VGA and CIF NR models performed worse than PSNR.
- The QCIF NR models occasionally performed better than PSNR, and occasionally performed worse than PSNR. It must be noted that PSNR is a FR model requiring full access to the source video and precise video registration / calibration. Note that statistics for NR models include the source video, which is a particularly easy quality assessment case for PSNR.
- The secondary analysis requires averaging over a well defined set of sequences while the tested system including all processing steps for the video sequences must remain exactly the same for all clips. Averaging over arbitrary sequences will lead to much worse results.

It should be noted that in case of new coding and transmission technologies, which were not included in this evaluation, the objective models can produce erroneous results. Here a subjective evaluation is required.

FURTHER INFORMATION

See Section 1 of this report for an overview of the MM testing procedure. See Section 9 and Appendicies I, III, and VI for detailed model performance results and plots. See Section 5 and Appendices IV, and V for details of the subjective experiment.

FINAL REPORT FROM THE VIDEO QUALITY EXPERTS GROUP ON THE VALIDATION OF OBJECTIVE MODELS OF MULTIMEDIA QUALITY ASSESSMENT, PHASE I

1 INTRODUCTION

The main purpose of the Video Quality Experts Group (VQEG) is to provide input to the relevant standardization bodies responsible for producing international Recommendations regarding the definition of an objective Video Quality Metric in the digital domain. To this end, VQEG initiated a program of work to validate objective quality models that may be applied to measure the perceptual quality of Multimedia (MM) services.

Multimedia in this context is defined as being of or relating to an application that can combine text, graphics, full-motion video, and sound into an integrated package that is digitally transmitted over a communications channel. Common applications of multimedia that are appropriate to this study include video teleconferencing, video on demand and Internet streaming media. The measurement tools evaluated by the MM group may be used to measure quality both in laboratory conditions using a FR method and in operational conditions using RRNR methods.

In this multimedia test, MM Phase I, video only test conditions were employed. Subsequent tests will involve audio-video test sequences. The performance of objective models is based on the comparison of the MOS obtained from controlled subjective tests and the MOSp predicted by the submitted models. The goal of the testing was to examine the performance of proposed video quality metrics across representative coding, transmission and decoding conditions. To this end, the tests were designed to enable assessment of models for mobile/PDA and broadband internet communications services. Any Recommendation(s) resulting from the VQEG MM testing will be deemed appropriate for services delivered at 4 Mbit/s or less presented on mobile/PDA and computer desktop monitors.

This Multimedia (MM) Phase I addresses three video resolutions: VGA, CIF, and QCIF. Forty-one subjective experiments provided data for model validation. Subjective experiments were performed using the Absolute Category Rating with Hidden Reference Removal (ACR-HR) methodology. The results of the experiments are given in terms of Differential Mean Opinion Score (DMOS) – a quantitative measure of the subjective quality of a video sequence as judged by a panel of human observers. The following organizations performed subjective testing (i.e., created HRCs or ran viewers): Acreo, CRC, France Telecom, FUB, IRCCyN, KDDI, Nortel, NTT, OPTICOM, Psytechnics, SwissQual, Symmetricom, Verizon, NTIA, and Yonsei University. The following organizations formed an independent lab group that supervised the MM experiments: Acreo, CRC, Ericson, Intel, France Telecom, FUB, IRCCyN, Nortel, NTIA, and Verizon.

The subjective experiments included a wide variety of source video sequences. Source video sequences from interlaced content were carefully de-interlaced. Proponents and ILG visually inspected all source video sequences, and only source video sequences judged to have “good” to “excellent” quality were retained. Some source video was donated by proponents and known to all proponents prior to model submission, while other source video was provided by the ILG and unknown to proponents. Where possible, the source video sequences in each experiment represented at least 6 of the following content types: home video, video conferencing, sports, advertisement, animation, music video, movies, and broadcast news. See section 6 for more

information on source video and scene selection.

A wide variety of compression, transmission errors, and live network conditions were examined. The VGA experiments included bit-rates from 128 kbits/s to 4 Mbits/s; CIF experiments included bit-rates from 64 kbits/s to 704 kbits/s; and QCIF experiments included bit-rates from 16 kbits/s to 320 kbits/s. All experiments included some video sequences containing only coding/decoding impairments. Most experiments also included some video sequences exhibiting simulated transmission errors and/or transmission errors from live networks. Ignoring anomalous events (e.g., transmission errors), each frame of each processed video sequences was limited to +/- 0.25 seconds temporal misalignment from the source video sequence. Most experiments focused on Windows Media 9 (VC-1), H.264, and Real Video. Other codecs examined include H.261, H.263, MPEG4, MPEG2, Cinepak, DivX, Sorenson3, and Theora. Pausing events were limited to 2 seconds duration, and systems exhibiting a steadily increasing delay were disallowed (e.g., a pause followed by resumed play with no loss of content). Only limited calibration problems were allowed, since ITU-T J.242 is separately addressing the issue of calibration. See section 6 for more information on degradations, and calibration limits.

All subjective experiments at a single resolution contained a common set of 30 video sequences. These common sequences spanned the range of quality desired, and served to provide consistency between experiments. The common set included secret sequences (i.e., video unknown to proponents), secret HRCs (i.e., systems unknown to proponents), and a wide range of content types. Each common set contained both 25 fps and 30 fps video.

Each of the 41 experiments examined either 25 fps video or 30 fps video. Due to a relative scarcity of 25 fps source video sequences and laboratories able to create 25 fps test conditions, approximately one-third (33%) of the experiments at each resolution contained 25 fps video, and approximately two-thirds (67%) of the experiments at each resolution contained 30 fps video.

Prior to subjective testing, proponents submitted objective models. The video sequences in each experiment were selected in secret by the ILG and vetted by proponents for any problems after model submission (e.g., quality below that specified in the MM Test Plan). Each proponent performed at least one subjective experiment, the design of which was made available to the ILG and other proponents prior to model submission. Each proponent created all HRCs for their own experiment, but did not also run the subjective test for their experiment. Labs swapped subjective tests, so they ran viewers through an experiment designed and created by another laboratory.

Proponents were able to submit for evaluation Full Reference (FR), Reduced Reference (RR), and No Reference (NR) models. The side-channels allowable for the RR models were:

- PDA/Mobile (QCIF): (1kbit/s, 10kbit/s)
- PC1 (CIF): (10kbit/s, 64kbit/s)
- PC2 (VGA): (10kbit/s, 64kbit/s, 128kbit/s)

Proponents could submit one model of each type for all image size conditions. Thus, any single proponent may have submitted up to a total of 13 different models (one FR model for QCIF, one FR model for CIF, one FR model for VGA; one NR model for QCIF, one NR model for CIF, one NR model for VGA; two RR models for QCIF, two RR models for CIF, three RR models for VGA).

FR and RR models were not required to predict the perceptual quality of the source (reference) video files used in subjective tests. NR models were required to predict the perceptual quality of both the source and processed video files used in subjective quality tests.

31 models were submitted, 6 were withdrawn, and 25 are reported on in this report. This report analyzes the following models:

Proponent	Video Resolution	Model Bit-Rate
NTT (Japan)	VGA & CIF & QCIF	FR
OPTICOM (Germany)	VGA & CIF & QCIF	FR
Psytechnics (UK)	VGA & CIF & QCIF	FR & NR
SwissQual (Switzerland)	VGA & CIF & QCIF	NR
Yonsei University (Korea)	VGA	FR RR128k (128 kbits/s) RR64k (64 kbits/s) RR10k (10kbits/s)
Yonsei University (Korea)	CIF	FR RR64k (64 kbits/s) RR10k (10 kbits/s
Yonsei University (Korea)	QCIF	FR RR10k RR1k

The intention of VQEG is that the MM Phase I data may not be used as evidence to standardize any objective video quality model which was not been tested within this phase. This comparison would not be fair, because another model could have been trained on the MM Phase I data.

PSNR results are presented for comparison purposes, only. Due to confidentiality agreements and usage limitations, most of the source video sequences and all of the processed video sequences cannot be redistributed.

This final report details the test method used in the subjective quality tests, selection of test material and conditions, and the evaluation metrics that were subsequently submitted for validation by the VQEG.

This report contains the following sections and Appendices:

Section 1: Summarizes the MM Test Phase I test.

Section 2: Definitions used in VQEG's Multimedia Test plan and this report.

Section 3: Acronyms used in VQEG's Multimedia Test Plan and this report.

Section 4: Identity of each test laboratory.

Section 5: Design overview: subjective testing methodology (ACR-HR), display specifications, test sessions, video PC-based playback mechanism, subjects, and viewing conditions.

Section 6: Limitations on source video sequences, HRCs, and processed video calibration.
Section 7: Objective quality model evaluation criteria.
Section 8: Common set analysis and interpretation.
Section 9: Official ILG data analysis.
Section 10: Secondary Data Analysis
Section 11: Conclusions.
Appendix I: Model descriptions.
Appendix II: Greater detail on each subjective testing facility.
Appendix III: Details on source scene selection and scene pools for each experiment.
Appendix IV: Details on HRC selection for each experiment.
Appendix V: Plots.
Appendix VI: Proponent Comments

2 LIST OF DEFINITIONS

Anomalous frame repetition is defined as an event where the HRC outputs a single frame repeatedly in response to an unusual or out of the ordinary event. Anomalous frame repetition includes but is not limited to the following types of events: an error in the transmission channel, a change in the delay through the transmission channel, limited computer resources impacting the decoder's performance, and limited computer resources impacting the display of the video signal.

Constant frame skipping is defined as an event where the HRC outputs frames with updated content at an effective frame rate that is fixed and less than the source frame rate.

Effective frame rate is defined as the number of unique frames (i.e., total frames – repeated frames) per second.

Frame rate is the number of (progressive) frames displayed per second (fps).

Handover :In cellular mobile systems, the process of transferring a phone call in progress from one cell transmitter and receiver and frequency pair to another cell transmitter and receiver using a different frequency pair without interruption of the call.

Intended frame rate (formerly absolute frame rate) is defined as the number of video frames per second physically stored for some representation of a video sequence. The intended frame rate may be constant or may change with time. Two examples of *constant* intended frame rates are a BetacamSP tape containing 25 fps and a VQEG FR-TV Phase I compliant 625-line YUV file containing 25 fps; these both have an intended frame rate of 25 fps. One example of a *variable* intended frame rate is a computer file containing only new frames; in this case the intended frame rate exactly matches the effective frame rate. The content of video frames is not considered when determining intended frame rate.

Live Network Conditions are defined as errors imposed upon the digital video bit stream as a result of live network conditions. Examples of error sources include packet loss due to heavy network traffic, increased delay due to transmission route changes, multi-path on a broadcast signal, and fingerprints on a DVD. Live network conditions tend to be unpredictable and unrepeatable.

Pausing with skipping (formerly frame skipping) is defined as events where the video pauses for some period of time and then restarts with some loss of video information. In pausing with skipping, the temporal delay through the system will vary about an average system delay, sometimes increasing and sometimes decreasing. One example of pausing with skipping is a pair of IP Videophones, where heavy network traffic causes the IP Videophone display to freeze briefly; when the IP Videophone display continues, some content has been lost. Another example is a videoconferencing system that performs constant frame skipping or variable frame skipping. Constant frame skipping and variable frame skipping are subsets of pausing with skipping. A processed video sequence containing pausing with skipping will be approximately the same duration as the associated original video sequence.

Pausing without skipping (formerly frame freeze) is defined as any event where the video pauses for some period of time and then restarts without losing any video information. Hence, the temporal delay through the system must increase. One example of pausing without skipping is a computer simultaneously downloading and playing an AVI file, where heavy network traffic causes the player to pause briefly and then continue playing. A processed video sequence containing pausing without skipping events will always be longer in duration than the associated original video sequence.

Refresh rate is defined as the rate at which the computer monitor is updated.

Simulated transmission errors are defined as errors imposed upon the digital video bit stream in a highly controlled environment. Examples include simulated packet loss rates and simulated bit errors. Parameters used to control simulated transmission errors are well defined.

Source frame rate (SFR) is the intended frame rate of the original source video sequences. The source frame rate is constant. For the MM test plan the SFR may be either 25 fps or 30 fps.

Transmission errors are defined as any error imposed on the video transmission. Example types of errors include simulated transmission errors and live network conditions.

Variable frame skipping is defined as an event where the HRC outputs frames with updated content at an effective frame rate that changes with time. The temporal delay through the system will increase and decrease with time, varying about an average system delay. A processed video sequence containing variable frame skipping will be approximately the same duration as the associated original video sequence.

3 LIST OF ACRONYMS

ACR	Absolute Category Rating
ACR-HR	Absolute Category Rating with Hidden Reference
ANOVA	ANalysis Of VAriance
ASCII	ANSI Standard Code for Information Interchange
AVI	Audio Video Interleave
BER	Bit error rates
BLER	Block error rates
CI	Confidence Interval
CIF	Common Intermediate Format (352 x 288 pixels)
CODEC	COder-DECoder
CRC	Communications Research Centre (Canada)
DVB-C	Digital Video Broadcasting-Cable
DMOS	Difference Mean Opinion Score
DMOSh	DMOS of the HRC (averaging over sources)
DMOSs	DMOS of the Source (averaging over HRCs)
DVD	Digital Versatile Disc
FR	Full Reference
GOP	Group Of Pictures
HRC	Hypothetical Reference Circuit
ILG	Independent Laboratory Group
IP	Internet Protocol
ITU	International Telecommunication Union
KDDI	Combined company formed from KDD and IDO Corporation
LCD	Liquid Crystal Display
LSB	Least Significant Bit
MM	MultiMedia
MOS	Mean Opinion Score
MOSp	Mean Opinion Score, predicted
MoSQuE	NTT's model name
MPEG	Moving Picture Experts Group

NR	No (or Zero) Reference
NTSC	National Television Standard Code (60 Hz TV)
NTT	Nippon Telegraph and Telephone
PAL	Phase Alternating Line standard (50 Hz TV)
PDA	Personal Digital Assistant
PS	Program Segment
PSNR	Peak Signal to Noise Ratio
PVS	Processed Video Sequence
QCIF	Quarter Common Intermediate Format (176 x 144 pixels)
RMSE	Root Mean Square Error
RR	Reduced Reference
RRNR	Reduced Reference / No Reference
SFR	Source Frame Rate
SMPTE	Society of Motion Picture and Television Engineers
SRC	Source Reference Channel or Circuit
TCO	Swedish acronym for "Swedish Confederation of Professional Employees". They own the company that administers the TCO Requirements for computer displays (www.tcodevelopment.com)
VGA	Video Graphics Array (640 x 480 pixels)
VQEG	Video Quality Experts Group
VQR	Video Quality Rating (as predicted by an objective model)
VTR	Video Tape Recorder
YUV	Color Space and file format

4 TEST LABORATORIES

Given the scope of the MM testing, both independent test laboratories and proponent laboratories were assigned subjective test responsibilities. A brief listing of the contributing laboratories follows. See also Appendix II.

4.1 Independent Laboratory Group (ILG)

Acreo, Sweden, http://www.acreo.se/

CRC, Communications Research Centre, Canada http://www.crc.ca/

Ericsson, Sweden, http://www.ericsson.com

FUB, Italy

Intel, USA, http://www.intel.com/

IRCCyN, University of Nantes, France, http://www2.irccyn.ec-nantes.fr/ivcdb/

Nortel, Canada, www.nortel.com

NTIA/ITS, U.S. Department of Commerce, USA, http://www.its.bldrdoc.gov/n3/video/index.php

Orange France Telecom, France, http://www.francetelecom.com

Verizon, USA, http://www.verizon.com

4.2 Proponent Laboratories

NTT, Japan, http://www.ntt.com

OPTICOM, Germany, http://www.pevq.org/

Psytechnics, UK, http://www.psytechnics.com

SwissQual, Switzerland, http://www.swissqual.com/

Yonsei University, Republic of Korea, http://www.yonsei.ac.kr/eng/

4.3 Other Laboratories

Symmetricom, USA

KDDI, Japan, http://www.kddi.com/english/index.html

5 DESIGN OVERVIEW: SUBJECTIVE EVALUATION PROCEDURE

This section provide an overview of the test method applied in the Multimedia Phase I tests to perform subjective testing and for model validation. For full details of the test procedure used in the Multimedia Phase I work, the interested reader is referred to the official test plan, available from http://www.its.bldrdoc.gov/vqeg/projects/multimedia/index.php.

5.1 Subjective Test Method: ACR Method with Hidden Reference

This section describes the test method according to which the VQEG multimedia (MM) subjective tests were performed. Tests used the absolute category rating scale (ACR) [ITU-T Rec. P.910] for collecting subjective judgments of video samples. ACR is a single-stimulus method in which a processed video segment is presented alone, without being paired with its unprocessed ("reference") version. The present test procedure includes a reference version of each video segment, not as part of a pair, but as a freestanding stimulus for rating like any other. During the data analysis the ACR scores were subtracted from the corresponding reference scores to obtain a DMOS. This procedure is known as "hidden reference" (henceforth referred to as ACR-HR). This choice was made due to the fact that ACR provides a reliable and standardized method that allows a large number of test conditions to be assessed in any single test session.

In the ACR test method, each test condition is presented singly for subjective assessment. The test presentation order is randomized via random number generator (with some restrictions as described in Section 5.4). The test format is shown in Figure 1. At the end of each test presentation, human judges ("subjects") provide a quality rating using the ACR rating scale shown in Figure 2. Note that the numerical values attached to each category are only used for data analysis and are not shown to subjects (see Figure 3).

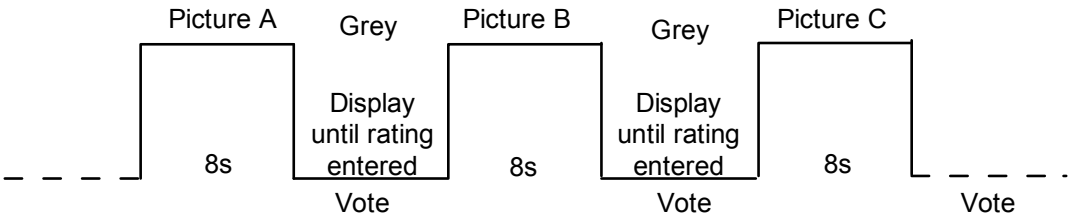


Figure 1 – ACR basic test cell.

5 Excellent

4 Good

3 Fair

2 Poor

1 Bad

Figure 2 – The ACR rating scale.

The length of the SRC and PVS were exactly 8 s.

Instructions to the subjects provide a more detailed description of the ACR procedure.

5.2 Viewing distance

The test instructions request subjects to maintain a specified viewing distance from the display device. The viewing distances were:

- QCIF: nominally 6-10 picture heights (H), and let the viewer choose within physical limits (natural for PDAs).
- CIF: 6-8H and let the viewer choose within physical limits.
- VGA: 4-6H and let the viewer choose within physical limits.

H=Picture Heights (picture is defined as the size of the video window).

5.3 Display Specification and Set-up

LCD displays were used in the test and the test laboratories were requested to use displays meeting the specifications below and to use a common set-up technique which is also specified below.

This MM test used LCD displays meeting the following specifications:

Monitor Feature	Specification
Diagonal Size	17-24 inches
Dot pitch	< 0.30
Resolution	Native resolution (no scaling allowed)
Gray to Gray Response Time (if specified by manufacturer, otherwise assume response time reported is white-black)	< 30 ms (<10 ms if based on white-black)
Color Temperature	6500K
Calibration	Yes
Calibration Method	Eye One / Video Essentials DVD
Bit Depth	8 bits/color
Refresh Rate	>= 60 Hz
Standalone/laptop	Standalone
Label	TCO ’03 or TCO ‘06 (TCO ’06 preferred)

The LCD was set-up using the following procedure:

- Use the autosetting to set the default values for luminance, contrast and colour shade of white.

- Adjust the brightness according to Rec. ITU-T P.910, but do not adjust the contrast (it might change balance of the color temperature).
- Set the gamma to 2.2.
- Set the color temperature to 6500 K (default value on most LCDs).
- The scan rate of the PC monitor must be at least 60 Hz.

Video sequences were displayed using a black border frame (grey value: 0) on a grey background (grey value: 128). The black border frame was of the following size:

- 36 lines/pixels VGA
- 18 lines/pixels CIF
- 9 lines/pixels QCIF

The black border frame was on all four sides of the video window.

5.4 Subjective Test Control Software

PCs were used to store and play the video content, using special purpose software, developed by Acreo (AcrVQWin version 1.0). This software was used by all test laboratories. The playback of a video clip was performed by pre-loading the clips in the memory of the PC's graphics card. This was done to ensure that no frame drops occurred and that the update of each played frame happened in synchronization with the display update. The tests included a mixture of 25 frames per second (fps) and 30 fps. The subjective results were stored directly on the same PCs that were used to present the video.

The most common LCD computer monitors have 60 Hertz (Hz) as their update frequency. The test plan, therefore, specified the monitor to be set to 60 Hz. Each frame was shown during two update frequency periods to obtain a frame rate of 30 fps. 25 fps was obtained using a modified 2-3 pulldown sequence. For example, each set of five frames was displayed according to the following number of screen updates: 2, 3, 2, 3 and 2.

To minimize waiting for the subjects, the next PVS video sequence was loaded during voting time using multi-threading programming techniques. The ACR rating scales were presented on the LCD after each video clip, using a dialog box as shown in Figure 3. A setup file was used to change the language of the text in the dialog box to that used by the testing laboratories in the different countries. Subjects provided their vote responses using the mouse of the PC. In each subjective test, the presentation order of test sequences was fully randomized between subjects with the exception that two PVSs originating from the same SRC were not allowed to be played next to each other, as specified in the test plan. After the vote was given and the OK button was pressed, the next PVS was automatically played. The software indicated when half of the PVSs had been rated, allowing the subjects to take a break.

Figure 3: The voting dialog in the subjective test software

The subjective test software (AcrVQWin) was controlled using a setup file, which the operator selected at startup. The setup file specified the particular PVSs and other startup parameters. Before the actual test, a practice session was performed to familiarize the viewer with the test procedure and the range of qualities used in the test. [1]

5.5 Subjects

Subjective experiments were distributed among several test laboratories. Some of the tests were performed by the ILG and some by the proponents. Between 1 and 3 tests were done by any given laboratory at one image resolution.

Exactly 24 valid viewers per experiment were used for data analysis. Only scores from valid viewers are reported in the results and used to validate objective models. A valid viewer means that after post-experiment results screening, their rating was accepted. Post-experiment results screening is used to discard data of viewers who may have voted randomly. The rejection criteria verify the level of consistency of the scores of one viewer according to the mean score of all observers over one individual experiment. The method for post-experiment results screening is described in Annex VI of the test plan (http://www.its.bldrdoc.gov/vqeg/projects/multimedia/index.php).

The following procedure was used to obtain ratings for 24 valid observers:

1. Conduct the experiment with 24 viewers.
2. Apply post-experiment screening to eventually discard viewers who may have voted randomly.
3. If n viewers were rejected, run n additional subjects.
4. Go back to step 2 and step 3 until valid results for 24 viewers are obtained.

Each individual subject could participate in one experiment only (i.e., one experiment at one image resolution). Only non-expert viewers participated in the subjective tests. The term non-expert is used in the sense that the viewers' work does not involve video picture quality and they are not experienced assessors. Subjects must not have had participated in a subjective video quality test over a period of the previous six months.

It was expected that prior to a test session, observers would be screened for normal visual acuity

or corrected-to-normal acuity and for normal color vision according to the method specified in ITU-T P.910 or ITU-R Rec. 500.

5.6 Viewing Conditions

Each test session involved only one subject per display assessing the test material. Subjects were seated directly in line with the center of the video display at a specified viewing distance (see Section 5.2). A requirement was that the test cabinet conformed to ITU-T Rec. P.910.

5.7 Experiment design

The length of the experiment was designed to be within 1 hour, including practice clips and a comfortable break. Each subjective experiment included 166 PVSs. They included both the common set of 30 PVSs inserted in each experiment and the hidden reference (hidden SRCs) sequences; i.e., each hidden SRC is one PVS. The common set of PVSs included "secret" PVSs and "secret" SRCs.

Randomization was applied across the 166 PVSs. The 166 PVSs were split into 2 sessions of 83 PVSs each. In this scenario, an experiment included the following steps:

1. Introduction and instructions to viewer.
2. Practice clips: these test clips allow the viewer to familiarize with the assessment procedure and software. They represented the range of distortions found in the experiment. The number of practice clips was 6. Each of the practice clips came from a different test. Ratings given to practice clips were not used for data analysis.
3. Assessment of 83 PVSs.
4. Short break.
5. Practice clips (this step was optional but advised to regain viewer's concentration after the break).
6. Assessment of 83 PVSs.

Each SRC was processed through each HRC. The test design was a full matrix of 8 by 17 SRC by HRC combinations. In addition to this the ILG created a common set of 30 PVSs (6 SRCs and 5 HRCs, one of which was the hidden reference).

The SRCs used in each experiment covered a variety of content categories and at least 6 categories of content were included in each experiment.

5.8 Randomization

For each subjective test, a randomization process was used to generate orders of presentation (playlists) of video sequences. See description of AcrVQWin above.

5.9 Data Collection

5.9.1 Results Data Format

The following format was designed to facilitate data analysis of the subjective data results file.

The subjective data for each test was stored in a Microsoft Excel spreadsheet containing the following columns in the following order: lab name, test identifier, test type, subject number, month, day, year, session, resolution, frame rate, age, gender, random order identifier, scene

identifier, HRC, ACR Score. Missing data values are indicated by the value -9999 to facilitate global search and replacement of missing values. Only data from valid viewers (i.e., viewers who passed the visual acuity and color tests, and whose data passed the consistency test) were used to create the final results spreadsheet.

5.9.2 Subjective Data Analysis

Difference scores were calculated for each processed video sequence (PVS). A PVS is defined as a SRCxHRC combination. The difference scores, known as Difference Mean Opinion Scores (DMOS), were produced for each PVS by subtracting the PVS's score from that of the corresponding hidden reference score for the SRC that had been used to produce the PVS. Subtraction was performed on a per subject basis. Difference scores were used to assess the performance of each full reference and reduced reference proponent model, applying the metrics defined in Section 7.4.

For evaluation of no-reference proponent models, the absolute (raw) subjective mean opinion score (MOS) was used. These MOS values were then used to evaluate the performance of NR models using the metrics specified in Section 8.4.

6 LIMITATIONS ON SOURCE SCENES, HRCS & CALIBRATION

Separate subjective tests were performed for different video sizes. One set of tests presented video in QCIF (176x 144 pixels). One set of tests presented CIF (352x288 pixels) video. One set of tests presented VGA (640x480). In the case of Rec. 601 video source, aspect ratio correction was performed on the video sequences prior to writing the AVI files (SRC) or processing the PVS.

Note that in all subjective tests 1 pixel of video was displayed as 1 pixel native display. No upsampling or downsampling of the video was allowed at the player.

6.1 Source Video Processing Overview

The test material was selected from a common pool of video sequences. Where the test sequences were in interlace format, then standard, agreed de-interlacing methods were applied to transform the video to progressive format. All source material was 25 or 30 frames per second progressive, and no more than one version of each source sequence for each resolution was allowed. Uncompressed AVI files were used for subjective and objective tests. The progressive test sequences used in the subjective tests were used by the models to produce objective scores.

All original SRC source sequences were 12 seconds duration (300 frames for 625-line source; 360 frames for 525-line source) for processing through each HRC. After each original 12s SRC was processed by the relevant HRC, the 12s output was then edited to produce an 8s PVS. For the original SRC, this was achieved by removing the first 2s and final 2s. For a PVS, the 8s edit was achieved by removing the first (2 + N) seconds and final (2 – N) seconds, where N is the temporal registration shift needed to meet the temporal registration limits. Only the middle 8s sequence was stored for use in subjective testing and for processing by objective models.

The source video sequences used for each experiment (named “scene pools”) were chosen in secret by the ILG.

6.2 Source Video Selection Criteria

Completely still video scenes were not used in any test. One scene in each common set contained still portions. See Appendix III for further details on scene selection.

In compliance with the MM test plan, scene pools were chosen to contain content from at least 6 of the 8 categories. Due to a shortage of 25 fps SRC content, some 25 fps scene pools had content from only 5 categories. This discrepancy was approved by proponents. More 30 fps SRC content was available than 25 fps SRC content, and in addition more laboratories could create 30 fps HRCs than 25 fps HRCs. Therefore, more 30 fps scene pools were created than 25 fps scene pools. In order to create robust, well rounded scene pools, the ILG identified further criteria to guide selection of SRCs for each scene pools. These criteria were as follows:

1. One scene that is very difficult to code.
2. One scene that is very easy to code.
3. One scene that contains high spatial detail
4. One scene that contains high motion and/or rapid scene cuts (e.g., object moves 20+ pixels at VGA resolution).
5. SRCs fairly evenly span the range of complexity: some low; some medium; and some high.

6. One scene with multiple objects moving in a random, unpredictable manner (e.g., CBCLePoint)
7. Some SRCs with high quality and high complexity; some SRCs with high quality but low complexity or medium quality with high complexity; and some SRCs with moderate quality and complexity.
8. One very colorful scene.
9. One scene that might challenge the model: fine detail that may be blurred by the codec in a manner that will not be perceived by viewers, a large black/white edge, a blurred background with the foreground in focus, a night scene, or a poorly lit scene.
10. One scene that might challenge the codec: SRC containing water or smoke or fire that moves in an unpredictable shifting manner, SRC that jiggles or bounces significantly as from a hand-held camera, flashing lights or other very fast events, or a graduated change in color or hue as from a sunset.
11. One scene that shows a close-up of a person's face or a person showing an obvious emotional response; this scene contains skin tones.
12. At least one scene with scene cuts and at least four scenes without scene cuts.
13. One scene that has some animation overlay or cartoon content.
14. If possible, a scene where most of the action is in a small portion of the total picture (e.g., NTIAfishmug1).
15. One scene with low contrast (e.g., soft edges like NTIAbells4); and one scene with high contrast (e.g., hard edges like SMPTEbirches1).
16. One scene with low brightness (e.g., NTIAbells4); and one scene with high brightness (e.g., NTIAoverview1).
17. If possible, at least one secret SRC.
18. No more than half of the SRCs were taken from any one source (e.g., ITU standard test sequences).
19. If possible, exactly one night scene or poorly lit scene.

Where possible, all scene pools conformed to the above 19 criteria. Where possible each SRC was used in only one scene pool at a given image resolution (VGA, CIF, QCIF). This was done to maximize the variety of source content in all tests. Occasionally, a SRC appeared in both a scene pool and the common set scene pool.

The following criteria were identified for selection of the common sets:

1. Both 25 fps and 30 fps represented.
2. Quality high enough that there is only a small chance that any SRC any will receive an MOS score less than 4.0.
3. One scene contains animation, because most test sets won't.
4. Includes other content types that are rare or represented in only a few scene pools. This was done to increase the number of content types in 25 fps experiments.
5. At least one secret scene.
6. A minimum of proponent material.
7. One scene that is very difficult to code.

8. One scene that is very easy to code.
9. SRCs span fairly evenly the range of complexity: some low, some medium, and some high.
10. One scene with multiple objects moving in a random, unpredictable manner (e.g., CBCLePoint)
11. One very colorful scene.
12. No scenes with unusual content that may challenge one model but not another and perhaps bias results.
13. One scene that may challenge the codec (see examples given for scene pool criteria, above).
14. One scene that shows a close-up of a person's face or an obvious emotional response, including skin tones.
15. At least one scene with scene cuts and at least one scene without scene cuts.
16. At least one secret SRC.
17. One SRC that contains a perfectly still portion, so that every experiment meets this constraint in the MM test plan.

The ILG sorted SRCs into the 8 categories identified in the MM test plan. SRCs that did not obviously fall into any category are listed in a 9th table. See Appendix III for these tables. The content source is identified, and each scene is briefly described. The right-most column of these tables identifies secret SRCs. A few of the SRCs listed were not used in any test.

Appendix III also identifies the video sequences used in each scene pool, the scene pool used in each test, and the frame rate of each test.

6.3 Hypothetical Reference Circuit (HRC) Limitations

The subjective tests were performed to investigate a range of HRC error conditions. The group agreed that these error conditions could include, but would not be limited to, the following:

- Compression errors (such as those introduced by varying bit-rate, codec type, frame rate and so on),
- Transmission errors,
- Post-processing effects,
- Live network conditions,
- Interlacing problems.

6.3.1 Video Bit-rates

The following bit rates were tested[1]:

_____

- PDA/Mobile (QCIF): 16 kbit/s to 320 kbit/s (e.g., 16, 32, 64, 128, 192, 320)
- PC1 (CIF): 64 kbit/s to 704 kbit/s (e.g., 64, 128, 192, 320, 448, 704)
- PC2 (VGA): 128kbit/s to 4Mbit/s (e.g., 128, 256, 320, 448, 704, ~1M, ~1.5M, ~2M, 3M,~4M)

6.3.2 Simulated Transmission Errors

A set of test conditions (HRC) included error profiles as follows:

- Packet-switched transport (e.g., 2G or 3G mobile video streaming, PC-based wireline video streaming),
- Circuit-switched transport (e.g., mobile video-telephony).

Packet-switched transmission

HRCs included packet loss with a range of packet loss ratios (PLR) representative of typical real-life scenarios. The PLR tested in the validation was from 0% to 12%.

In **mobile video streaming**, we considered the following scenarios:

1. Arrival of packets is delayed due to re-transmission over the air.
2. Arrival of packets is delayed, and the delay is too large: These packets are discarded by the video client.
3. Very bad radio conditions: Massive packet loss occurs.
4. Handovers: Packet loss can be caused by "handovers." Packets are lost in bursts and cause image artifacts.

In **PC-based wireline video streaming**, network congestion causes packet loss during IP transmission.

In order to cover different scenarios, we considered the following models of packet loss:

- Bursty packet loss. The packet loss pattern can be generated by a link simulator or by a bit or block error model, such as the Gilbert-Elliott model;
- Random packet loss;
- Periodic packet loss.

Choice of a specific PLR is not sufficient to characterize packet loss effects, as perceived quality will also be dependent on codecs, content, packet loss distribution (profiles) and which types of video frames were hit by the loss of packets. Different levels of loss ratio with different distribution profiles were selected in order to produce test material that spreads over a wide range of video quality. To confirm that test files do cover a wide range of quality, the generated test files (i.e., decoded video after simulation of transmission error) were:

1. Viewed by video experts to ensure that the visual degradations resulting from the simulated transmission error spread over a range of video quality over different content;
2. Checked to ensure that degradations remained within the limits stated by the test plan (e.g., in the case where packet loss caused loss of complete frames, it was verified that temporal misalignment remained within the limits stated by the test plan).

Circuit-switched transmission

HRCs included bit errors and/or block errors with a range of bit error rates (BER) or/and block error rates (BLER) representative of typical real-world scenarios. In circuit-switched transmission, e.g., video-telephony, no re-transmission is used. Bit or block errors occur in bursts.

In order to cover different scenarios, the following error levels were used:

Air interface block error rates: Normal uplink and downlink: 0.3%, normally not lower. High value uplink: 0.5%, high downlink: 1.0%. To make sure the models' algorithms will handle really bad conditions up to 2%-3% block errors on the downlink were used.

Bit stream errors: Block errors over the air cause bits to not be received correctly. Consequently, a video telephony (H.223) bit stream experiences cyclic redundancy check errors and chunks of the bit stream are lost.

6.3.3 Live Network Conditions

Simulated errors are an excellent means to test the behavior of a system under well defined conditions and to observe the effects of isolated distortions. In real live networks however usually a multitude of effects happen simultaneously when signals are transmitted, especially when radio interfaces are involved. Some effects, like handovers, can only be observed in live networks.

6.3.4 Pausing with Skipping and Pausing without Skipping

Anomalous frame repetition was not allowed during the first 1s or the final 1s of a video sequence. Other types of anomalous behavior are allowed provided they meet the following restrictions. The delay through the system before, after, and between anomalous behavior segments must vary around an average delay and must meet the temporal registration limits in section 6.4. The first 1s and final 1s of each video sequence cannot contain any anomalous behavior. At most 25% of any individual PVS's duration may exceed the temporal registration limits in section 6.4. These 25% must have at most a maximum temporal registration error of +3 seconds (added delay).

The detailed description of each test is provided in Appendix IV.

6.3.5 Frame Rates

For those codecs that only offer automatically set frame rate, this rate is decided by the codec. Some codecs have options to set the frame rate either automatically or manually. For those codecs that have options for manually setting the frame rate (and we choose to set it for the particular case), 5 fps will be considered the minimum frame rate for VGA and CIF, and 2.5 fps for PDA/Mobile.

Manually set frame rates (constant frame rate) included:

- QCIF: 2.5 – 30 fps
- CIF: 3 – 30 fps (C07, C08 and C09 have one HRC with 3 fps).
- VGA: 5 – 30 fps

Variable frame rates are acceptable for the HRCs. The first 1s and last 1s of each QCIF PVS was constrained to contain at least two unique frames, provided the source content was not still for those two seconds. The first 1s and last 1s of each CIF and VGA PVS contained at least four unique frames, provided the source content was not still for those two seconds.

Care was taken when creating the test sequences for display on a PC monitor because the refresh

rate can influence the reproduction quality of the video, and VQEG MM requires that the sampling rate and display output rate are compatible.

Given that a source frame rate of video is 30 fps, and the sampling rate is 30/X (e.g., 30/2 = sampling rate of 15fps), then 15 fps is called the frame rate. Then we upsample and repeat frames from the sampling rate of 15fps to obtain 30 fps for display output.

The intended frame rate of the source and the PVS were identical.

6.3.6 Pre-Processing

The HRC processing could include, typically prior to the encoding, one or more of the following:

- Filtering,
- Simulation of non-ideal cameras (e.g., mobile),
- Colour space conversion (e.g., from 4:2:2 to 4:2:0),
- Interlacing of previously deinterlaced source.

This processing was considered part of the HRC.

6.3.7 Post-Processing

The following post-processing effects could be used in the preparation of test material:

- Color space conversion
- De-blocking
- Decoder jitter
- Deinterlacing of codec output including when it has been interlaced prior to codec input.

6.3.8 Coding Schemes

Coding Schemes that could be used included, but were not limited to:

- Windows Media Video 9
- H.261
- H.263
- H.264 (MPEG-4 Part 10)
- Real Video (e.g., RV 10)
- MPEG1
- MPEG2
- MPEG4
- JPEG 2000 Part 3
- DiVX
- H.264/MPEG4 SVC
- Sorensen

- Cinepak
- VC1

6.3.9 A Note on Allowable Transmission Error Events

Pausing was allowed as a valid transmission error type. Other types of anomalous behavior were allowed provided they met the following restrictions. The delay through the system before, after, and between anomalous behavior segments was required to vary around an average delay and met the temporal registration limits. The first 1s and final 1s of each video sequence could not contain any anomalous behavior. At most 25% of any individual PVS's duration could exceed the temporal registration limits in section 7.4. These 25% must have at most a maximum temporal registration error of +3 seconds (added delay).

6.4 Processed Video Sequence Calibration: Limitations and Validation

6.4.1 Calibration Limitations

Measurements were only performed on the portions of PVSs that are not anomalously severely distorted (e.g., in the case of transmission errors or codec errors due to malfunction).

Models were required to include calibration and registration to handle the following technical criteria (Note: Deviation and shifts were defined as between a source sequence and its associated PVSs. Measurements of gain and offset were made on the first and last seconds of the sequences. If the first and last seconds were anomalously severely distorted, then another 2 second portion of the sequence was used.):

- maximum allowable deviation in *offset* is ±20
- maximum allowable deviation in *gain* is ±0.1
- maximum allowable *Horizontal Shift* is +/- 1 pixel
- maximum allowable *Vertical Shift* is +/- 1 pixel
- maximum allowable *Horizontal Cropping* is 12 pixels for VGA, 6 pixels for CIF, and 3 pixels for QCIF (for each side).
- maximum allowable *Vertical Cropping* is 12 pixels for VGA, 6 pixels for CIF, and 3 pixels for QCIF (for each side).
- no *Spatial Rotation or Vertical or Horizontal Re-scaling* is allowed
- no Spatial *Picture Jitter* is allowed. Spatial picture jitter is defined as a temporally varying horizontal and/or vertical shift.

Reduced Reference Models were required to include temporal registration if needed by the model. Temporal misalignment of no more than +/-0.25s was allowed, for 75% of clip duration. The rest of each clip could contain temporal misalignment up to +3s to -.25s (increased delay). This constraint was added due to concern about the subjective testing methodology and the visibility of impairments to viewers in these artificial settings (i.e. only seeing 8 second clips). The start frame of both the reference and its associated PVSs were matched as closely as possible.

6.4.2 Check of Calibration

Spatial offsets were rare. Spatial registration shifts ranged between +/- 1 pixel horizontally and

vertically. It was expected that no post-impairments were introduced to the outputs of the encoder before transmission. Calibration issues outside the allowable range were corrected prior to subjective testing, wherever possible or the PVS was replaced.

These calibration limits were checked by software provided by NTIA/ITS. The algorithm used is available in ITU-T Recommendation J.244, "Calibration methods for constant misalignment of spatial and temporal domains with constant gain and offset." Additionally, the temporal registration calibration algorithm from J.144 and BT.1683 in NTIA's General Model was used. The modifications to these standardized algorithms were all in response to the Multimedia test plan limitations. For example, the gain and offset were calculated for the first and last second only instead of using the whole PVS. These modifications made these algorithms less robust. Where the software indicated that a PVS did not conform to the test plan, a PVS was kept if it passed a visual inspection.

Proponents and the ILG had the opportunity to check calibration of all the PVSs before the subjective testing was started and after that no PVS could be removed from the data analysis due to calibration issues.

7 MODEL EVALUATION CRITERIA

This chapter describes the evaluation metrics and procedure used to assess the performance of an objective video quality model as an estimator of video picture quality in a variety of applications.

7.1 Evaluation Procedure

The performance of each objective quality model was characterized by three prediction attributes: accuracy, monotonicity and consistency.

The statistical metrics root mean square (rms) error, Pearson correlation, and outlier ratio together characterize the accuracy, monotonicity and consistency of a model's performance. These statistical metrics are named evaluation metrics in the following. The calculation of each evaluation metric is performed along with its 95% confidence intervals. To test for statistically significant differences among the performance of various models, a test based on the F-test was used on the rms error; tests based on approximations to the Gaussian distribution were constructed for the Pearson correlation coefficient and the Outlier Ratio.

The evaluation metrics were calculated using the objective model outputs and the results from viewer subjective rating of the test video clips. The objective model provides a single number (figure of merit) for every tested video clip. The same tested video clips get also a single subjective figure of merit. The subjective figure of merit for a video clip represents the average value of the scores provided by all subjects viewing the video clip.

The evaluation analysis is based on DMOS scores for the FR and RR models, and on MOS scores for the NR model. Discussion below regarding the DMOS scores was applied identically to MOS scores. For simplicity, only DMOS scores are mentioned for the rest of the chapter.

The objective quality model evaluation was performed in three steps. The first step is a mapping of the objective data to the subjective scale. The second calculates the evaluation metrics for the models and their confidence intervals. The third tests for statistical differences between the evaluation metrics value of different models..

7.2 PSNR

PSNR was calculated to provide a performance benchmark.

The NTIA PSNR calculation (NTIA_PSNR_search) used an exhaustive search method for computing PSNR. This algorithm performs an exhaustive search for the maximum PSNR over plus or minus the spatial uncertainty (in pixels) and plus or minus the temporal uncertainty (in frames). The processed video segment is fixed and the original video segment is shifted over the search range. For each spatial-temporal shift, a linear fit between the processed pixels and the original pixels is performed such that the mean square error of (original - gain*processed + offset) is minimized (hence maximizing PSNR). Thus, NTIA_PSNR_search should yield PSNR values that are greater than or equal to commonly used PSNR implementations if the exhaustive search covered enough spatial-temporal shifts. The spatial-temporal search range and the amount of image cropping were performed in accordance with the calibration requirements given in the MM test plan.

7.3 Data Processing

7.3.1 Validity Checks on SRCs and HRC After Subjective Testing

Several SRCs received an MOS score less than 4.0. The ILG examined these sequences and considered the implications of keeping or discarding these SRCs. The ILG decided to keep all SRCs for data analysis.

For data sets C11 and C14, a mistake was made in the common sets. For C11, common set PVS c00_328 was omitted and c00_306 used instead for subjective testing. For C14, common set PVS c00_528 was omitted and c00_501 included instead for subjective testing. These unintentional substitutions were discovered during analysis of the subjective data. For these two sequences, the missing MOS values were replaced with the average of that PVS from other CIF subjective tests. The replacement averaged MOS scores were used in the analysis. The unintended sequences and their associated MOS values were not used in the data analysis.

For test V08, HRCs 7, 8, and 9 were identified in the test design as H.264 with frame freezes. Unintentionally, HRCs 7, 8, and 9 were generated as lossless video with frame freezes inserted. The data rate of this impairment is outside the scope of the MM test plan, which is limited to 4 Mbits/s and less. Therefore, agreement was reached to discard HRCs 7, 8, and 9 from all data analysis. The raw data for HRCs 7, 8, and 9 are not published in this report. There were a total of 24 clips removed: 8 SRCs with the associated HRCs.

For test V13, HRC 16, the data bit rate is above the MM test plan limit of 4 Mbits/s. Because this was stated in the test design and no proponent objected, the HRC has been retained and was used for analysis.

7.3.2 Calculating DMOS Values

The data analysis was performed using the difference mean opinion score (DMOS) for FR and RR methods and using the MOS for NR models. DMOS values were calculated on a per subject per PVS basis. The appropriate hidden reference (SRC) was used to calculate the DMOS value for each PVS. DMOS values were calculated using the following formula:

DMOS = MOS (PVS) – MOS (SRC) + 5

In using this formula, higher DMOS values indicate better quality. Lower bound is 1 as MOS value but higher bound could be more than 5. Any DMOS values greater than 5 (i.e. where the processed sequence is rated better quality than its associated hidden reference sequence) was considered valid and included in the data analysis.

7.3.3 Mapping to the Subjective Scale

Subjective rating data often are compressed at the ends of the rating scales. It is not reasonable for objective models of video quality to mimic this weakness of subjective data. Therefore, a non-linear mapping step was applied before computing any of the performance metrics. A non-linear mapping function that has been found to perform well empirically is the cubic polynomial:

$$DMOSp = ax^3 + bx^2 + cx + d \quad (1)$$

where DMOSp is the predicted DMOS, and the VQR is the model's computed value for a clip-HRC combination. The weightings a, b and c and the constant d are obtained by fitting the function to the data [DMOS, VCR].

The mapping function maximizes the correlation between DMOSp and DMOS :

$$DMOSp = k(a'x^3 + b'x^2 + c'x) + d$$

with constant k = 1, d = 0

This function must be constrained to be monotonic within the range of possible values for our purposes. Then the root mean squared error is minimized over k and d.

a = k*a'

b = k*b'

c = k*c'

This non-linear mapping procedure has been applied to each model's outputs before the evaluation metrics are computed.

Proponents, in addition to the ILG, were allowed to compute the coefficients of the mapping functions for their models and submit the coefficients to ILGs. Proponents submitting coefficients were also required to submit their mapping tool (executable) to ILGs so that ILGs could use the mapping tool for other models. The ILG used the coefficients of the fitting function that produce the best correlation coefficient provided that it is a monotonic fit.

7.3.4 Analysis, Averaging Process and Aggregation Procedure

Primary analysis of model performance was calculated per processed video sequence per experiment.

Secondary analysis of model performance was also calculated and reported on averaged data, by averaging all SRC associated with each HRC ($DMOS_H$) per experiment. The common sequences (i.e., included in every experiment at one resolution) were not used for HRC analysis. This is in contrast to the primary data analysis, where the PVSs for each individual test and the common sequences were analyzed together. This secondary analysis used the same mapping as the primary analysis (e.g., computed on a per PVS basis). The evaluation of the objective metrics was performed in two steps. In the first step, the objective metrics were evaluated per experiment. In this case, the evaluation/statistical metrics were calculated for all tested objective metrics. A comparison analysis was then performed based on significance tests. In the second step, an aggregation of the performance results was performed by taking the average values for all three evaluation metrics for all experiments.

7.4 Evaluation Metrics

Once the mapping was applied to objective data, three evaluation metrics: root mean square error, Pearson correlation coefficient and outlier ratio were determined. The calculation of each evaluation metric was performed along with its 95% confidence interval.

7.4.1 Pearson Correlation Coefficient

The Pearson correlation coefficient R (see equation 2) measures the linear relationship between a model's performance and the subjective data. Its great virtue is that it is on a standard, comprehensible scale of -1 to 1 and it has been used frequently in similar testing.

$$R = \frac{\sum_{i=1}^{N} (Xi - \overline{X}) * (Yi - \overline{Y})}{\sqrt{\sum (Xi - \overline{X})^2} * \sqrt{\sum (Yi - \overline{Y})^2}} \quad (2)$$

Xi denotes the subjective score (DMOS(i) for FR/RR models and MOS(i) for NR models) and Yi the objective score (DMOSp(i) for FR/RR models and MOSp(i) for NR models).. N in equation (2) represents the total number of video clips considered in the analysis.

Therefore, in the context of this test, the value of N in equation (2) is:

- N=152 for FR/RR models (=166-14 since the evaluation for FR/RR discards the reference videos and there are 14 reference videos in each experiment).
- N=166 for NR models.
- Note, if any PVS in the experiment is discarded for data analysis, then the value of N changes accordingly.

The sampling distribution of Pearson's R is not normally distributed. "Fisher's z transformation" converts Pearson's R to the normally distributed variable z. This transformation is given by the following equation :

$$z = 0.5 \cdot \ln\left(\frac{1+R}{1-R}\right) \quad (3)$$

The statistic of z is approximately normally distributed and its standard deviation is defined by:

$$\sigma_z = \sqrt{\frac{1}{N-3}} \quad (4)$$

The 95% confidence interval (CI) for the correlation coefficient is determined using the Gaussian distribution, which characterizes the variable z and it is given by (5)

$$CI = \pm K1 * \sigma_z \quad (5)$$

NOTE1: For a Gaussian distribution, K1 = 1.96 for the 95% confidence interval. If N<30 samples are used then the Gaussian distribution must be replaced by the appropriate Student's t distribution, depending on the specific number of samples used.

Therefore, in the context of this test, K1 = 1.96.

The lower and upper bound associated to the 95% confidence interval (CI) for the correlation coefficient is computed for the Fisher's z value:

$$LowerBound = z - K1 * \sigma_z$$

$$UpperBound = z + K1 * \sigma_z$$

NOTE2**:** The values of Fisher's z of lower and upper bounds are then converted back to Pearson's R to get the CI of correlation R.

7.4.2 Root Mean Square Error

The accuracy of the objective metric is evaluated using the root mean square error (rmse) evaluation metric.

The difference between measured and predicted DMOS is defined as the absolute prediction error Perror:

$$Perror(i) = DMOS(i) - DMOS_p(i) \quad (6)$$

where the index i denotes the video sample.

NOTE: DMOS(i) and DMOSp(i) are used for FR/RR models. MOS(i) and MOSp(i) are used for NR models.

The root-mean-square error of the absolute prediction error Perror is calculated with the formula:

$$rmse = \sqrt{\left(\frac{1}{N-d}\sum_{N} Perror[i]^2\right)} \quad (7)$$

where N denotes the total number of video clips considered in the analysis, and d is the number of degrees of freedom of the mapping function (1).

In the case of a mapping using a 3rd-order monotonic polynomial function, d=4 (since there are 4 coefficients in the fitting function).

In the context of this test plan, the value of N in equation (7) is:

- N=152 for FR/RR models (since the evaluation discards the reference videos and there are 14 reference videos in each experiment)
- N=166 for NR models
- NOTE: if any PVS in the experiment is discarded for data analysis, then the value of N changes accordingly.

The root mean square error is approximately characterized by a χ^2 (n) [2], where n represents the degrees of freedom and it is defined by (8):

$$n = N - d \quad (8)$$

where N represents the total number of samples.

Using the χ^2 (n) distribution, the 95% confidence interval for the rmse is given by (9) [2]:

$$\frac{rmse * \sqrt{N-d}}{\sqrt{\chi^2_{0.025}(N-d)}} < rmse < \frac{rmse * \sqrt{N-d}}{\sqrt{\chi^2_{0.975}(N-d)}} \quad (9)$$

7.4.3 Outlier ratio (using standard error of the mean)

The consistency attribute of the objective metric is evaluated by the outlier ratio (OR) which represents the ratio of "outlier-points" to total points N:

$$OR = \frac{TotaNoOutliers}{N} \quad (10)$$

where an outlier is a point for which

$$| Perror(i) | > K2 * \frac{\sigma(DMOS(i))}{\sqrt{Nsubjs}} \quad (11)$$

where σ(DMOS(i)) represents the standard deviation of the individual scores associated with the video clip i, and Nsubjs is the number of viewers per video clip i. In this test plan, a number of 24 viewers (Nsubjs=24) per video clip was used.

NOTE1**:** DMOS(i) is used for FR/RR models. MOS(i) is used for NR models.

NOTE2**:** For a Gaussian distribution, K2 = 1.96 for the 95% confidence interval. If the mean (DMOS or MOS) is based on less than thirty samples (i.e. Nsubjs < 30), then the Gaussian distribution must be replaced by the appropriate Student's t distribution, depending on the specific number of samples in the mean. In the case of 24 viewers per video (i.e., the number of samples in the mean is 24), the number of degrees of freedom is df=23 and therefore the associated K2 = 2.069 is used for the 95% confidence interval.

Therefore, in the context of this test plan, K2 = 2.069.

The outlier ratio represents the proportion of outliers in N number of samples. Thus, the binomial distribution could be used to characterize the outlier ratio. The outlier ratio is represented by a distribution of proportions [2] characterized by the mean p (12) and standard deviation σ_p (13).

$$OR = p = \frac{TotalNoOutliers}{N} \quad (12)$$

$$\sigma_p = \sqrt{\frac{p*(1-p)}{N}} \quad (13)$$

where N is the total number of video clips considered in the analysis.

For N>30, the binomial distribution, which characterizes the proportion p, can be approximated with the Gaussian distribution . Therefore, the 95% confidence interval (CI) of the outlier ratio is given by (14)

$$CI = \pm 1.96*\sigma_p \quad (14)$$

NOTE**.** If the mean is based on less than thirty samples (ie., N < 30), then the Gaussian distribution must be replaced the appropriate Student's t distribution, depending on the specific number of samples in the mean [2].

7.5 Statistical Significance of the Results

7.5.1 Significance of the Difference between the Correlation Coefficients

The test is based on the assumption that the normal distribution is a good fit for the video quality scores' populations. The statistical significance test for the difference between the correlation coefficients uses the H_0 hypothesis that assumes that there is no significant difference between correlation coefficients. The H_1 hypothesis considers that the difference is significant, although not specifying better or worse.

The test uses the Fisher-z transformation (3) [2]. The normally distributed statistic Z_N (15) is determined for each comparison and evaluated against the 95% t-Student value for the two–tail test, which is the tabulated value t(0.05) =1.96.

$$Z_N = \frac{z1 - z2 - \mu_{(z1-z2)}}{\sigma_{(z1-z2)}} \quad (15)$$

$$\text{where } \mu_{(z1-z2)} = 0 \quad (16)$$

and

$$\sigma_{(z1-z2)} = \sqrt{\sigma_{z1}^2 + \sigma_{z2}^2} \quad (17)$$

σ_{z1} and σ_{z2} represent the standard deviation of the Fisher-z statistic for each of the compared correlation coefficients. The mean (16) is set to zero due to the H_0 hypothesis and the standard deviation of the difference metric z1-z2 is defined by (17).

The standard deviation of the Fisher-z statistic is given by (18):

$$\sigma_z = \sqrt{1/(N-3)} \quad (18)$$

where N represents the total number of samples used for the calculation of each of the two correlation coefficients.

Using (17) and (18), the standard deviation of the difference metric z1-z2 therefore becomes:

$$\sigma_{z1-z2} = \sqrt{\frac{1}{N_1 - 3} + \frac{1}{N_2 - 3}}$$

where N1=N2=N

7.5.2 Significance of the Difference between the Root Mean Square Errors

Considering the same assumption that the two populations are normally distributed, the comparison procedure is similar to the one used for the correlation coefficients. The H_0 hypothesis considers that there is no difference between rmse values. The alternative H_1 hypothesis is assuming that the lower prediction error value is statistically significantly lower. The statistic defined by (19) has a F-distribution with n1 and n2 degrees of freedom [2].

$$\zeta = \frac{(rmse_{\max})^2}{(rmse_{\min})^2} \tag{19}$$

$rmse_{max}$is the highest rmse and $rmse_{min}$is the lowest rmse involved in the comparison. The ζ statistic is evaluated against the tabulated value F(0.05, n1, n2) that ensures 95% significance level. The n1 and n2 degrees of freedom are given by N1-d, respectively and N2-d, with N1 and N2 representing the total number of samples for the compared average rmse (prediction errors) and d being the number of parameters in the fitting equation (1).

If ζ is higher than the tabulated value F(0.05, n1, n2) then there is a significant difference between the values of RMSE.

7.5.3 Significance of the Difference between the Outlier Ratios

As mentioned in paragraph 7.4.3, the outlier ratio could be described by a binomial distribution of parameters (p, 1-p), where p is defined by (12). In this case p is equivalent to the probability of success of the binomial distribution.

The distribution of differences of proportions from two binomially distributed populations with parameters (p1, 1-p1) and (p2, 1-p2) (where p1 and p2 correspond to the two compared outlier ratios) is approximated by a normal distribution for N1, N2 >30, with the mean:

$$\mu_{(p1-p2)} = \mu(p1) - \mu(p2) = p1 - p2 = 0 \tag{20}$$

and standard deviation:

$$\sigma_{p1-p2} = \sqrt{\frac{\sigma(p1)^2}{N1} + \frac{\sigma(p2)^2}{N2}} \tag{21}$$

The null hypothesis in this case considers that there is no difference between the population parameters p1 and p2, respectively p1=p2. Therefore, the mean (20) is zero and the standard deviation (21) becomes equation (22):

$$\sigma_{p1-p2} = \sqrt{p*(1-p)*(\frac{1}{N1} + \frac{1}{N2})} \tag{22}$$

where N1 and N2 represent the total number of samples of the compared outlier ratios p1 versus p2. The variable p is defined by equation (23):

$$p = \frac{N1*p1 + N2*p2}{N1+N2} \tag{23}$$

As for the hypothesis test of correlation coefficients, the normalized statistics Z_N is calculated as in (24).

$$Z_N = \frac{p1 - p2 - \mu_{(p1-p2)}}{\sigma_{(p1-p2)}} \tag{24}$$

Z_N is compared to the tabulated value of 1.96 for the 0.05 significance level of the two tailed test.

If the calculated $Z_N > 1.96$, then the compared outlier ratios p1 and p2 are statistically significantly different, with 0.05 significance level.

8 COMMON VIDEO CLIP ANALYSIS AND INTERPRETATION

The presence of a common set of video clips for each resolution (VGA, CIF, and QCIF) in each of the independent subjective experiments (13 tests for VGA, 14 tests for CIF and QCIF) provides a unique opportunity for assessing the reliability and repeatability of subjective experiments. It can also provide a benchmark for perceptual objective metrics, whose ultimate goal is to replace subjective viewing tests with a small number of viewers (e.g., 24).

The common clips at each resolution spanned the full range of perceptual quality on the ACR-HR scale. By computing a grand mean over all tests and viewers for each resolution (VGA, CIF, and QCIF), we can obtain 24 DMOS scores (i.e., the common set without the 6 reference SRCs) that get about as close to "Perceptual Quality Truth" as can ever be expected. These grand means are obtained by averaging 13x24=312 (VGA) or 14x24=336 (CIF or QCIF) viewers from all over the world. We can compare this grand "Perceptual Quality Truth" to what might be expected from one 24-viewer subjective test. The Pearson correlation coefficients (ρ) between the individual subjective experiments and the corresponding grand "Perceptual Quality Truth" have been computed to be:

VGA: $0.953 < \rho < 0.996$, median = 0.976

CIF: $0.939 < \rho < 0.990$, median = 0.981

QCIF: $0.943 < \rho < 0.982$, median = 0.971

This demonstrates that the majority of the subjective variance in a 24-viewer experiment results from actual perceived differences in quality, consistently perceived differences in quality across many labs, cultures, and resolutions. For the common set, the proportion of the grand variance that is explained by an individual 24-viewer experiment is given by ρ^2, and the proportion of unexplained error variance is given by $1- \rho^2$. The median error variance is thus estimated to be 4.74% for VGA ($1-0.976^2$), 3.76% for CIF ($1-0.981^2$), and 5.72% for QCIF ($1-0.971^2$).

These results provide strong evidence that all of the MM Phase I subjective experiments were conducted in the approved manner, and that each MM data set contains unbiased and non-discriminatory subjective scores. VQEG has a high level of confidence in the execution of the subjective testing. This confidence applies to both tests performed by proponents and tests performed by ILG. The high correlation between “Perceptual Quality Truth” and the individual subjective experiments confirms the reliability and repeatability of subjective experiments.

[Note: Each subjective test and each common set contained a carefully balanced set of scenes and a wide range of HRC quality. Experiments designed with less care may experience decreased accuracy.]

Similarly, if we compare the objective metrics in this report to the grand "Perceptual Quality Truth" as calculated above for the common set, we obtain *maximum* Pearson correlation coefficients of:

VGA: $\rho < 0.842$

CIF: $\rho < 0.796$

QCIF: $\rho < 0.800$

That is, each objective metric was compared to the grand “Perceptual Quality Truth”, and the highest Pearson correlation retained.

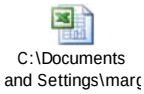
Therefore, none of the evaluated models reaches the accuracy of normative subjective testing.

The objective metrics in this report fail to explain a substantial portion of the subjective test variance. The best error variance for an objective metric for the common set is estimated to be 29.1% for VGA, 36.6% for CIF, and 36.0% for QCIF. This is 6.14 times the median error variance of a corresponding 24-viewer VGA subjective test (29.1/4.74), 9.73 times the median error variance of a corresponding 24-viewer CIF subjective test (36.6/3.76), and 6.29 times the median error variance of a corresponding 24-viewer CIF subjective test (36.0/5.72).

[Note: The VGA, CIF and QCIF common sets were designed to be a small part of a larger subjective experiment. When taken out of that context, the common sets are not suitable for analyzing whether an objective model is appropriate for standardization. Therefore, the statistics in this section should only be used for the intended purpose, which is (1) to analyze the repeatability and reliability of subjective testing, and (2) to determine whether the evaluated objective models can duplicate the precision of subjective testing.]

9 OFFICIAL ILG DATA ANALYSIS

The official ILG data analysis presented in this section is also available in the embedded Microsoft Excel document, here:



The Excel pages and contents of each are as follows:

VGA	Primary analysis for all VGA models.
CIF	Primary analysis for all CIF models.
QCIF	Primary analysis for all QCIF models.

Each of the above three pages includes for each experiment and each model Correlation, RMSE and Outlier Ratio. Below each of these three tables is the average performance for each model for that statistic. Below this are the significance testing for all three statistics, and significance testing comparing each model to PSNR using RMSE only.

Finally, each primary analysis page includes listing of the number of transmission error HRCs in each experiment, and plots the correlation versus the number of transmission error HRCs. The correlation numbers plotted are identical to those from the primary analysis at the top of the current MS-Excel page (i.e., correlation for each model, each experiment). The column "Error" identifies the number of HRCs that contained transmission errors for that experiment (e.g., VGA test V01, 3 of the 16 HRCs contained transmission errors). Every experiment contained 16 HRCs, except for V08 where three HRCs were eliminated. A plot is included for each model, where the Y-axis is correlation (per experiment) and the X-axis is the number of transmission error HRCs (per experiment). These plots relate the model's correlation to the frequency of transmission error HRCs.

VGA_Secondary	Secondary analysis for all VGA models.
CIF_Secondary	Secondary analysis for all CIF models.
QCIF_Secondary	Secondary analysis for all QCIF models.

Each of the above three pages includes for each experiment and each model Correlation, RMSE and Outlier Ratio, and the average performance for each model using each statistic.

All per-experiment analyses are high lit in light green. Results that have been aggregated (averaged) over all experiments are high lit in yellow.

9.1 VGA Primary Analysis

9.1.1 VGA Primary Analysis Metrics and Averages

Correlation

	FR Models					RR Models				NR Models		
Test	Psy_FR	Opt_FR	Yon_FR	NTT_FR	PSNR_DMOS	Yon_RR10k	Yon_RR64k	Yon_RR128k	PSNR_DMOS	Psy_NR	Swi_NR	PSNR_MOS
V01	0.884	0.787	0.827	0.879	0.825	0.878	0.878	0.879	0.825	0.659	0.416	0.849
V02	0.565	0.893	0.784	0.753	0.595	0.791	0.790	0.792	0.595	0.411	0.593	0.712
V03	0.749	0.778	0.612	0.801	0.726	0.756	0.757	0.758	0.726	0.597	0.430	0.838
V04	0.735	0.784	0.790	0.782	0.707	0.767	0.763	0.765	0.707	0.693	0.409	0.827
V05	0.892	0.939	0.926	0.920	0.825	0.930	0.930	0.930	0.825	0.733	0.575	0.840
V06	0.898	0.892	0.877	0.863	0.757	0.879	0.880	0.880	0.757	0.643	0.456	0.797
V07	0.843	0.883	0.856	0.806	0.764	0.861	0.858	0.859	0.764	0.621	0.344	0.804
V08	0.854	0.685	0.878	0.865	0.794	0.895	0.895	0.894	0.794	0.338	0.310	0.837
V09	0.778	0.758	0.692	0.654	0.583	0.648	0.651	0.652	0.583	0.190	0.555	0.780
V10	0.887	0.821	0.865	0.665	0.779	0.792	0.791	0.793	0.779	0.666	0.307	0.833
V11	0.863	0.859	0.795	0.598	0.773	0.818	0.815	0.814	0.773	0.584	0.372	0.782
V12	0.824	0.758	0.681	0.737	0.499	0.622	0.622	0.620	0.499	0.479	0.437	0.705
V13	0.918	0.887	0.887	0.891	0.635	0.799	0.804	0.805	0.635	0.725	0.456	0.715
Average=	0.822	0.825	0.805	0.786	0.713	0.803	0.803	0.803	0.713	0.565	0.435	0.794
Minimum	0.565	0.685	0.612	0.598	0.499	0.622	0.622	0.620	0.499	0.190	0.307	0.705
Maximum	0.918	0.939	0.926	0.920	0.825	0.930	0.930	0.930	0.825	0.733	0.593	0.849

RMSE

	FR Models					RR Models				NR Models		
Test	Psy_FR	Opt_FR	Yon_FR	NTT_FR	PSNR_DMOS	Yon_RR10k	Yon_RR64k	Yon_RR128k	PSNR_DMOS	Psy_NR	Swi_NR	PSNR_MOS
V01	0.505	0.665	0.607	0.514	0.610	0.516	0.517	0.515	0.610	0.848	1.025	0.595
V02	0.798	0.436	0.600	0.636	0.778	0.591	0.593	0.591	0.778	0.961	0.849	0.741
V03	0.669	0.635	0.799	0.605	0.694	0.662	0.660	0.659	0.694	0.789	0.888	0.537
V04	0.652	0.597	0.590	0.599	0.679	0.617	0.621	0.619	0.679	0.782	0.990	0.611
V05	0.486	0.369	0.406	0.420	0.607	0.396	0.396	0.395	0.607	0.727	0.874	0.580
V06	0.472	0.485	0.514	0.542	0.699	0.511	0.508	0.509	0.699	0.823	0.956	0.648
V07	0.556	0.485	0.535	0.612	0.667	0.527	0.531	0.530	0.667	0.815	0.976	0.618
V08	0.555	0.778	0.512	0.535	0.649	0.475	0.476	0.478	0.649	1.001	1.012	0.583
V09	0.575	0.597	0.661	0.693	0.744	0.698	0.695	0.694	0.744	1.067	0.904	0.681
V10	0.499	0.618	0.543	0.808	0.679	0.660	0.662	0.660	0.679	0.837	1.068	0.621
V11	0.575	0.583	0.691	0.913	0.722	0.656	0.660	0.662	0.722	0.836	0.957	0.642
V12	0.555	0.639	0.718	0.662	0.849	0.767	0.768	0.769	0.849	0.985	1.009	0.796
V13	0.464	0.542	0.541	0.533	0.905	0.705	0.698	0.696	0.905	0.833	1.076	0.845
Average=	0.566	0.571	0.593	0.621	0.714	0.599	0.599	0.598	0.714	0.870	0.968	0.654
Minimum	0.464	0.369	0.406	0.420	0.607	0.396	0.396	0.395	0.607	0.727	0.849	0.537
Maximum	0.798	0.778	0.799	0.913	0.905	0.767	0.768	0.769	0.905	1.067	1.076	0.845

Outlier Ratio

	FR Models					RR Models				NR Models		
Test	Psy_FR	Opt_FR	Yon_FR	NTT_FR	PSNR_DMOS	Yon_RR10k	Yon_RR64k	Yon_RR128k	PSNR_DMOS	Psy_NR	Swi_NR	PSNR_MOS
V01	0.566	0.592	0.599	0.553	0.618	0.526	0.533	0.539	0.618	0.711	0.813	0.572
V02	0.704	0.395	0.559	0.559	0.664	0.592	0.586	0.586	0.664	0.831	0.795	0.645
V03	0.572	0.566	0.697	0.533	0.625	0.678	0.664	0.664	0.625	0.747	0.771	0.614
V04	0.507	0.539	0.401	0.480	0.500	0.454	0.474	0.467	0.500	0.699	0.765	0.578
V05	0.368	0.309	0.388	0.276	0.533	0.382	0.362	0.362	0.533	0.633	0.783	0.560
V06	0.349	0.388	0.447	0.487	0.605	0.493	0.480	0.480	0.605	0.831	0.789	0.614
V07	0.487	0.414	0.507	0.467	0.539	0.500	0.500	0.493	0.539	0.723	0.807	0.572
V08	0.477	0.586	0.398	0.453	0.516	0.398	0.383	0.391	0.516	0.775	0.838	0.556
V09	0.651	0.658	0.645	0.638	0.711	0.645	0.638	0.645	0.711	0.801	0.783	0.651
V10	0.507	0.553	0.566	0.645	0.678	0.586	0.579	0.586	0.678	0.795	0.789	0.663
V11	0.533	0.474	0.618	0.599	0.605	0.625	0.605	0.599	0.605	0.753	0.807	0.590
V12	0.612	0.566	0.671	0.533	0.717	0.678	0.684	0.691	0.717	0.789	0.795	0.711
V13	0.480	0.487	0.546	0.579	0.684	0.678	0.697	0.678	0.684	0.753	0.855	0.705
Average=	0.524	0.502	0.542	0.523	0.615	0.556	0.553	0.552	0.615	0.757	0.799	0.618
Minimum	0.349	0.309	0.388	0.276	0.500	0.382	0.362	0.362	0.500	0.633	0.765	0.556
Maximum	0.704	0.658	0.697	0.645	0.717	0.678	0.697	0.691	0.717	0.831	0.855	0.711

9.1.2 VGA Statistical Significance using RMSE

Separate results for each model type: (FR models + PSNR on DMOS); (RR models + PSNR on DMOS); and (NR models + PSNR on MOS).

Statistical Equivalence to Top Performing Model

"1" indicates that this model is statistically equivalent to the top performing model.
"0" indicates that this model is statistically worse than the top performing model.

	FR Models					RR Models				NR Models		
Test	Psy_FR	Opt_FR	Yon_FR	NTT_FR	PSNR_DMOS	Yon_RR10k	Yon_RR64k	Yon_RR128k	PSNR_DMOS	Psy_NR	Swi_NR	PSNR_MOS
V01	1	0	0	1	0	1	1	1	0	0	0	1
V02	0	1	0	0	0	1	1	1	0	0	0	1
V03	1	1	0	1	0	1	1	1	1	0	0	1
V04	1	1	1	1	0	1	1	1	1	0	0	1
V05	0	1	1	0	0	1	1	1	0	0	0	1
V06	1	1	1	0	0	1	1	1	0	0	0	1
V07	0	1	1	0	0	1	1	1	0	0	0	1
V08	1	0	1	1	0	1	1	1	0	0	0	1
V09	1	1	0	0	0	1	1	1	1	0	0	1
V10	1	0	1	0	0	1	1	1	1	0	0	1
V11	1	1	0	0	0	1	1	1	1	0	0	1
V12	1	0	0	0	0	1	1	1	1	1	0	1
V13	1	0	0	0	0	1	1	1	0	0	0	1
Total=	10	8	6	4	0	13	13	13	6	1	0	13

Statistically Better than PSNR

"1" indicates that this model is statistically better than PSNR
"0" indicates that this model is not statistically better than PSNR

	FR Models				RR Models			NR Models	
Test	Psy_FR	Opt_FR	Yon_FR	NTT_FR	Yon_RR10k	Yon_RR64k	Yon_RR128k	Psy_NR	Swi_NR
V01	1	0	0	1	1	1	1	0	0
V02	0	1	1	1	1	1	1	0	0
V03	0	0	0	1	0	0	0	0	0
V04	0	1	1	0	0	0	0	0	0
V05	1	1	1	1	1	1	1	0	0
V06	1	1	1	1	1	1	1	0	0
V07	1	1	1	0	1	1	1	0	0
V08	1	0	1	1	1	1	1	0	0
V09	1	1	0	0	0	0	0	0	0
V10	1	0	1	0	0	0	0	0	0
V11	1	1	0	0	0	0	0	0	0
V12	1	1	1	1	0	0	0	0	0
V13	1	1	1	1	1	1	1	0	0
Total=	10	9	9	8	7	7	7	0	0

9.1.3 VGA Statistical Significance Using Outlier Ratio

Separate results for each model type: (FR models + PSNR on DMOS); (RR models + PSNR on DMOS); and (NR models + PSNR on MOS)

Statistical Equivalence to Top Performing Model using Outlier Ratio

"1" indicates that this model is statistically equivalent to the top performing model.
"0" indicates that this model is statistically worse than the top performing model.

	FR Models					RR Models				NR Models	
Test	Psy_FR	Opt_FR	Yon_FR	NTT_FR	PSNR_DMOS	Yon_RR10k	Yon_RR64k	Yon_RR128k	PSNR_DMOS	Psy_NR	Swi_NR
V01	1	1	1	1	1	1	1	1	1	1	1
V02	0	1	0	0	0	1	1	1	1	1	1
V03	1	1	0	1	1	1	1	1	1	1	1
V04	1	0	1	1	1	1	1	1	1	1	1
V05	1	1	0	1	0	1	1	1	0	1	0
V06	1	1	1	0	0	1	1	1	0	1	1
V07	1	1	1	1	0	1	1	1	1	1	1
V08	1	0	1	1	0	1	1	1	0	1	1
V09	1	1	1	1	1	1	1	1	1	1	1
V10	1	1	1	0	0	1	1	1	1	1	1
V11	1	1	0	0	0	1	1	1	1	1	1
V12	1	1	0	1	0	1	1	1	1	1	1
V13	1	1	1	1	0	1	1	1	1	1	1
Total=	12	11	8	9	4	13	13	13	10	13	12

Note: Comparison for NR models including PSNR_MOS is not available.

9.1.4 VGA Statistical Significance Using Correlation

Separate results for each model type: (FR models + PSNR on DMOS); (RR models + PSNR on DMOS); and (NR models + PSNR on MOS)

Statistical Equivalence to Top Performing Model using Correlation

"1" indicates that this model is statistically equivalent to the top performing model.
"0" indicates that this model is statistically worse than the top performing model.

	FR Models					RR Models				NR Models		
Test	Psy_FR	Opt_FR	Yon_FR	NTT_FR	PSNR_DMOS	Yon_RR10k	Yon_RR64k	Yon_RR128k	PSNR_DMOS	Psy_NR	Swi_NR	PSNR_MOS
V01	1	0	1	1	1	1	1	1	1	0	0	1
V02	0	1	0	0	0	1	1	1	0	0	1	1
V03	1	1	0	1	1	1	1	1	1	0	0	1
V04	1	1	1	1	1	1	1	1	1	0	0	1
V05	0	1	1	1	0	1	1	1	0	0	0	1
V06	1	1	1	1	0	1	1	1	0	0	0	1
V07	1	1	1	0	0	1	1	1	0	0	0	1
V08	1	0	1	1	0	1	1	1	0	0	0	1
V09	1	1	1	0	0	1	1	1	1	0	0	1
V10	1	0	1	0	0	1	1	1	1	0	0	1
V11	1	1	1	0	0	1	1	1	1	0	0	1
V12	1	1	0	1	0	1	1	1	1	0	0	1
V13	1	1	1	1	0	1	1	1	0	1	0	1
Total=	11	10	10	8	3	13	13	13	7	1	1	13

9.1.5 Number of VGA HRCs with Transmission Errors

Note: Official ILG Data Analysis Excel file also contains plots of number of transmission errors against experiment correlation.

		FR Models					RR Models				NR Models		
Test	Error	Psy_FR	Opt_FR	Yon_FR	NTT_FR	PSNR_DMOS	Yon_RR10k	Yon_RR64k	Yon_RR128k	PSNR_DMOS	Psy_NR	Swi_NR	PSNR_MOS
V01	3	0.884	0.787	0.827	0.879	0.825	0.878	0.878	0.879	0.825	0.659	0.416	0.849
V02	8	0.565	0.893	0.784	0.753	0.595	0.791	0.790	0.792	0.595	0.411	0.593	0.712
V03	8	0.749	0.778	0.612	0.801	0.726	0.756	0.757	0.758	0.726	0.597	0.430	0.838
V04	8	0.735	0.784	0.790	0.782	0.707	0.767	0.763	0.765	0.707	0.693	0.409	0.827
V05	0	0.892	0.939	0.926	0.920	0.825	0.930	0.930	0.930	0.825	0.733	0.575	0.840
V06	4	0.898	0.892	0.877	0.863	0.757	0.879	0.880	0.880	0.757	0.643	0.456	0.797
V07	4	0.843	0.883	0.856	0.806	0.764	0.861	0.858	0.859	0.764	0.621	0.344	0.804
V08	2	0.854	0.685	0.878	0.865	0.794	0.895	0.895	0.894	0.794	0.338	0.310	0.837
V09	8	0.778	0.758	0.692	0.654	0.583	0.648	0.651	0.652	0.583	0.190	0.555	0.780
V10	0	0.887	0.821	0.865	0.665	0.779	0.792	0.791	0.793	0.779	0.666	0.307	0.833
V11	0	0.863	0.859	0.795	0.598	0.773	0.818	0.815	0.814	0.773	0.584	0.372	0.782
V12	0	0.824	0.758	0.681	0.737	0.499	0.622	0.622	0.620	0.499	0.479	0.437	0.705
V13	0	0.918	0.887	0.887	0.891	0.635	0.799	0.804	0.805	0.635	0.725	0.456	0.715

9.2 CIF Primary Data Analysis

9.2.1 CIF Primary Analysis Metrics and Averages

Correlation

	FR Models					RR Models			NR Models		
Test	Psy_FR	Opt_FR	Yon_FR	NTT_FR	PSNR_DMOS	Yon_RR10k	Yon_RR64k	PSNR_DMOS	Psy_NR	Swi_NR	PSNR_MOS
C01	0.823	0.792	0.801	0.710	0.698	0.831	0.838	0.698	0.590	0.522	0.759
C02	0.856	0.828	0.728	0.838	0.696	0.717	0.718	0.696	0.728	0.650	0.812
C03	0.823	0.695	0.722	0.782	0.708	0.789	0.788	0.708	0.281	0.424	0.787
C04	0.796	0.810	0.734	0.791	0.698	0.699	0.694	0.698	0.387	0.516	0.801
C05	0.890	0.850	0.861	0.836	0.733	0.869	0.869	0.733	0.690	0.435	0.801
C06	0.892	0.877	0.874	0.864	0.796	0.904	0.904	0.796	0.717	0.663	0.835
C07	0.804	0.753	0.749	0.780	0.440	0.725	0.732	0.440	0.587	0.586	0.686
C08	0.826	0.844	0.855	0.728	0.648	0.844	0.849	0.648	0.546	0.658	0.719
C09	0.852	0.835	0.821	0.706	0.558	0.786	0.786	0.558	0.616	0.709	0.721
C10	0.769	0.737	0.809	0.723	0.639	0.791	0.792	0.639	0.478	0.378	0.737
C11	0.792	0.747	0.734	0.675	0.477	0.677	0.682	0.477	0.622	0.536	0.699
C12	0.788	0.779	0.748	0.811	0.636	0.734	0.733	0.636	0.586	0.461	0.730
C13	0.897	0.848	0.712	0.778	0.689	0.677	0.677	0.689	0.589	0.578	0.761
C14	0.891	0.923	0.836	0.853	0.768	0.883	0.886	0.768	0.684	0.617	0.850
Average=	0.836	0.808	0.785	0.777	0.656	0.780	0.782	0.656	0.579	0.552	0.764
Minimum	0.769	0.695	0.712	0.675	0.440	0.677	0.677	0.440	0.281	0.378	0.686
Maximum	0.897	0.923	0.874	0.864	0.796	0.904	0.904	0.796	0.728	0.709	0.850

RMSE

	FR Models					RR Models			NR Models		
Test	Psy_FR	Opt_FR	Yon_FR	NTT_FR	PSNR_DMOS	Yon_RR10k	Yon_RR64k	PSNR_DMOS	Psy_NR	Swi_NR	PSNR_MOS
C01	0.587	0.630	0.618	0.727	0.739	0.574	0.563	0.739	0.866	0.915	0.699
C02	0.466	0.505	0.617	0.491	0.646	0.628	0.627	0.646	0.705	0.781	0.599
C03	0.550	0.696	0.670	0.604	0.684	0.595	0.595	0.684	0.960	0.906	0.617
C04	0.525	0.508	0.589	0.531	0.621	0.620	0.624	0.621	0.855	0.794	0.555
C05	0.490	0.566	0.547	0.591	0.733	0.533	0.533	0.733	0.809	1.006	0.669
C06	0.495	0.526	0.530	0.550	0.662	0.468	0.467	0.662	0.780	0.838	0.615
C07	0.535	0.592	0.597	0.563	0.809	0.620	0.613	0.809	0.721	0.722	0.648
C08	0.503	0.479	0.462	0.612	0.679	0.479	0.472	0.679	0.777	0.699	0.645
C09	0.432	0.454	0.471	0.584	0.684	0.510	0.509	0.684	0.724	0.648	0.636
C10	0.663	0.700	0.609	0.715	0.797	0.634	0.632	0.797	0.900	0.948	0.692
C11	0.627	0.684	0.698	0.758	0.903	0.757	0.751	0.903	0.878	0.947	0.802
C12	0.561	0.571	0.605	0.533	0.703	0.618	0.619	0.703	0.798	0.874	0.674
C13	0.472	0.566	0.750	0.671	0.774	0.786	0.787	0.774	0.921	0.931	0.740
C14	0.460	0.390	0.557	0.530	0.650	0.476	0.472	0.650	0.836	0.902	0.604
Average=	0.526	0.562	0.594	0.604	0.720	0.593	0.590	0.720	0.824	0.851	0.657
Minimum	0.432	0.390	0.462	0.491	0.621	0.468	0.467	0.621	0.705	0.648	0.555
Maximum	0.663	0.700	0.750	0.758	0.903	0.786	0.787	0.903	0.960	1.006	0.802

Outlier Ratio

	FR Models					RR Models			NR Models		
Test	Psy_FR	Opt_FR	Yon_FR	NTT_FR	PSNR_DMOS	Yon_RR10k	Yon_RR64k	PSNR_DMOS	Psy_NR	Swi_NR	PSNR_MOS
C01	0.546	0.625	0.559	0.592	0.684	0.507	0.480	0.684	0.789	0.801	0.669
C02	0.408	0.467	0.526	0.408	0.572	0.592	0.579	0.572	0.614	0.687	0.590
C03	0.513	0.664	0.605	0.546	0.645	0.526	0.533	0.645	0.783	0.795	0.608
C04	0.480	0.434	0.500	0.441	0.599	0.539	0.553	0.599	0.765	0.741	0.627
C05	0.454	0.493	0.421	0.533	0.645	0.421	0.428	0.645	0.759	0.831	0.705
C06	0.447	0.454	0.493	0.480	0.579	0.382	0.368	0.579	0.699	0.747	0.639
C07	0.454	0.507	0.507	0.474	0.638	0.520	0.507	0.638	0.614	0.602	0.633
C08	0.493	0.355	0.395	0.507	0.638	0.349	0.355	0.638	0.663	0.633	0.645
C09	0.434	0.434	0.454	0.526	0.605	0.507	0.467	0.605	0.651	0.639	0.657
C10	0.605	0.605	0.520	0.579	0.664	0.586	0.572	0.664	0.789	0.843	0.663
C11	0.579	0.586	0.546	0.691	0.678	0.572	0.566	0.678	0.747	0.735	0.693
C12	0.539	0.599	0.546	0.513	0.546	0.546	0.533	0.546	0.741	0.687	0.614
C13	0.625	0.599	0.743	0.691	0.789	0.763	0.750	0.789	0.771	0.837	0.735
C14	0.513	0.355	0.493	0.553	0.566	0.454	0.461	0.566	0.801	0.723	0.602
Average=	0.507	0.513	0.522	0.538	0.632	0.519	0.511	0.632	0.728	0.736	0.648
Minimum	0.408	0.355	0.395	0.408	0.546	0.349	0.355	0.546	0.614	0.602	0.590
Maximum	0.625	0.664	0.743	0.691	0.789	0.763	0.750	0.789	0.801	0.843	0.735

9.2.2 CIF Statistical Significance Using RMSE

Separate results for each model type: (FR models + PSNR on DMOS); (RR models + PSNR on DMOS); and (NR models + PSNR on MOS)

Statistical Equivalence to Top Performing Model

"1" indicates that this model is statistically equivalent to the top performing model.
"0" indicates that this model is statistically worse than the top performing model.

	FR Models					RR Models			NR Models		
Test	Psy_FR	Opt_FR	Yon_FR	NTT_FR	PSNR_DMOS	Yon_RR10k	Yon_RR64k	PSNR_DMOS	Psy_NR	Swi_NR	PSNR_MOS
C01	1	1	1	0	0	1	1	0	0	0	1
C02	1	1	0	1	0	1	1	1	0	0	1
C03	1	0	0	1	0	1	1	0	0	0	1
C04	1	1	0	1	0	1	1	1	0	0	1
C05	1	0	1	0	0	1	1	0	0	0	1
C06	1	1	1	1	0	1	1	0	0	0	1
C07	1	1	1	1	0	1	1	0	1	1	1
C08	1	1	1	0	0	1	1	0	0	1	1
C09	1	1	1	0	0	1	1	0	1	1	1
C10	1	0	1	0	0	1	1	0	0	0	1
C11	1	1	1	0	0	1	1	0	1	0	1
C12	1	1	1	1	0	1	1	1	0	0	1
C13	1	0	0	0	0	1	1	1	0	0	1
C14	0	1	0	0	0	1	1	0	0	0	1
Total=	13	10	9	6	0	14	14	4	3	3	14

Statistically Better than PSNR

"1" indicates that this model is statistically better than PSNR
"0" indicates that this model is not statistically better than PSNR

	FR Models				RR Models		NR Models	
Test	Psy_FR	Opt_FR	Yon_FR	NTT_FR	Yon_RR10k	Yon_RR64k	Psy_NR	Swi_NR
C01	1	1	1	0	1	1	0	0
C02	1	1	0	1	0	0	0	0
C03	1	0	0	0	1	1	0	0
C04	1	1	0	1	0	0	0	0
C05	1	1	1	1	1	1	0	0
C06	1	1	1	1	1	1	0	0
C07	1	1	1	1	1	1	0	0
C08	1	1	1	0	1	1	0	0
C09	1	1	1	1	1	1	0	0
C10	1	0	1	0	1	1	0	0
C11	1	1	1	1	1	1	0	0
C12	1	1	1	1	0	0	0	0
C13	1	1	0	1	0	0	0	0
C14	1	1	1	1	1	1	0	0
Total=	14	12	10	10	10	10	0	0

9.2.3 CIF Statistical Significance Using Outlier Ratio

Separate results for each model type: (FR models + PSNR on DMOS); (RR models + PSNR on DMOS); and (NR models + PSNR on MOS)

Statistical Equivalence to Top Performing Model using Outlier Ratio

"1" indicates that this model is statistically equivalent to the top performing model.
"0" indicates that this model is statistically worse than the top performing model.

	FR Models					RR Models			NR Models		
Test	Psy_FR	Opt_FR	Yon_FR	NTT_FR	PSNR_DMOS	Yon_RR10k	Yon_RR64k	PSNR_DMOS	Psy_NR	Swi_NR	PSNR_MOS
C01	1	1	1	1	0	1	1	0	0	0	1
C02	1	1	0	1	0	1	1	1	1	0	1
C03	1	0	1	1	0	1	1	0	0	0	1
C04	1	1	1	1	0	1	1	1	0	0	1
C05	1	1	1	0	0	1	1	0	0	0	1
C06	1	1	1	1	0	1	1	0	0	0	1
C07	1	1	1	1	0	1	1	0	1	1	1
C08	0	1	1	0	0	1	1	0	0	1	1
C09	1	1	1	1	0	1	1	0	1	1	1
C10	1	1	1	1	0	1	1	1	0	0	1
C11	1	1	1	0	0	1	1	0	1	0	1
C12	1	1	1	1	1	1	1	1	0	0	1
C13	1	1	0	1	0	1	1	1	0	0	1
C14	0	1	0	0	0	1	1	0	0	0	1
Total=	12	13	11	10	1	14	14	5	4	3	14

9.2.4 CIF Statistical Significance Using Correlation

Separate results for each model type: (FR models + PSNR on DMOS); (RR models + PSNR on DMOS); and (NR models + PSNR on MOS)

Statistical Equivalence to Top Performing Model using Correlation

"1" indicates that this model is statistically equivalent to the top performing model.
"0" indicates that this model is statistically worse than the top performing model.

	FR Models					RR Models			NR Models		
Test	Psy_FR	Opt_FR	Yon_FR	NTT_FR	PSNR_DMOS	Yon_RR10k	Yon_RR64k	PSNR_DMOS	Psy_NR	Swi_NR	PSNR_MOS
C01	1	1	1	0	0	1	1	0	0	0	1
C02	1	1	0	1	0	1	1	1	1	0	1
C03	1	0	0	1	0	1	1	1	0	0	1
C04	1	1	1	1	0	1	1	1	0	0	1
C05	1	1	1	1	0	1	1	0	0	0	1
C06	1	1	1	1	0	1	1	0	0	0	1
C07	1	1	1	1	0	1	1	0	1	1	1
C08	1	1	1	0	0	1	1	0	0	1	1
C09	1	1	1	0	0	1	1	0	1	1	1
C10	1	1	1	1	0	1	1	0	0	0	1
C11	1	1	1	0	0	1	1	0	1	0	1
C12	1	1	1	1	0	1	1	1	0	0	1
C13	1	1	0	0	0	1	1	1	0	0	1
C14	1	1	0	0	0	1	1	0	0	0	1
Total=	14	13	10	8	0	14	14	5	4	3	14

9.2.5 Number of CIF HRCs with Transmission Errors

Note: Official ILG Data Analysis Excel file also contains plots of number of transmission errors against experiment correlation.

		FR Models					RR Models			NR Models		
Test	Error	Psy_FR	Opt_FR	Yon_FR	NTT_FR	PSNR_DMOS	Yon_RR10k	Yon_RR64k	PSNR_DMOS	Psy_NR	Swi_NR	PSNR_MOS
C01	4	0.823	0.792	0.801	0.710	0.698	0.831	0.838	0.698	0.590	0.522	0.759
C02	8	0.856	0.828	0.728	0.838	0.696	0.717	0.718	0.696	0.728	0.650	0.812
C03	8	0.823	0.695	0.722	0.782	0.708	0.789	0.788	0.708	0.281	0.424	0.787
C04	8	0.796	0.810	0.734	0.791	0.698	0.699	0.694	0.698	0.387	0.516	0.801
C05	0	0.890	0.850	0.861	0.836	0.733	0.869	0.869	0.733	0.690	0.435	0.801
C06	4	0.892	0.877	0.874	0.864	0.796	0.904	0.904	0.796	0.717	0.663	0.835
C07	4	0.804	0.753	0.749	0.780	0.440	0.725	0.732	0.440	0.587	0.586	0.686
C08	4	0.826	0.844	0.855	0.728	0.648	0.844	0.849	0.648	0.546	0.658	0.719
C09	4	0.852	0.835	0.821	0.706	0.558	0.786	0.786	0.558	0.616	0.709	0.721
C10	0	0.769	0.737	0.809	0.723	0.639	0.791	0.792	0.639	0.478	0.378	0.737
C11	3	0.792	0.747	0.734	0.675	0.477	0.677	0.682	0.477	0.622	0.536	0.699
C12	10	0.788	0.779	0.748	0.811	0.636	0.734	0.733	0.636	0.586	0.461	0.730
C13	10	0.897	0.848	0.712	0.778	0.689	0.677	0.677	0.689	0.589	0.578	0.761
C14	0	0.891	0.923	0.836	0.853	0.768	0.883	0.886	0.768	0.684	0.617	0.850

9.3 QCIF Primary Data Analysis

9.3.1 QCIF Primary Analysis Metrics & Averages

Correlation

	FR Models					RR Models			NR Models		
Test	Psy_FR	Opt_FR	Yon_FR	NTT_FR	PSNR_DMOS	Yon_RR1k	Yon_RR10k	PSNR_DMOS	Psy_NR	Swi_NR	PSNR_MOS
Q01	0.886	0.787	0.838	0.857	0.656	0.835	0.837	0.656	0.777	0.522	0.716
Q02	0.879	0.776	0.744	0.829	0.675	0.779	0.815	0.675	0.828	0.733	0.771
Q03	0.664	0.754	0.609	0.755	0.703	0.736	0.784	0.703	0.632	0.675	0.751
Q04	0.764	0.813	0.694	0.833	0.648	0.676	0.669	0.648	0.629	0.635	0.821
Q05	0.908	0.902	0.861	0.864	0.783	0.877	0.886	0.783	0.780	0.714	0.824
Q06	0.943	0.896	0.870	0.849	0.734	0.868	0.877	0.734	0.750	0.754	0.793
Q07	0.915	0.873	0.846	0.839	0.675	0.842	0.864	0.675	0.790	0.737	0.677
Q08	0.877	0.858	0.843	0.858	0.540	0.818	0.842	0.540	0.660	0.728	0.639
Q09	0.869	0.881	0.805	0.777	0.561	0.788	0.810	0.561	0.771	0.722	0.700
Q10	0.806	0.889	0.713	0.823	0.752	0.730	0.769	0.752	0.624	0.550	0.814
Q11	0.726	0.724	0.656	0.711	0.555	0.643	0.645	0.555	0.538	0.420	0.664
Q12	0.885	0.894	0.799	0.813	0.680	0.816	0.840	0.680	0.732	0.521	0.798
Q13	0.833	0.834	0.721	0.796	0.587	0.754	0.779	0.587	0.699	0.557	0.717
Q14	0.671	0.890	0.587	0.867	0.715	0.638	0.658	0.715	0.605	0.618	0.799
Average=	0.830	0.841	0.756	0.819	0.662	0.771	0.791	0.662	0.701	0.635	0.749
Minimum	0.664	0.724	0.587	0.711	0.540	0.638	0.645	0.540	0.538	0.420	0.639
Maximum	0.943	0.902	0.870	0.867	0.783	0.877	0.886	0.783	0.828	0.754	0.824

RMSE

	FR Models					RR Models			NR Models		
Test	Psy_FR	Opt_FR	Yon_FR	NTT_FR	PSNR_DMOS	Yon_RR1k	Yon_RR10k	PSNR_DMOS	Psy_NR	Swi_NR	PSNR_MOS
Q01	0.534	0.710	0.627	0.593	0.869	0.633	0.630	0.869	0.707	0.959	0.784
Q02	0.422	0.559	0.592	0.495	0.653	0.556	0.513	0.653	0.556	0.673	0.631
Q03	0.750	0.660	0.796	0.658	0.714	0.680	0.623	0.714	0.804	0.765	0.684
Q04	0.548	0.494	0.611	0.470	0.647	0.625	0.631	0.647	0.780	0.775	0.573
Q05	0.402	0.416	0.489	0.484	0.598	0.462	0.446	0.598	0.642	0.718	0.580
Q06	0.325	0.433	0.481	0.514	0.662	0.485	0.468	0.662	0.696	0.692	0.641
Q07	0.412	0.498	0.544	0.557	0.754	0.552	0.516	0.754	0.624	0.687	0.748
Q08	0.463	0.496	0.519	0.496	0.812	0.555	0.521	0.812	0.779	0.711	0.797
Q09	0.481	0.459	0.576	0.611	0.804	0.598	0.569	0.804	0.659	0.716	0.739
Q10	0.558	0.432	0.662	0.535	0.622	0.645	0.603	0.622	0.795	0.850	0.590
Q11	0.658	0.660	0.722	0.673	0.796	0.733	0.731	0.796	0.855	0.921	0.759
Q12	0.441	0.425	0.569	0.551	0.694	0.547	0.514	0.694	0.750	0.940	0.664
Q13	0.577	0.576	0.723	0.631	0.845	0.686	0.655	0.845	0.838	0.973	0.817
Q14	0.671	0.413	0.732	0.451	0.632	0.696	0.681	0.632	0.812	0.801	0.613
Average=	0.517	0.516	0.617	0.551	0.721	0.604	0.578	0.721	0.735	0.799	0.687
Minimum	0.325	0.413	0.481	0.451	0.598	0.462	0.446	0.598	0.556	0.673	0.573
Maximum	0.750	0.710	0.796	0.673	0.869	0.733	0.731	0.869	0.855	0.973	0.817

Outlier Ratio

	FR Models					RR Models			NR Models		
Test	Psy_FR	Opt_FR	Yon_FR	NTT_FR	PSNR_DMOS	Yon_RR1k	Yon_RR10k	PSNR_DMOS	Psy_NR	Swi_NR	PSNR_MOS
Q01	0.461	0.612	0.513	0.493	0.678	0.559	0.533	0.678	0.687	0.801	0.699
Q02	0.408	0.507	0.507	0.507	0.612	0.533	0.500	0.612	0.530	0.687	0.608
Q03	0.612	0.500	0.651	0.513	0.546	0.434	0.428	0.546	0.687	0.693	0.584
Q04	0.539	0.454	0.513	0.428	0.559	0.592	0.566	0.559	0.729	0.747	0.584
Q05	0.349	0.349	0.382	0.408	0.553	0.349	0.276	0.553	0.657	0.711	0.627
Q06	0.243	0.395	0.289	0.461	0.553	0.382	0.342	0.553	0.608	0.657	0.602
Q07	0.368	0.414	0.414	0.487	0.579	0.428	0.428	0.579	0.675	0.657	0.699
Q08	0.388	0.467	0.447	0.428	0.671	0.447	0.441	0.671	0.741	0.639	0.687
Q09	0.408	0.342	0.507	0.579	0.678	0.467	0.480	0.678	0.651	0.681	0.663
Q10	0.480	0.408	0.632	0.474	0.493	0.618	0.605	0.493	0.651	0.699	0.560
Q11	0.572	0.684	0.546	0.618	0.586	0.546	0.526	0.586	0.783	0.711	0.614
Q12	0.480	0.382	0.559	0.546	0.599	0.487	0.493	0.599	0.675	0.747	0.620
Q13	0.520	0.553	0.724	0.579	0.750	0.625	0.592	0.750	0.753	0.831	0.747
Q14	0.579	0.395	0.632	0.434	0.487	0.605	0.592	0.487	0.729	0.651	0.578
Average=	0.458	0.461	0.523	0.497	0.596	0.505	0.486	0.596	0.682	0.708	0.634
Minimum	0.243	0.342	0.289	0.408	0.487	0.349	0.276	0.487	0.530	0.639	0.560
Maximum	0.612	0.684	0.724	0.618	0.750	0.625	0.605	0.750	0.783	0.831	0.747

9.3.2 QCIF Statistical Significance Using RMSE

Separate results for each model type: (FR models + PSNR on DMOS); (RR models + PSNR on DMOS); and (NR models + PSNR on MOS)

Statistical Equivalence to Top Performing Model

"1" indicates that this model is statistically equivalent to the top performing model.
"0" indicates that this model is statistically worse than the top performing model.

	FR Models					RR Models			NR Models		
Test	Psy_FR	Opt_FR	Yon_FR	NTT_FR	PSNR_DMOS	Yon_RR1k	Yon_RR10k	PSNR_DMOS	Psy_NR	Swi_NR	PSNR_MOS
Q01	1	0	0	1	0	1	1	0	1	0	1
Q02	1	0	0	0	0	1	1	0	1	0	1
Q03	1	1	0	1	1	1	1	0	0	1	1
Q04	0	1	0	1	0	1	1	1	0	0	1
Q05	1	1	0	0	0	1	1	0	1	0	1
Q06	1	0	0	0	0	1	1	0	1	1	1
Q07	1	0	0	0	0	1	1	0	1	1	0
Q08	1	1	1	1	0	1	1	0	1	1	1
Q09	1	1	0	0	0	1	1	0	1	1	1
Q10	0	1	0	0	0	1	1	1	0	0	1
Q11	1	1	1	1	0	1	1	1	1	0	1
Q12	1	1	0	0	0	1	1	0	1	0	1
Q13	1	1	0	1	0	1	1	0	1	0	1
Q14	0	1	0	1	0	1	1	1	0	0	1
Total=	11	10	2	7	1	14	14	4	10	5	13

Statistically Better than PSNR

"1" indicates that this model is statistically better than PSNR
"0" indicates that this model is not statistically better than PSNR

	FR Models				RR Models		NR Models	
Test	Psy_FR	Opt_FR	Yon_FR	NTT_FR	Yon_RR1k	Yon_RR10k	Psy_NR	Swi_NR
Q01	1	1	1	1	1	1	0	0
Q02	1	1	0	1	1	1	0	0
Q03	0	0	0	0	0	1	0	0
Q04	1	1	0	1	0	0	0	0
Q05	1	1	1	1	1	1	0	0
Q06	1	1	1	1	1	1	0	0
Q07	1	1	1	1	1	1	1	0
Q08	1	1	1	1	1	1	0	0
Q09	1	1	1	1	1	1	0	0
Q10	0	1	0	1	0	0	0	0
Q11	1	1	0	1	0	0	0	0
Q12	1	1	1	1	1	1	0	0
Q13	1	1	1	1	1	1	0	0
Q14	0	1	0	1	0	0	0	0
Total=	11	13	8	13	9	10	1	0

9.3.3 QCIF Statistical Significance Using Outlier Ratio

Separate results for each model type: (FR models + PSNR on DMOS); (RR models + PSNR on DMOS); and (NR models + PSNR on MOS)

Statistical Equivalence to Top Performing Model using Outlier Ratio

"1" indicates that this model is statistically equivalent to the top performing model.
"0" indicates that this model is statistically worse than the top performing model.

	FR Models					RR Models			NR Models	
Test	Psy_FR	Opt_FR	Yon_FR	NTT_FR	PSNR_DMOS	Yon_RR1k	Yon_RR10k	PSNR_DMOS	Psy_NR	Swi_NR
Q01	1	0	1	1	0	1	1	0	1	0
Q02	1	1	1	1	0	1	1	0	1	0
Q03	0	1	0	1	1	1	1	0	1	1
Q04	1	1	1	1	0	1	1	1	1	1
Q05	1	1	1	1	0	1	1	0	1	1
Q06	1	0	1	0	0	1	1	0	1	1
Q07	1	1	1	0	0	1	1	0	1	1
Q08	1	1	1	1	0	1	1	0	1	1
Q09	1	1	0	0	0	1	1	0	1	1
Q10	1	1	0	1	1	0	0	1	1	1
Q11	1	0	1	1	1	1	1	1	1	1
Q12	1	1	0	0	0	1	1	0	1	1
Q13	1	1	0	1	0	1	1	0	1	1
Q14	0	1	0	1	1	0	1	1	1	1
Total=	12	11	8	10	4	12	13	4	14	12

Note: Comparison including PSNR_MOS not available.

9.3.4 QCIF Statistical Significance Using Correlation

Separate results for each model type: (FR models + PSNR on DMOS); (RR models + PSNR on DMOS); and (NR models + PSNR on MOS)

Statistical Equivalence to Top Performing Model using Correlation

"1" indicates that this model is statistically equivalent to the top performing model.
"0" indicates that this model is statistically worse than the top performing model.

	FR Models					RR Models			NR Models		
Test	Psy_FR	Opt_FR	Yon_FR	NTT_FR	PSNR_DMOS	Yon_RR1k	Yon_RR10k	PSNR_DMOS	Psy_NR	Swi_NR	PSNR_MOS
Q01	1	0	1	1	0	1	1	0	1	0	1
Q02	1	0	0	1	0	1	1	0	1	0	1
Q03	1	1	0	1	1	1	1	1	0	1	1
Q04	1	1	0	1	0	1	1	1	0	0	1
Q05	1	1	1	1	0	1	1	0	1	0	1
Q06	1	0	0	0	0	1	1	0	1	1	1
Q07	1	1	0	0	0	1	1	0	1	1	0
Q08	1	1	1	1	0	1	1	0	1	1	1
Q09	1	1	0	0	0	1	1	0	1	1	1
Q10	0	1	0	0	0	1	1	1	0	0	1
Q11	1	1	1	1	0	1	1	1	1	0	1
Q12	1	1	0	0	0	1	1	0	1	0	1
Q13	1	1	0	1	0	1	1	0	1	0	1
Q14	0	1	0	1	0	1	1	1	0	0	1
Total=	12	11	4	9	1	14	14	5	10	5	13

9.3.5 Number of QCIF HRCs With Transmission Errors

Note: Official ILG Data Analysis Excel file also contains plots of number of transmission errors against experiment correlation.

		FR Models					RR Models			NR Models		
Test	Error	Psy_FR	Opt_FR	Yon_FR	NTT_FR	PSNR_DMOS	Yon_RR1k	Yon_RR10k	PSNR_DMOS	Psy_NR	Swi_NR	PSNR_MOS
Q01	3	0.886	0.787	0.838	0.857	0.656	0.835	0.837	0.656	0.777	0.522	0.716
Q02	8	0.879	0.776	0.744	0.829	0.675	0.779	0.815	0.675	0.828	0.733	0.771
Q03	8	0.664	0.754	0.609	0.755	0.703	0.736	0.784	0.703	0.632	0.675	0.751
Q04	8	0.764	0.813	0.694	0.833	0.648	0.676	0.669	0.648	0.629	0.635	0.821
Q05	4	0.908	0.902	0.861	0.864	0.783	0.877	0.886	0.783	0.780	0.714	0.824
Q06	0	0.943	0.896	0.870	0.849	0.734	0.868	0.877	0.734	0.750	0.754	0.793
Q07	4	0.915	0.873	0.846	0.839	0.675	0.842	0.864	0.675	0.790	0.737	0.677
Q08	4	0.877	0.858	0.843	0.858	0.540	0.818	0.842	0.540	0.660	0.728	0.639
Q09	4	0.869	0.881	0.805	0.777	0.561	0.788	0.810	0.561	0.771	0.722	0.700
Q10		0.806	0.889	0.713	0.823	0.752	0.730	0.769	0.752	0.624	0.550	0.814
Q11	10	0.726	0.724	0.656	0.711	0.555	0.643	0.645	0.555	0.538	0.420	0.664
Q12	12	0.885	0.894	0.799	0.813	0.680	0.816	0.840	0.680	0.732	0.521	0.798
Q13	4	0.833	0.834	0.721	0.796	0.587	0.754	0.779	0.587	0.699	0.557	0.717
Q14		0.671	0.890	0.587	0.867	0.715	0.638	0.658	0.715	0.605	0.618	0.799

10 SECONDARY DATA ANALYSIS

10.1 Explanation and Warnings

10.1.1 Procedure and Purpose of this Analysis

This secondary analysis was performed by averaging the mapped model output values per experiment and per HRC. The mapped values were calculated using the coefficients from the primary analysis. The purpose of this analysis is to show in how far a model can be used to evaluate a system under test if the only variable is the content which must be controlled by the experimenter. This closely resembles the applications of codec validation and system fine tuning.

10.1.2 Remarks for this Analysis

Averaging per HRC has mainly two effects:

1. It is clear, that all models will gain from this averaging process since the "measurement noise" will be reduced. This effect is typically in the range of a 0.1 better correlation compared to the primary analysis.
2. The averaging per HRC eliminates the SRC dependency from both the model outputs as well as the subjective data. It is therefore expected, that models which are unable to properly predict the differences between SRCs will gain excessively from this step.

10.1.3 Validity of the Secondary Analysis

It is important to note that results of this analysis are only valid for

- the averaging of the scores for a well balanced set of SRCs and
- for averaging within one HRC. If eight random sequences were averaged instead of those from the same HRC, the results would be completely different (significantly worse and depending on the random selection).

These two requirements must be kept in mind when choosing a perceptual model for a specific application, based on the performance of the model in this secondary analysis.

For codec tuning and validation, it is easy to meet these requirements since typically full control over the entire system under test is granted.

The situation is however different for monitoring applications, where the regular programme material must be used for the measurement. In this case typically both requirements are violated, since it is generally neither possible to ensure balanced content per HRC, nor is it possible to ensure that all recordings were made using the same HRC. The HRC is defined by the entire signal processing between the very high quality SRC and the final PVS. It includes various compression steps, postprocessing, filtering, potential transmission errors, error concealment etc. The HRC will typically be the same for the duration of one video clip or movie, but, as soon as the next clip/movie starts, any component which forms part of the HRC will most likely change and thus the HRC is not the same anymore, although the codec settings used for the transmission may have remained unchanged. For mobile applications this is even worse since moving the receiver to a different location may also lead to a changed HRC as well.

Please note, that MPEG defines the decoders only. Two different encoders using identical settings may produce streams of very different video quality. These form different HRCs.

Due to the averaging of eight scores per HRC, only very few data points are left for analysis (16 for FR and 17 for NR models).

10.2 Official ILG Secondary Data Analysis

Secondary data analysis is calculated on a per-HRC basis, where the per-clip fitted data is averaged. The common set is not included in the secondary data analysis, because most common set HRCs are available for only 1 scene.

10.2.1 VGA Secondary Data Analysis Metrics and Averages

Correlation

	FR Models					RR Models				NR Models		
Test	Psy_FR	Opt_FR	Yon_FR	NTT_FR	PSNR_DMOS	Yon_RR10k	Yon_RR64k	Yon_RR128k	PSNR_DMOS	Psy_NR	Swi_NR	PSNR_MOS
V01	0.958	0.935	0.867	0.976	0.933	0.925	0.923	0.923	0.933	0.819	0.701	0.940
V02	0.676	0.960	0.919	0.909	0.698	0.884	0.883	0.884	0.698	0.554	0.891	0.665
V03	0.907	0.958	0.591	0.972	0.910	0.882	0.881	0.883	0.910	0.805	0.763	0.924
V04	0.976	0.937	0.886	0.955	0.863	0.839	0.835	0.837	0.863	0.868	0.830	0.844
V05	0.969	0.993	0.984	0.982	0.948	0.982	0.982	0.982	0.948	0.892	0.982	0.946
V06	0.992	0.971	0.982	0.950	0.938	0.982	0.982	0.982	0.938	0.812	0.934	0.918
V07	0.742	0.934	0.867	0.820	0.723	0.914	0.908	0.911	0.723	0.715	0.732	0.805
V08	0.907	0.547	0.883	0.921	0.743	0.938	0.935	0.931	0.743	0.599	0.851	0.792
V09	0.863	0.867	0.807	0.663	0.720	0.780	0.769	0.768	0.720	-0.002	0.718	0.827
V10	0.932	0.960	0.869	0.825	0.740	0.745	0.741	0.743	0.740	0.809	0.722	0.760
V11	0.975	0.965	0.877	0.921	0.876	0.922	0.918	0.917	0.876	0.846	0.869	0.690
V12	0.907	0.907	0.780	0.761	0.709	0.710	0.700	0.693	0.709	0.526	0.548	0.670
V13	0.929	0.955	0.924	0.934	0.718	0.811	0.814	0.814	0.718	0.825	0.717	0.567
Average=	0.903	0.914	0.864	0.891	0.809	0.870	0.867	0.867	0.809	0.698	0.789	0.796
Minimum	0.676	0.547	0.591	0.663	0.698	0.710	0.700	0.693	0.698	-0.002	0.548	0.567
Maximum	0.992	0.993	0.984	0.982	0.948	0.982	0.982	0.982	0.948	0.892	0.982	0.946

RMSE

Note: the scene averaging process may introduce a gain and shift which may impact RMSE (i.e., higher values than expected).
Note: a linear fit is not used to remove gain and level bais, due to the impact of the reduced degrees of freedom on RMSE.

	FR Models					RR Models				NR Models		
Test	Psy_FR	Opt_FR	Yon_FR	NTT_FR	PSNR_DMOS	Yon_RR10k	Yon_RR64k	Yon_RR128k	PSNR_DMOS	Psy_NR	Swi_NR	PSNR_MOS
V01	0.27	0.40	0.47	0.22	0.38	0.347	0.350	0.350	0.379	0.589	0.776	0.459
V02	0.59	0.23	0.37	0.34	0.57	0.367	0.368	0.366	0.572	0.744	0.494	0.961
V03	0.32	0.23	0.58	0.19	0.31	0.347	0.345	0.343	0.309	0.511	0.634	0.757
V04	0.30	0.30	0.38	0.26	0.44	0.435	0.439	0.437	0.441	0.453	0.765	0.795
V05	0.24	0.13	0.21	0.17	0.36	0.188	0.189	0.187	0.358	0.441	0.585	0.656
V06	0.14	0.22	0.21	0.27	0.38	0.179	0.179	0.177	0.384	0.546	0.692	0.897
V07	0.39	0.22	0.31	0.35	0.40	0.235	0.243	0.238	0.400	0.494	0.610	0.844
V08	0.32	0.59	0.37	0.29	0.46	0.293	0.296	0.301	0.461	0.657	0.659	0.829
V09	0.33	0.33	0.39	0.49	0.48	0.432	0.437	0.438	0.480	0.860	0.644	0.584
V10	0.23	0.27	0.32	0.44	0.42	0.433	0.436	0.435	0.425	0.446	0.596	0.730
V11	0.18	0.21	0.38	0.40	0.38	0.316	0.324	0.326	0.378	0.544	0.738	1.570
V12	0.39	0.43	0.58	0.57	0.70	0.642	0.646	0.650	0.704	0.854	0.869	0.855
V13	0.30	0.32	0.34	0.33	0.72	0.509	0.505	0.504	0.715	0.560	0.650	0.734
Average=	0.308	0.298	0.378	0.332	0.462	0.363	0.366	0.366	0.462	0.592	0.670	0.821
Minimum	0.138	0.128	0.208	0.170	0.309	0.179	0.179	0.177	0.309	0.441	0.494	0.459
Maximum	0.586	0.589	0.584	0.569	0.715	0.642	0.646	0.650	0.715	0.860	0.869	1.570

Outlier Ratio

Note: averaging produces 24*8 viewers per sample, resulting in worse Outlier Ratio values for HRC analysis when compared to primary analysis
Note: a linear fit is not used to remove gain and level bais.

	FR Models					RR Models				NR Models		
Test	Psy_FR	Opt_FR	Yon_FR	NTT_FR	PSNR_DMOS	Yon_RR10k	Yon_RR64k	Yon_RR128k	PSNR_DMOS	Psy_NR	Swi_NR	PSNR_MOS
V01	0.625	0.938	0.688	0.625	0.875	0.813	0.875	0.875	0.875	0.941	0.765	0.941
V02	0.813	0.688	0.750	0.875	0.813	0.688	0.750	0.750	0.813	0.941	0.882	0.941
V03	0.563	0.438	0.938	0.500	0.750	0.625	0.625	0.625	0.750	0.882	0.941	0.941
V04	0.625	0.563	0.688	0.625	0.688	0.750	0.688	0.688	0.688	0.647	0.941	0.882
V05	0.688	0.500	0.625	0.375	0.688	0.688	0.688	0.688	0.688	0.882	0.941	1.000
V06	0.438	0.875	0.688	0.688	0.813	0.375	0.438	0.438	0.813	0.882	1.000	1.000
V07	0.688	0.750	0.563	0.438	0.875	0.625	0.625	0.625	0.875	0.824	0.882	0.882
V08	0.692	0.692	0.692	0.769	1.000	0.692	0.692	0.692	1.000	0.714	0.786	0.857
V09	0.750	0.688	0.813	0.813	0.938	0.875	0.875	0.875	0.938	0.882	0.941	0.882
V10	0.688	0.688	0.750	0.688	0.813	0.813	0.813	0.813	0.813	0.765	0.882	0.765
V11	0.375	0.625	0.750	0.563	0.688	0.625	0.625	0.625	0.688	1.000	0.882	1.000
V12	1.000	0.625	0.750	0.750	0.688	0.813	0.813	0.813	0.688	0.882	0.882	0.941
V13	0.750	0.813	0.813	0.875	0.875	1.000	1.000	1.000	0.875	0.941	0.941	1.000
Average=	0.669	0.683	0.731	0.660	0.808	0.722	0.731	0.731	0.808	0.860	0.898	0.926
Minimum	0.375	0.438	0.563	0.375	0.688	0.375	0.438	0.438	0.688	0.647	0.765	0.765
Maximum	1.000	0.938	0.938	0.875	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

10.2.2 CIF Secondary Data Analysis Metrics and Averages

Correlation

	FR Models					RR Models			NR Models		
Test	Psy_FR	Opt_FR	Yon_FR	NTT_FR	PSNR_DMOS	Yon_RR10k	Yon_RR64k	PSNR_DMOS	Psy_NR	Swi_NR	PSNR_MOS
C01	0.872	0.921	0.933	0.940	0.825	0.923	0.929	0.825	0.918	0.800	0.756
C02	0.978	0.939	0.884	0.937	0.807	0.752	0.748	0.807	0.941	0.826	0.720
C03	0.912	0.821	0.904	0.879	0.776	0.907	0.903	0.776	0.282	0.572	0.831
C04	0.925	0.950	0.814	0.947	0.825	0.740	0.741	0.825	0.747	0.786	0.840
C05	0.941	0.977	0.969	0.951	0.872	0.933	0.931	0.872	0.895	0.947	0.684
C06	0.960	0.965	0.978	0.969	0.931	0.972	0.971	0.931	0.902	0.951	0.940
C07	0.911	0.922	0.910	0.918	0.909	0.886	0.907	0.909	0.821	0.905	0.666
C08	0.884	0.940	0.941	0.933	0.896	0.947	0.945	0.896	0.858	0.907	0.655
C09	0.924	0.919	0.936	0.893	0.818	0.925	0.925	0.818	0.855	0.925	0.876
C10	0.719	0.778	0.921	0.783	0.658	0.871	0.872	0.658	0.766	0.674	0.683
C11	0.920	0.879	0.855	0.797	0.638	0.740	0.742	0.638	0.800	0.602	0.546
C12	0.905	0.910	0.863	0.956	0.870	0.879	0.872	0.870	0.882	0.778	0.608
C13	0.953	0.961	0.607	0.935	0.677	0.500	0.496	0.677	0.808	0.694	0.708
C14	0.982	0.990	0.981	0.969	0.939	0.966	0.968	0.939	0.930	0.837	0.770
Average=	0.913	0.919	0.892	0.915	0.817	0.853	0.854	0.817	0.815	0.800	0.735
Minimum	0.719	0.778	0.607	0.783	0.638	0.500	0.496	0.638	0.282	0.572	0.546
Maximum	0.982	0.990	0.981	0.969	0.939	0.972	0.971	0.939	0.941	0.951	0.940

RMSE

Note: the scene averaging process may introduce a gain and shift which may impact RMSE (i.e., higher values than expected).
Note: a linear fit is not used to remove gain and level bais, due to the impact of the reduced degrees of freedom on RMSE.

	FR Models					RR Models			NR Models		
Test	Psy_FR	Opt_FR	Yon_FR	NTT_FR	PSNR_DMOS	Yon_RR10k	Yon_RR64k	PSNR_DMOS	Psy_NR	Swi_NR	PSNR_MOS
C01	0.442	0.402	0.448	0.408	0.576	0.378	0.364	0.576	0.508	0.694	0.786
C02	0.200	0.277	0.394	0.256	0.457	0.471	0.474	0.457	0.357	0.561	1.576
C03	0.280	0.365	0.336	0.310	0.428	0.293	0.298	0.428	0.686	0.619	0.770
C04	0.289	0.252	0.412	0.254	0.404	0.453	0.453	0.404	0.610	0.585	0.764
C05	0.321	0.284	0.326	0.308	0.530	0.344	0.349	0.530	0.476	0.796	1.243
C06	0.315	0.311	0.320	0.286	0.467	0.254	0.259	0.467	0.507	0.618	0.804
C07	0.263	0.283	0.264	0.241	0.484	0.279	0.252	0.484	0.389	0.422	0.741
C08	0.298	0.252	0.226	0.260	0.452	0.212	0.217	0.452	0.439	0.444	1.211
C09	0.234	0.254	0.243	0.294	0.485	0.249	0.247	0.485	0.413	0.412	0.561
C10	0.530	0.513	0.354	0.489	0.587	0.389	0.389	0.587	0.625	0.716	0.971
C11	0.352	0.446	0.479	0.561	0.742	0.597	0.595	0.742	0.629	0.801	2.453
C12	0.290	0.293	0.372	0.239	0.458	0.361	0.363	0.458	0.428	0.611	1.647
C13	0.273	0.319	0.665	0.452	0.631	0.715	0.717	0.631	0.601	0.687	0.815
C14	0.223	0.151	0.269	0.253	0.435	0.242	0.235	0.435	0.497	0.716	1.380
Average=	0.308	0.314	0.365	0.329	0.510	0.374	0.372	0.510	0.512	0.620	1.123
Minimum	0.200	0.151	0.226	0.239	0.404	0.212	0.217	0.404	0.357	0.412	0.561
Maximum	0.530	0.513	0.665	0.561	0.742	0.715	0.717	0.742	0.686	0.801	2.453

Outlier Ratio

Note: averaging produces 24*8 viewers per sample, resulting in worse Outlier Ratio values for HRC analysis when compared to primary analysis
Note: a linear fit is not used to remove gain and level bais.

	FR Models					RR Models			NR Models		
Test	Psy_FR	Opt_FR	Yon_FR	NTT_FR	PSNR_DMOS	Yon_RR10k	Yon_RR64k	PSNR_DMOS	Psy_NR	Swi_NR	PSNR_MOS
C01	1.000	1.000	0.813	0.813	0.938	0.938	0.938	0.938	0.941	0.941	0.941
C02	0.438	0.625	0.750	0.563	0.813	0.750	0.688	0.813	0.882	0.706	0.765
C03	0.625	0.688	0.750	0.625	0.875	0.625	0.625	0.875	0.941	0.824	0.882
C04	0.750	0.688	0.688	0.500	0.750	0.875	0.813	0.750	0.824	0.824	0.882
C05	0.750	0.750	0.750	0.625	0.813	0.688	0.688	0.813	0.706	0.941	0.824
C06	0.813	0.813	0.813	0.688	0.938	0.500	0.500	0.938	0.882	0.941	0.941
C07	0.563	0.688	0.938	0.500	0.875	0.750	0.750	0.875	0.765	0.941	0.765
C08	0.625	0.688	0.625	0.625	0.875	0.438	0.563	0.875	0.882	0.824	0.765
C09	0.563	0.563	0.688	0.625	0.813	0.875	0.875	0.813	0.824	0.765	0.706
C10	0.938	0.938	0.813	0.875	0.875	0.688	0.688	0.875	0.941	0.941	0.941
C11	0.750	0.750	0.813	0.813	0.750	0.875	0.875	0.750	0.941	0.882	0.882
C12	0.563	0.688	0.625	0.438	0.750	0.750	0.750	0.750	0.706	0.882	0.765
C13	0.625	0.750	1.000	0.938	0.938	0.938	0.938	0.938	0.824	0.941	1.000
C14	0.438	0.625	0.750	0.625	0.875	0.750	0.688	0.875	1.000	0.882	0.941
Average=	0.674	0.732	0.772	0.661	0.848	0.746	0.741	0.848	0.861	0.874	0.857
Minimum	0.438	0.563	0.625	0.438	0.750	0.438	0.500	0.750	0.706	0.706	0.706
Maximum	1.000	1.000	1.000	0.938	0.938	0.938	0.938	0.938	1.000	0.941	1.000

10.2.3 QCIF Secondary Data Analysis and Averages

Correlation

	FR Models					RR Models			NR Models		
Test	Psy_FR	Opt_FR	Yon_FR	NTT_FR	PSNR_DMOS	Yon_RR1k	Yon_RR10k	PSNR_DMOS	Psy_NR	Swi_NR	PSNR_MOS
Q01	0.926	0.964	0.962	0.941	0.771	0.964	0.959	0.771	0.962	0.749	0.796
Q02	0.951	0.927	0.911	0.923	0.805	0.867	0.902	0.805	0.964	0.890	0.796
Q03	0.753	0.878	0.709	0.906	0.932	0.909	0.960	0.932	0.897	0.870	0.919
Q04	0.910	0.938	0.863	0.963	0.844	0.820	0.786	0.844	0.767	0.812	0.772
Q05	0.970	0.983	0.988	0.951	0.950	0.983	0.987	0.950	0.939	0.957	0.944
Q06	0.985	0.986	0.990	0.956	0.913	0.975	0.970	0.913	0.938	0.974	0.932
Q07	0.967	0.961	0.982	0.956	0.969	0.983	0.988	0.969	0.951	0.945	0.774
Q08	0.953	0.965	0.983	0.956	0.921	0.987	0.987	0.921	0.933	0.937	0.807
Q09	0.952	0.953	0.978	0.935	0.931	0.973	0.984	0.931	0.954	0.913	0.790
Q10	0.927	0.951	0.785	0.933	0.891	0.704	0.772	0.891	0.913	0.776	0.835
Q11	0.853	0.761	0.781	0.942	0.806	0.662	0.748	0.806	0.888	0.837	0.629
Q12	0.976	0.995	0.994	0.968	0.944	0.987	0.990	0.944	0.910	0.901	0.790
Q13	0.901	0.921	0.840	0.902	0.766	0.891	0.901	0.766	0.840	0.719	0.663
Q14	0.850	0.939	0.741	0.952	0.904	0.800	0.813	0.904	0.855	0.810	0.892
Average=	0.920	0.937	0.893	0.942	0.882	0.893	0.910	0.882	0.908	0.864	0.810
Minimum	0.753	0.761	0.709	0.902	0.766	0.662	0.748	0.766	0.767	0.719	0.629
Maximum	0.985	0.995	0.994	0.968	0.969	0.987	0.990	0.969	0.964	0.974	0.944

RMSE

Note: the scene averaging process may introduce a gain and shift which may impact RMSE (i.e., higher values than expected).
Note: a linear fit is not used to remove gain and level bais, due to the impact of the reduced degrees of freedom on RMSE.

	FR Models					RR Models			NR Models		
Test	Psy_FR	Opt_FR	Yon_FR	NTT_FR	PSNR_DMOS	Yon_RR1k	Yon_RR10k	PSNR_DMOS	Psy_NR	Swi_NR	PSNR_MOS
Q01	0.338	0.238	0.266	0.301	0.588	0.240	0.253	0.588	0.317	0.715	0.761
Q02	0.246	0.336	0.381	0.296	0.498	0.406	0.329	0.498	0.257	0.425	0.748
Q03	0.548	0.417	0.575	0.366	0.418	0.428	0.275	0.418	0.495	0.448	0.588
Q04	0.355	0.276	0.405	0.230	0.439	0.445	0.464	0.439	0.606	0.608	1.122
Q05	0.241	0.184	0.246	0.269	0.375	0.225	0.168	0.375	0.373	0.469	0.755
Q06	0.156	0.190	0.212	0.260	0.447	0.242	0.221	0.447	0.370	0.371	0.636
Q07	0.227	0.265	0.276	0.234	0.565	0.316	0.205	0.565	0.330	0.433	0.755
Q08	0.291	0.275	0.298	0.272	0.683	0.320	0.248	0.683	0.430	0.407	0.719
Q09	0.262	0.248	0.260	0.295	0.603	0.299	0.208	0.603	0.334	0.445	0.749
Q10	0.416	0.232	0.548	0.291	0.361	0.553	0.488	0.361	0.439	0.552	0.617
Q11	0.313	0.365	0.401	0.282	0.454	0.453	0.414	0.454	0.449	0.548	1.417
Q12	0.222	0.121	0.266	0.328	0.450	0.249	0.190	0.450	0.487	0.789	0.956
Q13	0.442	0.403	0.562	0.454	0.698	0.515	0.468	0.698	0.628	0.820	1.325
Q14	0.599	0.261	0.639	0.274	0.382	0.596	0.573	0.382	0.552	0.537	0.569
Average=	0.333	0.272	0.381	0.296	0.497	0.378	0.322	0.497	0.433	0.541	0.837
Minimum	0.156	0.121	0.212	0.230	0.361	0.225	0.168	0.361	0.257	0.371	0.569
Maximum	0.599	0.417	0.639	0.454	0.698	0.596	0.573	0.698	0.628	0.820	1.417

Outlier Ratio

Note: averaging produces 24*8 viewers per sample, resulting in worse Outlier Ratio values for HRC analysis when compared to primary analysis
Note: a linear fit is not used to remove gain and level bais.

	FR Models					RR Models			NR Models		
Test	Psy_FR	Opt_FR	Yon_FR	NTT_FR	PSNR_DMOS	Yon_RR1k	Yon_RR10k	PSNR_DMOS	Psy_NR	Swi_NR	PSNR_MOS
Q01	0.750	0.688	0.750	0.750	0.625	0.625	0.563	0.625	0.706	1.000	1.000
Q02	0.563	0.625	0.813	0.813	0.813	0.750	0.875	0.813	0.706	0.941	0.765
Q03	0.875	0.813	0.875	0.688	0.875	0.875	0.688	0.875	0.941	0.765	0.824
Q04	0.813	0.750	0.813	0.500	0.875	0.875	0.688	0.875	1.000	0.824	0.882
Q05	0.688	0.688	0.750	0.563	0.813	0.625	0.625	0.813	0.706	0.824	0.941
Q06	0.625	0.375	0.625	0.563	0.875	0.688	0.563	0.875	0.765	0.765	0.941
Q07	0.625	0.750	0.813	0.688	0.875	0.875	0.688	0.875	0.765	0.941	0.941
Q08	0.625	0.750	0.813	0.688	0.938	0.813	0.813	0.938	0.765	0.765	0.941
Q09	0.750	0.438	0.750	0.750	1.000	0.875	0.688	1.000	0.824	0.882	0.941
Q10	0.750	0.563	0.813	0.625	0.813	0.938	0.875	0.813	0.706	0.882	0.706
Q11	0.750	0.750	0.688	0.750	0.688	0.688	0.625	0.688	0.765	1.000	0.941
Q12	0.688	0.438	0.563	0.563	0.938	0.625	0.563	0.938	0.882	0.941	0.765
Q13	1.000	1.000	1.000	0.813	0.875	0.813	0.750	0.875	1.000	0.941	0.882
Q14	0.563	0.563	0.750	0.688	0.938	0.813	0.813	0.938	0.647	0.882	0.941
Average=	0.719	0.656	0.772	0.674	0.853	0.777	0.701	0.853	0.798	0.882	0.887
Minimum	0.563	0.375	0.563	0.500	0.625	0.625	0.563	0.625	0.647	0.765	0.706
Maximum	1.000	1.000	1.000	0.813	1.000	0.938	0.875	1.000	1.000	1.000	1.000

11 CONCLUSIONS

The data analysis in its entirety having been presented and discussed previously, this section focuses on what went well with testing, and lessons learned for future testing. See the Executive Summary for a summarized interpretation of the MM Phase I results.

The MM experiments successfully evaluated a very large number of video sequences, with the assistance of both proponents and ILG. The high lab-to-lab correlations on the common video sequences provide strong evidence that all of the MM Phase I subjective experiments were conducted in the approved manner, and that each MM data set contains unbiased and non-discriminatory subjective scores. VQEG has a high level of confidence in the execution of the subjective testing. This confidence applies to both tests performed by proponents and tests performed by ILG. The common set of sequences was a valuable aspect of the testing.

Three aspects of the testing could perhaps have been improved. First, there was an extended delay between model submission and completion of data analysis. Some of the delay resulted from problems coordinating a large number of laboratories through a series of deadlines (i.e., events where data must pass from one organization to another before work could continue). Second, the distribution of HRCs with respect to impairments was an uncontrolled variable in the MM Phase I testing. This led to some imbalances that complicate interpretation of the results (e.g., coding algorithms that are only associated with one HRC; or a coding algorithm that was tested extensively with coding-only but never with transmission errors). Third, the calibration limits led to unexpected problems (e.g., ambiguities on whether specific frame-delay patterns were valid, how to check calibration values, and the extended time required for these validation checks.)

Despite these small problems, the MM Phase I test was a huge success. Forty-one subjective tests provide the largest data set of its kind ever produced. The algorithms validated in this test can be assumed to have been tested more extensively than any other video quality algorithm.

12 REFERENCES

[1] J. Jonsson and K. Brunnström, "Getting Started With ArcVQWin", acr022250, Acreo AB, Kista, Sweden , (2007).

[2] M. Spiegel, “Theory and problems of statistics”, McGraw Hill, 1998.

Appendix I Model Descriptions

Appendix I.1 Proponent A, NTT

The NTT model (MoSQuE 1.0) calculates subjective assessment values accurately using a precise alignment process and a video quality algorithm reflects human visual characteristics in order to consider the influence of codecs, bit-rate, frame-rate, and video quality distorted by packet loss. The alignment process is divided into the macro alignment process and the micro alignment process. The macro alignment process filters the video sequences to consider the influence of video capturing and post-processing of the decoder and matches pixels between reference video sequences and processed video sequences in the spatial temporal directions. The micro alignment process matches frames between reference video sequences and processed video sequences to consider the influence of video frame skipping and freezing after the macro alignment process has finished.

The video quality algorithm calculates the objective video quality that reflects human visual characteristics by using (i) a spatial degradation parameter based on four parameters, which reflect the presence of overall noise, spurious edges, localized motion distortion, and localized spatial distortions caused by packet loss, respectively, and (ii) a temporal degradation parameter, which reflects frame-rate freezing and variation.

Appendix I.2 Proponent B, OPTICOM

PEVQ is a very robust model which is designed to predict the effects of transmission impairments on the video quality as perceived by a human subject. Its main targets are mobile applications and IPTV. PEVQ is built on PVQM, a TV quality measure developed by John Beerends and Andries Hekstra from KPN. The key features of PEVQ are:

- (fast and reliable) temporal alignment of the input sequences based on multi dimensional feature correlation analysis with limits that reach far beyond those tested by VQEG, especially with regard to the amount of time clipping, frame freezing and frame skipping which can be handled.
- Full frame spatial alignment
- Color alignment algorithm based on cumulative histograms
- Enhanced framerate estimation and rating
- Detection and perceptually correct weighting of frame freezes and frame skips.
- Only four indicators are used to detect the video quality. Those indicators operate in different domains (temporal, spatial, chrominance) and are motivated by the Human Visual System. Perceptual masking properties of the HVS are modelled at several stages of the algorithm. These indicators are integrated using a sophisticated spatial and temporal integration algorithm.

In its first stage the algorithm all the alignment steps are performed and information frozen or

skipped frames is collected. In the second step the now synchronized and equalized images are compared for visual differences in the luminance as well as in the chrominance domain, taking masking effects and motion into account. This results in a set of indicators which all describe certain quality aspects. The last step is finally the integration of the indicators by non-linear functions in order to derive the final MOS.

Due to the low number of indicators and the resulting low degree of freedom the model can hardly be over trained and is very robust. PEVQ can be efficiently implemented without sacrificing the prediction accuracy and is already widely used in the market.

Appendix I.3 Proponent C, Psytechnics

Description of the Psytechnics FR model

The Psytechnics' full-reference video model is an objective measurement algorithm that predicts overall subjective video quality on a scale from 1 to 5, with 1 representing the worst quality (or highest quality difference between reference and processed videos) and 5 representing the best quality (or lowest quality difference between reference and processed videos).

The model first spatio-temporally registers the reference and processed videos. For each frame of the processed video, the alignment procedure identifies the temporally matching frame in the reference video with its associated spatial shifting. The alignment procedure is designed to cope with time-varying spatial and temporal misalignment between reference and processed videos. Each pair of reference-processed frames is then processed by several modules producing parameters relevant to the perceptual spatial quality, which can be affected for example by digital compression and transmission errors. Additional parameters relevant to the perceptual temporal quality of the video, which can be affected for example by frame freezing, are also extracted from the alignment procedure. All computed parameters are then pooled together in an integration function that produces an overall quality prediction for the processed video.

The model was submitted to the VQEG Multimedia Test as a command line executable. The Psytechnics' video model was designed to be fast enough to provide a practical tool to the industry. Although a single-threaded version of the software was submitted to the VQEG Multimedia Test, a multi-threaded version of the software is now available and can produce the quality prediction score faster than real-time, even for VGA resolution. For example, processing of a pair (source/processed) of 8-sec videos takes about 2.2 seconds (QCIF) , 2.7 seconds (CIF), and 5.5 seconds (VGA) on a PC with dual-core 3 GHz CPU and hard disk in RAID 0 configuration. These durations include the time spent on file reading from disk.

Description of the Psytechnics NR model

The Psytechnics' no-reference video model is an objective measurement algorithm that predicts overall subjective video quality on a scale from 1 to 5, with 1 representing the worst quality and 5 representing the best quality.

In the no-reference video model, each video frame is processed through several modules producing parameters relevant to the perceptual spatial quality, which can be affected for example by digital compression and transmission errors. The model also computes parameters relevant to the perceptual temporal quality of the video, which can be affected for example by frame freezing. All computed parameters are then pooled together in an integration function that produces an overall quality prediction for the processed video.

The NR model was submitted to the VQEG Multimedia Test as a command line executable. The

code was not optimized in any way and many parameters not used in the calculation of the MOS prediction are computed. Therefore it is difficult to estimate the true speed of the current version of the executable.

Appendix I.4 Proponent D, Yonsei University

The RR models first extract features that represent human perception of degradation from the source video sequence. At the receiver, using these features, a video quality metric is computed. The models are very efficient and can be implemented in real time. The FR models use additional features to obtain improved performance.

Appendix I.5 Proponent E, SwissQual

SwissQual's no-reference model is organized in two stages. The first stage analyses the temporal behaviour with respect to freezing events and calculates a perceptually weighted jerkiness value.

The second stage is focussed on the spatial domain. It detects different typical degradations as usual for compression techniques as well as events classified as un-natural, as for example incoherent motion as a result of package loss.

Since, SwissQual's model is supposed to handle asynchronous captured video sequences by means of analogue devices (such as camera devices) and resulting smearing effects the model calculates indicators are derived after applying a fuzzy analysis in the spatial domain. A set of those quality indicators will be calculated for each frame.

Finally, the individual quality indicators are weighted and aggregated over all frames. The resulting raw value is transformed into a common 1 to 5 scale.

Appendix II Subjective Testing Facilities

Appendix II.1 KDDI

Tests Conducted: KDDI's QCIF and CIF Tests, C01 C02 C05 Q01 Q04 Q06

Display

Display Manufacturer	Samsung
Display Model	940BX
Display Screen Size	19 inch
Display Resolution	1280x1024
Display Scanning Rate	60Hz
Display Pixel Pitch	0.294mm
Display Response Time (Black-White)	5ms
Display Colour Temperature	6500K
Display Bit Depth	8bits/color
Display Type (Standalone / Laptop)	Standalone
Display Label (TCO stamp)	TCO'03

Display Calibration

Calibration Tool	EyeOne
Luminance Value (video display window peak white)	200cd/m2
Luminance Value (background display region)	20cd/m2
Brightness Value	1000:1
Contrast Value	300cd/m2
Gamma Value	2.2

Test Computer

Computer Manufacturer	DELL
Model	Optiplex 745
Processor	Intel Core2Duo E6700 (2.66GHz)
SDRAM	1.0GBytes
HDD	Seagate ST3250820AS (SATA 250GB 7200rpm)
Connection to Display	DVI
Graphics card	ATI Radeon X1300 256MB

Test Environment and Procedure

Viewing Distance	4-8H
Viewing Angle	0°

Visual Acuity Test Method	Landolt Ring Test
Colour Vision Test Method	Ishihara Test
Room illumination (ambient light level [lux])	Low
Background luminance of wall behind the monitor	Low

Appendix II.2 NTT

Tests Conducted: NTT's VGA Tests, V05 V06 V08

Display

Display Manufacturer	EIZO
Display Model	M170
Display Screen Size	17"
Display Resolution	1280 x 1024
Display Scanning Rate	60 Hz
Display Pixel Pitch	0.264 mm
Display Response Time (Black-White)	12 ms
Display Colour Temperature	6500 K
Display Bit Depth	8 bits/colour
Display Type (Standalone / Laptop)	Standalone
Display Label (TCO stamp)	TCO´03

Display Calibration

Calibration Tool	GretagMacbeth Eye-One Monitor
Luminance Value (video display window peak white)	180 cd/m^2
Luminance Value (background display region)	According calibration with EyeOne
Brightness Value	According calibration with EyeOne
Contrast Value	According calibration with EyeOne
Gamma Value	2.2

Test Computer

Computer Manufacturer	HP
Model	XW8400 Workstation
Processor	Intel Xeon 5130 2 GHz

SDRAM	2 GB
HDD	3ware 9650SE RAID0 using three disks (Maxtor MaXLine Pro 500 7H500F0 500GB)
Connection to Display	DVI
Graphics card	Sapphire RADEON X1600 PRO

Test Environment and Procedure

Viewing Distance	6H
Viewing Angle	12.68° × 9.53° for the images
Visual Acuity Test Method	Snellen Type Plastic Eye chart
Colour Vision Test Method	Ishihara´s test for Colour Deficiency Concise Edition 2006
Room illumination (ambient light level [lux])	About 20 lux
Background luminance of wall behind the monitor	5 cd/m^2

Tests Conducted: NTT's CIF Tests, C06 C07 C08

Display

Display Manufacturer	EIZO
Display Model	M170
Display Screen Size	17"
Display Resolution	1280 x 1024
Display Scanning Rate	60 Hz
Display Pixel Pitch	0.264 mm
Display Response Time (Black-White)	12 ms
Display Colour Temperature	6500 K
Display Bit Depth	8 bits/colour
Display Type (Standalone / Laptop)	Standalone
Display Label (TCO stamp)	TCO´03

Display Calibration

Calibration Tool	GretagMacbeth Eye-One Monitor
Luminance Value (video display window peak white)	180 cd/m^2

Luminance Value (background display region)	According calibration with EyeOne
Brightness Value	According calibration with EyeOne
Contrast Value	According calibration with EyeOne
Gamma Value	2.2

Test Computer

Computer Manufacturer	HP
Model	XW8400 Workstation
Processor	Intel Xeon 5130 2 GHz
SDRAM	2 GB
HDD	3ware 9650SE RAID0 using three disks (Maxtor MaXLine Pro 500 7H500F0 500GB)
Connection to Display	DVI
Graphics card	Sapphire RADEON X1600 PRO

Test Environment and Procedure

Viewing Distance	8H
Viewing Angle	8.73° × 7.15° for the images
Visual Acuity Test Method	Snellen Type Plastic Eye chart
Colour Vision Test Method	Ishihara´s test for Colour Deficiency Concise Edition 2006
Room illumination (ambient light level [lux])	About 20 lux
Background luminance of wall behind the monitor	5 cd/m^2

Tests Conducted: NTT's QCIF Tests, Q05 Q07 Q09

Display

Display Manufacturer	EIZO
Display Model	M170
Display Screen Size	17"
Display Resolution	1280 x 1024
Display Scanning Rate	60 Hz
Display Pixel Pitch	0.264 mm
Display Response Time (Black-White)	12 ms

Display Colour Temperature	6500 K
Display Bit Depth	8 bits/colour
Display Type (Standalone / Laptop)	Standalone
Display Label (TCO stamp)	TCO´03

Display Calibration

Calibration Tool	GretagMacbeth Eye-One Monitor
Luminance Value (video display window peak white)	180 cd/m^2
Luminance Value (background display region)	According calibration with EyeOne
Brightness Value	According calibration with EyeOne
Contrast Value	According calibration with EyeOne
Gamma Value	2.2

Test Computer

Computer Manufacturer	HP
Model	XW8400 Workstation
Processor	Intel Xeon 5130 2 GHz
SDRAM	2 GB
HDD	3ware 9650SE RAID0 using three disks (Maxtor MaXLine Pro 500 7H500F0 500GB)
Connection to Display	DVI
Graphics card	Sapphire RADEON X1600 PRO

Test Environment and Procedure

Viewing Distance	8H
Viewing Angle	8.73° × 7.15° for the images
Visual Acuity Test Method	Snellen Type Plastic Eye chart
Colour Vision Test Method	Ishihara´s test for Colour Deficiency Concise Edition 2006
Room illumination (ambient light level [lux])	About 20 lux
Background luminance of wall behind the monitor	5 cd/m^2

Appendix II.3 OPTICOM

Tests Conducted: OPTICOM's VGA Test, V01

Display

Display Manufacturer	Samsung
Display Model	Syncmaster 214T
Display Screen Size	21.3" (53.3cm) diag , 174x130mm (active)
Display Resolution	1600x1200
Display Scanning Rate	60Hz (default ACRVQWIN)
Display Pixel Pitch	0.27mm
Display Response Time (Black-White)	8ms (spec: gray to gray)
Display Colour Temperature	6500K
Display Bit Depth	8 bit
Display Type (Standalone / Laptop)	Standalone
Display Label (TCO stamp)	TCO '03

Display Calibration

Calibration Tool	Eye One Display
Luminance Value (video display window peak white)	120cd/ m^2
Luminance Value (background display region)	18cd/ m^2 (Background 108 in ACRVQWIN), measured 16-20 by Barko TMF6
Brightness Value	120cd/m^2
Contrast Value	Monitor specification max 1000:1 Measured using Barko TMF6: 120/0.33=360
Gamma Value	2.2 (see test plan)

Test Computer

Computer Manufacturer	OEM
Model	OEM
Processor	Intel Core 2 Duo E6400
SDRAM	DDR2 2x1024MB
HDD	Seagate ST3400620NS 400GB (2HDD Raid0)
Connection to Display	DVI
Graphics card	Gainward Bliss GF 7600GS 512MB DDR2

Test Environment and Procedure

Viewing Distance	4-6H (see test plan)
Viewing Angle	Near 0 degree horizontal and vertical
Visual Acuity Test Method	Far (OCU 46016 Optitypeboard letters) Near (Nieden eye chart)
Colour Vision Test Method	Ishihara 38 plates (OCU 47640)
Room illumination (ambient light level [lux])	Low
Background luminance of wall behind the monitor	15-16cd/ m^2 (Measured Barko TMF6)

Appendix II.4 Psytechnics

Tests Conducted: Psytechnics' QCIF/CIF/VGA Tests q01/c01/v01

Display

Display Manufacturer	BenQ
Display Model	FP241WZ
Display Screen Size	24'' (widescreen)
Display Resolution	1920x1200
Display Scanning Rate	60Hz
Display Pixel Pitch	0.270mm
Display Response Time (Black-White)	16ms (6ms GTG) / 12ms MPRT
Display Colour Temperature	6500K
Display Bit Depth	8 bits/colour
Display Type (Standalone / Laptop)	Stand-alone
Display Label (TCO stamp)	TCO 06

Display Calibration

Calibration Tool	Spyder 2 pro
Luminance Value (video display window peak white)	140.6 cd/m2
Luminance Value (background display region)	Grey level set to 108
Brightness Value	50% of maximum value
Contrast Value	Default

Gamma Value	2.2

Test Computer

Computer Manufacturer	DELL
Model	Dimension XPS
Processor	Dual processor Pentium 4 3.4Ghz & 3.4Ghz
SDRAM	2046 MB DDR
HDD	Western Digital 160GB
Connection to Display	DVI
Graphics card	ATI RADEON X800 XT 256MB 400MhZ DAC

Test Environment and Procedure

Viewing Distance	According to MM test plan
Viewing Angle	0 degrees – viewer seated in front of middle of the display
Visual Acuity Test Method	HOTV LogMAR (Logarithmic Visual Acuity) chart for near vision
Colour Vision Test Method	Ishihara plates
Room illumination (ambient light level [lux])	Low (exact measure unavailable)
Background luminance of wall behind the monitor	20 Lux

Appendix II.5 SwissQual

Tests Conducted: SwissQual's VGA Test, V04

Display

Display Manufacturer	Samsung
Display Model	SyncMaster 214T
Display Screen Size	21"
Display Resolution	1600 x 1200
Display Scanning Rate	75Hz
Display Pixel Pitch	0.27mm
Display Response Time (Black-White)	16ms
Display Colour Temperature	6500K

Display Bit Depth	8bit
Display Type (Standalone / Laptop)	Standalone
Display Label (TCO stamp)	TCO 03

Display Calibration

Calibration Tool	Eye One
Luminance Value (video display window peak white)	according calibration with EyeOne
Luminance Value (background display region)	according calibration with EyeOne
Brightness Value	according calibration with EyeOne
Contrast Value	according calibration with EyeOne
Gamma Value	according calibration with EyeOne

Test Computer

Computer Manufacturer	Dell
Model	Precision 490
Processor	Intel Xeon Dual Core
SDRAM	3GB
HDD	SATA 300GB
Connection to Display	DVI
Graphics card	NVIDIA Quadro FX 3450

Test Environment and Procedure

Viewing Distance	70 – 80cm
Viewing Angle	0.24 – 0.27 rad
Visual Acuity Test Method	Reading table with numbers
Colour Vision Test Method	Ishihara tables
Room illumination (ambient light level [lux])	about 20 lux
Background luminance of wall behind the monitor	according to standard

Appendix II.6 Symmetricom

Tests Conducted: Symmetricom's CIF Test, c01

Display

Display Manufacturer	Dell
Display Model	Ultrasharp 1707 FP
Display Screen Size	17"
Display Resolution	1280x1024
Display Scanning Rate	60Hz
Display Pixel Pitch	0.264mm
Display Response Time (Black-White)	8ms
Display Colour Temperature	6500K
Display Bit Depth	8 bit/color
Display Type (Standalone / Laptop)	Standalone
Display Label (TCO stamp)	TCO '03

Display Calibration

Calibration Tool	Eye-One
Luminance Value (video display window peak white)	180 cd/m^2
Luminance Value (background display region)	40 cd/m^2
Brightness Value	according calibration with Eye-One
Contrast Value	according calibration with Eye-One
Gamma Value	2.2

Test Computer

Computer Manufacturer	OEM
Model	OEM
Processor	Intel P4 3GHz
SDRAM	1GB
HDD	160GB
Connection to Display	DVI
Graphics card	nVidia 7600GS

Test Environment and Procedure

Viewing Distance	6-8H
Viewing Angle	Near 0°
Visual Acuity Test Method	Snellen chart
Colour Vision Test Method	Ishihara plates
Room illumination (ambient light level [lux])	Low
Background luminance of wall behind the monitor	Low

Appendix II.7 Yonsei University

Tests Conducted: Yonsei's, QCIF/ CIF/ VGA Tests, q02, q08, c04, c09, v02, v03

Display

Display Manufacturer	SAMSUNG
Display Model	SYNCMASTER 216TW
Display Screen Size	21"
Display Resolution	1680 x 1050
Display Scanning Rate	60 Hz
Display Pixel Pitch	0.270 mm
Display Response Time (Black-White)	8 ms
Display Colour Temperature	6500K
Display Bit Depth	8bit
Display Type (Standalone / Laptop)	Standalone
Display Label (TCO stamp)	TCO´03 CALA36+

Display Calibration

Calibration Tool	Eye one
Luminance Value (video display window peak white)	according calibration with Eye-one
Luminance Value (background display region)	according calibration with Eye-one
Brightness Value	according calibration with Eye-one
Contrast Value	according calibration with Eye-one
Gamma Value	2.2

Test Computer

Computer Manufacturer	OEM
Model	OEM
Processor	Intel Pentium 4 2.80 Ghz
SDRAM	512 MB
HDD	WD102288-DOCJA1(Western digital 115GB)
Connection to Display	DVI
Graphics card	NVIDIA Geforce 7300GT

Test Environment and Procedure

Viewing Distance	6-8H
Viewing Angle	Near 0°
Visual Acuity Test Method	Snellen Type Plastic Eye char (Korean edition)
Colour Vision Test Method	Ishihara's test
Room illumination (ambient light level [lux])	According to standard
Background luminance of wall behind the monitor	According to standard

Appendix II.8 Orange France Telecom

Tests Conducted: Francetelecom's QCIF Tests, Q10 and Q14

Display

Display Manufacturer	EIZO
Display Model	FlexScan L778
Display Screen Size	19"
Display Resolution	1280 x 1024
Display Scanning Rate	60 Hz
Display Pixel Pitch	0.294 x 0.294 mm
Display Response Time (Black-White)	16ms
Display Colour Temperature	6500 °K
Display Bit Depth	8 bit
Display Type (Standalone / Laptop)	Standalone
Display Label (TCO stamp)	

Display Calibration

Calibration Tool	Minolta CS 1000
Luminance Value (video display window peak white)	130 cd/m2
Luminance Value (background display region)	7 cd/m2
Brightness Value	130 cd/m2 (30% of max withe)
Contrast Value	520:1
Gamma Value	2.2

Test Computer

Computer Manufacturer	HP
Model	HP Workstation XW 800
Processor	2 x Intel Xeon (TM) 3.06 GHz
SDRAM	1 G Ecc DDR 266 MHz
HDD	Seagate Ultra SCSI, 73G 15 000 rpm
Connection to Display	DVI-D standard 1.0
Graphics card	Nvidia quadro4 380 XGL 128 Mb

Test Environment and Procedure

Viewing Distance	40 cm (i.e. : 1H)
Viewing Angle	0°
Visual Acuity Test Method	Graham-Field
Colour Vision Test Method	Ishihara
Room illumination (ambient light level [lux])	20 Lux
Background luminance of wall behind the monitor	7 cd/m2

Appendix II.9 IRCCyN

Tests Conducted: IRCCyN's VGA Test, V11

Display

Display Manufacturer	Apple

Display Model	M9178
Display Screen Size	23''
Display Resolution	1920x1200
Display Scanning Rate	60Hz
Display Pixel Pitch	0.258
Display Response Time (Black-White)	16ms
Display Colour Temperature	6500K
Display Bit Depth	8 bits
Display Type (Standalone / Laptop)	Standalone
Display Label (TCO stamp)	TCO'03

Display Calibration

Calibration Tool	Eye One
Luminance Value (video display window peak white)	180cd/ m^2
Luminance Value (background display region)	30cd/m²
Brightness Value	according calibration with EyeOne
Contrast Value	according calibration with EyeOne
Gamma Value	2.2

Test Computer

Computer Manufacturer	Dell
Model	Precision 380
Processor	Intel Pentium Extreme Edition (double core) 3.2GHz
SDRAM	3.5G
HDD	204Go
Connection to Display	DVI
Graphics card	Sapphire X1950XT

Test Environment and Procedure

Viewing Distance	4H
Viewing Angle	0°

Visual Acuity Test Method	Monoyer's plates
Colour Vision Test Method	Ishihara´s test for Colour Deficiency
Room illumination (ambient light level [lux])	Low
Background luminance of wall behind the monitor	About 10cd/m²

Tests Conducted: IRCCyN's VGA Test, V12

Display

Display Manufacturer	Apple
Display Model	M9178
Display Screen Size	23''
Display Resolution	1920x1200
Display Scanning Rate	60Hz
Display Pixel Pitch	0.258
Display Response Time (Black-White)	16ms
Display Colour Temperature	6500K
Display Bit Depth	8 bits
Display Type (Standalone / Laptop)	Standalone
Display Label (TCO stamp)	TCO'03

Display Calibration

Calibration Tool	Eye One
Luminance Value (video display window peak white)	180cd/ m^2
Luminance Value (background display region)	30cd/m²
Brightness Value	according calibration with EyeOne
Contrast Value	according calibration with EyeOne
Gamma Value	2.2

Test Computer

Computer Manufacturer	Dell
Model	Precision 380
Processor	Intel Pentium Extreme Edition (double core) 3.2GHz

SDRAM	3.5G
HDD	204Go
Connection to Display	DVI
Graphics card	Sapphire X1950XT

Test Environment and Procedure

Viewing Distance	4H
Viewing Angle	0°
Visual Acuity Test Method	Monoyer's plates
Colour Vision Test Method	Ishihara´s test for Colour Deficiency
Room illumination (ambient light level [lux])	Low
Background luminance of wall behind the monitor	About 10cd/m²

Test Conducted: IRCCyN's QCIF Test, Q13

Display

Display Manufacturer	Apple
Display Model	M9178
Display Screen Size	23''
Display Resolution	1920x1200
Display Scanning Rate	60Hz
Display Pixel Pitch	0.258
Display Response Time (Black-White)	16ms
Display Colour Temperature	6500K
Display Bit Depth	8 bits
Display Type (Standalone / Laptop)	Standalone
Display Label (TCO stamp)	TCO'03

Display Calibration

Calibration Tool	Eye One
Luminance Value (video display window peak white)	180cd/ m^2
Luminance Value (background display region)	30cd/m²

Brightness Value	according calibration with EyeOne
Contrast Value	according calibration with EyeOne
Gamma Value	2.2

Test Computer

Computer Manufacturer	Dell
Model	Precision 380
Processor	Intel Pentium Extreme Edition (double core) 3.2GHz
SDRAM	3.5G
HDD	204Go
Connection to Display	DVI
Graphics card	Sapphire X1950XT

Test Environment and Procedure

Viewing Distance	10H
Viewing Angle	0°
Visual Acuity Test Method	Monoyer's plates
Colour Vision Test Method	Ishihara´s test for Colour Deficiency
Room illumination (ambient light level [lux])	Low
Background luminance of wall behind the monitor	About 10cd/m²

Appendix II.10 Verizon

Test Conducted: NTIA/Verizon VGA Test: V09

Display

Display Manufacturer	Dell
Display Model	3007WFP
Display Screen Size	30 in (display size 25.25 by 15.78 inches)
Display Resolution	2560 x 1600
Display Scanning Rate	60 Hz
Display Pixel Pitch	0.2505 mm x 0.2505 mm
Display Response Time (Black-White)	12-14 ms

Display Colour Temperature	6500 nominal
Display Bit Depth	8
Display Type (Standalone / Laptop)	standalone
Display Label (TCO stamp)	Active matrix - TFT LCD

Display Calibration

Calibration Tool	ProTelevision Technologies PM5639/80
Luminance Value (video display window peak white)	400 CD/m2 nominal
Luminance Value (background display region)	1.0 CD/m2 measured
Brightness Value	300 CD/m2 nominal
Contrast Value	700-1000/1 nominal
Gamma Value	Factory setup

Test Computer

Computer Manufacturer	Dell
Model	XPS 600
Processor	Intel Pentium D 3.2Ghz
SDRAM	4 X Micro Technology DDR2 PC2-5300 (333Mhz) 512MB
HDD	2 X Western Digital Raptor WD800 80GB SATA
Connection to Display	dual DVI ports or S-Video port
Graphics card	Nvidia GeForce 7800 GTX 256MB PCI-e

Test Environment and Procedure

Viewing Distance	28 inches
Viewing Angle	90 degrees
Visual Acuity Test Method	Bernell Snellen chart
Colour Vision Test Method	Ishihara test
Room illumination (ambient light level [lux])	2.8 CD/m2
Background luminance of wall behind the monitor	2.4 CD/m2

Test Conducted: NTIA/Verizon's CIF Test, C11

Display

Display Manufacturer	Dell
Display Model	3007WFP
Display Screen Size	30 in (display size 25.25 by 15.78 inches)
Display Resolution	2560 x 1600
Display Scanning Rate	60 Hz
Display Pixel Pitch	0.2505 mm x 0.2505 mm
Display Response Time (Black-White)	12-14 ms
Display Colour Temperature	6500 nominal
Display Bit Depth	8
Display Type (Standalone / Laptop)	standalone
Display Label (TCO stamp)	Active matrix - TFT LCD

Display Calibration

Calibration Tool	ProTelevision Technologies PM5639/80
Luminance Value (video display window peak white)	400 CD/m2 nominal
Luminance Value (background display region)	1.0 CD/m2 measured
Brightness Value	300 CD/m2 nominal
Contrast Value	700-1000/1 nominal
Gamma Value	Factory setup

Test Computer

Computer Manufacturer	Dell
Model	XPS 600
Processor	Intel Pentium D 3.2Ghz
SDRAM	4 X Micro Technology DDR2 PC2-5300 (333Mhz) 512MB
HDD	2 X Western Digital Raptor WD800 80GB SATA
Connection to Display	dual DVI ports or S-Video port
Graphics card	Nvidia GeForce 7800 GTX 256MB PCI-e

Test Environment and Procedure

Viewing Distance	22 inches
Viewing Angle	90 degrees
Visual Acuity Test Method	Bernell Snellen chart
Colour Vision Test Method	Ishihara test
Room illumination (ambient light level [lux])	2.8 CD/m2
Background luminance of wall behind the monitor	2.4 CD/m2

Appendix II.11 CRC-Nortel

Test Conducted: CRC/Nortel's CIF Test, C12

Display

Display Manufacturer	Viewsonic
Display Model	STATION 1: VX922 s/n: PXU070554972 STATION 2: VX922 s/n: PXU070554979 STATION 3: VX922 s/n: PXU070554973
Display Screen Size	19"
Display Resolution	1280 x 1024
Display Scanning Rate	60
Display Pixel Pitch	0.2944 mm
Display Response Time (Black-White)	2ms
Display Colour Temperature	6500K
Display Bit Depth	32bit color
Display Type (Standalone / Laptop)	Standalone
Display Label (TCO stamp)	TCO 99

Display Calibration

Calibration Tool	Eye-One Display2
Luminance Value (video display window peak white)	STATION 1: 119.8 cd/m2 STATION 2: 119.9 cd/m2 STATION 3: 118.6 cd/m2
Luminance Value (background display region)	STATION 1: 15.8 lux STATION 2: 14.9 lux

	STATION 3: 14.6 lux
Brightness Value	STATION 1: 40% of maximum STATION 2: 35% of maximum STATION 3: 30% of maximum
Contrast Value	STATION 1: 100% STATION 2: 100% STATION 3: 100%
Gamma Value	ALL STATIONS: 2.3

Test Computer

Computer Manufacturer	ALL STATIONS : Generic
Model	STATION 1: Asus P5N32-SLI Deluxe STATION 2: Asus P5WDG2 WS STATION 3: Asus P5WDG2 WS
Processor	STATION 1: Intel pentium-D 840 3.2GHz STATION 2: Intel CORE2 E6400 2.13GHz STATION 3: Intel CORE2 E6400 2.13GHz
SDRAM	STATION 1: 4GB PC4200 DDR2 STATION 2: 2GB PC2-5400 DDR2 STATION 3: 2GB PC2-5400 DDR2
HDD	STATION 1: WD Raptor 74GB 10K (x4 Raid 0) STATION 2: WD Raptor 150GB 10K (x3 Raid 0) STATION 3: WD Raptor 150GB 10K (x3 Raid 0)
Connection to Display	ALL STATIONS: DVI-D
Graphics card	STATION 1: Asus Radeon X850-XT 256MB pci-e STATION 2: Asus Radeon X1650 256MB pci-e STATION 3: Asus Radeon X1650 256MB pci-e

Test Environment and Procedure

Viewing Distance	7H (58.8cm)
Viewing Angle	0 degrees (perpendicular)
Visual Acuity Test Method	Good-Lite Near Vision Chart
Colour Vision Test Method	Ishihara
Room illumination (ambient light level [lux])	STATION 1: 2.5 lux STATION 2: 3.2 lux STATION 3: 2.2 lux
Background luminance of wall behind the monitor	STATION 1: 21.0 lux STATION 2: 21.8 lux STATION 3: 20.3 lux

Test Conducted: CRC/Nortel's QCIF Test, Q11

Display

Display Manufacturer	Viewsonic
Display Model	STATION 1: VX922 s/n: PXU070554972 STATION 2: VX922 s/n: PXU070554979 STATION 3: VX922 s/n: PXU070554973
Display Screen Size	19"
Display Resolution	1280 x 1024
Display Scanning Rate	60
Display Pixel Pitch	0.2944 mm
Display Response Time (Black-White)	2ms
Display Colour Temperature	6500K
Display Bit Depth	32bit color
Display Type (Standalone / Laptop)	Standalone
Display Label (TCO stamp)	TCO 99

Display Calibration

Calibration Tool	Eye-One Display2
Luminance Value (video display window peak white)	STATION 1: 119.8 cd/m2 STATION 2: 119.9 cd/m2 STATION 3: 118.6 cd/m2

Luminance Value (background display region)	STATION 1: 15.8 lux STATION 2: 14.9 lux STATION 3: 14.6 lux
Brightness Value	STATION 1: 40% of maximum STATION 2: 35% of maximum STATION 3: 30% of maximum
Contrast Value	STATION 1: 100% STATION 2: 100% STATION 3: 100%
Gamma Value	ALL STATIONS: 2.3

Test Computer

Computer Manufacturer	ALL STATIONS : Generic
Model	STATION 1: Asus P5N32-SLI Deluxe STATION 2: Asus P5WDG2 WS STATION 3: Asus P5WDG2 WS
Processor	STATION 1: Intel pentium-D 840 3.2GHz STATION 2: Intel CORE2 E6400 2.13GHz STATION 3: Intel CORE2 E6400 2.13GHz
SDRAM	STATION 1: 4GB PC4200 DDR2 STATION 2: 2GB PC2-5400 DDR2 STATION 3: 2GB PC2-5400 DDR2
HDD	STATION 1: WD Raptor 74GB 10K (x4 Raid 0) STATION 2: WD Raptor 150GB 10K (x3 Raid 0) STATION 3: WD Raptor 150GB 10K (x3 Raid 0)
Connection to Display	ALL STATIONS: DVI-D
Graphics card	STATION 1: Asus Radeon X850-XT 256MB pci-e STATION 2: Asus Radeon X1650 256MB pci-e STATION 3: Asus Radeon X1650 256MB pci-e

Test Environment and Procedure

Viewing Distance	8H (33.6cm)
Viewing Angle	0 degrees (perpendicular)
Visual Acuity Test Method	Good-Lite Near Vision Chart
Colour Vision Test Method	Ishihara
Room illumination (ambient light level [lux])	STATION 1: 2.5 lux STATION 2: 3.2 lux STATION 3: 2.2 lux
Background luminance of wall behind the monitor	STATION 1: 21.0 lux STATION 2: 21.8 lux STATION 3: 20.3 lux

Appendix II.12 Acreo

Test Conducted: Acreo's QCIF & CIF Tests, q12 and c14

Display

Display Manufacturer	Samsung
Display Model	SyncMaster 215TW
Display Screen Size	21.3"
Display Resolution	1650 x 1050
Display Scanning Rate	60 Hz, 65.3 kHz, 144.25 MHz
Display Pixel Pitch	0.273 mm
Display Response Time (Black-White)	8 ms
Display Colour Temperature	6500 K, measured in CIE 1976 u′= 0.196, v′= 0.467
Display Bit Depth	8 bits/colour
Display Type (Standalone / Laptop)	Standalone
Display Label (TCO stamp)	TCO´03 + TCO´06

Display Calibration

Calibration Tool	PhotoResearch PR705 Spectroradiometer
Luminance Value (video display window peak	Set to 200 cd/m^2

white)	
Luminance Value (background display region)	Grey level 108 corresponding to 24 cd/m^2
Brightness Value	64
Contrast Value	73
Gamma Value	About 2.2, but dependent of measured used

Test Computer

Computer Manufacturer	Dell
Model	Precision Workstation 530MT
Processor	Intel Xeon 1.7 GHz
SDRAM	1 GB
HDD	C: 40 GB Western Digital WD400BB-75AUA1 D: 120 GB Western Digital WD1200BB-CAA1
Connection to Display	DVI
Graphics card	Matrox Parhelia 400 MHz 256 MB

Test Environment and Procedure

Viewing Distance	8 times the picture height i.e. 31 cm for the QCIF and 62 cm for the CIF
Viewing Angle	8.73° × 7.15° for the images
Visual Acuity Test Method	Snellen letter test chart designed for reading at 40 cm
Colour Vision Test Method	Ishihara´s test for Colour Deficiency Concise Edition 2007 with 14 plates
Room illumination (ambient light level [lux])	E_{vh} about 20 lux at about 20 cm in front of the screen
Background luminance of wall behind the monitor	2 – 3 cd/m^2

Appendix II.13 FUB

Test Conducted: FUB's CIF Test, C13; and FUB's VGA Tests, V10 & V13

Display

Display Manufacturer	Samsung
Display Model	SyncMaster192v
Display Screen Size	19"
Display Resolution	1280 x 1024
Display Scanning Rate	60 Hz
Display Pixel Pitch	0.294 x 0.294 mm
Display Response Time (Black-White)	16ms
Display Colour Temperature	6500 °K
Display Bit Depth	8 bit
Display Type (Standalone / Laptop)	Standalone
Display Label (TCO stamp)	TCO 99

Display Calibration

Calibration Tool	Minolta CS 1000
Luminance Value (video display window peak white)	249 cd/m2
Luminance Value (background display region)	7 cd/m2
Brightness Value	249 cd/m2 (30% of max withe)
Contrast Value	510:1
Gamma Value	2.2

Test Computer

Computer Manufacturer	OEM
Model	OEM
Processor	Intel Pentium D 3.2 GHz
SDRAM	2 G DDR-2
HDD	WD Raptor SATA II, 73G 10.000 rpm
Connection to Display	DVI-D standard 1.0
Graphics card	Nvidia GeForce 6600 LE , 512 M

Test Environment and Procedure

Viewing Distance	4H for VGA 6H for CIF
Viewing Angle	0°
Visual Acuity Test Method	Snellen Chart
Colour Vision Test Method	Ishihara
Room illumination (ambient light level [lux])	low
Background luminance of wall behind the monitor	7 cd/m2

Appendix III SRC Associated with Each Individual Experiment

Appendix III.1 Scene Descriptions and Classifications

The ILG sorted SRC into the 8 categories identified in the MM test plan. The SRC category tables used by the ILG follow. SRC that did not obviously fall into any category are listed in a 9th table. The content source is identified, and each scene is briefly described. The right-most column of these tables identifies secret SRC.

Category 1: Videoconferencing

	Clip	Description	Source	Frame Rate	Secret?
1	VQEGSusie	Static headshot of woman talking on phone	CRC	30 fps	
2	NTIAcatjoke	Man telling joke, bright wall behind him, some fast motion	NTIA	30 fps	
3	NTIAcchart1	Man with color chart, against grey textured wall	NTIA	30 fps	
4	NTIAcchart2	Man with color chart, against grey textured wall	NTIA	30 fps	
5	NTIAcchart3pp	Man with color chart, against grey textured wall	NTIA	30 fps	
6	NTIAoverview1	Man in white shirt sips coffee, against grey textured wall	NTIA	30 fps	
7	NTIArfdev1	Man explains Rf device, some detail on walls behind him.	NTIA	30 fps	
8	NTIArfdev2	Man explains Rf device, some detail on walls behind him.	NTIA	30 fps	
9	NTIAschart1	Camera zooms in slowly as elderly woman tells story, with quilt hanging in BG.	NTIA	30 fps	
10	NTIAschart2	Tighter shot as elderly woman tells story, with quilt hanging in BG.	NTIA	30 fps	
11	NTIAspectrum1	Close-up of man's face and colorful chart, with zoom out in mid sequence.	NTIA	30 fps	
12	ANSIwashdc	Close up of map, hand, pencil.	NTIA	30 fps	
13	NTIApghtalk1a	Two men in hard-hats talking to each other and the camera, gesturing animatedly	NTIA	30 fps	Secret
14	NTIAoverview2	Man in white shirt speaks, against grey textured wall	NTIA	30 fps	
15	NTIAspectrum2	Zoomed out view of man and colorful chart	NTIA	30 fps	
16	NTIAwboard1	Man and whiteboard, slow pan and zoom.	NTIA	30 fps	
17	ANSIvtc2mp	Static shot of teacher and world map.	NTIA	30 fps	
18	NTIAfire04	Fire fighters receiving instruction before being deployed.	NTIA	30 fps	Secret

19	CRCbench	Static shot of woman speaking from park bench	CRC	30 fps	Secret
20	CRCheadshot	Static headshot of woman speaking, with Canadian flag in BG	CRC	30 fps	Secret
21	CRChouseoffer	Static medium shot of woman speaking, with Canadian flag in BG	CRC	30 fps	secret
22	NTIAwboard2	Close-up of man's hand writing on whiteboard	NTIA	30 fps	
23	ANSI3inrow	Camera pans between two poorly lit people at table.	NTIA	30 fps	
24	ANSI5row1	Five sit at table, reflections in tabletop, under poor lighting.	NTIA	30 fps	
25	ANSIboblec	Instructor at the blackboard, some small pan and zoom.	NTIA	30 fps	
26	ANSIvtc1nw	Static shot, poorly lit newsreader.	NTIA	30 fps	
27	ANSIvtc2zm	Closer view of world map and pointer, some camera tilt.	NTIA	30 fps	
28	SRCvisio	Man with hands-free phone, looking slightly down	FT	25 fps	
29	NTTBlock21	Man demonstrating building Lego giraffe	NTT	30 fps	
30	NTTBlock23	Man demonstrating building Lego pyramid	NTT	30 fps	
31	NTTCount31	Man counting, raising & lowering hand	NTT	30 fps	
32	NTTCount33	Man counting, raising & lowering hand	NTT	30 fps	
33	NTTTalk14	Close-up of woman's face, talking	NTT	30 fps	
34	NTTMix	Cuts from three NTT clips: woman talking, man counting, & building Lego giraffe WARNING: different versions for each resolution	NTT	30 fps	
35	NTIAwashdcStill	Close up of map, hand, pencil – digitally still portion inserted	NTIA	30 fps	Secret
36	NTIASusieStill	Static headshot of woman talking on phone – digitally still portion inserted	NTIA	30 fps	Secret
37	TW01	Close-up of man's head & shoulders, talking	TW	25 fps	Secret
38	TW02	Close-up of man's head & shoulders, talking	TW	25 fps	Secret
39	TW03	Zoom on calendar, stretching definition, but perhaps "graphics with pointer"	TW	25 fps	Secret
40	TW04	Zoom out from desk, could occur during videoconferencing	TW	25 fps	Secret

Category 2: Movies

	Clips	Description	Source	Frame Rate	Secret?
1	KBSwanggunC	historical drama, static headshot, high detail	KBS/YONSEI	30 fps	
2	KBSwanggunD	historical drama, zooming and panning with 1 scene cut	KBS/YONSEI	30 fps	
3	KBSwanggunE	historical drama, long slow zoom to closeup of detailed face	KBS/YONSEI	30 fps	
4	KBSwinterA	camera tilts downward to show distant person between rows of wintery pines	KBS/YONSEI	30 fps	
5	KDDI3D01	Static shot of woman on garden path, walking away	KDDI	30 fps	
6	KDDI3D02	Static shot of woman in tulip patch, turning and disappearing	KDDI	30 fps	
7	KDDI3D04	Camera follows woman walking through tulip garden	KDDI	30 fps	
8	KDDISD13	Woman walks horse through woods, as camera zooms in	KDDI	30 fps	
9	KDDISD18	Couple stand at poolside, pool has gridlines at bottom	KDDI	30 fps	
10	NTIAbpit5	Overhead rotating shot of child in ballpit	NTIA	30 fps	
11	PSYdrink01	Complex camera shot, from overhead view of cobblestone street, to tabletop. VGA & CIF only	Psythechnics	25 fps	
12	PSYinter01	Slow zoom onto boardroom scene VGA & CIF only	Psythechnics	25 fps	
13	KBSwanggunB	historical drama, 2 scene cuts, close, far and medium views	KBS/YONSEI	30 fps	
14	KBSwanggunF	historical drama, trucking / zooming of procession	KBS/YONSEI	30 fps	
15	KBSwinterB	as above, with cut to snow fight at reduced speed playback	KBS/YONSEI	30 fps	
16	KDDI3D05	Closeup of woman in tulip garden, with trees in BG	KDDI	30 fps	
17	KDDI3D06	More distant shot of woman in tulip garden, standing on stone pavement	KDDI	30 fps	
18	KDDISD16	Camera follows actions of woman examining a vase	KDDI	30 fps	
19	NTIAbpit1	Camera pans over 2 kids in ballpit, seen through mesh	NTIA	30 fps	
20	NTIAbpit2	Camera tilts and zooms in tightly to colored balls	NTIA	30 fps	
21	NTIAcargas	Camera zooms in slowly as car pulls up for	NTIA	30 fps	

		gas			
22	NTIAfiremovie1	Scene cuts between burning fire and fire fighters, ending with water spray extinguishing the fire	NTIA	30 fps	Secret
23	NTIAhose	Fire fighter training session, practicing unrolling hoses. The rolling hose raises a small dust cloud. Foreground is in focus, and background is out of focus	NTIA	30 fps	Secret
24	PSYfesti01	Static shot of fairgrounds, complex motion but low contrast	Psythechnics	25 fps	
25	PSYmovie01	Camera pedestals as car drives away on scenic road VGA & CIF only. Animation overlay.	Psythechnics	25 fps	
26	KDDISD08	jerky aerial shot of car speeding down highway	KDDI	30 fps	
27	KDDISD19	Poolside party, 2 scene cuts	KDDI	30 fps	
28	NTIAbpit3	Camera follows child crawling through balls	NTIA	30 fps	
29	NTIAbpit4	Like ballpit1, but further out with only 1 child	NTIA	30 fps	
30	NTIAstreet1	Skewed Vegas skyline as shot from moving car	NTIA	30 fps	
31	NTIAduckmovie	Sequence contains water movement, then a 1/5 second period of digitally perfect stillness	NTIA	30 fps	Secret
32	PSYfesti02	Static shot of 2 park rides against light sky	Psythechnics	25 fps	
33	NTIAstore1	Camera pedestals and zooms across dark storefront scene	NTIA	30 fps	
34	KBSwanggunG	Close-up on young man's face, scene cut to zoom on old man	KBS	30 fps	
34	SVTPrincessRun	Lady running through green woods, subdued lighting	SVT	25 fps	
35	SVTParkJoy	Small group of happy people run on path across stream, with woods in background, subdued lighting	SVT	25 fps	
36	SVTIntoTree	Arial point of view, zoom into tree next to building	SVT	25 fps	

Category 3: Sports

	Clip	Description	Source	Frame Rate	Secret?
1	KBSsoccerB	soccer match, 2 scene cuts, tight-wide-tight, (1st cut at 28f). Animation overlay.	KBS/YONSEI	30 fps	
2	KDDISD14	camera pans and zooms in on woman horseback riding	KDDI	30 fps	
3	ITUFootball	quick camera pans, tight shots of football action	CRC	30 fps	
4	VQEGTableTennis	zoom then scene cut to static shot with textured BG	CRC	30 fps	
5	NTIAplayerout	Football players escorted out of stadium after game.Fans line sides of path, reaching & waving. Some camera bounce	NTIA	30 fps	Secret
6	NTIAstadpan	High in stadium panning across a football game and crowd.	NTIA	30 fps	Secret
7	PSYfootb01	Camera pans and zooms from behind soccer net	Psythechnics	25 fps	
8	KBSsoccerA	soccer match, wideshot, slow panning, shadows on field. Animation overlay.	KBS/YONSEI	30 fps	
9	KBSsoccerC	soccer match, wideshot, slow panning, changing luminance. Animation overlay.	KBS/YONSEI	30 fps	
10	KBSsoccerD	soccer match, scene cuts with graphic fly-bys (1st cut inside 1 sec). Animation overlay.	KBS/YONSEI	30 fps	
11	KBSsoccerE	soccer match, wide view cuts to tighter shot. Animation overlay.	KBS/YONSEI	30 fps	
12	KDDI3D09	dance troop, 2 scene cuts (1st cut at 23f)	KDDI	30 fps	
13	KDDI3D10	dance troop, 2 scene cuts (2nd cut 22f before end)	KDDI	30 fps	
14	KDDISD01	camera zooms in on woman swimming in pool	KDDI	30 fps	
15	CRCvolleyball	camera pans to follow action	CRC	30 fps	Secret
16	NTIAflag	Football game from high on stands showing stadium and pre-game show. Zooms in on a giant US flag	NTIA	30 fps	Secret
17	PSYccski02	camera trucks quickly to follow skiers in wintery scene	Psythechnics	25 fps	
18	PSYskidh01	camera follows downhill skier, sideview	Psythechnics	25 fps	
19	PSYskidh02	camera follows downhill skier, rearview	Psythechnics	25 fps	
20	PSYskidh03	camera follows downhill skier, frontview	Psythechnics	25 fps	
21	NTIAstadsc	two shots of football stadium during game. Shows camera crew and end of field; then switches to view from field goal	NTIA	30 fps	Secret

		watching players warm up on the field.			
22	PSYccski01	low angle shot, some very visible judder	Psythechnics		
23	CRCvolleyball25fps	camera pans to follow action	CRC	25 fps	Secret
24	NTIAstadpan25fps	High in stadium panning across a football game and crowd.	NTIA	25 fps	secret
25	NTIAplayerout25fps	Football players escorted out of stadium after game.Fans line sides of path, reaching & waving. Some camera bounce	NTIA	25 fps	Secret
26	NTIAstadsc25fps	two shots of football stadium during game. Shows camera crew and end of field; then switches to view from field goal watching players warm up on the field.	NTIA	25 fps	Secret
27	CUhockey1	Hockey game, distant shot through white net, small figures QCIF hides netting. Quality acceptable for QCIF only. Animation overlay.	CU	30 fps	Secret
28	CUhockey2	Hockey game, medium distance through net QCIF & CIF only. Animation overlay.	CU	30 fps	Secret
29	CUhockey3	Hockey game, close then far distance through net, QCIF hides netting; QCIF & CIF only. Animation overlay.	CU	30 fps	Secret
30	CUbbshoot	Basketball shoot, then follow action across court QCIF & CIF only. Animation overlay.	CU	30 fps	Secret
31	CUbbfoul	Replay of basketball foul, then animation change to free throw. QCIF & CIF only. Animation overlay.	CU	30 fps	Secret
32	SVTCrowdRun	Crowd running a race, all people small; probably not well suited to QCIF	SVT	25 fps	
33	ITUarrividerci2	Soccer, detail and fast motion	ITU	25 fps	Secret
34	ITUBicycleRace	Bicycle Race, fast motion. Animation overlay.	ITU	25 fps	Secret
35	ITUccraceA	Cross country race, two cuts of lady with red jersey finishing the race, blurred background. Animation overlay.	ITU	25 fps	Secret
36	ITUccraceB	Cross country race, group of men run past, fast pan following, blurred background. Animation overlay.	ITU	25 fps	Secret
37	ITUf1raceA	Car race, QCIF & CIF only, very fast motion. Animation overlay.	ITU	25 fps	Secret
38	ITUf1raceB	Car race, QCIF & CIF only, fast motion; animation overlay on screen longer. Animation overlay.	ITU	25 fps	Secret

39	NTIAftballslow	A variant of the ITU Football scene. A segment is shown twice, the second time being a slow-motion replay. This slow motion portion effectively contains a reduced frame rate, as seen in cartoons.	NTIA	30fps	Secret

Category 4: Music video

		Description	Source	Frame Rate	Secret?
1	KBSgayoA	variety show, zoom & pan of trombone player. Animation overlay.	KBS/YONSEI	30 fps	
2	KBSgayoD	variety show, slow pan and zoom of singer against detailed BG. Animation overlay.	KBS/YONSEI	30 fps	
3	KBSmubankA	music video show, complex camera motion, medium shots of 2 hosts	KBS/YONSEI	30 fps	
4	KBSmubankD	music video show, complex camera motion, host in wading pool. Animation overlay.	KBS/YONSEI	30 fps	
5	KBSmubankE	music video show, two shots with scene cut / flash effect. Animation overlay.	KBS/YONSEI	30 fps	
6	NTIAmusic3	Camera zooms in for close-up of banjo picking. Animation overlay.	NTIA	30 fps	
7	KBSgayoB	variety show, singer and dancers, 1 scene cut to tighter shot. Animation overlay.	KBS/YONSEI	30 fps	
8	KBSgayoC	variety show, wide panning shot of dancers on stage. Animation overlay.	KBS/YONSEI	30 fps	
9	KBSgayoE	variety show, closeup of signer againsted blurred BG. Animation overlay.	KBS/YONSEI	30 fps	
10	NTIAbells4	Close-up shots of bells being rung	NTIA	30 fps	
11	NTIAbells5	Camera zooms in on conductor	NTIA	30 fps	
12	NTIAdrmside	Handheld shot of drummer in action, higher angle view	NTIA	30 fps	
13	NTIAguitar3	Slow zoom towards guitarist's fingers	NTIA	30 fps	
14	NTIAmusic2	Handheld camera on banjo player	NTIA	30 fps	
15	SMPTEbirches2	mostly pan down of birch trees, musicians always present	NTIA	30 fps	Secret
16	SMPTEdivatext2	diva with text, zoom in close with pan to window. Animation overlay.	NTIA	30 fps	Secret
17	NTIAbell7	Dark shot of bell players as camera zooms	NTIA	30 fps	
18	NTIAdrmfeet	Handheld shot of drummer in action, low to medium angle	NTIA	30 fps	
19	NTIAguitar1	Slow zoom towards guitarist sitting against wall hanging	NTIA	30 fps	
20	NTIAguitar2	Slow zoom towards guitarist sitting against wall hanging	NTIA	30 fps	
21	NTIAmusic1	Handheld pan between three musicians	NTIA	30 fps	
22	NTIApathsong	Man sings to camera in outdoor scene	NTIA	30 fps	
23	SMPTEdivatext1	diva with text, zoom in from far away.	NTIA	30 fps	secret

		Animation overlay.			
24	KBSmorningBp	talk show. Animation overlay.	KBS/YONSEI	30 fps	
25	KBSmubankBp	music video show. Animation overlay.	KBS/YONSEI	30 fps	
26	KBSmubankCp	music video show. Animation overlay.	KBS/YONSEI	30 fps	
27	KBSmubankFp	music video show. Animation overlay.	KBS/YONSEI	30 fps	
28	CUtubaspin1	Basketball half time music show, tuba player spins while playing; then zoom out while musician runs off field. QCIF & CIF only. Animation overlay.	CU	30 fps	Secret
29	CUtubaspin2	Basketball half time music show, tuba player spins while playing; stops (8s) as musician begins to run off field. QCIF & CIF only. Animation overlay.	CU	30 fps	Secret

Category 5: Advertisement

	Clip	Description	Source	Frame Rate	Secret?
1	NTIAtea1p	Panning shots of ornate interiors, 2 crossfades, 1 scene cut. Animation overlay.	NTIA	30 fps	
2	NTIAtea2	Panning shots of ornate interiors, picture in picture, 2 crossfades	NTIA	30 fps	
3	NTIAtea3	Panning shots of ornate interiors, 1 scene cut, 2 crossfades	NTIA	30 fps	
4	OPT013	Fast clips: elephants, rafting, filming Quality of some portions lower than others.	OPTICOM	25 fps	
5	OPT014p	Fast clips, mostly black & white, some bombs & tanks	OPTICOM	25 fps	
6	OPT015p	Fast clips: elephant, Africa, fire, fireworks; letterbox Quality of some portions lower than others. WARNING: needs scene cut adjustment	OPTICOM	25 fps	
7	OPT016p	Fast clips of animals, letterbox Quality of some portions lower than others. WARNING: needs scene cut adjustment	OPTICOM	25 fps	
8	CUpsa1	Public service announcement, girl & beach & water; soft edges, some noise; QCIF only	CU	30 fps	Secret
9	CUpsa2	Public service announcement, wilderness. QCIF only. Animation overlay.	CU	30 fps	Secret
10	CUpresents1	Fast paced opening credits, appearance of an advertisement, lots of animation & processing QCIF & CIF only. Animation overlay.	CU	30 fps	Secret
11	CUpresents2	Fast paced opening credits, soft focus scoreboard in background; fast paced cuts of sporting event clips, animation overlay QCIF & CIF only. Animation overlay.	CU	30 fps	Secret
12	CUpresents3	Fast paced opening credits, soft focus scoreboard in background; fast paced cuts of sporting event clips, animation overlay; cuts briefly to woman holding sign QCIF & CIF only. Animation overlay.	CU	30 fps	secret
13	CUpresents4	Fast paced opening credits, soft focus scoreboard in background; fast paced cuts of sporting event clips, animation overlay; ends with text in front of buffalo QCIF & CIF only. Animation overlay.	CU	30 fps	Secret

Category 6: Animation

		Description	Source	Frame Rate	Secret?
1	CBCBetesPasBetesP	Colorful animated creatures with scene cuts	CRC	30 fps	
2	KBSnewsG	weather segment, animated intro cross fades to weather reporter	KBS/YONSEI	30 fps	
3	CBCLePoint	Colorful letters move at different speeds	CRC	30 fps	
4	NTIAbrick2	Snail eats mushroom	NTIA	30 fps	
5	NTIAbrick	Snail does not eat mushroom	NTIA	30 fps	
6	ITUUnGenerique	Text scrolls vertically, 2 speeds	CRC	30 fps	
7	OPT006	computer graphics walkthrough of office building, fade	OPTICOM	25 fps	
8	OPT008	computer graphics walkthrough outside office building	OPTICOM	25 fps	
9	IRCCyNanim1	Dark background animation of spinning “planets” with people & etc. superimposed. QCIF only	IRCCyN	25 fps	Secret
10	IRCCyNGob2	Professional appearing cartoon, depicting surprised man and exterior market. QCIF only	IRCCyN	25 fps	Secret
11	IRCCyNanim13	Brown, black & white, birds flying, morphing. CIF only	IRCCyN	25 fps	Secret
12	IRCCyNGob3	old lady & cat, drowsy. CIF only	IRCCyN	25 fps	secret

Category 7: Broadcast news

		Description	Source	Frame Rate	Secret?
1	KBSnewsH	weather segment, reporter and changing weather maps. Animation overlay.	KBS/YONSEI	30 fps	
2	KBSnewsA	news show, male newscaster, with cut to flaming vehicle video. Animation overlay.	KBS/YONSEI	30 fps	
3	KBSnewsC	news show, reporter on scene, 2 scene cuts. Animation overlay.	KBS/YONSEI	30 fps	
4	KBSnewsD	news show, male newscaster, no scene cuts	KBS/YONSEI	30 fps	
5	KBSnewsF	news show, female newscaster, no scene cuts	KBS/YONSEI	30 fps	
6	NTIAdirtywin	passenger view through windshield, bouncy video	NTIA	30 fps	
7	NTIAheli02	Daytime footage from helicopter, looking down at a parking	NTIA	30 fps	Secret
8	NTIAfishrob1	Simulated robbery from surveillance camera. Shot with fish eye lens.	NTIA	30 fps	Secret
9	NTIArbtnews1	Simulated news coverage of experimental rescue robots	NTIA	30 fps	Secret
11	NTIArbtnews2	Simulated news coverage of experimental rescue robots. Includes a very fast event of a window shattering.	NTIA	30 fps	Secret
12	NTIAffgear	A firefighter puts on equipment. Includes a zoom out	NTIA	30 fps	Secret
13	NTIAfire06	Inside fire-truck, driving, looking out of front windshied	NTIA	30 fps	Secret
14	NTIAnstopbf	Slow camera zoom towards policeman standing beside stopped car	NTIA	30 fps	
15	NTIAnstopm	Slow camera zoom as policeman approaches stopped vehicle	NTIA	30 fps	
16	NTIAfcnstop	Two police cars pulling over a van at night. Some noise present due to night conditions. Dark scene with quickly flashing lights that glint on the lens.	NTIA	30 fps	Secret
18	NTIAhcuff	Close-up, handcuffing someone	NTIA	30 fps	
19	KBSnewsBp	news show footage	KBS/YONSEI	30 fps	
20	KBSnewsEp	news show	KBS/YONSEI	30 fps	
21	NTIAfcnstop25fps	Two police cars pulling over a van at night. Some noise present due to night conditions. Dark scene with quickly flashing lights that glint on the lens.	NTIA	25 fps	Secret
22	NTIAheli0225fps	Daytime footage from helicopter, looking	NTIA	25 fps	Secret

		down at a parking			
23	FTnews	Purple background, head & shoulders of man talking, overlaid text on right side. Animation overlay	FT	25 fps	
24	CUelecnews	Cuts from an electricity conservation student news cast QCIF only	CU	30 fps	Secret
25	CUbcancer2	Crowds walking, student newscast of fundraiser, QCIF only	CU	30 fps	Secret
26	TW09	Follow cars driving, fine gravel texture	TW	25 fps	Secret
27	TW10	Follow car driving, some jiggle	TW	25 fps	Secret

Category 8: Home video

		Description	Source	Frame Rate	Secret?
1	NTIAcollage1	Spinning feathers & cloth, brightly colored	NTIA	30 fps	Secret
2	CRCFlamingoHilton	handheld slow zoom out from strobing neon lights	CRC	30 fps	
3	CRCcarrousel	handheld slow zoom to bright carrousel	CRC	30 fps	Secret
4	NTIAfish1	fish in pond, 2 crossfades (3rd crossfade starts 2f before end)	NTIA	30 fps	
5	NTIAfish2	closer view of fish, 2 crossfades	NTIA	30 fps	
6	NTIAfish3	closer view of fish, 3 crossfades (3rd crossfade in last 10f)	NTIA	30 fps	
7	NTIApool	view of pool table and pool shot	NTIA	30 fps	
8	NTIAtwoducks	2 ducks walk into water and swim away	NTIA	30 fps	
9	NTIAcartalk1	boy in car speaks animatedly, fast arm & head motion	NTIA	30 fps	
10	NTIAdiner	medium shot of man at diner table	NTIA	30 fps	
11	NTIAfish5	zoom in in fish in a pond, no scene cuts	NTIA	30 fps	
12	NTIAflower1	camera pans and zooms in an garden, some shake	NTIA	30 fps	
13	NTIAmagic1	girl does magic trick in front of fireplace	NTIA	30 fps	
14	NTIAtea4	camera sweeps ornate room, changing luminance, some shake	NTIA	30 fps	
15	NTIAcollage4	medley of footage, each showing portions of a collage of brightly colored items.Scene cuts	NTIA	30 fps	Secret
16	NTIAcollage5	medley of footage, each showing portions of a collage of brightly colored items.Scene cuts	NTIA	30 fps	Secret
17	NTIAlowrider	Camera outside car window, by tire, as driving	NTIA	30 fps	secret
18	NTIAtowtruck1	Night shot of tow truck with flashing lights	NTIA	30 fps	Secret
19	NTIAchicken	Fast pan then zoom in on a car with a chicken inside.	NTIA	30 fps	Secret
20	YONSEIzooA	zoo scene, slow zoom out from rhino	KBS/YONSEI	30 fps	
21	CRCCaesarsPalace	handheld pan / zoom to flaming torch, at night	CRC	30 fps	
22	NTIAmlion	handheld zoom into warning sign, some shake	NTIA	30 fps	
23	NTIApond	camera pans from statue to pond, some shake	NTIA	30 fps	

24	NTIAtwogeese	2 geese walk through brown reeds	NTIA	30 fps	
25	NTIAwfall	zoom in on distant waterfall	NTIA	30 fps	
26	NTIAcartalk2	boy in car speaks animatedly, fast arm & head motion, different angle, lower light	NTIA	30 fps	
27	NTIAflower2	camera pans and zooms in an garden, some shake	NTIA	30 fps	
28	NTIAmagic3	girl does magic trick in front of fireplace	NTIA	30 fps	
29	NTIAfence	Camera carried while walking, look sideways, walking past fence at night. Fence looks like vertical bars moving past quickly.	NTIA	30 fps	Secret
30	NTIAtowtruck2	Pan at night along road, starting at a tow truck with flashing lights then following a car. Some noise present due to night conditions	NTIA	30 fps	Secret
31	YONSEIzooC	Warm tan alligator in water, against warm tan rocks. Slight water motion; nearly still	Yonsei	30 fps	
31	CRCcarrousel25fps	handheld slow zoom to bright carrousel.	CRC	25 fps	Secret
32	SQChildrenPlaying	Children sitting on large tree branch (side/back), with other children playing in the background WARNING: VGA version may have damage	SQ	25 fps	
33	SQMix	Three cuts of a boy playing; includes a piece of 'ChildrenPlaying' and 'LivingRoom CIF & QCIF differe from VGA	SQ	25 fps	
34	SQLivingRoom	Zoom in on boy playing with toy on livingroom floor	SQ	25 fps	
35	TW05	Two ladies walking in park by lake, pan; bright contrast	TW	25 fps	Secret
36	TW06	Follow bird walking on gravel	TW	25 fps	Secret
37	TW07	Pan following small tractor, gravel & grass fine texture	TW	25 fps	Secret
38	TW08	Zoom out of sunset over lake	TW	25 fps	Secret
39	TW12	Walking shot holding camera, lens flare and grass texture	TW	25 fps	Secret

SRC that did not seem to fall into any of the MM test plan's categories.

		Description	Source	Frame Rate	Secret?
1	KBSleeparkC	talk show, medium shot 2 hosts sitting, with detailed FG and BG	KBS/YONSEI	30 fps	
2	KDDI3D12	amusement park ride, distant shot, high detail in motion	KDDI	30 fps	
3	KDDISD11	camera slowly pans past glass ships inside bottles	KDDI	30 fps	
4	ITUMobileCalendar	Toy train, bobbing calendar, brighly colored wallpaper	CRC	30 fps	
5	ITUParkRide	Amusement park ride, fast complex motion	CRC	30 fps	
6	ITUFlowerGarden	Pan across flower garden and houses	CRC	30 fps	
7	KBSleeparkA	talk show, camera pans audience, 2 scene cuts to hosts (last cut 18f before end)	KBS/YONSEI	30 fps	
8	KBSleeparkB	talk show, medium shot of two hosts standing and talking	KBS/YONSEI	30 fps	
9	KBSleeparkD	talk show, medium 2 shot, with cut to closeup	KBS/YONSEI	30 fps	
10	YONSEIzooB	nature shot, a bush sways in the wind	KBS/YONSEI	30 fps	
11	KDDI3D03	nature shot, camera pans across tulip bed	KDDI	30 fps	
12	KDDI3D11	amusement park ride, medium close, high motion content	KDDI	30 fps	
13	KDDISD03	seal juggles ball at zoo	KDDI	30 fps	
14	KDDISD15	horses running in field, with scene cut	KDDI	30 fps	
15	ITUPopple	Fuzzy toy bird in red revolving cage, camera zooms in	CRC	30 fps	
16	VQEGAutumnLeaves	Camera zooms out slowly from distant waterfall	CRC	30 fps	
17	VQEGTempete	Camera tilts and zooms out while leaves blow in rustic scene	CRC	30 fps	
18	NTIAfire21	Outside, watch/follow vehicle driving by. Simulated security footage	NTIA	30 fps	Secret
19	NTIApghtruck2a	pan and slight zoom, following a fire patrol vehicle driving down a street	NTIA	30 fps	Secret
20	NTIAfisheye2	Simulated hallway surveillance footage, shot with a fish eye lens	NTIA	30 fps	Secret
21	SMPTEbirches1	pan down of birch trees	NTIA	30 fps	Secret
22	SMPTEbirches3	pan down slightly, musicians at the end of the picture	NTIA	30 fps	Secret

23	SMPTEbicycles	(? sports)distant shot of two ladies on bicycles, then zoom in on a bicycle wheel	NTIA	30 fps	secret
24	KDDI3D08	nature shot, brook, with scene cut	KDDI	30 fps	
25	KDDI3D14	amusement park ride, roller coaster	KDDI	30 fps	
26	KDDISD02	static nature shot, fish in tank	KDDI	30 fps	
27	KDDISD05	nature shot, flying over coastine	KDDI	30 fps	
28	CRCRedflower	static shot of flowers, slight breeze	CRC	30 fps	
29	VQEGSailboat	Nearly static shot of tall ship in port, small changes in luminance	CRC	30 fps	
30	CRCfence	A chain link fence gate swings open slowly	CRC	30 fps	Secret
31	CRCmobike	motorcyclist rides away	CRC	30 fps	Secret
32	NTIApghvansd	Mountain road, white van drives toward & past, man walking across street	NTIA	30 fps	Secret
33	ITUMobileCalendar	Toy train, bobbing calendar, brighly colored wallpaper	NTIA	25 fps	Secret
34	OPT001p	lasar demonstration with mist, two scene cuts	OPTICOM	25 fps	
35	OPT003	Zoom back over hardware, down over lasar	OPTICOM	25 fps	
36	OPT004	Glassworks demo, sand	OPTICOM	25 fps	
37	OPT005p	Shots of computer screen & board of lights,	OPTICOM	25 fps	
38	OPT009	Car crash test, water balloon pop; slow motion; scene change effect	OPTICOM	25 fps	
39	OPT010	Glass breaks in slow motion	OPTICOM	25 fps	
40	OPT011	Milky liquid pouring into wine glass, missing, slow motion	OPTICOM	25 fps	
41	OPT017	Family, LCD screen, security system	OPTICOM	25 fps	
42	OPT019	Mist, spinning liquid into fibers	OPTICOM	25 fps	
43	OPT020	Slow pan over equipment Grainy video due to low lighting	OPTICOM	25 fps	
44	OPT021	Shots of a train, gravel	OPTICOM	25 fps	
45	ITUCalMobA625	Calendar-Mobile 625-line, traditional pan/zoom section	ITU	25 fps	Secret
46	ITUCalMobB625	Calendar-Mobile 625-line, pan only	ITU	25 fps	Secret
47	ITUPopple625	Spinning red cage, blue background; 625-line	ITU	25 fps	Secret
48	ITUFlowerGarden625	Flower garden & windmill; washed out / white sky	ITU	25 fps	Secret

Note: SRC below with extra characters appended (e.g., CUpresents3NTT) contain the same SRC

content as listed in the above table, and only differ by the method used to de-interlace and rescale the video from the original into QCIF, CIF, or VGA.

Appendix III.2 SRC in Each Common Set

Following are the SRC in each common set.

QCIF Common Set

IRCCyNanim1_qcif
CUbbshoot_qcif
NTIASusieStill_qcif
CUbcancer2_qcif
KBSgayoB_qcif
CUpresents1_qcif

CIF Common Set

IRCCyNanim13_cif
CUpresents3NTT_cif
NTTTalk14_cif
KBSmubankA_cif
NTIAWashdcStill_cif
CUbbfoulirccyn_cif

VGA Common Set

NTIAstadpan_vga
SVTcrowdrunP_vga
KBSnewsGpsy1_vga
KBSgayoD_vga
NTIAduckmovie_vga
OPT013_vga

Appendix III.3 SRC in Each Experiment's Scene Pool

Following are the SRC in each experiment's scene pool.

QCIF Scene Pools

qcif.A – 25fps

IRCCyNGob2psy1_QCIF
OPT016p_qcif
ITUBicycleRace_qcif
PSYskidh02_qcif
TW01_qcif
SQLiving_Room_qcif
CRCCarrousel25fps_qcif
OPT010_qcif

qcif.D – 25fps

OPT015p_qcif
OPT021irccyn2_qcif
ITUf1raceB_qcif
NTIAftballslow_qcif
TW06_qcif
TW04_qcif
FTnews_qcif
NTIAplayerout25fps_qcif

qcif.G – 25fps

NTIAfcnstop25fps_qcif
TW09_qcif
ITUf1raceA_qcif
ITUarrividerci2_qcif
OPT006_qcif
SQLiving_Room_qcif
FTnews_qcif
PSYdrink01_qcif

qcif.I – 25fps

OPT020_qcif
PSYfootb01_qcif
ITUccraceA_qcif
OPT013_qcif
TW08_qcif
NTIAstadpan25fps_qcif
IRCCyNGob2psy1_QCIF
TW03_qcif

qcif.J – 30fps

CRCbench_qcif
KBSwanggunD_qcif
NTIAplayerout_qcif
KBSleeparkA_qcif
KBSnewsH_qcif
NTIAtwoducks_qcif
NTIAguitar3_qcif
KDDISD08_qcif

qcif.K – 30fps

NTIAtea1p_qcif
KBSnewsG_qcif
NTIAstadpan_qcif
NTIAoverview2_qcif
KBSwinterA_qcif
KBSgayoA_qcif
KDDI3D11_qcif
KDDISD03_qcif

qcif.L – 30fps

NTIAcollage1_qcif
CRCcarrousel_qcif
ITUpopple_qcif

NTIAspectrum1_qcif
KBSnewsF_qcif
NTIAbells5_qcif
KDDISD01_qcif
KDDISD19_qcif

qcif.P – 30fps

NTIAcartalk1_qcif
KDDI3D02_qcif
NTIApghtruck2a_qcif.vai
KBSwanggunB_qcif
KDDISD14_qcif
KBSmubankBp_qcif
NTIAffgear_qcif
ANSIvtc2mp_qcif

qcif.S – 30fps

NTIArfdev2_qcif
NTIArbtnews1_qcif
NTIAbpit5_qcif
KBSgayoE_qcif
KBSleeparkC_qcif
NTIAtwogeese_qcif
NTIApghvansd_qcif
SMPTEbicycles_qcif

qcif.T – 30fps

KBSmubankE_qcif
NTIAcatjoke_qcif
NTIAtowtruck1_qcif
KBSwanggunC_qcif
KDDI3D10_qcif
NTIApghtruck2a_qcif
KDDISD15_qcif

KBSnewsD_qcif

qcif.U – 30fps

CRCvolleyball_qcif
NTIAfcnstop_qcif
KBSwanggunG_qcif
NTIAmusic3_qcif
CUpresents4_qcif
NTIAschart2_qcif
NTIAfish5_qcif
KBSnewsEp_qcif

qcif.V – 30fps

NTIAtea4_qcif
CRCheadshot_qcif
KDDISD11_qcif
KBSsoccerD_qcif
KBSmubankBp_qcif
NTIAbpit2_qcif
KBSnewsH_qcif
NTIArbtnews2_qcif

qcif.W – 30fps

NTIAplayerout_qcif
KBSleeparkD_qcif
KBSmubankD_qcif
KBSnewsG_qcif
KBSgayoB_qcif
KDDISD16_qcif
YONSEIzooCpsy1_qcif
KDDI3D04_qcif

qcif.X – 30fps

NTIAfiremovie1_qcif

CRCvolleyball_qcif
NTIAcchart3pp_qcif
CRCcarrousel_qcif
CRCbench_qcif
NTIAcollage5_qcif
NTIAheli02_qcif
SMPTEbirches1_qcif

CIF Scene Pools

cif.B – 25fps

SQChildrenPlaying_cif
ITUccraceA_cif
SVTPrincessRunPP_cif
NTIAftballslow_cif
IRCCyNGob3irccyn_CIF
TW02_cif
PSYinter01_cif
NTIAstadpan25fps_cif

cif.E – 25fps

SVTParkJoyPP_cif
FTvisio_cif
OPT015p_cif
PSYccski01_cif
NTIAheli0225fps_cif
PSYfesti01_cif
OPT009_cif
TW07_cif

cif.G – 25fps

NTIAfcnstop25fps_cif

TW09_cif
ITUf1raceA_cif
ITUarrividerci2_cif
IRCCyNGob3irccyn_CIF
SQLiving_Room_cif
FTnews_cif
PSYdrink01_cif

cif.H – 25fps

OPT020_cif
PSYccski02_cif
CRCvolleyball25fps_cif
FTvisio_cif
OPT016p_cif
SVTCrowdRunP_cif
NTIAheli0225fps_cif
OPT008_cif

cif.J – 30fps

CRCbench_cif
KBSwanggunD_cif
NTIAplayerout_cif
KBSleeparkANTT_cif
KBSnewsH_cif
NTIAtwoducks_cif
NTIAguitar3_cif
KDDISD08_cif

cif.L – 30fps

NTIAcollage1_cif
CRCcarrousel_cif
ITUpopple_cif
NTIAspectrum1_cif
KBSnewsF_cif

NTIAbells5_cif
KDDISD01_cif
KDDISD19_cif

cif.M – 30fps

CRChouseoffer_cif
NTIAbrick2_cif
NTIAheli02_cif
NTIAmagic1_cif
KBSsoccerB_cif
KDDISD16_cif
CRCmobike_cif
KBSmubankA_cif

cif.N – 30fps

NTIAfiremovie1_cif
NTIAfcnstop_cif
CBCLePoint_cif
NTIAwfall_cif
SMPTEbirches2_cif
KDDI3D09psy1_cif
NTIAfish1_cif
CRCredflower_cif

cif.O – 30fps

NTIApghtalk1a_cif
CRCheadshot_cif
ITUungenerique_cif
CRCFlamingoHilton_cif
KBSnewsA_cif
KBSnewsBp_cif
CRCvolleyball_cif
NTIAbpit1opt1p_cif

cif.Q – 30fps

NTIAhose_cif
NTIAstadsc_cif
KBSmorningBp_cif
CBCBetesPasBetesP_cif
NTIA nstopbf_cif
NTTBlock_2-1_cif
KBS soccerD_cif
YonseizooA_cif

cif.R – 30fps

KBSmubankCp_cif
KBSsoccerC_cif
KDDI3D01psy1_cif
ITUMobileCalendar_cif
NTIAdrumfeet_cif
NTIAfishrob1_cif
CRCCaesarsPalace_cif
NTIAcollage5_cif

cif.U – 30fps

CRCvolleyball_cif
NTIAfcnstop_cif
KBSwanggunG_cif
NTIAmusic3_cif
CUpresents4_cif
NTIAschart2_cif
NTIAfish5_cif
KBSnewsEp_cif

cif.W – 30fps

NTIAplayerout_cif
KBSleeparkD_cif

KBSmubankD_cif
KBSnewsG_cif
KBSgayoB_cif
KDDISD16_cif
YONSEIzooC_cif
KDDI3D04_cif

cif.X – 30fps

NTIAfiremovie1_cif
CRCvolleyball_cif
NTIAcchart3pp_cif
CRCcarrousel_cif
CRCbench_cif
NTIAcollage5_cif
NTIAheli02_cif
SMPTEbirches1_cif

VGA Scene Pools

vga.C – 25fps

ITUpopple625_vga
PSYskidh03_vga
OPT004_vga
PSYfesti02_vga
TW05p_vga
SVTCrowdRunP_vga
SVTcloseuplegs2_vga
TW02_vga

vga.E – 25fps

SVTParkJoyPP_vga

FTvisio_vga
OPT015p_vga
PSYccski01_vga
NTIAheli0225fps_vga
PSYfesti01_vga
OPST009opt1_vga
TW07_vga

vga.F – 25fps

SVTIntoTree_vga
ITUccraceB_vga
SVTFirstGirls2_vga
TW10_vga
TW08_vga
OPT01p_vga
ITUCalMobB625_vga
NTIAftballslow_vga

vga.H – 25fps

OPT020_vga
PSYccski02_vga
CRCvolleyball25fps_vga
FTvisio_vga
OPT016p_vga
SVTCrowdRunP_vga
NTIAheli0225fps_vga
SVTOldTownCrossPP_vga

vga.K – 30fps

NTIAtea1p_vga
KBSnewsGpsy1_vga
NTIAstadpan_vga
NTIAoverview2_vga
KBSwinterA_vga

KBSgayoA_vga
KDDI3D11_vga
KDDISD03_vga

vga.L – 30fps

NTIAcollage1_vga
CRCcarrousel_vga
ITUpopple_vga
NTIAspectrum1_vga
KBSnewsF_vga
NTIAbells5_vga
KDDISD01_vga
KDDISD19_vga

vga.M – 30fps

CRChouseoffer_vga
NTIAbrick2_vga
NTIAheli02_vga
NTIAmagic1_vga
KBSnewsEp_vga
KDDISD16_vga
CRCmobike_vga
KBSmubankA_vga

vga.N – 30fps

NTIAfiremovie1_vga
NTIAfcnstop_vga
CBCLePoint_vga
NTIAwfall_vga
SMPTEbirches2_vga
KDDI3D09psy1_vga
NTIAfish1_vga
CRCredflower_vga

vga.O – 30fps

NTIApghtalk1a_vga
CRCheadshot_vga
ITUungenerique_vga
CRCFlamingoHilton_vga
KBSnewsAopt1_vga
KBSnewsBpopt1_vga
CRCvolleyball_vga
NTIAbpit1_vga

vga.P – 30fps

NTIAcartalk1_vga
KDDI3D02irccyn_vga
NTIApghtruck2a_vga.vai
KBSwanggunB_vga
KDDISD14opt2_vga
KBSmubankBp_vga
NTIAffgear_vga
ANSIvtc2mp_vga

vga.Q – 30fps

NTIAhose_vga
NTIAstadsc_vga
KBSmubankA_vga
CBCBetesPasBetesP_vga
NTIA nstopm_vga
NTTBlock_2-3_vga
KDDISD15ps1_vga
YonseizooA_vga

vga.R – 30fps

KBSmubankCp_vga
NTIAtea3_vga
KDDI3D01psy1_vga

NTIAplayerout_vga
NTIAdrmfeet_vga
NTIAfishrob1_vga
CRCCaesarsPalace_vga
NTIAcollage5_vga

vga.S – 30fps

NTIArfdev2_vga
NTIArbtnews1_vga
NTIAbpit5_vga
KBSgayoE_vga
KBSleeparkCpsy1_vga
NTIAtwogeese_vga
NTIApghvansd_vga
SMPTEbicycles_vga

Appendix III.4 Mapping of Scene Pools to Subjective Experiment

The following table shows the mapping of scene pools to subjective tests:

VGA Tests

Test Name	Scene Pool	Frame Rate	
		30fps	25fps
V01	vga.C		X
V02	vga.K	X	
V03	vga.Q	X	
V04	vga.N	X	
V05	vga.P	X	
V06	vga.O	X	
V07	vga.H		X
V08	vga.M	X	
V09	vga.R	X	
V10	vga.E		X
V11	vga.F		X
V12	vga.S	X	
V13	vga.L	X	

25fps Scene Pools: C, E, F, H

30fps Scene Pools: K, M, N, O, P, Q, R, S, L

CIF Tests

Test Name	Scene Pool	Frame Rate 30fps	25fps
C01	cif.E		X
C02	cif.J	X	
C03	cif.M	X	
C04	cif.Q	X	
C05	cif.N	X	
C06	cif.L	X	
C07	cif.O	X	
C08	cif.W	X	
C09	cif.R	X	
Q10	cif.H		X
C11	cif.U	X	
C12	cif.X	X	
C13	cif.B		X
C14	cif.G		X

25fps Scene Pools: B, E, G, H

30fps Scene Pools: J, M, N, O, Q, R, U, W, X, L

QCIF Tests

Test Name	Scene Pool	Frame Rate	
		30fps	25fps
Q01	qcif.A		X
Q02	qcif.J	X	
Q03	qcif.K	X	
Q04	qcif.U	X	
Q05	qcif.L	X	
Q06	qcif.W	X	
Q07	qcif.V	X	
Q08	qcif.P	X	
Q09	qcif.T	X	
Q10	qcif.S	X	
Q11	qcif.X	X	
Q12	qcif.D		X
Q13	qcif.I		X
Q14	qcif.G		X

25fps Scene Pools: A, D, G, I

30fps Scene Pools: J, K, P, S, T, U, V, W, X, L

Appendix IV HRCs Associated with Each Individual Experiment

This appendix contains the individual experiment designs. Bit rates are specified in kb/s, and frame rates in fps. Only codec type, not the specific model and implementation is listed. The packet loss rates (PLR) given below are nominal random packet loss rates in percent, without error correction or concealment. Manufacturers are intentionally not identified.

Test	Lab	HRC #	Codec	Bit Rate	Frame Rate	PLR	Other
V01	Psytechnics	0	None	N/A	25	0	reference
V01	Psytechnics	1	MPEG-4	2000	25	0	
V01	Psytechnics	2	VC1	1000	25	0	
V01	Psytechnics	3	MPEG-4	1000	25	0	
V01	Psytechnics	4	H.264	1000	25	0	
V01	Psytechnics	5	VC1	512	25	0	
V01	Psytechnics	6	RV10	512	25	0	
V01	Psytechnics	7	MPEG-4	512	25	0	
V01	Psytechnics	8	H.264	512	25	0	
V01	Psytechnics	9	VC1	320	12.5	0	
V01	Psytechnics	10	RV10	320	12.5	0	
V01	Psytechnics	11	MPEG-4	320	12.5	0	
V01	Psytechnics	12	VC1	128	5	0	
V01	Psytechnics	13	RV10	128	5	0	
V01	Psytechnics	14	MPEG-4	2000	25	2	random
V01	Psytechnics	15	MPEG-4	2000	25	2	bursty
V01	Psytechnics	16	MPEG-4	2000	25	5	bursty

Test	Lab	HRC #	Codec	Bit Rate	Frame Rate	PLR	Other
V02	NTT	0	None	N/A	30	0	reference
V02	NTT	1	MPEG-4	2000	30	0	
V02	NTT	2	MPEG-4	1000	30	0	
V02	NTT	3	MPEG-4	1000	15	0	
V02	NTT	4	MPEG-4	1000	15	0	
V02	NTT	5	MPEG-4	320	10	0	
V02	NTT	6	MPEG-4	128	15	0	
V02	NTT	7	MPEG-4	128	10	0	

V02	NTT	8	MPEG-4	128	5	0
V02	NTT	9	MPEG-4	4096	30	1
V02	NTT	10	MPEG-4	4096	30	2
V02	NTT	11	MPEG-4	4096	30	3
V02	NTT	12	MPEG-4	1024	30	1
V02	NTT	13	MPEG-4	1024	30	2
V02	NTT	14	MPEG-4	1024	30	4
V02	NTT	15	MPEG-4	320	30	2
V02	NTT	16	MPEG-4	320	30	4

Test	Lab	HRC #	Codec	Bit Rate	Frame Rate	PLR	Other
V03	NTT	0	None	N/A	30	0	reference
V03	NTT	1	RV10	4096	30	0	
V03	NTT	2	RV10	1024	30	0	
V03	NTT	3	RV10	1024	15	0	
V03	NTT	4	RV10	320	15	0	
V03	NTT	5	RV10	320	10	0	
V03	NTT	6	RV10	128	15	0	
V03	NTT	7	RV10	128	10	0	
V03	NTT	8	RV10	128	5	0	
V03	NTT	9	RV10	4096	30	1	
V03	NTT	10	RV10	4096	30	2	
V03	NTT	11	RV10	4096	30	4	
V03	NTT	12	RV10	1024	30	1	
V03	NTT	13	RV10	1024	30	2	
V03	NTT	14	RV10	1024	30	4	
V03	NTT	15	RV10	320	15	2	
V03	NTT	16	RV10	320	15	4	

Test	Lab	HRC #	Codec	Bit Rate	Frame Rate	PLR	Other
V04	NTT	0	None	N/A	30	0	reference
V04	NTT	1	H.264	4096	30	0	
V04	NTT	2	H.264	1024	30	0	
V04	NTT	3	H.264	1024	15	0	

V04	NTT	4	H.264	320	15	0
V04	NTT	5	H.264	320	10	0
V04	NTT	6	H.264	128	15	0
V04	NTT	7	H.264	128	10	0
V04	NTT	8	H.264	128	5	0
V04	NTT	9	H.264	4096	30	1
V04	NTT	10	H.264	4096	30	2
V04	NTT	11	H.264	4096	30	4
V04	NTT	12	H.264	1024	30	1
V04	NTT	13	H.264	1024	30	2
V04	NTT	14	H.264	1024	30	4
V04	NTT	15	H.264	1024	15	2
V04	NTT	16	H.264	1024	15	4

Test	Lab	HRC #	Codec	Bit Rate	Frame Rate	PLR	Other
V05	Yonsei	0	None	N/A	30	0	reference
V05	Yonsei	1	H.264	128	15	0	QuickTime 7.1
V05	Yonsei	2	H.264	320	15	0	QuickTime 7.1
V05	Yonsei	3	H.264	704	30	0	QuickTime 7.1
V05	Yonsei	4	H.264	1500	30	0	QuickTime 7.1
V05	Yonsei	5	H.264	3000	30	0	QuickTime 7.1
V05	Yonsei	6	MPEG-4	128	15	0	QuickTime 7.1
V05	Yonsei	7	MPEG-4	320	15	0	QuickTime 7.1
V05	Yonsei	8	MPEG-4	704	30	0	QuickTime 7.1
V05	Yonsei	9	MPEG-4	1500	30	0	QuickTime 7.1
V05	Yonsei	10	MPEG-4	3000	30	0	QuickTime 7.1
V05	Yonsei	11	RV10	128	15	0	Real Producer 11
V05	Yonsei	12	RV10	704	30	0	Real Producer 11
V05	Yonsei	13	RV10	3000	30	0	Real Producer 11
V05	Yonsei	14	VC1	320	15	0	Media Encoder 9
V05	Yonsei	15	VC1	704	30	0	Media Encoder 9
V05	Yonsei	16	VC1	1500	30	0	Media Encoder 9

Test	Lab	HRC #	Codec	Bit Rate	Frame Rate	PLR	Other

V06	Yonsei	0	None	N/A	30	0	reference
V06	Yonsei	1	MPEG-4	128	15	5	QuickTime 7.1
V06	Yonsei	2	MPEG-4	320	15	2	QuickTime 7.1
V06	Yonsei	3	MPEG-4	704	30	0	QuickTime 7.1
V06	Yonsei	4	MPEG-4	1500	30	0	QuickTime 7.1
V06	Yonsei	5	MPEG-4	3000	30	1	QuickTime 7.1
V06	Yonsei	6	H.264	128	15	0	QuickTime 7.1
V06	Yonsei	7	H.264	320	15	0	QuickTime 7.1
V06	Yonsei	8	H.264	1500	30	0	QuickTime 7.1
V06	Yonsei	9	H.264	3000	30	0	QuickTime 7.1
V06	Yonsei	10	H.264	704	30	7	QuickTime 7.1
V06	Yonsei	11	RV10	128	15	0	Real Producer 11
V06	Yonsei	12	RV10	704	30	0	Real Producer 11
V06	Yonsei	13	RV10	3000	30	0	Real Producer 11
V06	Yonsei	14	VC1	320	15	0	Media Encoder 9
V06	Yonsei	15	VC1	704	30	0	Media Encoder 9
V06	Yonsei	16	VC1	1500	30	0	Media Encoder 9

Test	Lab	HRC #	Codec	Bit Rate	Frame Rate	PLR	Other
V07	OPTICOM	0	None	N/A	25	0	reference
V07	OPTICOM	1	H.264	1024	25	0	
V07	OPTICOM	2	H.264	512	25	0	
V07	OPTICOM	3	H.264	512	12.5	0	
V07	OPTICOM	4	H.264	256	12.5	0	
V07	OPTICOM	5	H.264	256	8.33	0	
V07	OPTICOM	6	MPEG-4	1024	25	0	
V07	OPTICOM	7	MPEG-4	1024	12.5	0	
V07	OPTICOM	8	MPEG-4	512	12.5	0	
V07	OPTICOM	9	MPEG-4	512	8.33	0	
V07	OPTICOM	10	MPEG-4	256	8.33	0	
V07	OPTICOM	11	JPEG2000	1024	25	0	
V07	OPTICOM	12	JPEG2000	1024	12.5	0	
V07	OPTICOM	13	MPEG-4	1024	12.5	1	
V07	OPTICOM	14	MPEG-4	1024	12.5	3	
V07	OPTICOM	15	MPEG-4	1024	12.5	1	

V07	OPTICOM	16	MPEG-4	1024	12.5	0.5

Test	Lab	HRC #	Codec	Bit Rate	Frame Rate	PLR	Other
V08	SwissQual	0	None	N/A	25	0	reference
V08	SwissQual	1	H.264	2048	25	0	
V08	SwissQual	2	H.264	512	25	0	
V08	SwissQual	3	H.264	256	25	0	
V08	SwissQual	4	H.264	512	12.5	0	
V08	SwissQual	5	H.264	512	8.3	0	
V08	SwissQual	6	H.264	256	12.5	0	
V08	SwissQual	7	H.264	512	25	0.5	5% freeze
V08	SwissQual	8	H.264	512	25	1	10% freeze
V08	SwissQual	9	H.264	512	25	2.5	25% freeze
V08	SwissQual	10	H.264	512	25	0.125	
V08	SwissQual	11	H.264	512	25	0.5	
V08	SwissQual	12	MPEG-4	2048	25	0	
V08	SwissQual	13	MPEG-4	512	25	0	
V08	SwissQual	14	MPEG-4	256	25	0	
V08	SwissQual	15	MPEG-4	512	8.3	0	
V08	SwissQual	16	H.264	512	25	0	5 key frames/sec

Test	Lab	HRC #	Codec	Bit Rate	Frame Rate	PLR	Other
V09	NTIA	0	None	N/A	29.9	0	Reference
V09	NTIA	1	H.264	512	7.1	0	
V09	NTIA	2	H.264	768	7.5	0.3	3% random packet loss, error concealment
V09	NTIA	3	H.264	256	18.3	0.5	5% random packet loss, error concealment
V09	NTIA	4	H.264	384	18.1	0	
V09	NTIA	5	MPEG-4	VBR (3800)	14.1	0.2	up to 4 Mb/s including FEC bandwidth, burst packet loss, CIF
V09	NTIA	6	H.261	VBR (1800)	14.5	0.2	up to 2 Mb/s including FEC bandwidth, burst error
V09	NTIA	7	H.264	VBR (1800)	26.9	0	up to 2 Mb/s including FEC bandwidth
V09	NTIA	8	H.264	1536	29.7	0.1	1/2 clips have 0.2% randomly distributed

							packet loss
V09	NTIA	9	H.264	704	29.3	0.1	1/2 clips have 0.2% randomly distributed packet loss
V09	NTIA	10	H.264	448	27	0	on one clip, decoder cannot keep up with frame rate & displays bars briefly
V09	NTIA	11	MPEG-2	1000	28.6	0.25	1/2 clips have 0.5% randomly distributed packet loss
V09	NTIA	12	MPEG-2	512	20.1	0	
V09	NTIA	13	H.261	2000	20	0	
V09	NTIA	14	H.261	384	7.1	2	burst errors, CIF resolution
V09	NTIA	15	H.263	256	22.5	0	QCIF resolution
V09	NTIA	16	H.264	256	6.3	0	CIF resolution

Test	Lab	HRC #	Codec	Bit Rate	Frame Rate	PLR	Other
V10	FUB	0	None	N/A	25	0	Reference
V10	FUB	1	H.264	128	5	0	Deblocking =yes
V10	FUB	2	H.264	256	5	0	Deblocking = yes
V10	FUB	3	H.264	256	12.5	0	Deblocking = yes
V10	FUB	4	H.264	384	5	0	Deblocking = yes
V10	FUB	5	H.264	384	12.5	0	Deblocking = yes
V10	FUB	6	H.264	384	25	0	Deblocking = yes
V10	FUB	7	H.264	512	5	0	Deblocking =no
V10	FUB	8	H.264	512	5	0	Deblocking = yes
V10	FUB	9	H.264	512	12.5	0	Deblocking =no
V10	FUB	10	H.264	512	12.5	0	Deblocking = yes
V10	FUB	11	H.264	512	25	0	Deblocking = no
V10	FUB	12	H.264	512	25	0	Deblocking = yes
V10	FUB	13	H.264	750	12.5	0	Deblocking = yes
V10	FUB	14	H.264	750	25	0	Deblocking = no
V10	FUB	15	H.264	750	25	0	Deblocking = yes
V10	FUB	16	H.264	1024	25	0	Deblocking = yes

Test	Lab	HRC #	Codec	Bit Rate	Frame Rate	PLR	Other
V11	IRCCyN	0	None	N/A	25	0	Reference

V11	IRCCyN	1	x264	128	5	0	Pre- & post-processing
V11	IRCCyN	2	x264	128	10	0	Pre- & post-processing
V11	IRCCyN	3	x264	256	12.5	0	Pre- & post-processing
V11	IRCCyN	4	x264	256	10	0	Pre- & post-processing
V11	IRCCyN	5	x264	512	12.5	0	Pre- & post-processing
V11	IRCCyN	6	x264	512	25	0	Pre- & post-processing
V11	IRCCyN	7	H.264	256	25	0	Pre- & post-processing
V11	IRCCyN	8	H.264	512	12.5	0	Pre- & post-processing
V11	IRCCyN	9	H.264	512	25	0	Pre- & post-processing
V11	IRCCyN	10	H.264	704	25	0	
V11	IRCCyN	11	H.264	1000	25	0	
V11	IRCCyN	12	SVC	256	12.5	0	Pre- & post-processing
V11	IRCCyN	13	SVC	256	25	0	Pre- & post-processing
V11	IRCCyN	14	SVC	704	25	0	Pre- & post-processing
V11	IRCCyN	15	SVC	704	25	0	Pre- & post-processing
V11	IRCCyN	16	SVC	1000	25	0	Pre- & post-processing

Test	Lab	HRC #	Codec	Bit Rate	Frame Rate	PLR	Other
V12	NTIA	0	None	N/A	27.9	0	Reference
V12	NTIA	1	MPEG-2	512	4.6	0	
V12	NTIA	2	H.264	256	6	0	
V12	NTIA	3	MPEG-2	≈448	28.6	0	QCIF resolution
V12	NTIA	4	MPEG-2	≈1000	27.6	0	
V12	NTIA	5	MPEG-2	3000	19.7	0	
V12	NTIA	6	H.264	256	9.9	0	
V12	NTIA	7	Cinepak	2500	28.1	0	RGB colorspace conversion
V12	NTIA	8	WMV9	2000	9.9	0	
V12	NTIA	9	DivX	704	9.9	0	noise filter
V12	NTIA	10	DivX	≈320	20.1	0	VBR
V12	NTIA	11	MPEG-4	128	20.1	0	SIF resolution
V12	NTIA	12	Sorenson	320	4.6	0	RGB colorspace conversion
V12	NTIA	13	Chain of codecs	* (128)	11.2	0	Multiple transformations
V12	NTIA	14	H.263	128 to 448	4.8	0	VBR
V12	NTIA	15	H.264	256	15	0	240x180 resolution
V12	NTIA	16	Theora	643 to 2014	27.9	0	VBR

Test	Lab	HRC #	Codec	Bit Rate	Frame Rate	PLR	Other
V13	FUB	0	None	N/A	30	0	Reference
V13	FUB	1	H.264	384	15	0	Deblocking = yes
V13	FUB	2	H.264	512	5	0	Deblocking = yes
V13	FUB	3	H.264	512	15	0	Deblocking = no
V13	FUB	4	H.264	512	15	0	Deblocking = yes
V13	FUB	5	H.264	512	30	0	Deblocking = no
V13	FUB	6	H.264	512	30	0	Deblocking = yes
V13	FUB	7	H.264	750	15	0	Deblocking = yes
V13	FUB	8	H.264	750	30	0	Deblocking = no
V13	FUB	9	H.264	750	30	0	Deblocking = yes
V13	FUB	10	H.264	1024	15	0	Deblocking = yes
V13	FUB	11	H.264	1024	30	0	Deblocking = yes
V13	FUB	12	H.264	1024	30	0	Deblocking = no
V13	FUB	13	H.264	2048	30	0	Deblocking = yes
V13	FUB	14	H.264	2048	30	0	Deblocking = no
V13	FUB	15	H.264	4096	30	0	Deblocking = yes
V13	FUB	16	H.264	6000	30	0	Deblocking = yes

Test	Lab	HRC #	Codec	Bit Rate	Frame Rate	PLR	Other
C01	Psytechnics	0	none	N/A	25	0	reference
C01	Psytechnics	1	RV10	448	25	0	
C01	Psytechnics	2	VC1	448	25	0	
C01	Psytechnics	3	MPEG-4	448	25	0	
C01	Psytechnics	4	VC1	320	25	0	
C01	Psytechnics	5	RV10	320	12.5	0	
C01	Psytechnics	6	VC1	320	12.5	0	
C01	Psytechnics	7	MPEG-4	320	12.5	0	
C01	Psytechnics	8	RV10	128	12.5	0	
C01	Psytechnics	9	VC1	128	12.5	0	
C01	Psytechnics	10	MPEG-4	128	12.5	0	
C01	Psytechnics	11	VC1	64	5	0	
C01	Psytechnics	12	MPEG-4	64	5	0	

C01	Psytechnics	13	MPEG-4	704	25	1	periodic
C01	Psytechnics	14	MPEG-4	704	25	2	bursty
C01	Psytechnics	15	MPEG-4	704	25	5	bursty
C01	Psytechnics	16	MPEG-4	704	25	10	bursty

Test	Lab	HRC #	Codec	Bit Rate	Frame Rate	PLR	Other
C02	NTT	0	none	N/A	30	0	reference
C02	NTT	1	MPEG-4	704	30	0	
C02	NTT	2	MPEG-4	320	30	0	
C02	NTT	3	MPEG-4	320	15	0	
C02	NTT	4	MPEG-4	128	15	0	
C02	NTT	5	MPEG-4	128	10	0	
C02	NTT	6	MPEG-4	64	15	0	
C02	NTT	7	MPEG-4	64	10	0	
C02	NTT	8	MPEG-4	64	5	0	
C02	NTT	9	MPEG-4	704	30	1	
C02	NTT	10	MPEG-4	704	30	2	
C02	NTT	11	MPEG-4	704	30	4	
C02	NTT	12	MPEG-4	320	30	1	
C02	NTT	13	MPEG-4	320	30	2	
C02	NTT	14	MPEG-4	320	30	4	
C02	NTT	15	MPEG-4	128	30	2	
C02	NTT	16	MPEG-4	128	30	4	

Test	Lab	HRC #	Codec	Bit Rate	Frame Rate	PLR	Other
C03	NTT	0	none	N/A	30	0	reference
C03	NTT	1	RV10	704	30	0	
C03	NTT	2	RV10	320	30	0	
C03	NTT	3	RV10	320	15	0	
C03	NTT	4	RV10	128	15	0	
C03	NTT	5	RV10	128	10	0	
C03	NTT	6	RV10	64	15	0	
C03	NTT	7	RV10	64	10	0	
C03	NTT	8	RV10	64	5	0	

C03	NTT	9	RV10	704	30	1
C03	NTT	10	RV10	704	30	2
C03	NTT	11	RV10	704	30	4
C03	NTT	12	RV10	320	30	1
C03	NTT	13	RV10	320	30	2
C03	NTT	14	RV10	320	30	4
C03	NTT	15	RV10	320	15	2
C03	NTT	16	RV10	320	15	4

Test	Lab	HRC #	Codec	Bit Rate	Frame Rate	PLR	Other
C04	NTT	0	none	N/A	30	0	reference
C04	NTT	1	H.264	704	30	0	
C04	NTT	2	H.264	320	30	0	
C04	NTT	3	H.264	320	15	0	
C04	NTT	4	H.264	128	15	0	
C04	NTT	5	H.264	128	10	0	
C04	NTT	6	H.264	64	15	0	
C04	NTT	7	H.264	64	10	0	
C04	NTT	8	H.264	64	5	0	
C04	NTT	9	H.264	704	30	1	
C04	NTT	10	H.264	704	30	2	
C04	NTT	11	H.264	704	30	4	
C04	NTT	12	H.264	320	30	1	
C04	NTT	13	H.264	320	30	2	
C04	NTT	14	H.264	320	30	4	
C04	NTT	15	H.264	320	15	2	
C04	NTT	16	H.264	320	15	4	

Test	Lab	HRC #	Codec	Bit Rate	Frame Rate	PLR	Other
C05	Yonsei	0	none	N/A	30	0	reference
C05	Yonsei	1	H.264	64	10	0	
C05	Yonsei	2	H.264	128	15	0	
C05	Yonsei	3	H.264	320	15	0	
C05	Yonsei	4	H.264	320	30	0	

C05	Yonsei	5	H.264	704	30	0
C05	Yonsei	6	MPEG-4	64	10	0
C05	Yonsei	7	MPEG-4	128	15	0
C05	Yonsei	8	MPEG-4	320	15	0
C05	Yonsei	9	MPEG-4	320	30	0
C05	Yonsei	10	MPEG-4	704	30	0
C05	Yonsei	11	RV10	128	15	0
C05	Yonsei	12	RV10	320	30	0
C05	Yonsei	13	RV10	704	30	0
C05	Yonsei	14	RV10	64	10	0
C05	Yonsei	15	RV10	320	15	0
C05	Yonsei	16	RV10	704	30	0

Test	Lab	HRC #	Codec	Bit Rate	Frame Rate	PLR	Other
C06	Yonsei	0	none	N/A	30	0	reference
C06	Yonsei	1	MPEG-4	64	10	0	
C06	Yonsei	2	MPEG-4	128	15	5	
C06	Yonsei	3	MPEG-4	320	15	2	
C06	Yonsei	4	MPEG-4	320	30	1	
C06	Yonsei	5	MPEG-4	704	30	0	
C06	Yonsei	6	H.264	64	10	0	
C06	Yonsei	7	H.264	128	15	0	
C06	Yonsei	8	H.264	320	15	0	
C06	Yonsei	9	H.264	320	30	7	
C06	Yonsei	10	H.264	704	30	0	
C06	Yonsei	11	RV10	64	10	0	
C06	Yonsei	12	RV10	320	15	0	
C06	Yonsei	13	RV10	704	30	0	
C06	Yonsei	14	VC1	64	10	0	
C06	Yonsei	15	VC1	128	15	0	
C06	Yonsei	16	VC1	704	30	0	

Test	Lab	HRC #	Codec	Bit Rate	Frame Rate	PLR	Other
C07	KDDI	0	none	N/A	30	0	reference

C07	KDDI	1	H.264	64	5	0
C07	KDDI	2	H.264	64	5	0
C07	KDDI	3	H.264	128	5	0
C07	KDDI	4	H.264	256	5	0
C07	KDDI	5	H.264	256	10	0
C07	KDDI	6	H.264	384	10	0
C07	KDDI	7	H.264	384	15	0
C07	KDDI	8	MPEG-4	128	5	0
C07	KDDI	9	MPEG-4	256	5	0
C07	KDDI	10	MPEG-4	512	10	0
C07	KDDI	11	MPEG-4	768	10	0
C07	KDDI	12	MPEG-4	768	15	0
C07	KDDI	13	H.264	256	10	1
C07	KDDI	14	H.264	256	10	2
C07	KDDI	15	MPEG-4	512	10	1
C07	KDDI	16	MPEG-4	512	10	2

Test	Lab	HRC #	Codec	Bit Rate	Frame Rate	PLR	Other
C08	KDDI	0	none	N/A	30	0	reference
C08	KDDI	1	H.264	64	5	0	
C08	KDDI	2	H.264	64	5	0	
C08	KDDI	3	H.264	128	5	0	
C08	KDDI	4	H.264	256	5	0	
C08	KDDI	5	H.264	256	10	0	
C08	KDDI	6	H.264	384	10	0	
C08	KDDI	7	H.264	384	15	0	
C08	KDDI	8	MPEG-4	128	5	0	
C08	KDDI	9	MPEG-4	256	5	0	
C08	KDDI	10	MPEG-4	512	10	0	
C08	KDDI	11	MPEG-4	768	10	0	
C08	KDDI	12	MPEG-4	768	15	0	
C08	KDDI	13	H.264	256	10	1	
C08	KDDI	14	H.264	256	10	2	
C08	KDDI	15	MPEG-4	512	10	1	
C08	KDDI	16	MPEG-4	512	10	2	

Test	Lab	HRC #	Codec	Bit Rate	Frame Rate	PLR	Other
C09	KDDI	0	none	N/A	30	0	reference
C09	KDDI	1	H.264	64	5	0	
C09	KDDI	2	H.264	64	5	0	
C09	KDDI	3	H.264	128	5	0	
C09	KDDI	4	H.264	256	5	0	
C09	KDDI	5	H.264	256	10	0	
C09	KDDI	6	H.264	384	10	0	
C09	KDDI	7	H.264	384	15	0	
C09	KDDI	8	MPEG-4	128	5	0	
C09	KDDI	9	MPEG-4	256	5	0	
C09	KDDI	10	MPEG-4	512	10	0	
C09	KDDI	11	MPEG-4	768	10	0	
C09	KDDI	12	MPEG-4	768	15	0	
C09	KDDI	13	H.264	256	10	1	
C09	KDDI	14	H.264	256	10	2	
C09	KDDI	15	MPEG-4	512	10	1	
C09	KDDI	16	MPEG-4	512	10	2	

Test	Lab	HRC #	Codec	Bit Rate	Frame Rate	PLR	Other
C10	Symmetricom	0	none	N/A	25	0	reference
C10	Symmetricom	1	H.264	64	12.5	0	
C10	Symmetricom	2	H.264	128	12.5	0	
C10	Symmetricom	3	H.264	256	12.5	0	
C10	Symmetricom	4	H.264	256	25	0	
C10	Symmetricom	5	H.264	512	25	0	
C10	Symmetricom	6	WMV9	64	12.5	0	
C10	Symmetricom	7	WMV9	128	12.5	0	
C10	Symmetricom	8	WMV9	256	12.5	0	
C10	Symmetricom	9	WMV9	256	25	0	
C10	Symmetricom	10	WMV9	512	25	0	
C10	Symmetricom	11	MPEG-4	256	25	0	
C10	Symmetricom	12	MPEG-4	512	25	0	

C10	Symmetricom	13	JPEG2000	256	8.33	0
C10	Symmetricom	14	JPEG2000	256	12.5	0
C10	Symmetricom	15	JPEG2000	512	12.5	0
C10	Symmetricom	16	JPEG2000	768	25	0

Test	Lab	HRC #	Codec	Bit Rate	Frame Rate	PLR	Other
C11	NTIA	0	none	N/A	29.7	0	reference
C11	NTIA	1	H.264	704	29.7	0	
C11	NTIA	2	MPEG-1	320	29.7	0	
C11	NTIA	3	MPEG-1	448	29.7	0	
C11	NTIA	4	DivX	448	29.7	0	
C11	NTIA	5	DivX	192	29.7	0	
C11	NTIA	6	Cinepak	320	29.7	0	RGB Colorspace conversion
C11	NTIA	7	Sorenson 3	64	29.7	0	RGB Colorspace conversion
C11	NTIA	8	MPEG-4-ISO	128	29.7	0	
C11	NTIA	9	H.264	384	automatic (24.4)	1	error concealment
C11	NTIA	10	H.264	128	automatic (9.5)	0	
C11	NTIA	11	H.263	256	automatic (5.9)	0	
C11	NTIA	12	H.264	256	automatic (13.4)	0.2	burst error with FEC
C11	NTIA	13	H.263	128	automatic (8.0)	0	
C11	NTIA	14	H.263	384	automatic (14.5)	0	
C11	NTIA	15	H.264	704	automatic (29.4)	0.25	1/2 clips 0.5% packet loss
C11	NTIA	16	H.261	384	automatic (21.4)	0	QCIF resolution

Test	Lab	HRC #	Codec	Bit Rate	Frame Rate	PLR	Other
C12	CRC	0	none	N/A	30	0	reference
C12	CRC	1	H.264	768	30	0	
C12	CRC	2	H.264	768	20	0	
C12	CRC	3	H.264	256	20	0	
C12	CRC	4	H.264	256	15	0	
C12	CRC	5	H.264	128	20	0	
C12	CRC	6	H.264	128	15	0	
C12	CRC	7	H.264	768	30	0.5	
C12	CRC	8	H.264	768	30	1	

C12	CRC	9	H.264	768	30	2
C12	CRC	10	H.264	768	30	4
C12	CRC	11	H.264	768	30	8
C12	CRC	12	H.264	256	20	0.5
C12	CRC	13	H.264	256	20	1
C12	CRC	14	H.264	256	20	2
C12	CRC	15	H.264	256	20	4
C12	CRC	16	H.264	256	20	8

Test	Lab	HRC #	Codec	Bit Rate	Frame Rate	PLR	Deblocking
C13	FUB	0	none	N/A	30	0	reference
C13	FUB	1	H.264	64	5	0	yes
C13	FUB	2	H.264	96	5	0	yes
C13	FUB	3	H.264	128	5	0	yes
C13	FUB	4	H.264	192	5	0	yes
C13	FUB	5	H.264	256	5	0	yes
C13	FUB	6	H.264	128	25	0	yes
C13	FUB	7	H.264	192	25	0	yes
C13	FUB	8	H.264	256	25	0	yes
C13	FUB	9	H.264	384	25	0	yes
C13	FUB	10	H.264	128	5	0	no
C13	FUB	11	H.264	192	5	0	no
C13	FUB	12	H.264	256	5	0	no
C13	FUB	13	H.264	384	5	0	no
C13	FUB	14	H.264	128	25	0	no
C13	FUB	15	H.264	192	25	0	no
C13	FUB	16	H.264	256	25	0	no

Test	Lab	HRC #	Codec	Bit Rate	Frame Rate	PLR	Other
C14	Acreo	0	none	N/A	25	0	reference
C14	Acreo	1	MPEG-4	300	25	0	
C14	Acreo	2	MPEG-4	300	25	2	
C14	Acreo	3	MPEG-4	200	12.5	0	
C14	Acreo	4	MPEG-4	200	12.5	2	

C14	Acreo	5	MPEG-4	200	12.5	6
C14	Acreo	6	MPEG-4	90	8.33	2
C14	Acreo	7	MPEG-4	90	8.33	6
C14	Acreo	8	MPEG-4	90	8.33	12
C14	Acreo	9	H.264	300	25	0
C14	Acreo	10	H.264	300	25	2
C14	Acreo	11	H.264	200	12.5	0
C14	Acreo	12	H.264	200	12.5	2
C14	Acreo	13	H.264	200	12.5	6
C14	Acreo	14	H.264	90	8.33	2
C14	Acreo	15	H.264	90	8.33	6
C14	Acreo	16	H.264	90	8.33	12

Test	Lab	HRC #	Codec	Bit Rate	Frame Rate	PLR	Other
Q01	Psytechnics	0	none	N/A	25	0	reference
Q01	Psytechnics	1	H.264	320	25	0	
Q01	Psytechnics	2	MPEG-4	320	25	0	
Q01	Psytechnics	3	MPEG-4	320	12.5	0	
Q01	Psytechnics	4	H.263	320	12.5	0	
Q01	Psytechnics	5	H.264	128	12.5	0	
Q01	Psytechnics	6	MPEG-4	128	12.5	0	
Q01	Psytechnics	7	H.263	128	12.5	0	
Q01	Psytechnics	8	H.264	64	12.5	0	
Q01	Psytechnics	9	MPEG-4	64	12.5	0	
Q01	Psytechnics	10	H.263	64	12.5	0	
Q01	Psytechnics	11	H.264	32	5	0	
Q01	Psytechnics	12	MPEG-4	32	5	0	
Q01	Psytechnics	13	H.263	32	5	0	
Q01	Psytechnics	14	MPEG-4	320	12.5	2	periodic PL
Q01	Psytechnics	15	MPEG-4	320	12.5	5	bursty PL
Q01	Psytechnics	16	MPEG-4	320	12.5	1	periodic PL

Test	Lab	HRC #	Codec	Bit Rate	Frame Rate	PLR	Other
Q02	NTT	0	none	N/A	30	0	reference

Q02	NTT	1	MPEG-4	320	30	0
Q02	NTT	2	MPEG-4	128	30	0
Q02	NTT	3	MPEG-4	128	10	0
Q02	NTT	4	MPEG-4	64	10	0
Q02	NTT	5	MPEG-4	64	2.5	0
Q02	NTT	6	MPEG-4	32	10	0
Q02	NTT	7	MPEG-4	32	2.5	0
Q02	NTT	8	MPEG-4	16	2.5	0
Q02	NTT	9	MPEG-4	128	30	1
Q02	NTT	10	MPEG-4	128	30	2
Q02	NTT	11	MPEG-4	128	30	4
Q02	NTT	12	MPEG-4	64	30	1
Q02	NTT	13	MPEG-4	64	30	2
Q02	NTT	14	MPEG-4	64	30	4
Q02	NTT	15	MPEG-4	32	30	2
Q02	NTT	16	MPEG-4	32	30	4

Test	Lab	HRC #	Codec	Bit Rate	Frame Rate	PLR	Other
Q03	NTT	0	none	N/A	30	0	reference
Q03	NTT	1	RV10	320	30	0	
Q03	NTT	2	RV10	128	30	0	
Q03	NTT	3	RV10	128	10	0	
Q03	NTT	4	RV10	64	10	0	
Q03	NTT	5	RV10	64	2.5	0	
Q03	NTT	6	RV10	32	10	0	
Q03	NTT	7	RV10	32	2.5	0	
Q03	NTT	8	RV10	16	2.5	0	
Q03	NTT	9	RV10	128	30	1	
Q03	NTT	10	RV10	128	30	2	
Q03	NTT	11	RV10	128	30	4	
Q03	NTT	12	RV10	64	30	1	
Q03	NTT	13	RV10	64	30	2	
Q03	NTT	14	RV10	64	30	4	
Q03	NTT	15	RV10	64	10	2	
Q03	NTT	16	RV10	64	10	4	

Test	Lab	HRC #	Codec	Bit Rate	Frame Rate	PLR	Other
Q04	NTT	0	none	N/A	30	0	reference
Q04	NTT	1	H.264	320	30	0	
Q04	NTT	2	H.264	128	30	0	
Q04	NTT	3	H.264	128	10	0	
Q04	NTT	4	H.264	64	10	0	
Q04	NTT	5	H.264	64	2.5	0	
Q04	NTT	6	H.264	32	10	0	
Q04	NTT	7	H.264	32	2.5	0	
Q04	NTT	8	H.264	16	2.5	0	
Q04	NTT	9	H.264	128	30	1	
Q04	NTT	10	H.264	128	30	2	
Q04	NTT	11	H.264	128	30	4	
Q04	NTT	12	H.264	64	30	1	
Q04	NTT	13	H.264	64	30	2	
Q04	NTT	14	H.264	64	30	4	
Q04	NTT	15	H.264	64	10	2	
Q04	NTT	16	H.264	64	10	4	

Test	Lab	HRC #	Codec	Bit Rate	Frame Rate	PLR	Other
Q05	Yonsei	0	none	N/A	30	0	reference
Q05	Yonsei	1	MPEG-4	32	5	5	
Q05	Yonsei	2	MPEG-4	32	10	0	
Q05	Yonsei	3	MPEG-4	64	10	2	
Q05	Yonsei	4	MPEG-4	128	15	0	
Q05	Yonsei	5	MPEG-4	320	30	1	
Q05	Yonsei	6	H.264	32	5	0	
Q05	Yonsei	7	H.264	32	10	0	
Q05	Yonsei	8	H.264	64	10	0	
Q05	Yonsei	9	H.264	128	15	7	Darwin streaming server capture
Q05	Yonsei	10	H.264	320	30	0	
Q05	Yonsei	11	RV10	16	5	0	
Q05	Yonsei	12	RV10	64	10	0	

Q05	Yonsei	13	RV10	320	30	0
Q05	Yonsei	14	VC1	32	10	0
Q05	Yonsei	15	VC1	128	15	0
Q05	Yonsei	16	VC1	320	30	0

Test	Lab	HRC #	Codec	Bit Rate	Frame Rate	PLR	Other
Q06	Yonsei	0	none	N/A	30	0	reference
Q06	Yonsei	1	H.264	32	5	0	
Q06	Yonsei	2	H.264	32	10	0	
Q06	Yonsei	3	H.264	64	10	0	
Q06	Yonsei	4	H.264	128	15	0	
Q06	Yonsei	5	H.264	320	30	0	
Q06	Yonsei	6	MPEG-4	32	5	0	
Q06	Yonsei	7	MPEG-4	32	10	0	
Q06	Yonsei	8	MPEG-4	64	10	0	
Q06	Yonsei	9	MPEG-4	128	15	0	
Q06	Yonsei	10	MPEG-4	320	30	0	
Q06	Yonsei	11	RV10	16	5	0	
Q06	Yonsei	12	RV10	32	10	0	
Q06	Yonsei	13	RV10	320	30	0	
Q06	Yonsei	14	VC1	32	5	0	
Q06	Yonsei	15	VC1	64	10	0	
Q06	Yonsei	16	VC1	320	30	0	

Test	Lab	HRC #	Codec	Bit Rate	Frame Rate	PLR	Other
Q07	KDDI	0	none	N/A	30	0	reference
Q07	KDDI	1	H.264	16	3	0	
Q07	KDDI	2	H.264	16	5	0	
Q07	KDDI	3	H.264	32	5	0	
Q07	KDDI	4	H.264	64	5	0	
Q07	KDDI	5	H.264	64	10	0	
Q07	KDDI	6	H.264	128	10	0	
Q07	KDDI	7	H.264	128	15	0	
Q07	KDDI	8	MPEG-4	32	5	0	

Q07	KDDI	9	MPEG-4	64	5	0
Q07	KDDI	10	MPEG-4	128	10	0
Q07	KDDI	11	MPEG-4	256	10	0
Q07	KDDI	12	MPEG-4	256	15	0
Q07	KDDI	13	H.264	128	10	1
Q07	KDDI	14	H.264	128	10	2
Q07	KDDI	15	MPEG-4	128	10	1
Q07	KDDI	16	MPEG-4	128	10	2

Test	Lab	HRC #	Codec	Bit Rate	Frame Rate	PLR	Other
Q08	KDDI	0	none	N/A	30	0	reference
Q08	KDDI	1	H.264	16	3	0	
Q08	KDDI	2	H.264	16	5	0	
Q08	KDDI	3	H.264	32	5	0	
Q08	KDDI	4	H.264	64	5	0	
Q08	KDDI	5	H.264	64	10	0	
Q08	KDDI	6	H.264	128	10	0	
Q08	KDDI	7	H.264	128	15	0	
Q08	KDDI	8	MPEG-4	32	5	0	
Q08	KDDI	9	MPEG-4	64	5	0	
Q08	KDDI	10	MPEG-4	128	10	0	
Q08	KDDI	11	MPEG-4	256	10	0	
Q08	KDDI	12	MPEG-4	256	15	0	
Q08	KDDI	13	H.264	128	10	1	
Q08	KDDI	14	H.264	128	10	2	
Q08	KDDI	15	MPEG-4	128	10	1	
Q08	KDDI	16	MPEG-4	128	10	2	

Test	Lab	HRC #	Codec	Bit Rate	Frame Rate	PLR	Other
Q09	KDDI	0	none	N/A	30	0	reference
Q09	KDDI	1	H.264	16	3	0	
Q09	KDDI	2	H.264	16	5	0	
Q09	KDDI	3	H.264	32	5	0	
Q09	KDDI	4	H.264	64	5	0	

Q09	KDDI	5	H.264	64	10	0
Q09	KDDI	6	H.264	128	10	0
Q09	KDDI	7	H.264	128	15	0
Q09	KDDI	8	MPEG-4	32	5	0
Q09	KDDI	9	MPEG-4	64	5	0
Q09	KDDI	10	MPEG-4	128	10	0
Q09	KDDI	11	MPEG-4	256	10	0
Q09	KDDI	12	MPEG-4	256	15	0
Q09	KDDI	13	H.264	128	10	1
Q09	KDDI	14	H.264	128	10	2
Q09	KDDI	15	MPEG-4	128	10	1
Q09	KDDI	16	MPEG-4	128	10	2

Test	Lab	HRC #	Codec	Bit Rate	Frame Rate	PLR	Other
Q10	FT	0	none	N/A	30	0	reference
Q10	FT	1					
Q10	FT	2					
Q10	FT	3					
Q10	FT	4					
Q10	FT	5					
Q10	FT	6					
Q10	FT	7					
Q10	FT	8					
Q10	FT	9					
Q10	FT	10					
Q10	FT	11					
Q10	FT	12					
Q10	FT	13					
Q10	FT	14					
Q10	FT	15					
Q10	FT	16					

Test	Lab	HRC #	Codec	Bit Rate	Frame Rate	PLR	Other
Q11	CRC	0	none	N/A	30	0	reference

Q11	CRC	1	MPEG-4	256	20	0
Q11	CRC	2	MPEG-4	256	15	0
Q11	CRC	3	MPEG-4	128	15	0
Q11	CRC	4	MPEG-4	128	10	0
Q11	CRC	5	MPEG-4	64	10	0
Q11	CRC	6	MPEG-4	64	7.5	0
Q11	CRC	7	MPEG-4	256	20	0.5
Q11	CRC	8	MPEG-4	256	20	1
Q11	CRC	9	MPEG-4	256	20	2
Q11	CRC	10	MPEG-4	256	20	4
Q11	CRC	11	MPEG-4	256	20	8
Q11	CRC	12	MPEG-4	128	15	0.5
Q11	CRC	13	MPEG-4	128	15	1
Q11	CRC	14	MPEG-4	128	15	2
Q11	CRC	15	MPEG-4	128	15	4
Q11	CRC	16	MPEG-4	128	15	8

Test	Lab	HRC #	Codec	Bit Rate	Frame Rate	PLR	Other
Q12	Acreo	0	none	N/A	25	0	reference
Q12	Acreo	1	MPEG-4	200	25	0	
Q12	Acreo	2	MPEG-4	200	25	2	
Q12	Acreo	3	MPEG-4	90	12.5	0	
Q12	Acreo	4	MPEG-4	90	12.5	2	
Q12	Acreo	5	MPEG-4	90	12.5	6	
Q12	Acreo	6	MPEG-4	40	8.33	2	
Q12	Acreo	7	MPEG-4	40	8.33	6	
Q12	Acreo	8	MPEG-4	40	8.33	12	
Q12	Acreo	9	H.264	200	25	0	
Q12	Acreo	10	H.264	200	25	2	
Q12	Acreo	11	H.264	90	12.5	0	
Q12	Acreo	12	H.264	90	12.5	2	
Q12	Acreo	13	H.264	90	12.5	6	
Q12	Acreo	14	H.264	40	8.33	2	
Q12	Acreo	15	H.264	40	8.33	6	
Q12	Acreo	16	H.264	40	8.33	12	

Test	Lab	HRC #	Codec	Bit Rate	Frame Rate	PLR	Other
Q13	NTIA	0	none	N/A	25	0	reference
Q13	NTIA	1	H.264	32	12	0	
Q13	NTIA	2	MPEG-4 AVC	16	12	0	noise reduction, color correct
Q13	NTIA	3	MPEG-1	320	25	0	conturing & de-noising
Q13	NTIA	4	MPEG-1	192	25	0	
Q13	NTIA	5	DivX	128	8	0	noise reduction
Q13	NTIA	6	DivX	32	5	0	
Q13	NTIA	7	Cinepak	128	8	0	
Q13	NTIA	8	Sorenson 3	16	8	0	RGB conversion required
Q13	NTIA	9	MPEG-4-ISO	64	12	0	
Q13	NTIA	10	H.264	256	automatic(10)	0.5	error concealment
Q13	NTIA	11	H.264	320	automatic(12)	0.5	error concealment
Q13	NTIA	12	H.264	128	automatic(8)	0	
Q13	NTIA	13	H.263	128	automatc(8)	0	
Q13	NTIA	14	H.264	320	automatic(12)	0	
Q13	NTIA	15	H.261	256	automatic(12)	2	1/2 clips have burst errors
Q13	NTIA	16	H.261	64	automatic(10)	2	1/2 clips have burst errors

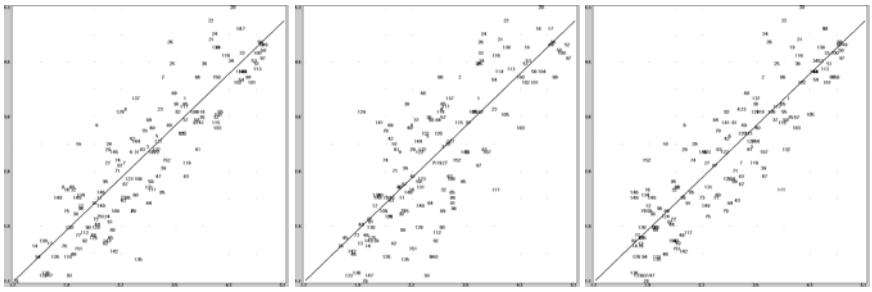
Test	Lab	HRC #	Codec	Bit Rate	Frame Rate	PLR	Other
Q14	FT	0	none	N/A	25	0	reference
Q14	FT	1					
Q14	FT	2					
Q14	FT	3					
Q14	FT	4					
Q14	FT	5					
Q14	FT	6					
Q14	FT	7					
Q14	FT	8					
Q14	FT	9					
Q14	FT	10					
Q14	FT	11					
Q14	FT	12					

Q14 FT 13
Q14 FT 14
Q14 FT 15
Q14 FT 16

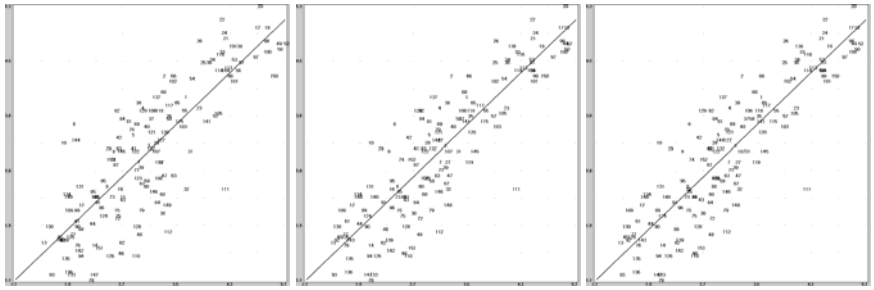
Appendix V Plots

Appendix V.1 VGA Plots

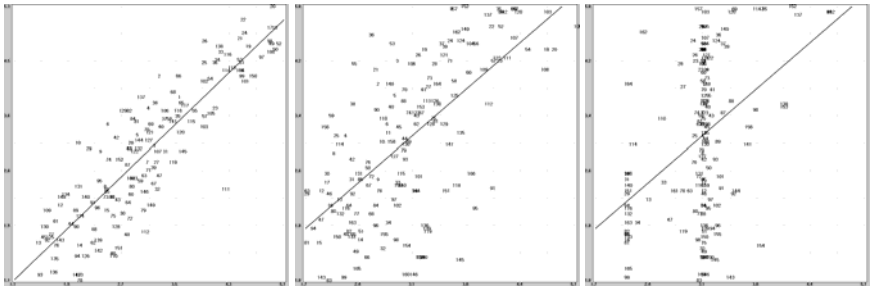
(a) NTT (b) OPTICOM (c) Psytechnics(FR)



(d) Yonsei_FR (e) Yonsei_RR10k (f) Yonsei_RR64k



(g) Yonsei_RR128k (h) Psytechnics(NR) (i) SwissQual(NR)



(j) PSNR(NTT) (k) PSNR(NTIA) (l) PSNR(Yonsei)

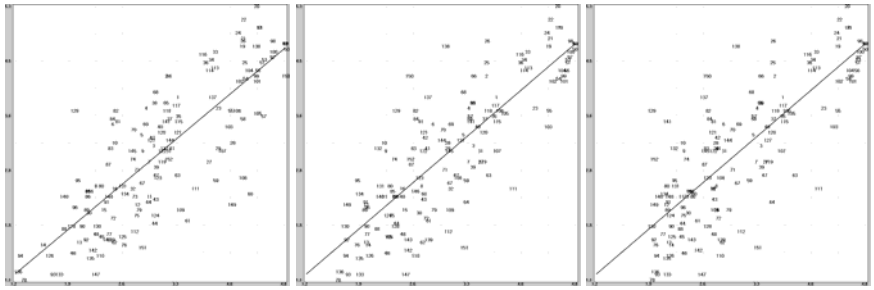
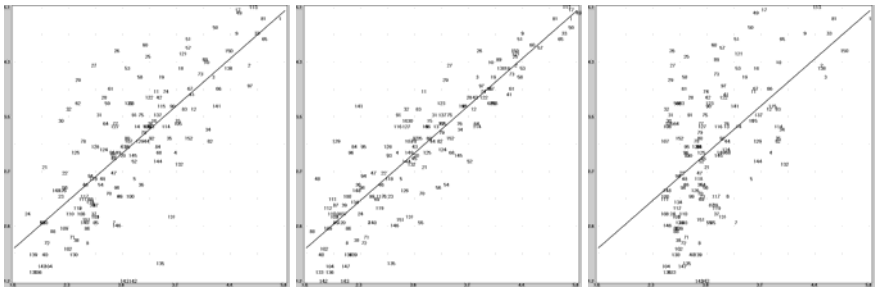
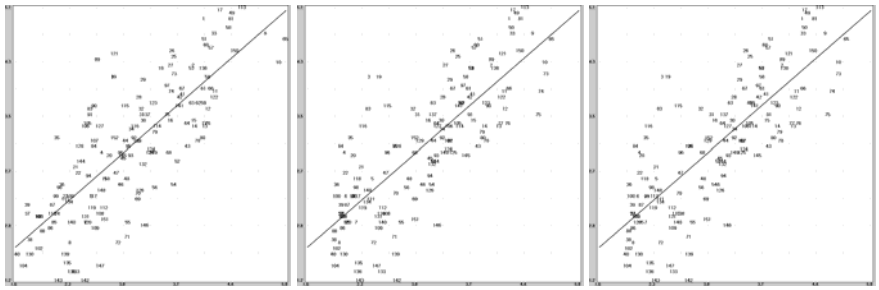


Fig.V.1 VGA01

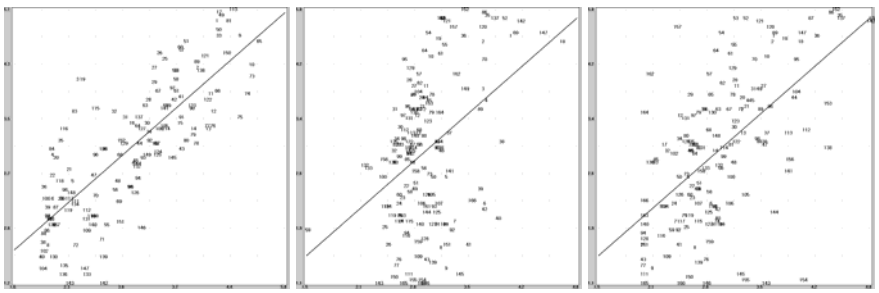
(a) NTT (b) OPTICOM (c) Psytechnics(FR)



(d) Yonsei_FR (e) Yonsei_RR10k (f) Yonsei_RR64k



(g) Yonsei_RR128k (h) Psytechnics(NR) (i) SwissQual(NR)



(j) PSNR(NTT) (k) PSNR(NTIA) (l) PSNR(Yonsei)

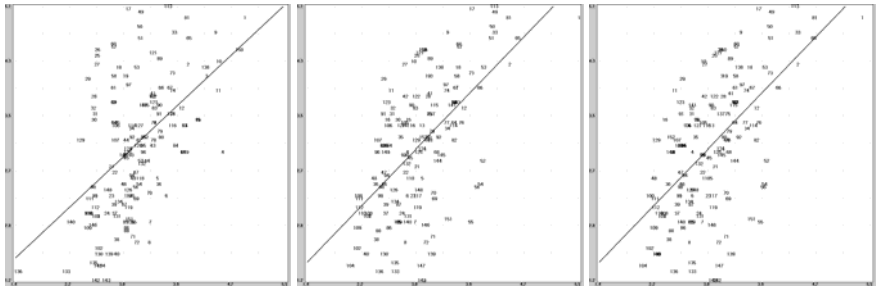
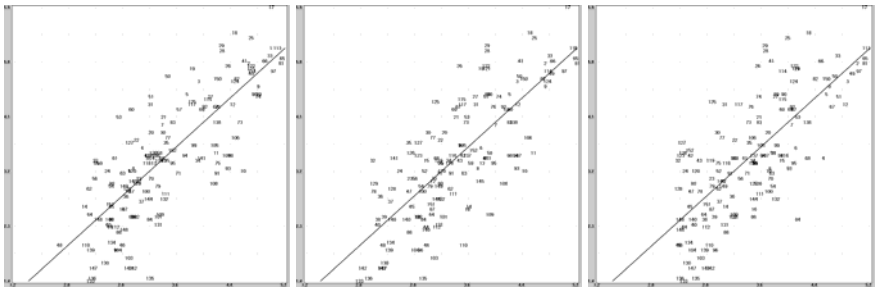
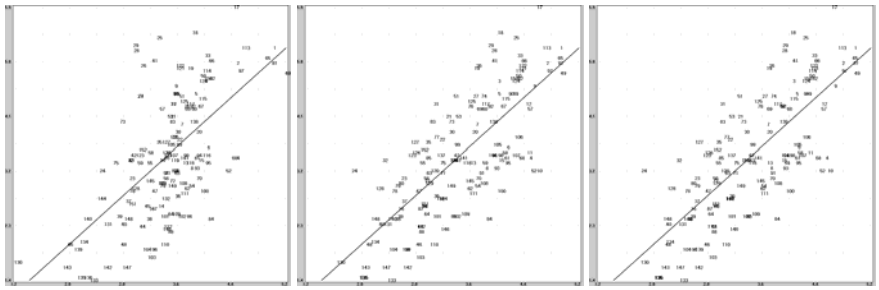


Fig.V.2 VGA02

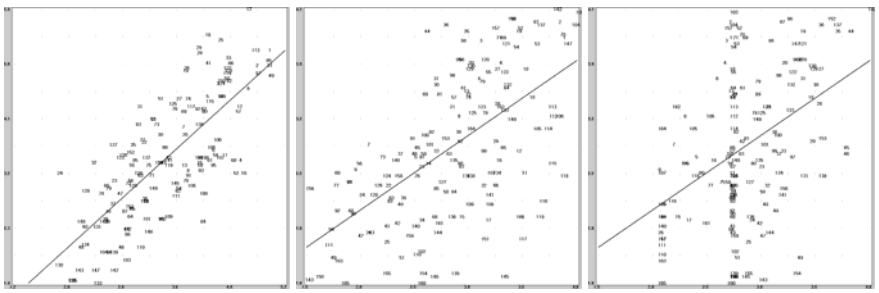
(a) NTT (b) OPTICOM (c) Psytechnics(FR)



(d) Yonsei_FR (e) Yonsei_RR10k (f) Yonsei_RR64k



(g) Yonsei_RR128k (h) Psytechnics(NR) (i) SwissQual(NR)



(j) PSNR(NTT) (k) PSNR(NTIA) (l) PSNR(Yonsei)

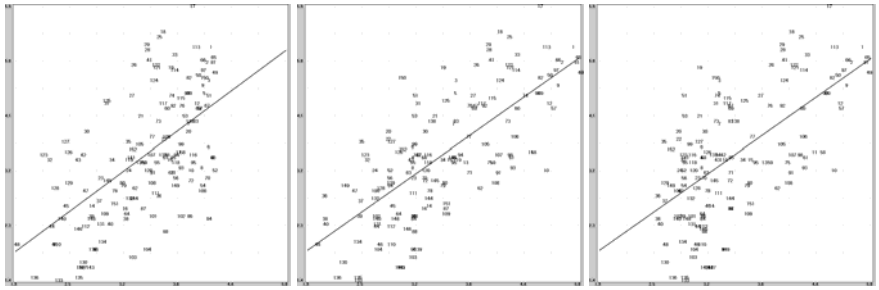
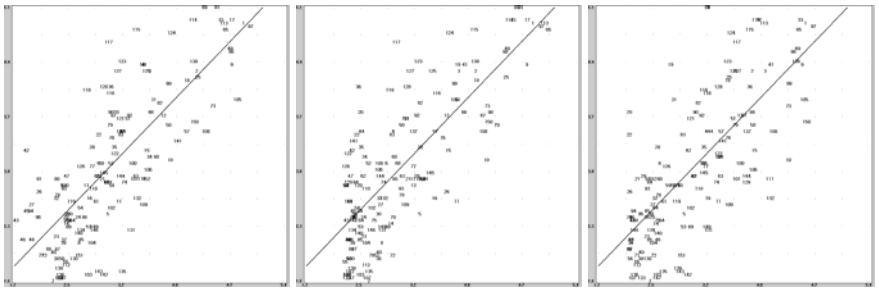
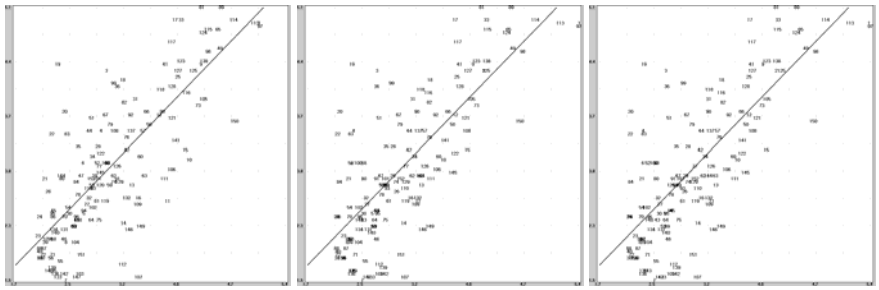


Fig. V.3 VGA03

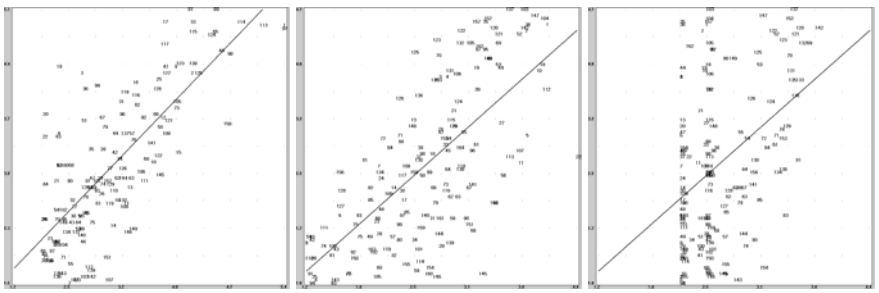
(a) NTT (b) OPTICOM (c) Psytechnics(FR)



(d) Yonsei_FR (e) Yonsei_RR10k (f) Yonsei_RR64k



(g) Yonsei_RR128k (h) Psytechnics(NR) (i) SwissQual(NR)



(j) PSNR(NTT) (k) PSNR(NTIA) (l) PSNR(Yonsei)

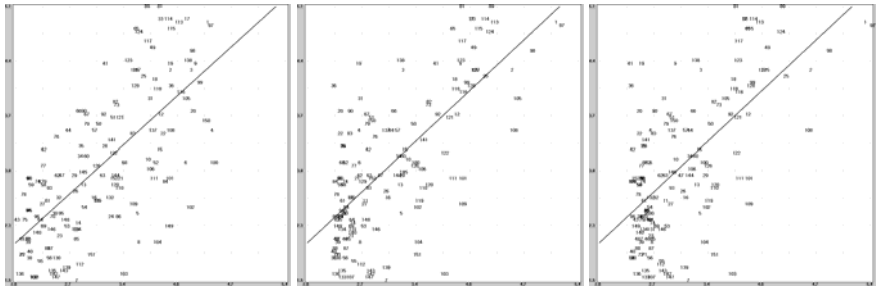
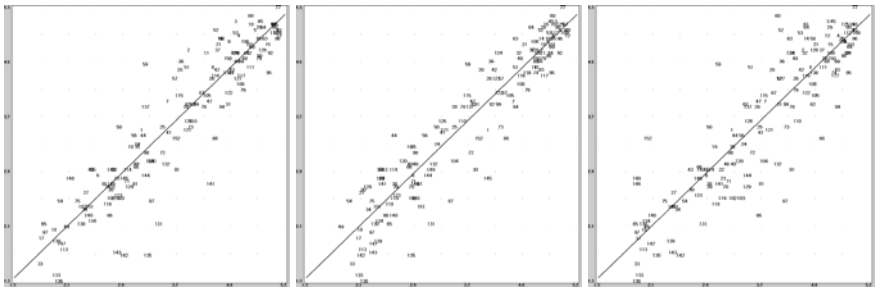
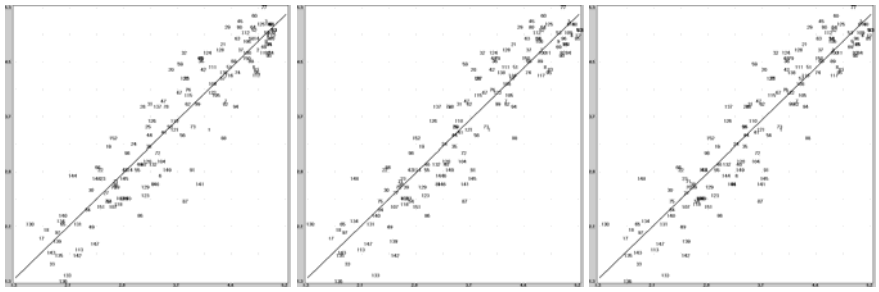


Fig. V.4 VGA04

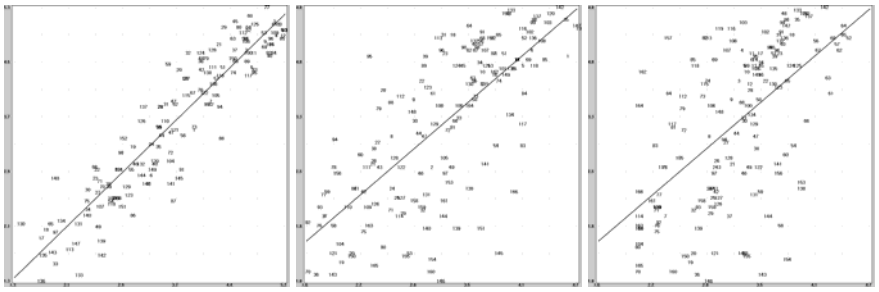
(a) NTT (b) OPTICOM (c) Psytechnics(FR)



(d) Yonsei_FR (e) Yonsei_RR10k (f) Yonsei_RR64k



(g) Yonsei_RR128k (h) Psytechnics(NR) (i) SwissQual(NR)



(j) PSNR(NTT) (k) PSNR(NTIA) (l) PSNR(Yonsei)

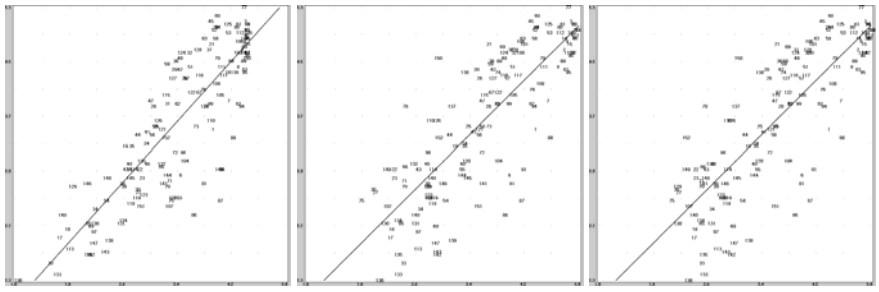
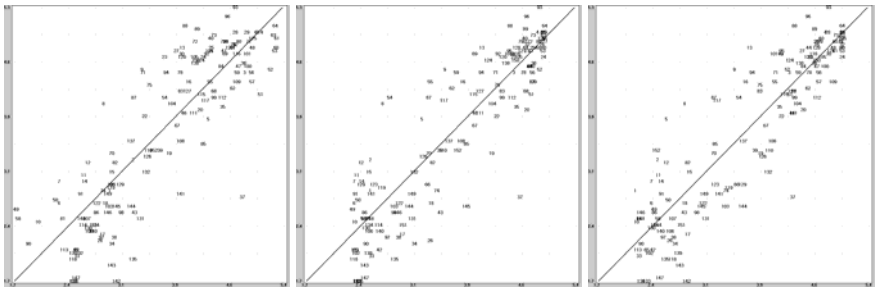
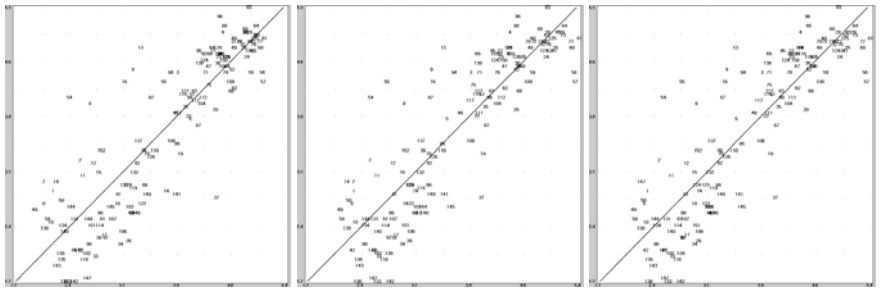


Fig. V.5 VGA05

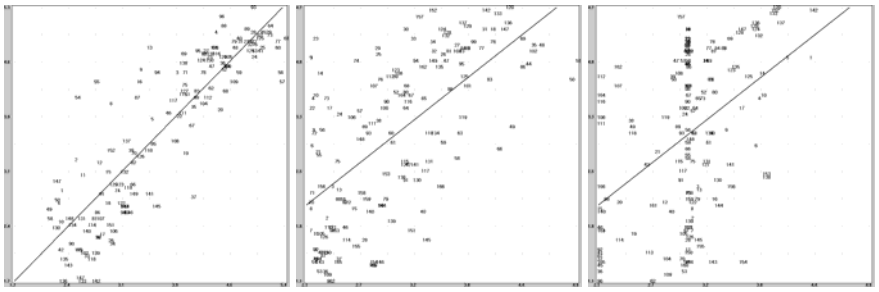
(a) NTT (b) OPTICOM (c) Psytechnics(FR)



(d) Yonsei_FR (e) Yonsei_RR10k (f) Yonsei_RR64k



(g) Yonsei_RR128k (h) Psytechnics(NR) (i) SwissQual(NR)



(j) PSNR(NTT) (k) PSNR(NTIA) (l) PSNR(Yonsei)

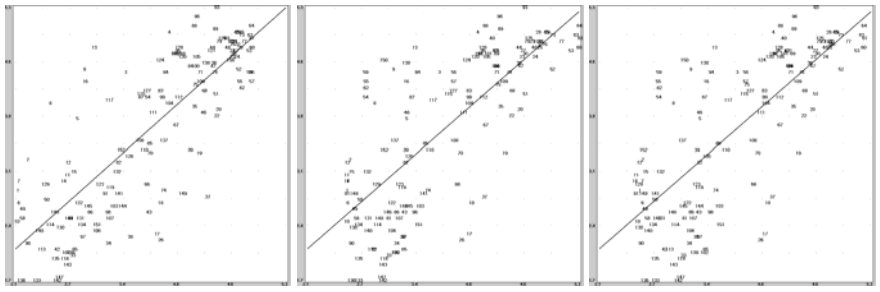
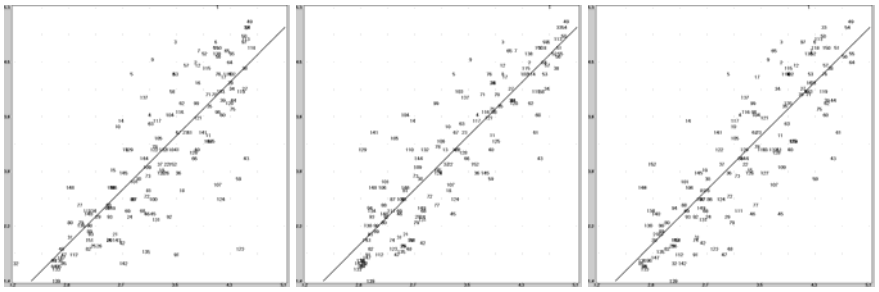
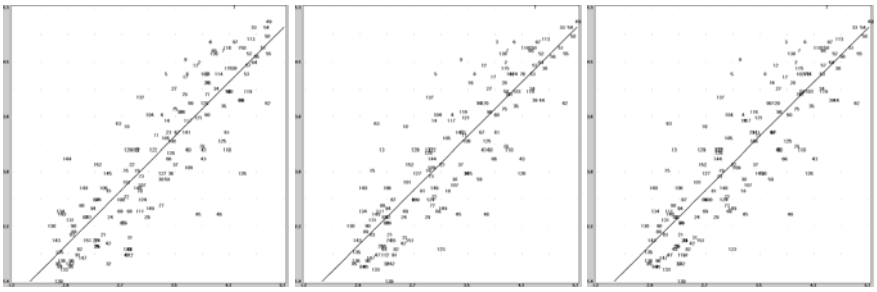


Fig. V.6 VGA06

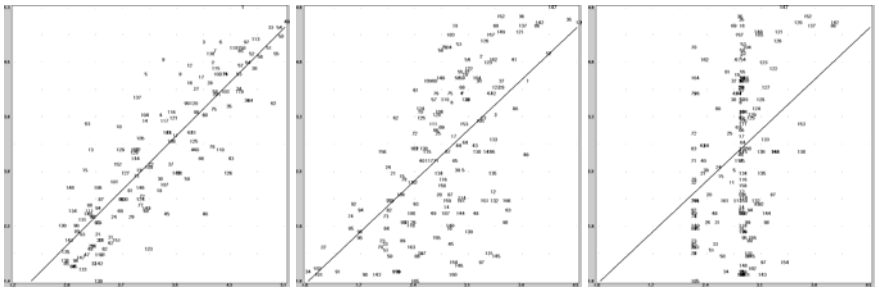
(a) NTT (b) OPTICOM (c) Psytechnics(FR)



(d) Yonsei_FR (e) Yonsei_RR10k (f) Yonsei_RR64k



(g) Yonsei_RR128k (h) Psytechnics(NR) (i) SwissQual(NR)



(j) PSNR(NTT) (k) PSNR(NTIA) (l) PSNR(Yonsei)

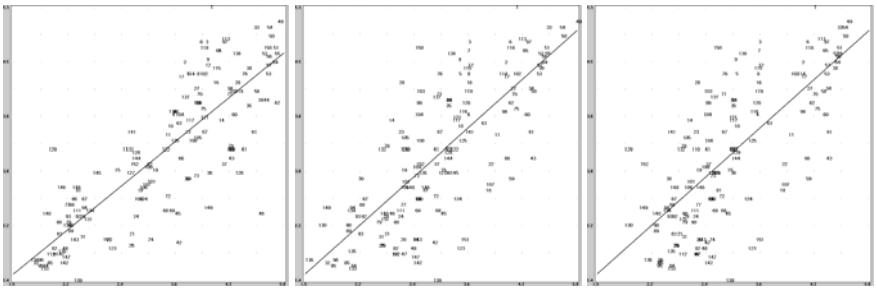
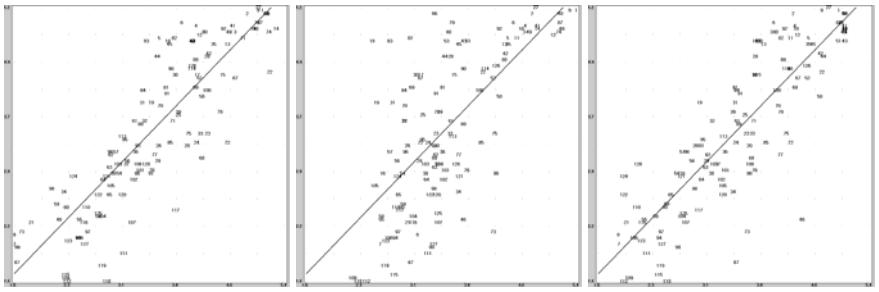
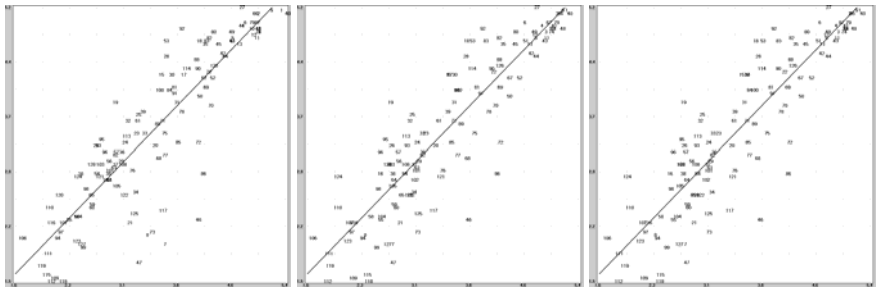


Fig. V.7 VGA07

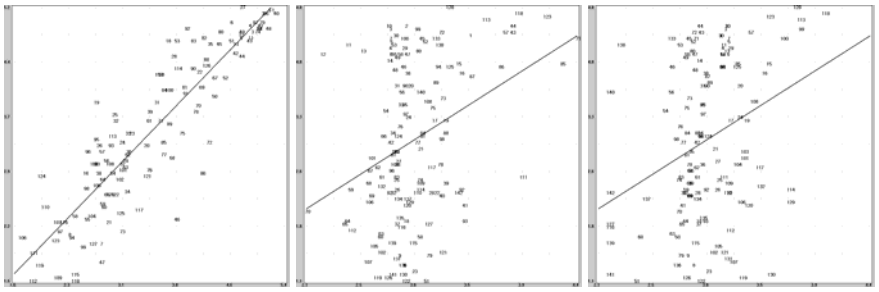
(a) NTT (b) OPTICOM (c) Psytechnics(FR)



(d) Yonsei_FR (e) Yonsei_RR10k (f) Yonsei_RR64k



(g) Yonsei_RR128k (h) Psytechnics(NR) (i) SwissQual(NR)



(j) PSNR(NTT) (k) PSNR(NTIA) (l) PSNR(Yonsei)

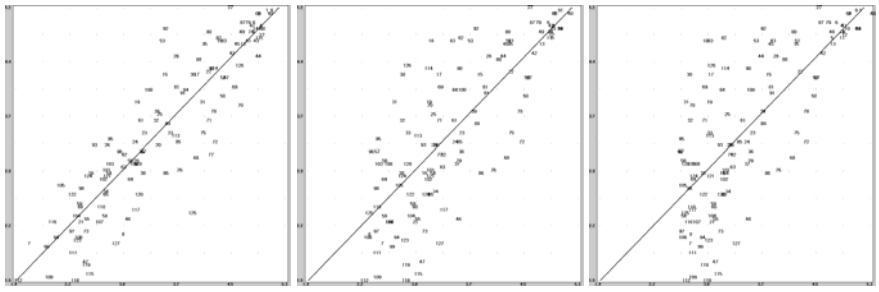
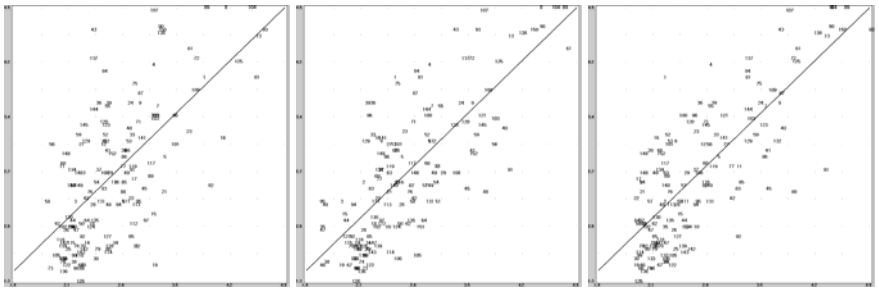
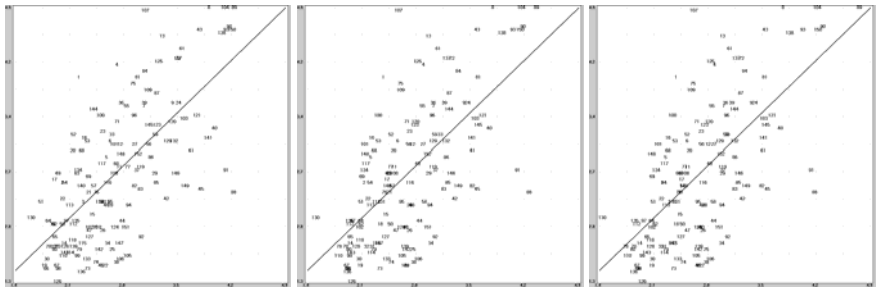


Fig.V.8 VGA08

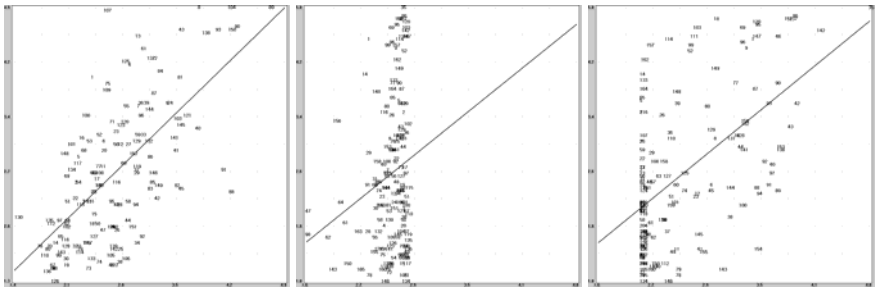
(a) NTT (b) OPTICOM (c) Psytechnics(FR)



(d) Yonsei_FR (e) Yonsei_RR10k (f) Yonsei_RR64k



(g) Yonsei_RR128k (h) Psytechnics(NR) (i) SwissQual(NR)



(j) PSNR(NTT) (k) PSNR(NTIA) (l) PSNR(Yonsei)

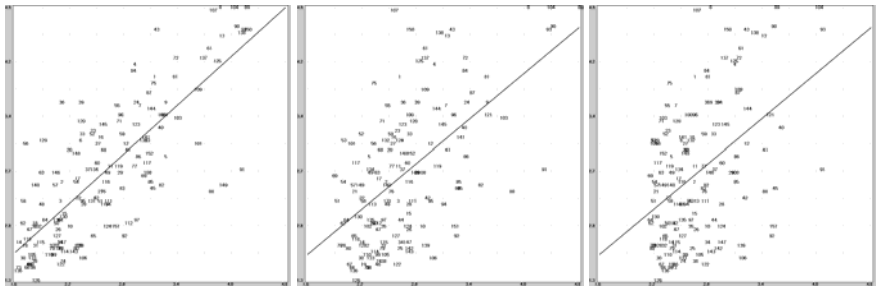
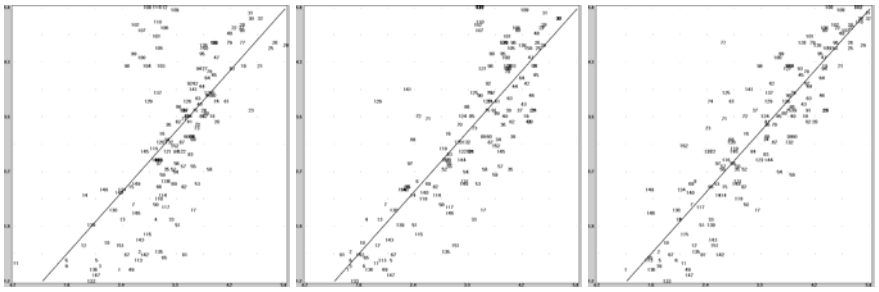
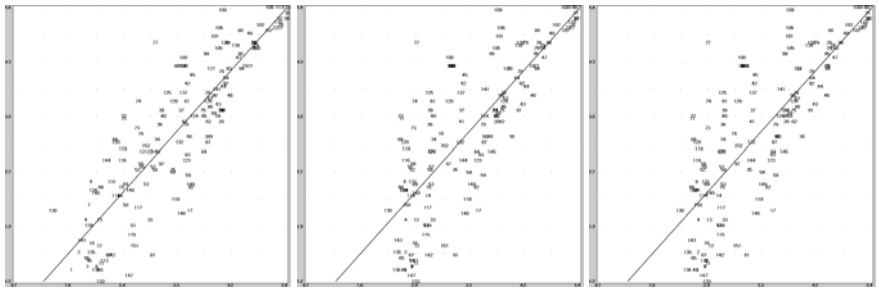


Fig.V.9 VGA09

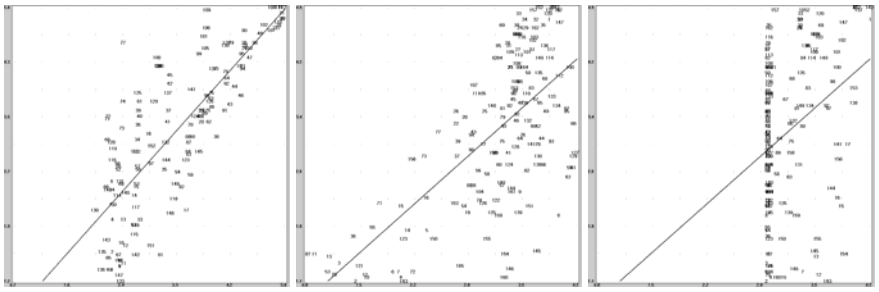
(a) NTT (b) OPTICOM (c) Psytechnics(FR)



(d) Yonsei_FR (e) Yonsei_RR10k (f) Yonsei_RR64k



(g) Yonsei_RR128k (h) Psytechnics(NR) (i) SwissQual(NR)



(j) PSNR(NTT) (k) PSNR(NTIA) (l) PSNR(Yonsei)

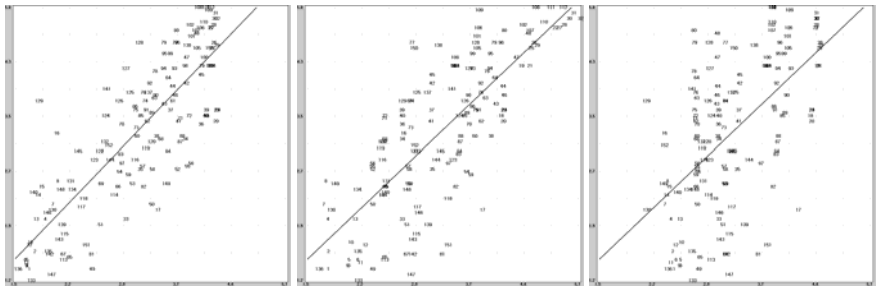
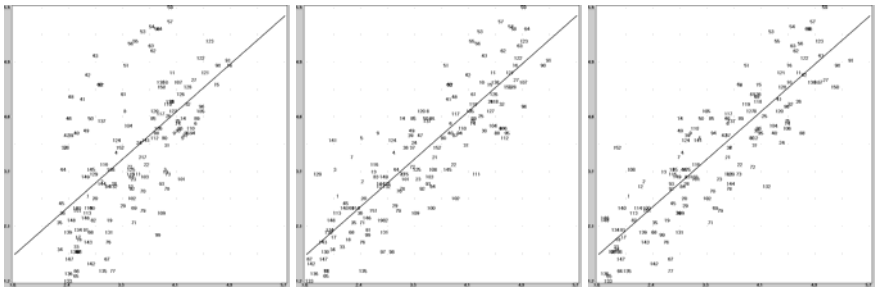
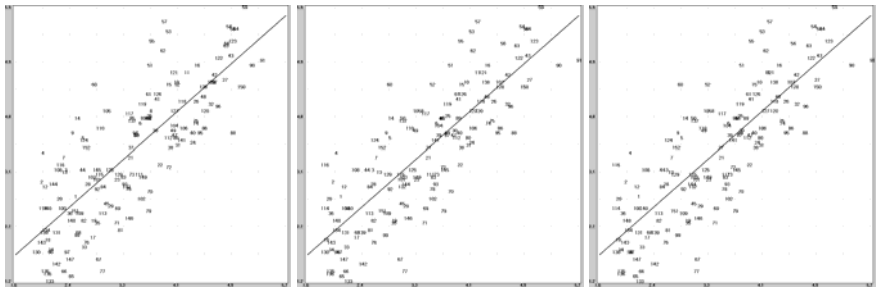


Fig.V.10 VGA10

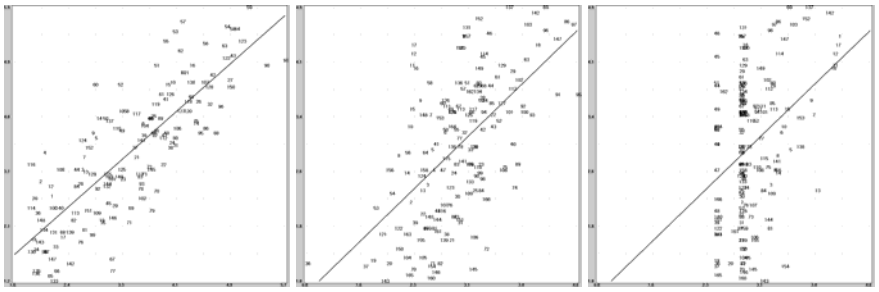
(a) NTT (b) OPTICOM (c) Psytechnics(FR)



(d) Yonsei_FR (e) Yonsei_RR10k (f) Yonsei_RR64k



(g) Yonsei_RR128k (h) Psytechnics(NR) (i) SwissQual(NR)



(j) PSNR(NTT) (k) PSNR(NTIA) (l) PSNR(Yonsei)

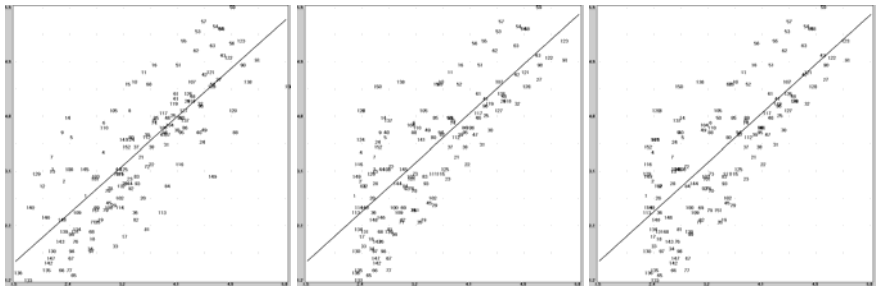
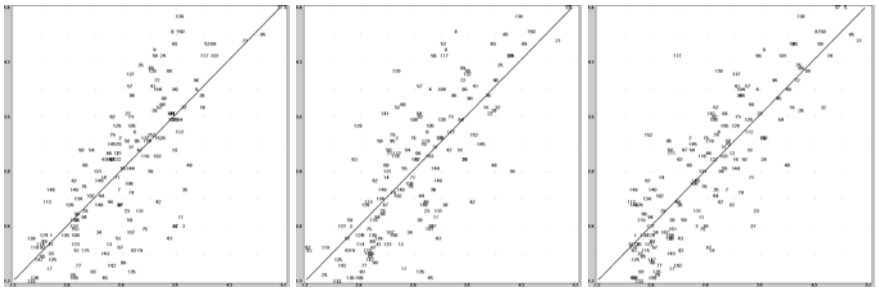
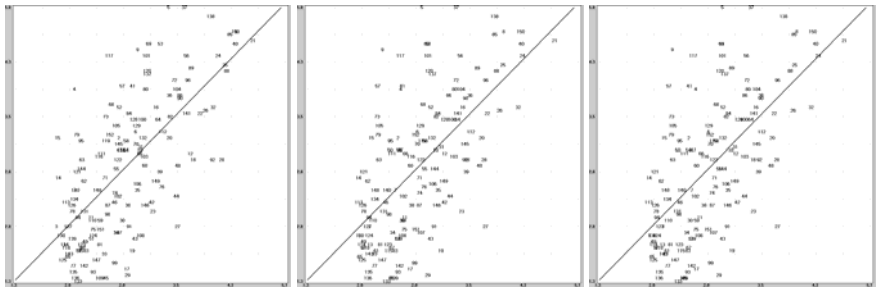


Fig.V.11 VGA11

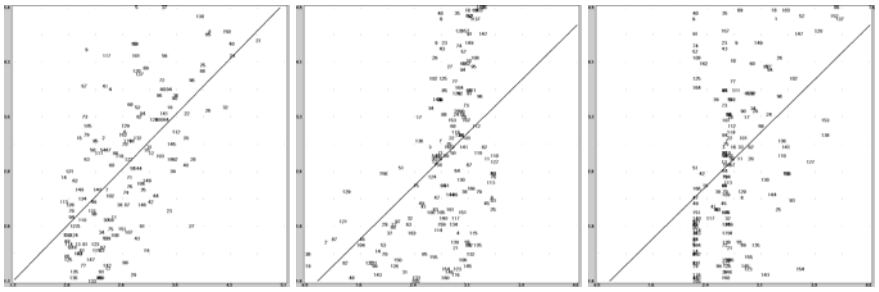
(a) NTT (b) OPTICOM (c) Psytechnics(FR)



(d) Yonsei_FR (e) Yonsei_RR10k (f) Yonsei_RR64k



(g) Yonsei_RR128k (h) Psytechnics(NR) (i) SwissQual(NR)



(j) PSNR(NTT) (k) PSNR(NTIA) (l) PSNR(Yonsei)

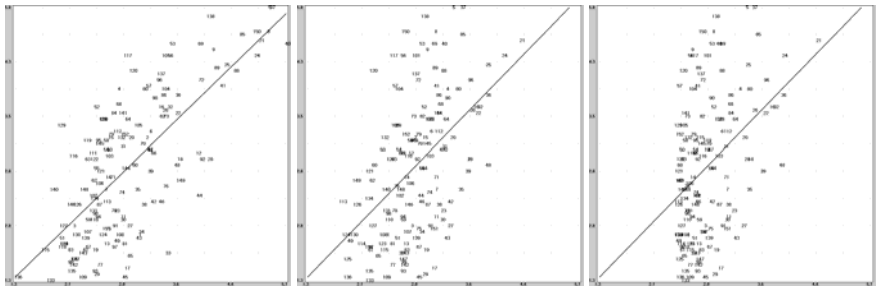
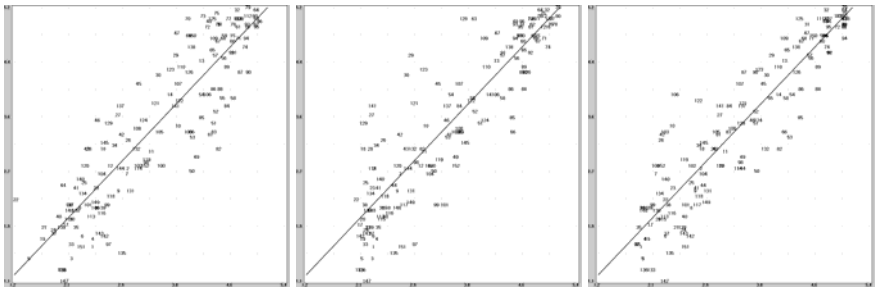
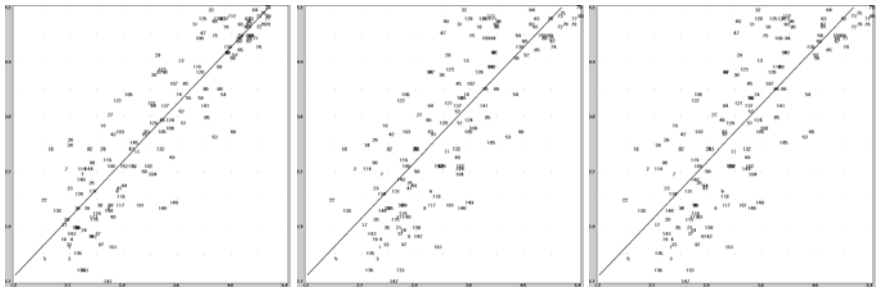


Fig.V.12 VGA12

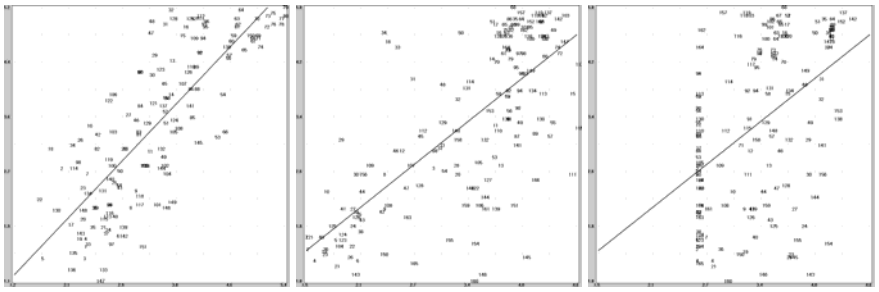
(a) NTT (b) OPTICOM (c) Psytechnics(FR)



(d) Yonsei_FR (e) Yonsei_RR10k (f) Yonsei_RR64k



(g) Yonsei_RR128k (h) Psytechnics(NR) (i) SwissQual(NR)



(j) PSNR(NTT) (k) PSNR(NTIA) (l) PSNR(Yonsei)

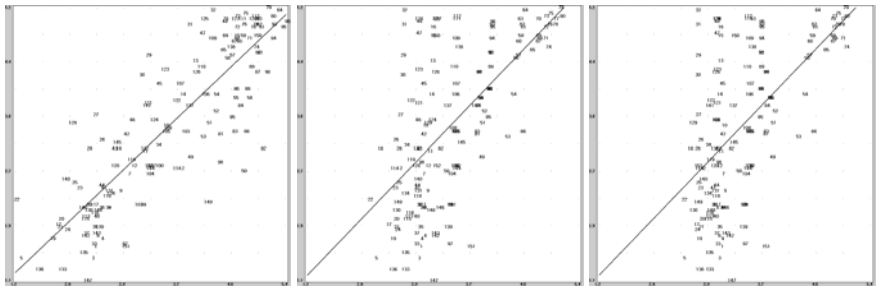
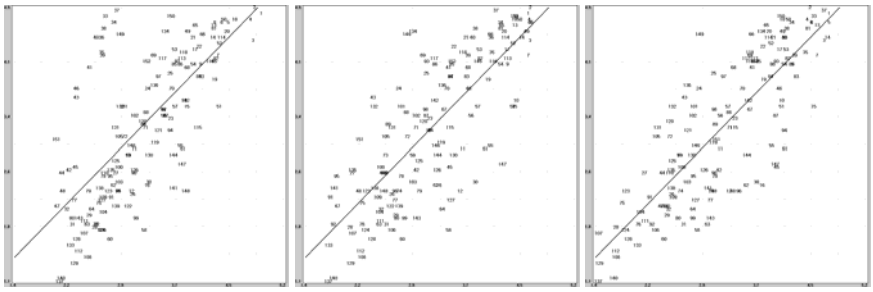


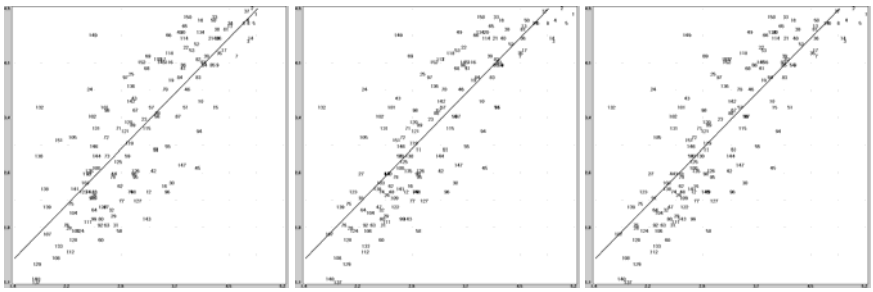
Fig. V.13 VGA13

Appendix V.2 CIF Plots

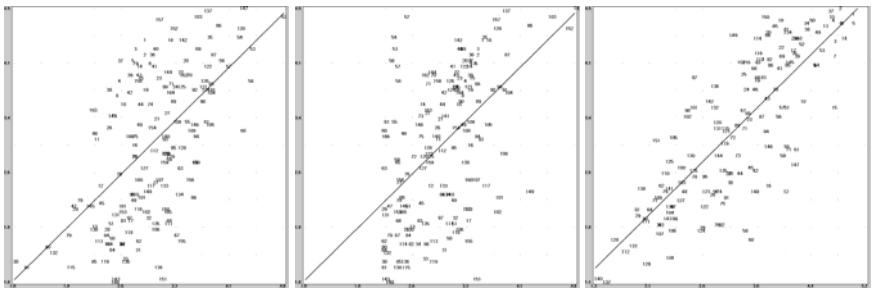
(a) NTT (b) OPTICOM (c) Psytechnics(FR)



(d) Yonsei_FR (e) Yonsei_RR10k (f) Yonsei_RR64k



(g) Psytechnics(NR) (h) SwissQual(NR) (i) PSNR(NTT)



(j) PSNR(NTIA) (k) PSNR(Yonsei)

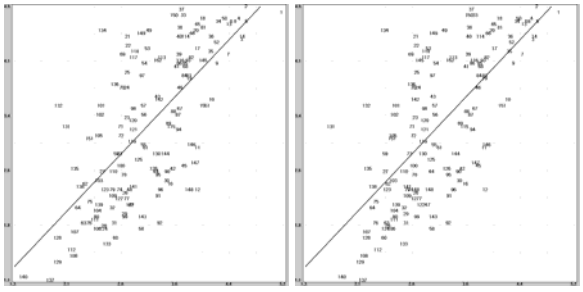
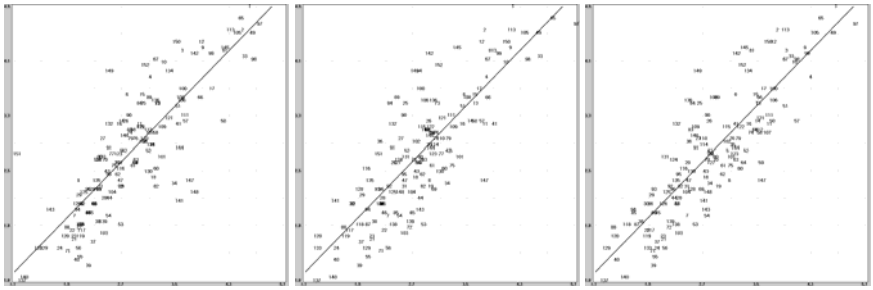
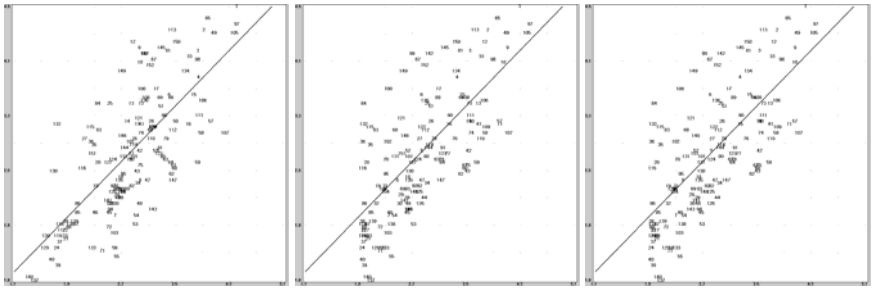


Fig. V.14 CIF01

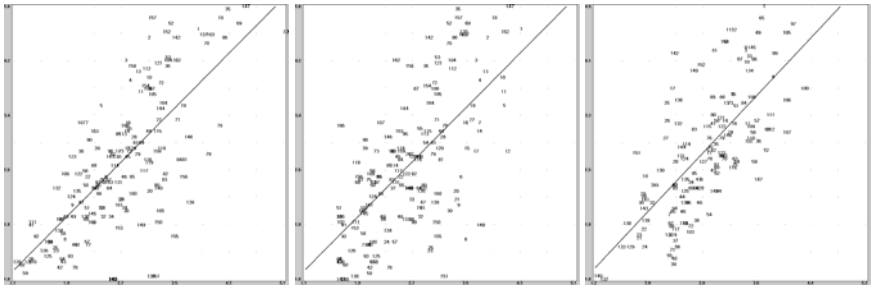
(a) NTT (b) OPTICOM (c) Psytechnics(FR)



(d) Yonsei_FR (e) Yonsei_RR10k (f) Yonsei_RR64k



(g) Psytechnics(NR) (h) SwissQual(NR) (i) PSNR(NTT)



(j) PSNR(NTIA) (k) PSNR(Yonsei)

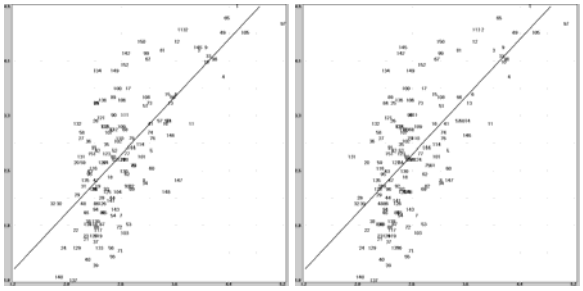
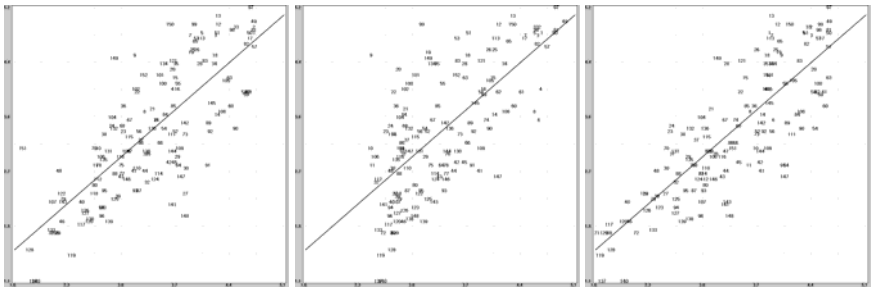
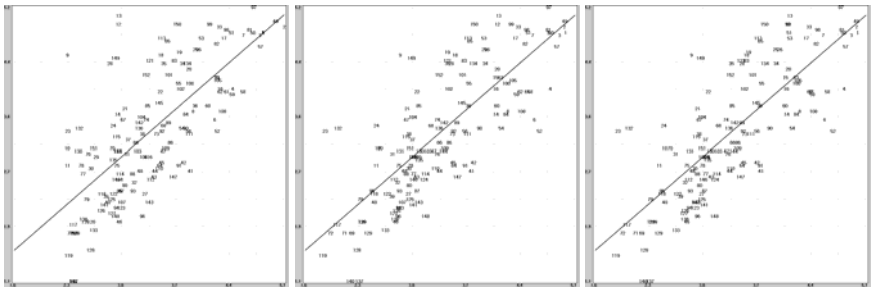


Fig. V.15 CIF02

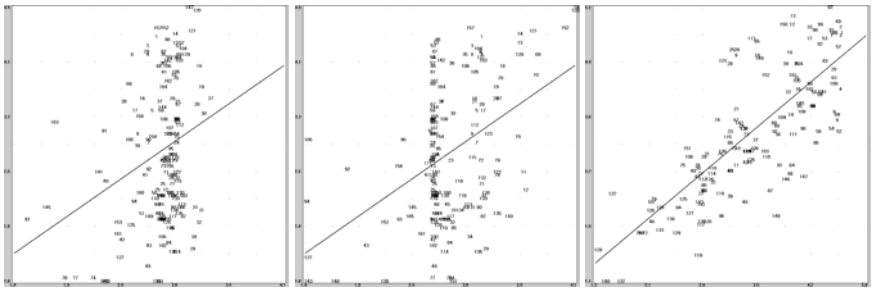
(a) NTT (b) OPTICOM (c) Psytechnics(FR)



(d) Yonsei_FR (e) Yonsei_RR10k (f) Yonsei_RR64k



(g) Psytechnics(NR) (h) SwissQual(NR) (i) PSNR(NTT)



(j) PSNR(NTIA) (k) PSNR(Yonsei)

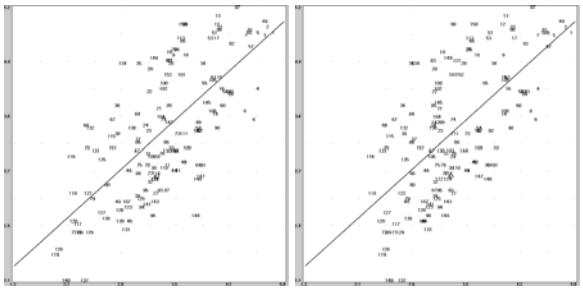
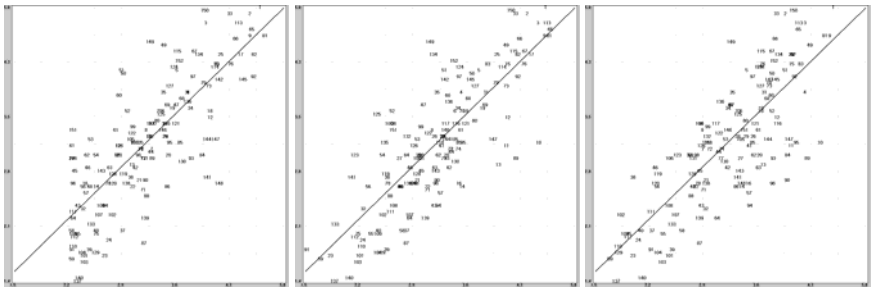
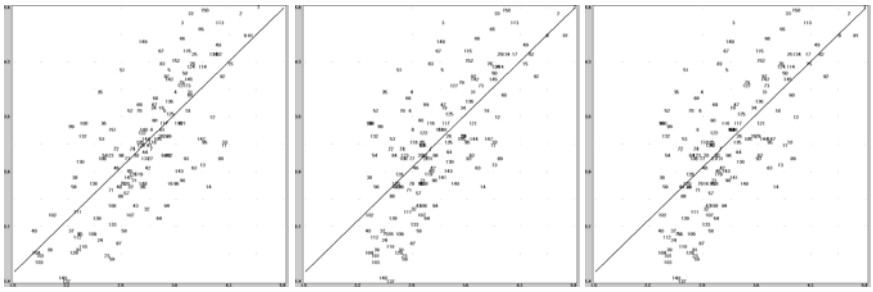


Fig. V.16 CIF03

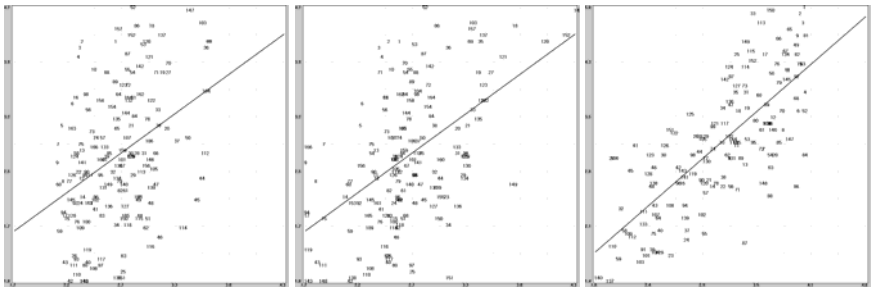
(a) NTT (b) OPTICOM (c) Psytechnics(FR)



(d) Yonsei_FR (e) Yonsei_RR10k (f) Yonsei_RR64k



(g) Psytechnics(NR) (h) SwissQual(NR) (i) PSNR(NTT)



(j) PSNR(NTIA) (k) PSNR(Yonsei)

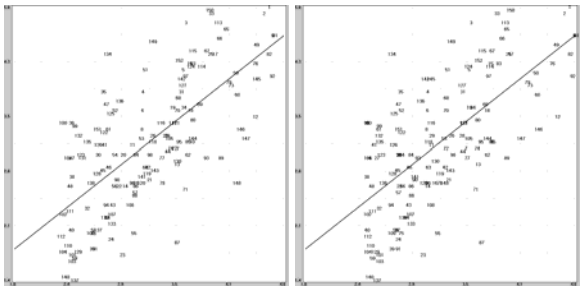
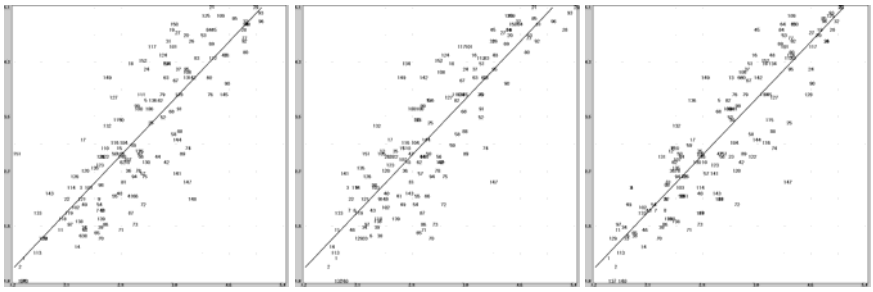
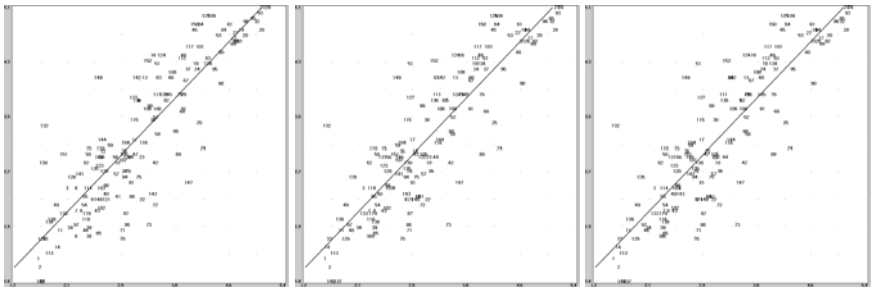


Fig. V.17 CIF04

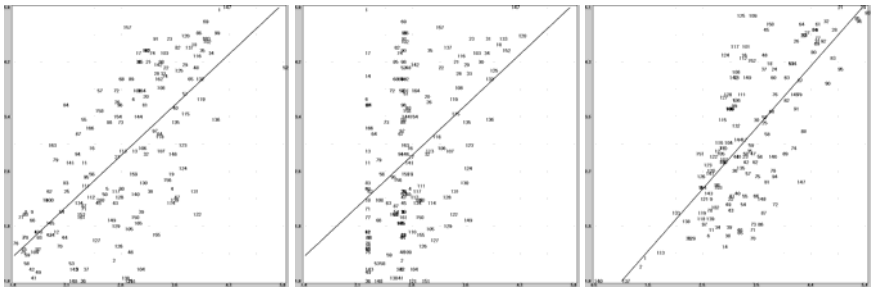
(a) NTT (b) OPTICOM (c) Psytechnics(FR)



(d) Yonsei_FR (e) Yonsei_RR10k (f) Yonsei_RR64k



(g) Psytechnics(NR) (h) SwissQual(NR) (i) PSNR(NTT)



(j) PSNR(NTIA) (k) PSNR(Yonsei)

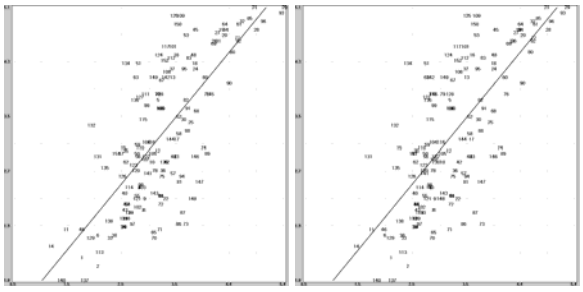
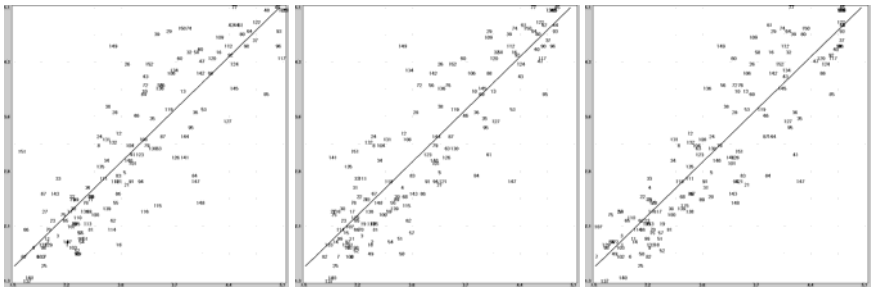
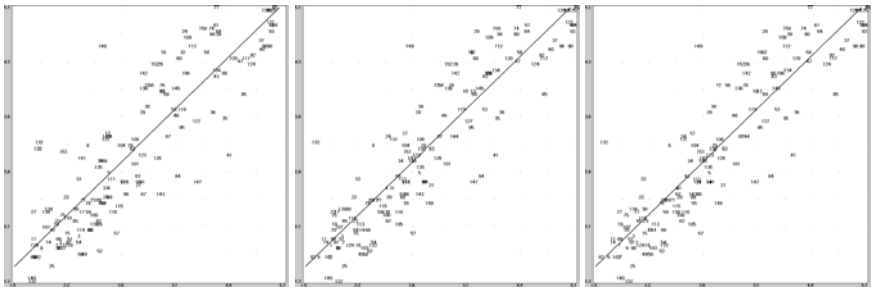


Fig. V.18. CIF05

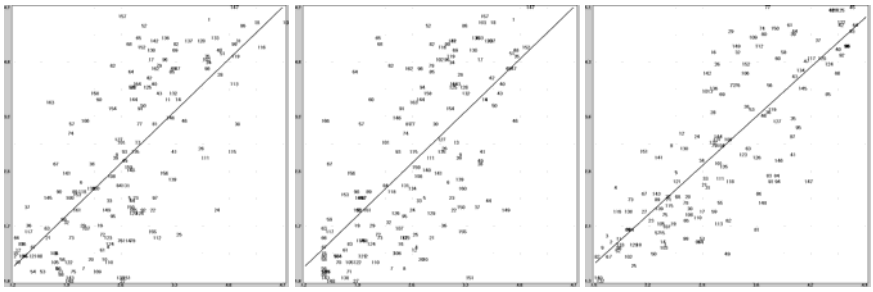
(a) NTT (b) OPTICOM (c) Psytechnics(FR)



(d) Yonsei_FR (e) Yonsei_RR10k (f) Yonsei_RR64k



(g) Psytechnics(NR) (h) SwissQual(NR) (i) PSNR(NTT)



(j) PSNR(NTIA) (k) PSNR(Yonsei)

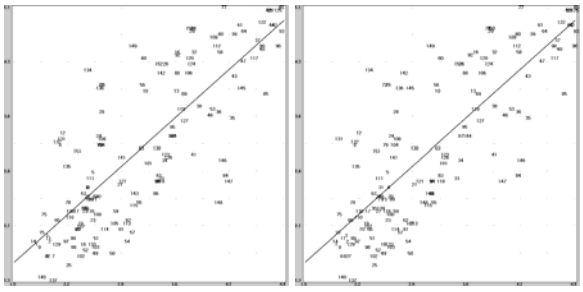
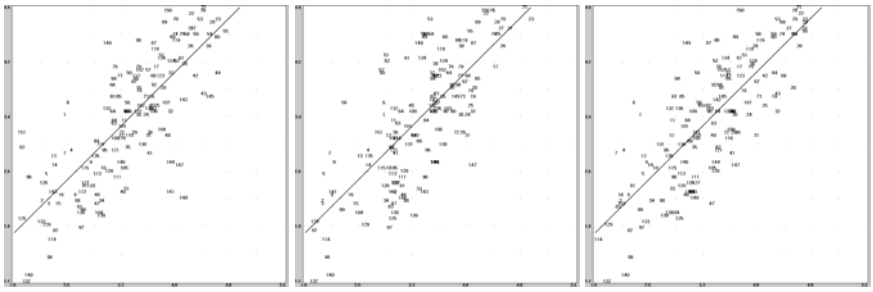
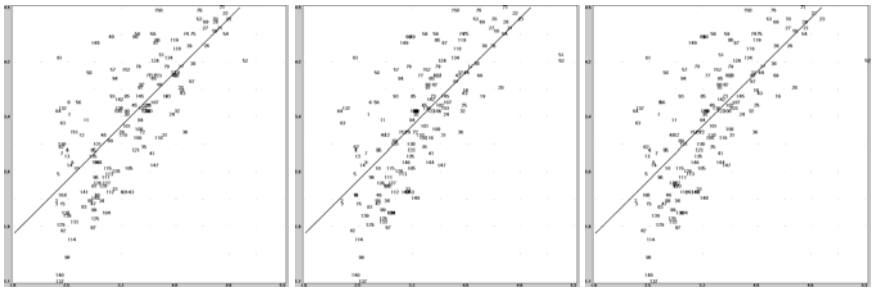


Fig. V.19 CIF06

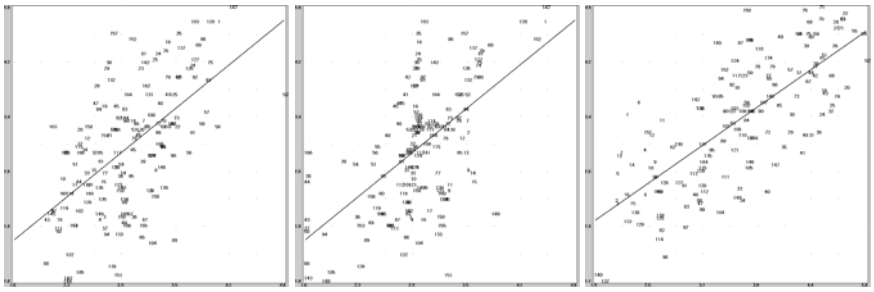
(a) NTT (b) OPTICOM (c) Psytechnics(FR)



(d) Yonsei_FR (e) Yonsei_RR10k (f) Yonsei_RR64k



(g) Psytechnics(NR) (h) SwissQual(NR) (i) PSNR(NTT)



(j) PSNR(NTIA) (k) PSNR(Yonsei)

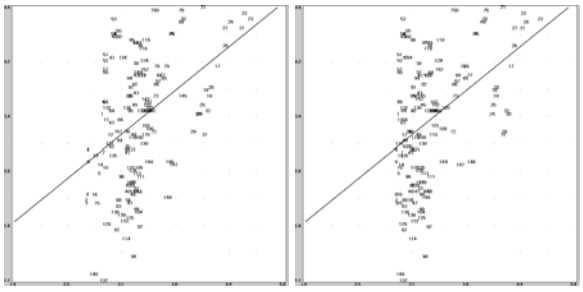
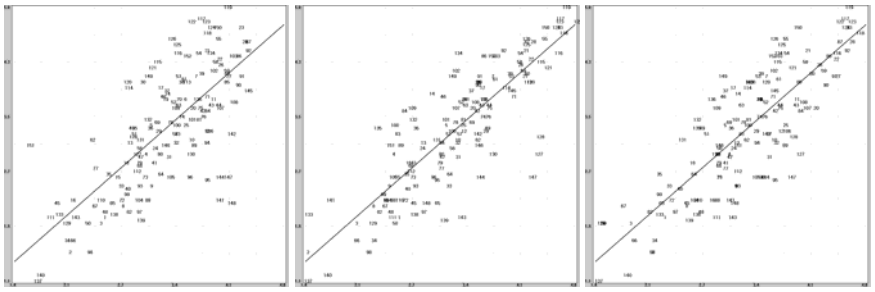
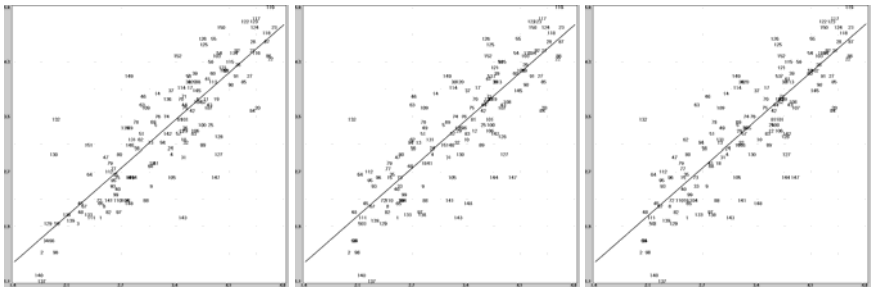


Fig. V.20 CIF07

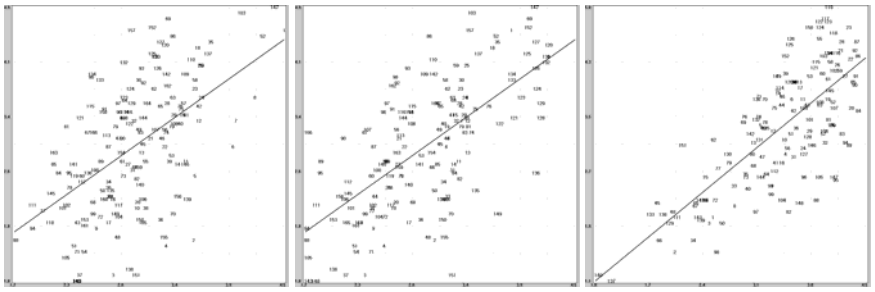
(a) NTT (b) OPTICOM (c) Psytechnics(FR)



(d) Yonsei_FR (e) Yonsei_RR10k (f) Yonsei_RR64k



(g) Psytechnics(NR) (h) SwissQual(NR) (i) PSNR(NTT)



(j) PSNR(NTIA) (k) PSNR(Yonsei)

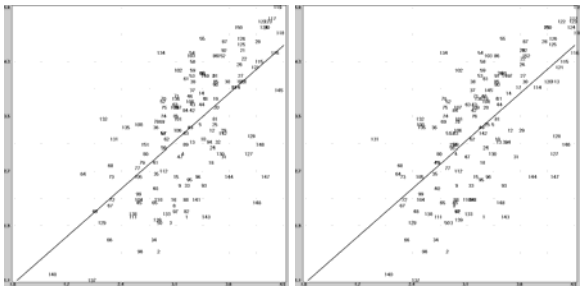
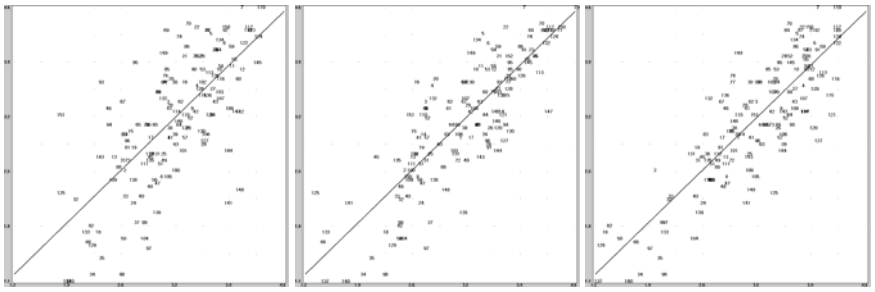
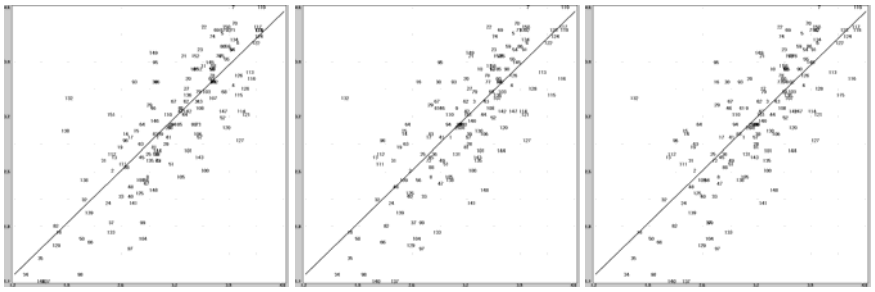


Fig. V.21 CIF08

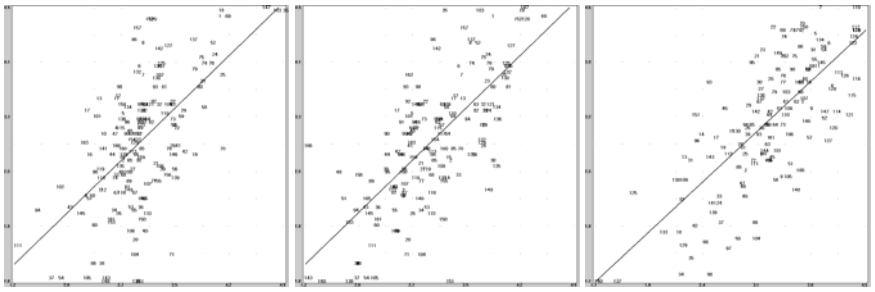
(a) NTT (b) OPTICOM (c) Psytechnics(FR)



(d) Yonsei_FR (e) Yonsei_RR10k (f) Yonsei_RR64k



(g) Psytechnics(NR) (h) SwissQual(NR) (i) PSNR(NTT)



(j) PSNR(NTIA) (k) PSNR(Yonsei)

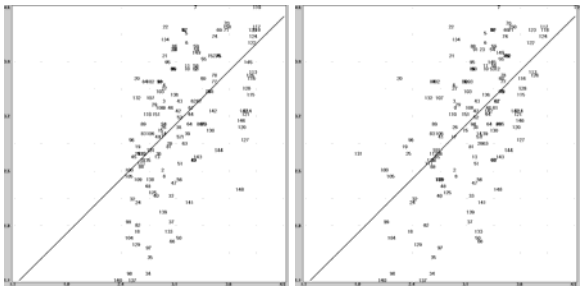
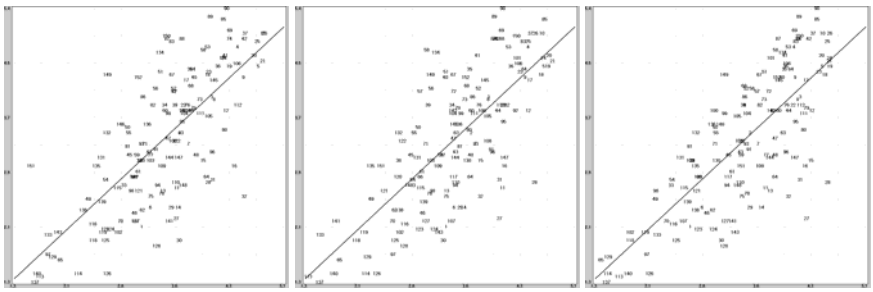
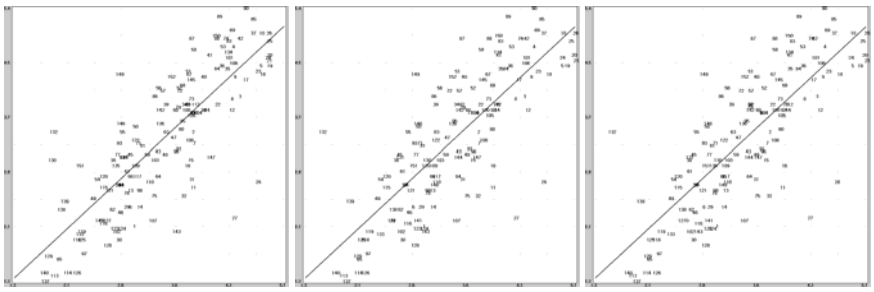


Fig. V.22 CIF09

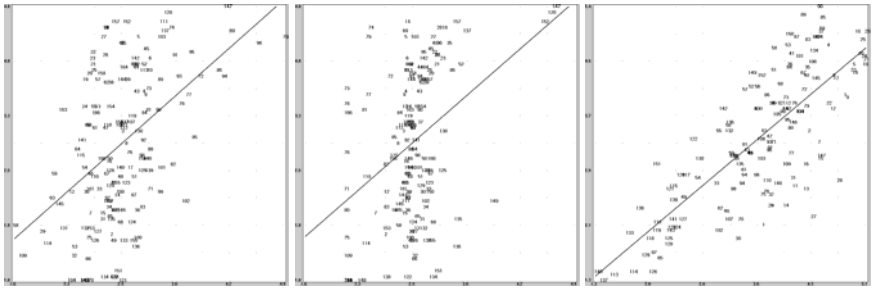
(a) NTT (b) OPTICOM (c) Psytechnics(FR)



(d) Yonsei_FR (e) Yonsei_RR10k (f) Yonsei_RR64k



(g) Psytechnics(NR) (h) SwissQual(NR) (i) PSNR(NTT)



(j) PSNR(NTIA) (k) PSNR(Yonsei)

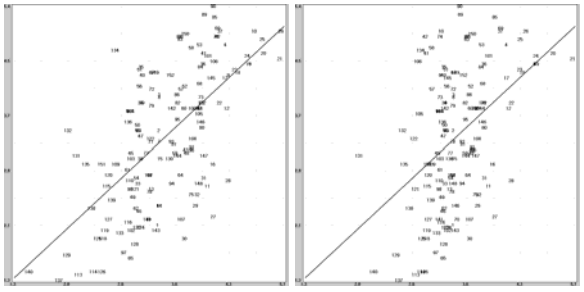
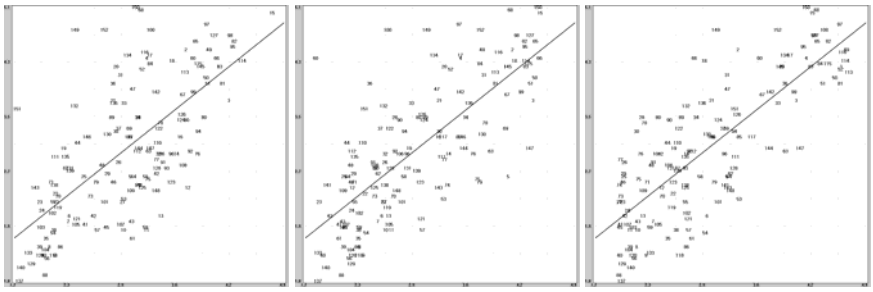
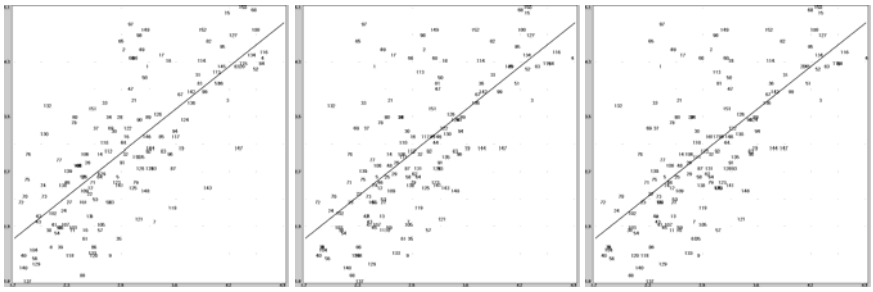


Fig. V.23 CIF10

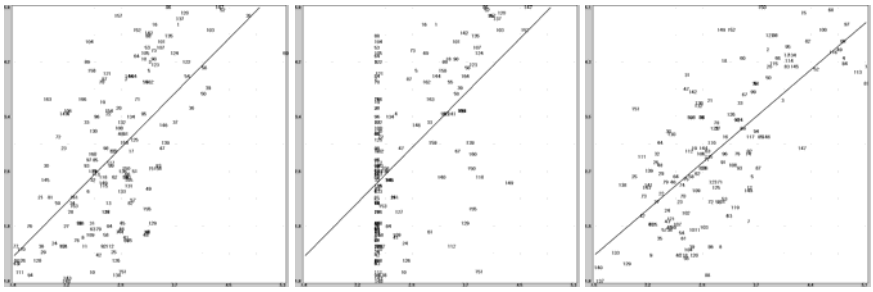
(a) NTT (b) OPTICOM (c) Psytechnics(FR)



(d) Yonsei_FR (e) Yonsei_RR10k (f) Yonsei_RR64k



(g) Psytechnics(NR) (h) SwissQual(NR) (i) PSNR(NTT)



(j) PSNR(NTIA) (k) PSNR(Yonsei)

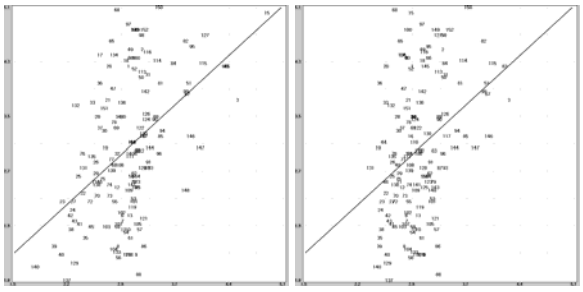
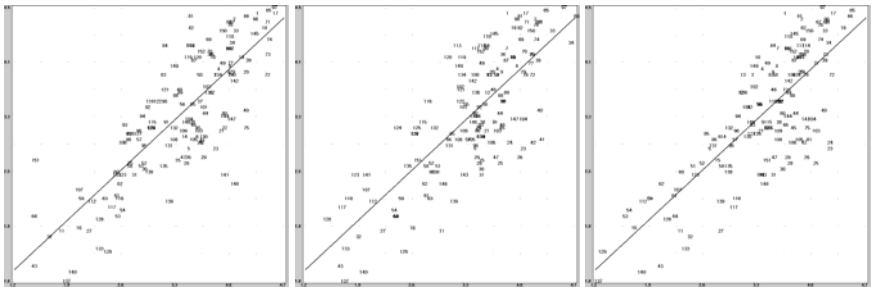
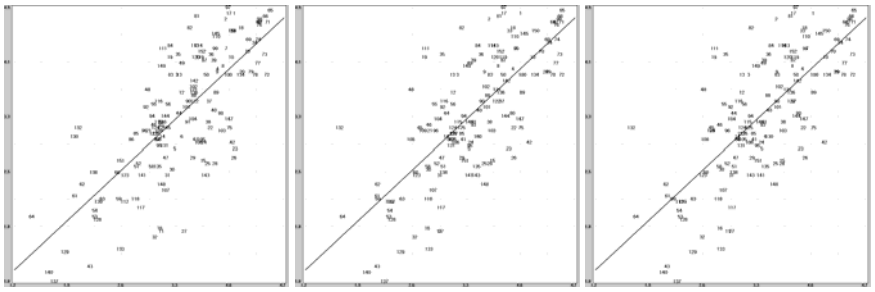


Fig. V.24 CIF11

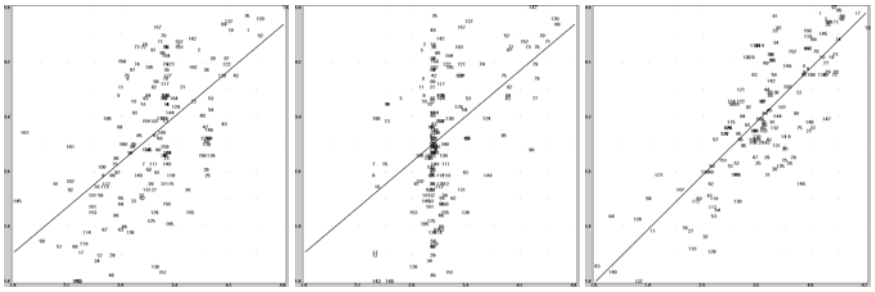
(a) NTT (b) OPTICOM (c) Psytechnics(FR)



(d) Yonsei_FR (e) Yonsei_RR10k (f) Yonsei_RR64k



(g) Psytechnics(NR) (h) SwissQual(NR) (i) PSNR(NTT)



(j) PSNR(NTIA) (k) PSNR(Yonsei)

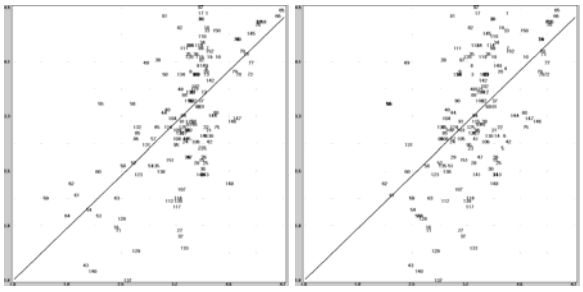
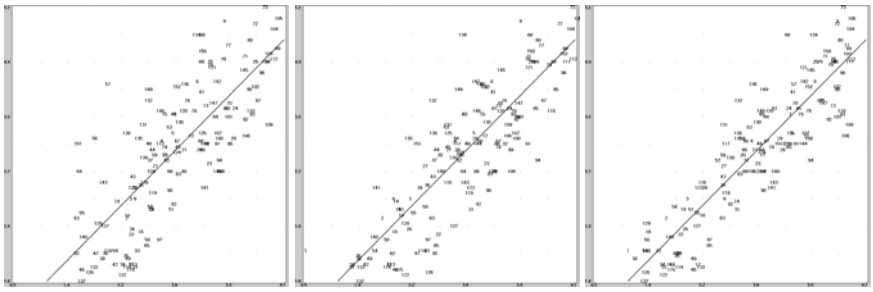
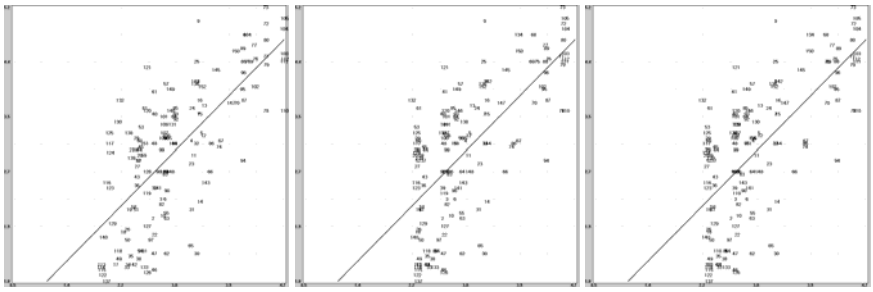


Fig. V.25 CIF12

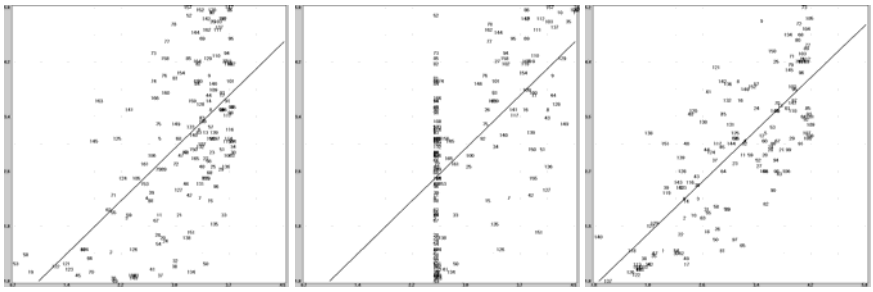
(a) NTT (b) OPTICOM (c) Psytechnics(FR)



(d) Yonsei_FR (e) Yonsei_RR10k (f) Yonsei_RR64k



(g) Psytechnics(NR) (h) SwissQual(NR) (i) PSNR(NTT)



(j) PSNR(NTIA) (k) PSNR(Yonsei)

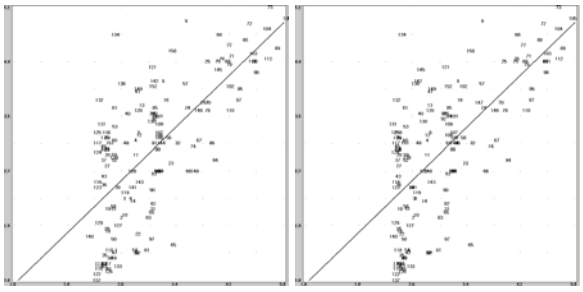
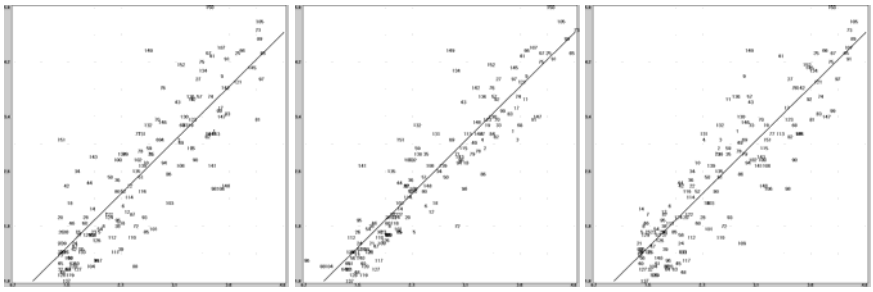
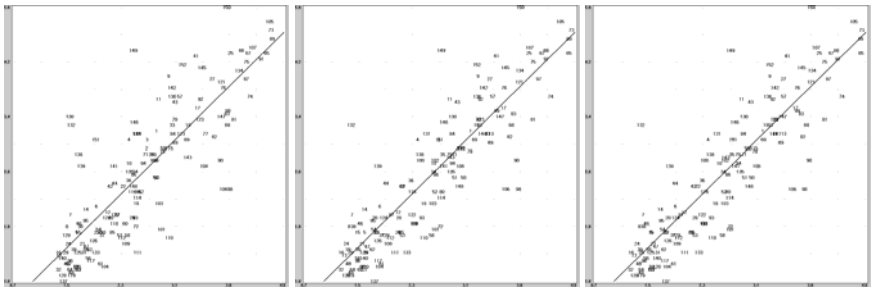


Fig. V.26 CIF13

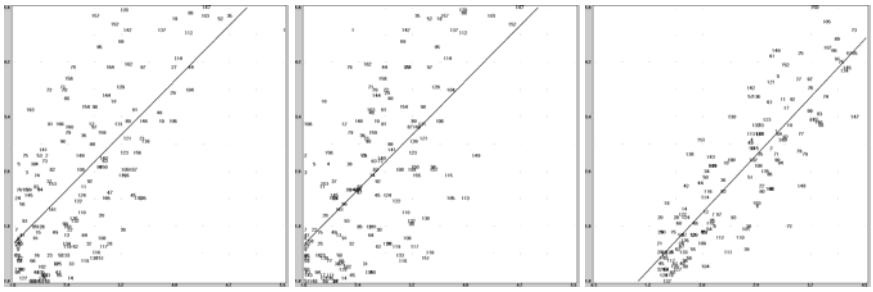
(a) NTT (b) OPTICOM (c) Psytechnics(FR)



(d) Yonsei_FR (e) Yonsei_RR10k (f) Yonsei_RR64k



(g) Psytechnics(NR) (h) SwissQual(NR) (i) PSNR(NTT)



(j) PSNR(NTIA) (k) PSNR(Yonsei)

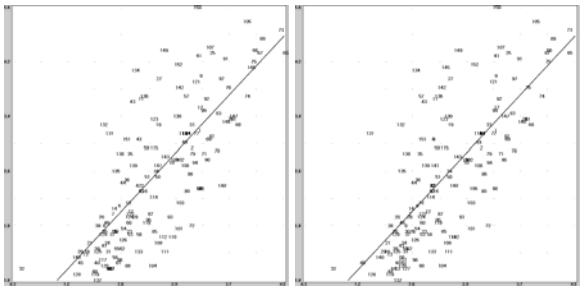
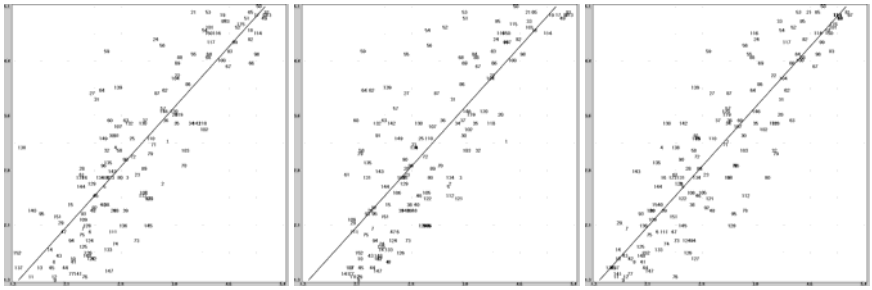


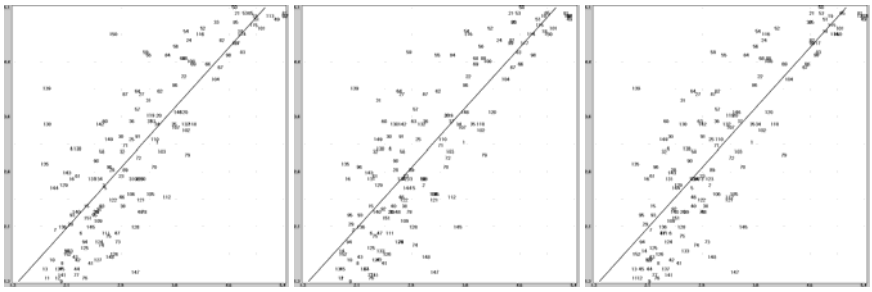
Fig. V.27 CIF14

Appendix V.3 QCIF Plots

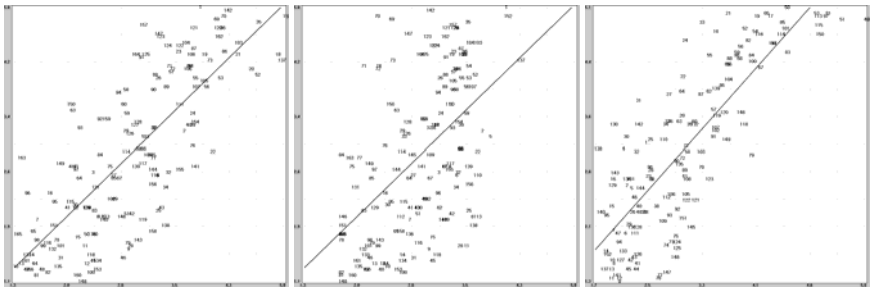
(a) NTT (b) OPTICOM (c) Psytechnics(FR)



(d) Yonsei_FR (e) Yonsei_RR1k (f) Yonsei_RR10k



(g) Psytechnics(NR) (h) SwissQual(NR) (i) PSNR(NTT)



(j) PSNR(NTIA) (k) PSNR(Yonsei)

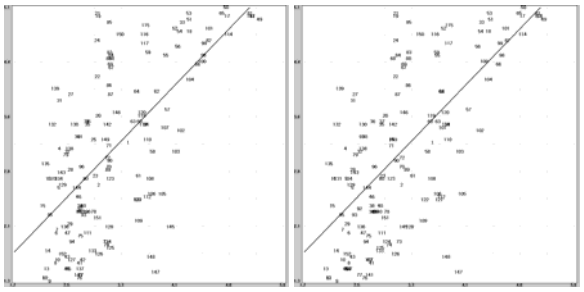
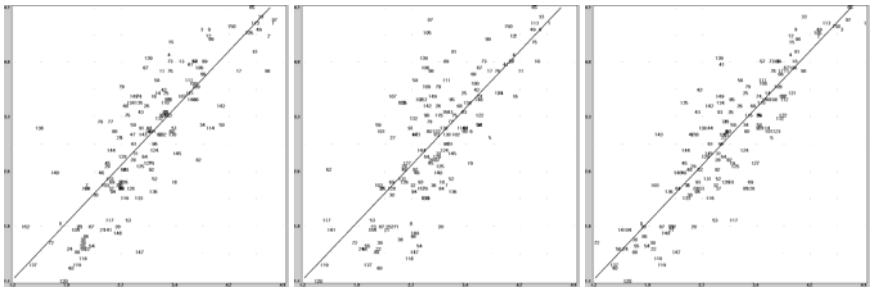
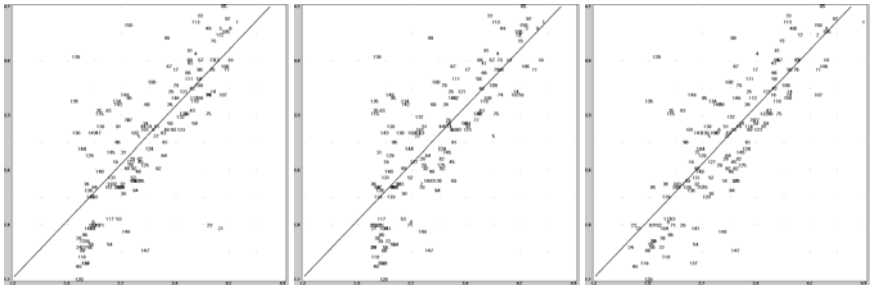


Fig. V.28 QCIF01

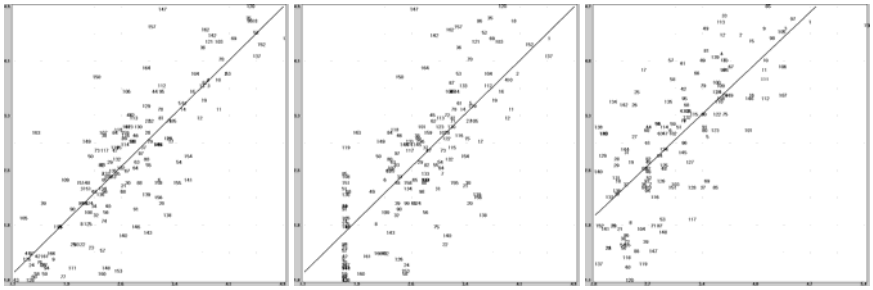
(a) NTT (b) OPTICOM (c) Psytechnics(FR)



(d) Yonsei_FR (e) Yonsei_RR1k (f) Yonsei_RR10k



(g) Psytechnics(NR) (h) SwissQual(NR) (i) PSNR(NTT)



(j) PSNR(NTIA) (k) PSNR(Yonsei)

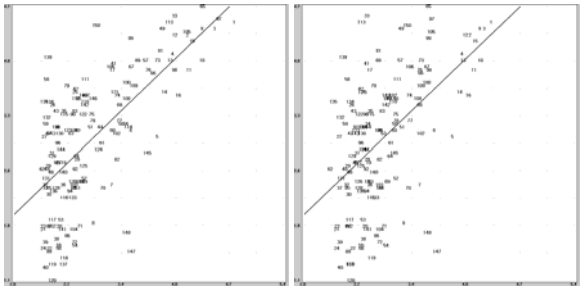
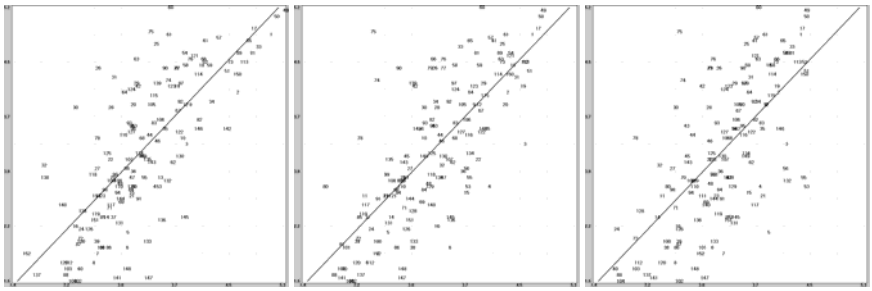
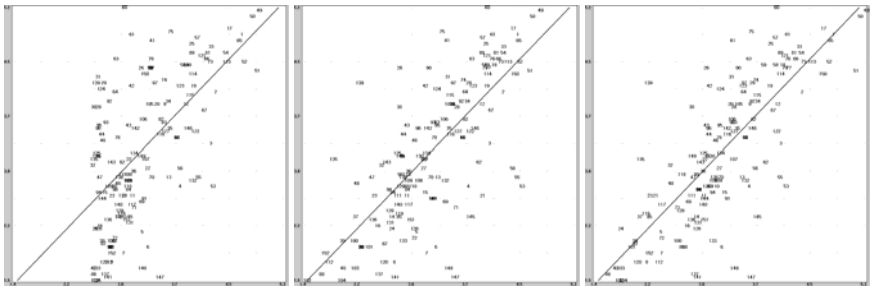


Fig. V.29 QCIF02

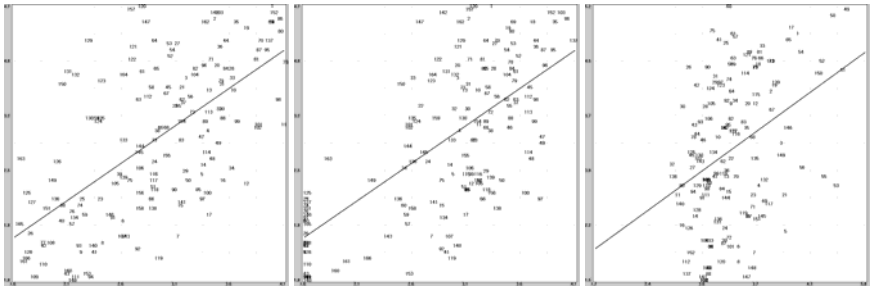
(a) NTT (b) OPTICOM (c) Psytechnics(FR)



(d) Yonsei_FR (e) Yonsei_RR1k (f) Yonsei_RR10k



(g) Psytechnics(NR) (h) SwissQual(NR) (i) PSNR(NTT)



(j) PSNR(NTIA) (k) PSNR(Yonsei)

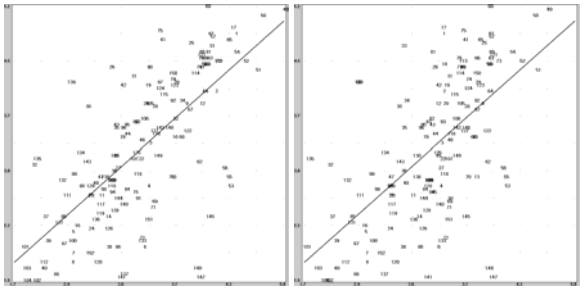
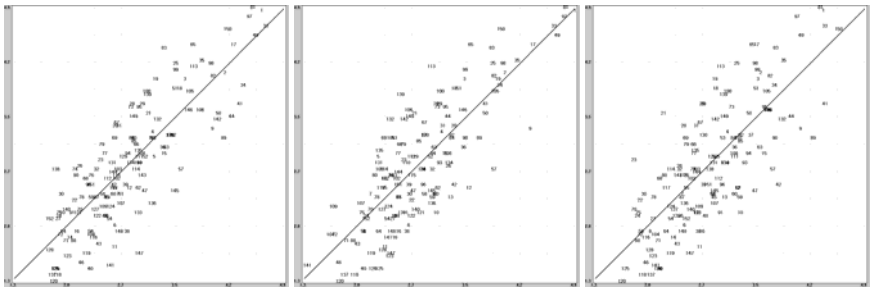
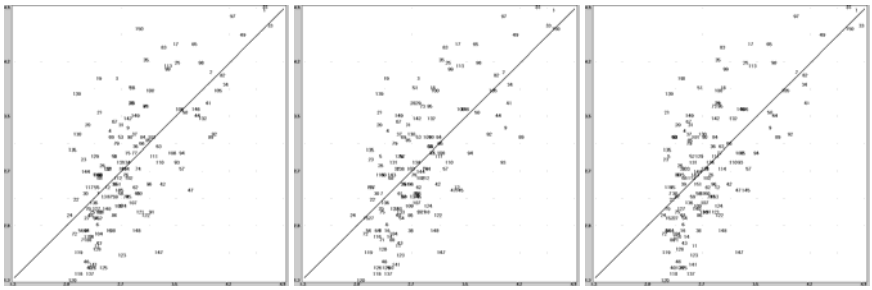


Fig. V.30 QCIF03

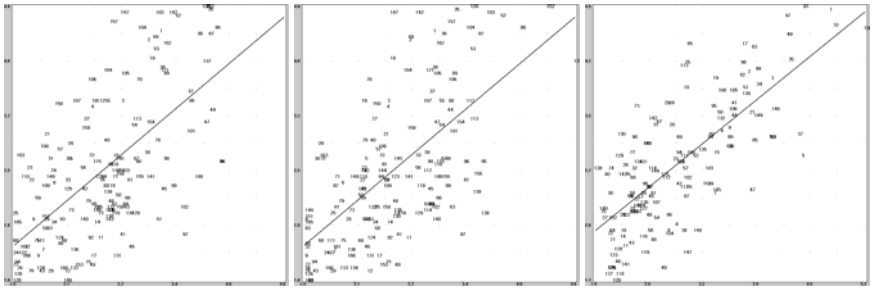
(a) NTT (b) OPTICOM (c) Psytechnics(FR)



(d) Yonsei_FR (e) Yonsei_RR1k (f) Yonsei_RR10k



(g) Psytechnics(NR) (h) SwissQual(NR) (i) PSNR(NTT)



(j) PSNR(NTIA) (k) PSNR(Yonsei)

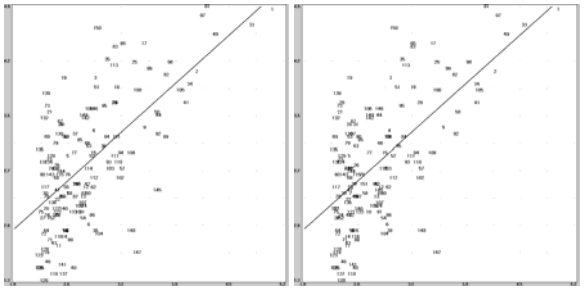
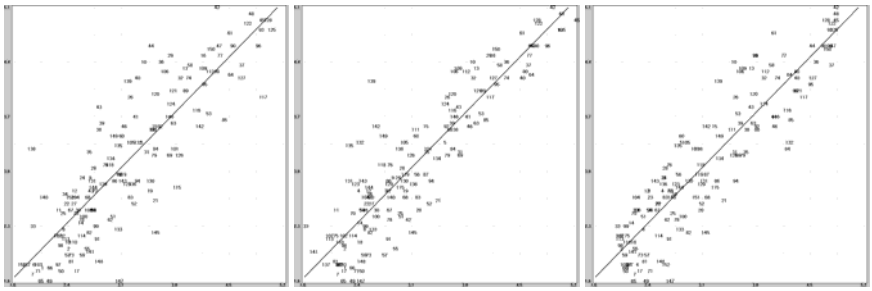
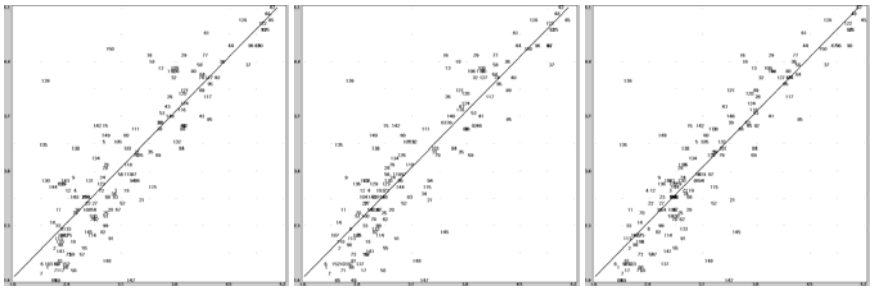


Fig. V.31 QCIF04

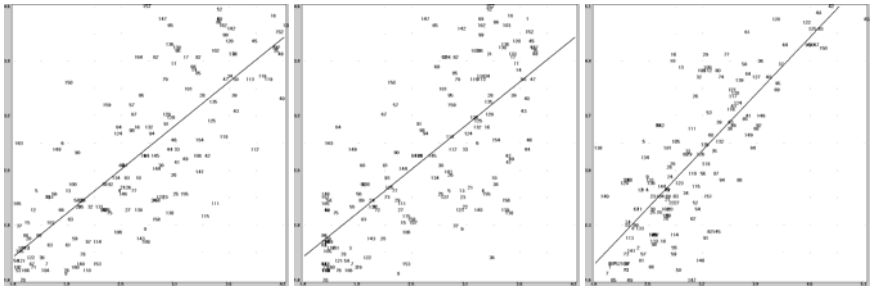
(a) NTT (b) OPTICOM (c) Psytechnics(FR)



(d) Yonsei_FR (e) Yonsei_RR1k (f) Yonsei_RR10k



(g) Psytechnics(NR) (h) SwissQual(NR) (i) PSNR(NTT)



(j) PSNR(NTIA) (k) PSNR(Yonsei)

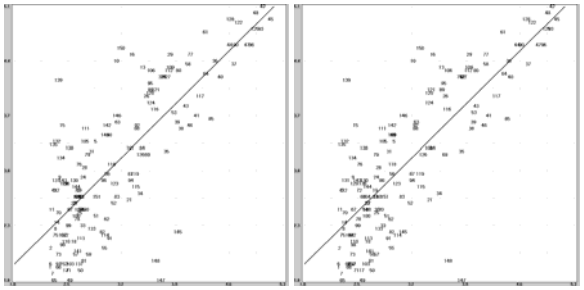
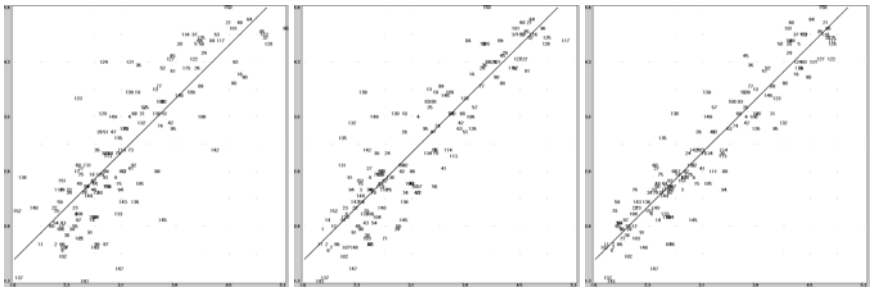
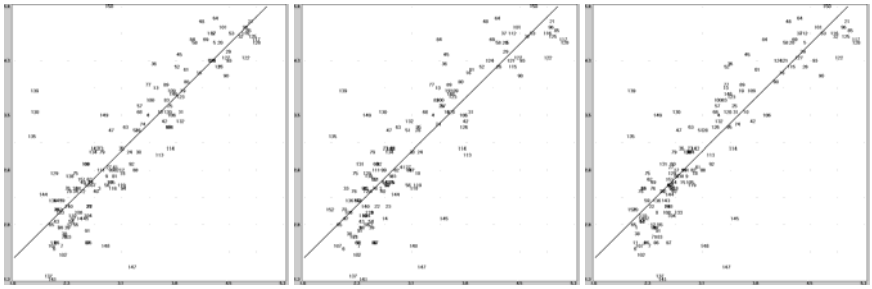


Fig. V.32 QCIF05

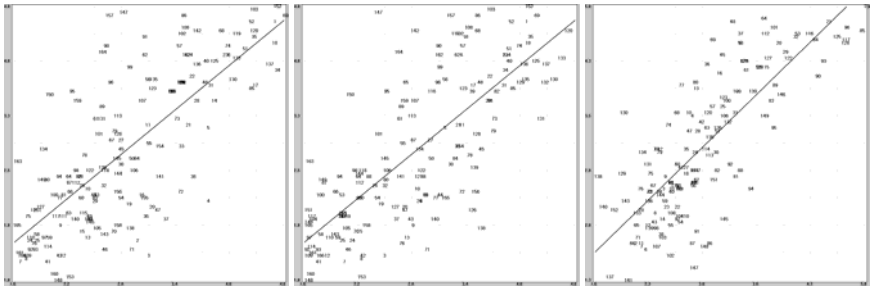
(a) NTT (b) OPTICOM (c) Psytechnics(FR)



(d) Yonsei_FR (e) Yonsei_RR1k (f) Yonsei_RR10k



(g) Psytechnics(NR) (h) SwissQual(NR) (i) PSNR(NTT)



(j) PSNR(NTIA) (k) PSNR(Yonsei)

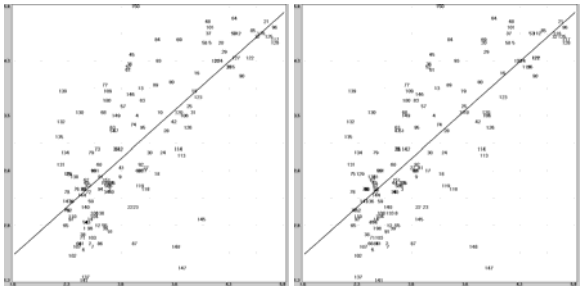
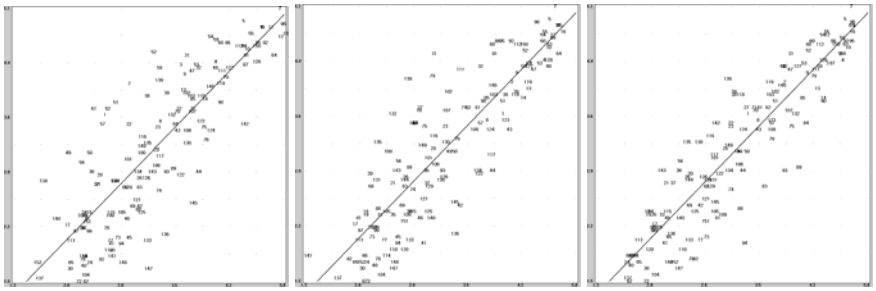
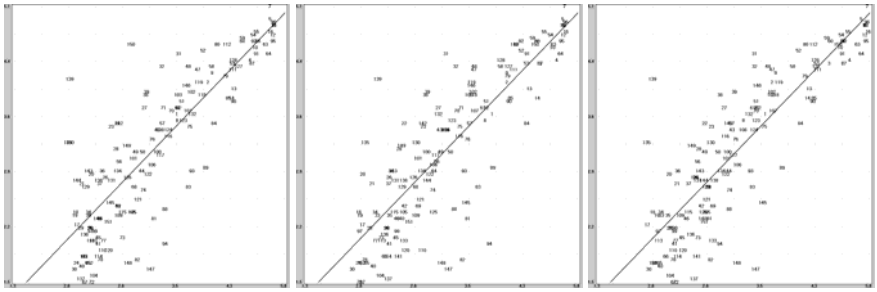


Fig. V.33 QCIF06

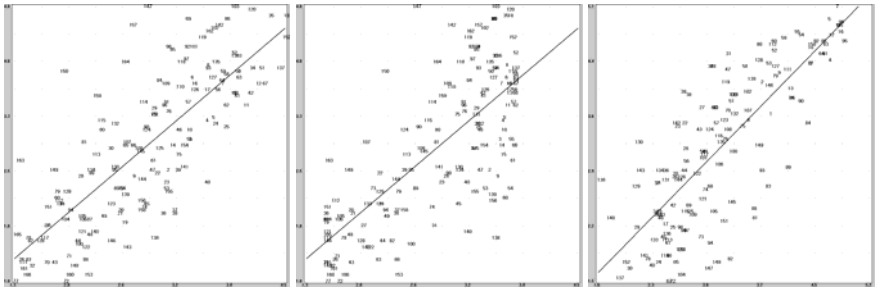
(a) NTT (b) OPTICOM (c) Psytechnics(FR)



(d) Yonsei_FR (e) Yonsei_RR1k (f) Yonsei_RR10k



(g) Psytechnics(NR) (h) SwissQual(NR) (i) PSNR(NTT)



(j) PSNR(NTIA) (k) PSNR(Yonsei)

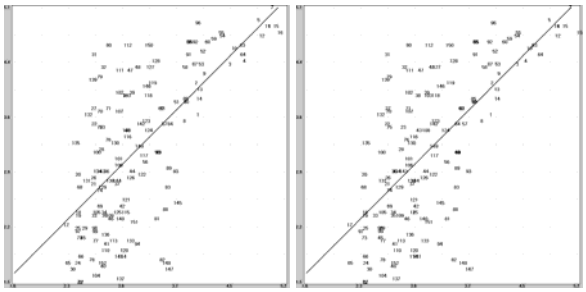
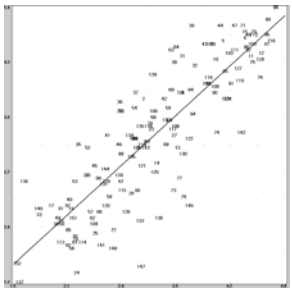
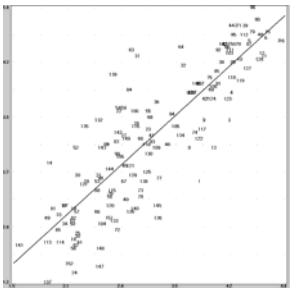
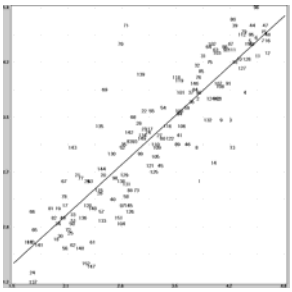
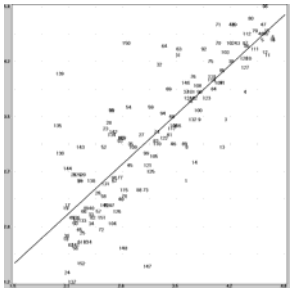
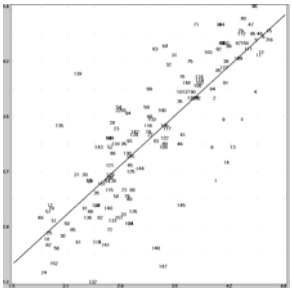
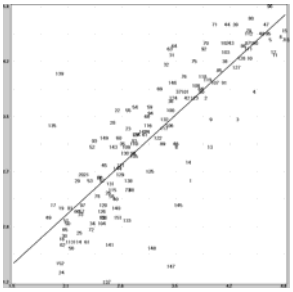


Fig. V.34 QCIF07

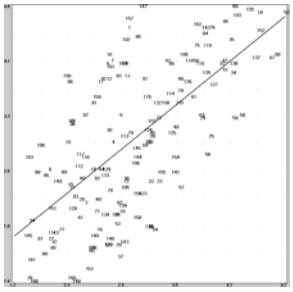
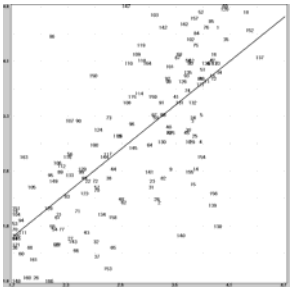
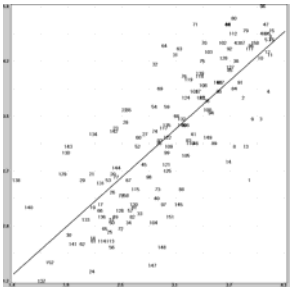
(a) NTT (b) OPTICOM (c) Psytechnics(FR)

(d) Yonsei_FR (e) Yonsei_RR1k (f) Yonsei_RR10k

(g) Psytechnics(NR) (h) SwissQual(NR) (i) PSNR(NTT)

(j) PSNR(NTIA) (k) PSNR(Yonsei)

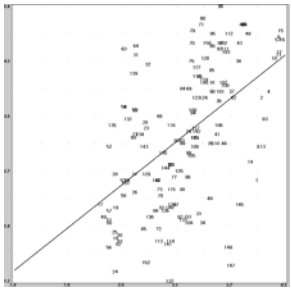
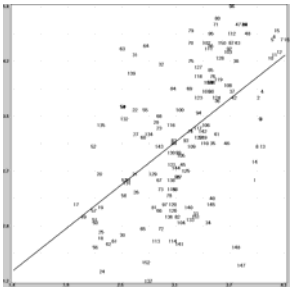
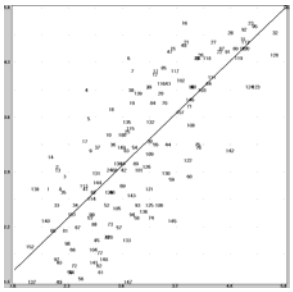
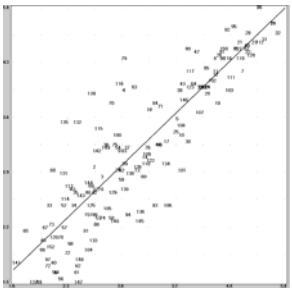
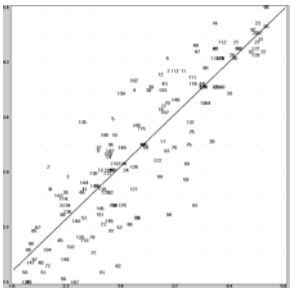
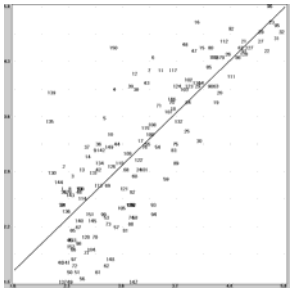
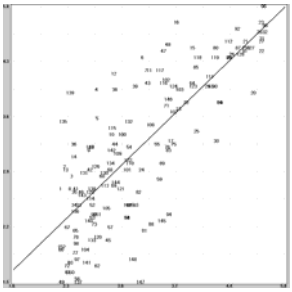
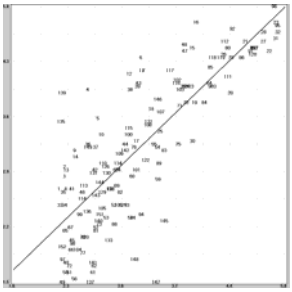



Fig. V.35 QCIF08

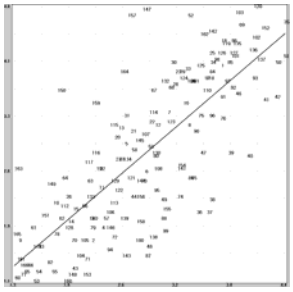
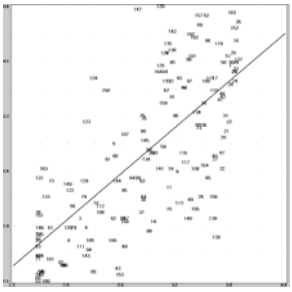
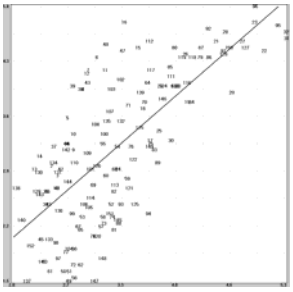
(a) NTT (b) OPTICOM (c) Psytechnics(FR)

(d) Yonsei_FR (e) Yonsei_RR1k (f) Yonsei_RR10k

(g) Psytechnics(NR) (h) SwissQual(NR) (i) PSNR(NTT)

(j) PSNR(NTIA) (k) PSNR(Yonsei)

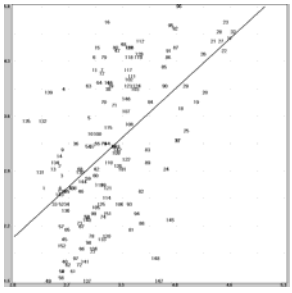
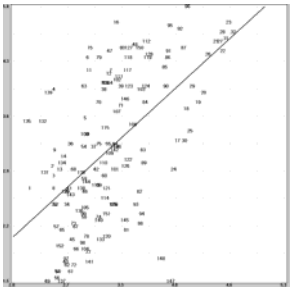
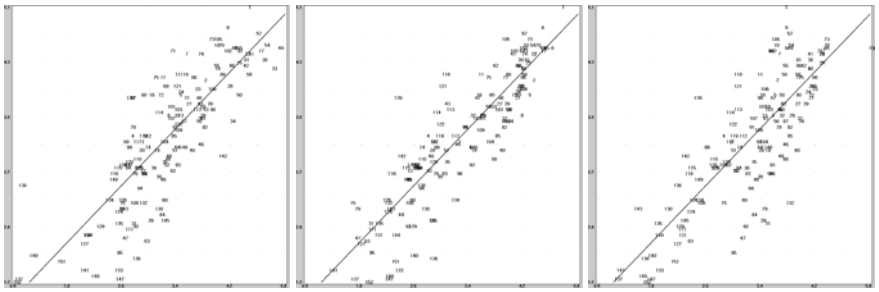
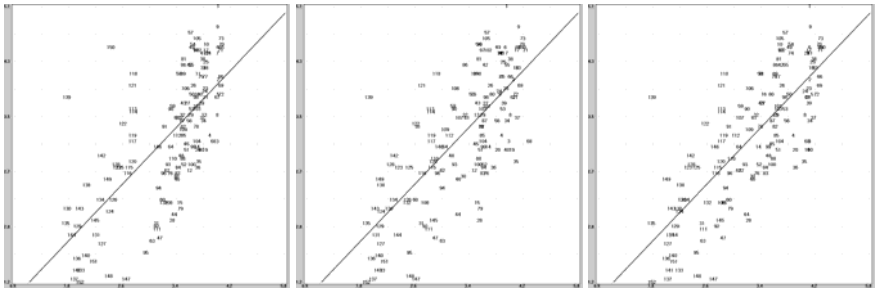



Fig. V.36 QCIF09

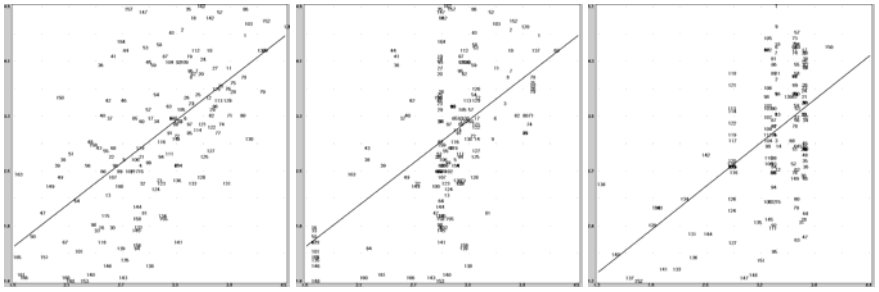
(a) NTT (b) OPTICOM (c) Psytechnics(FR)



(d) Yonsei_FR (e) Yonsei_RR1k (f) Yonsei_RR10k



(g) Psytechnics(NR) (h) SwissQual(NR) (i) PSNR(NTT)



(j) PSNR(NTIA) (k) PSNR(Yonsei)

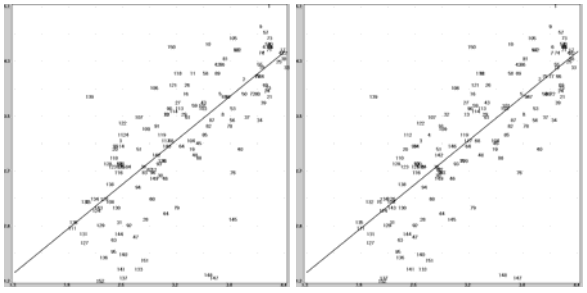
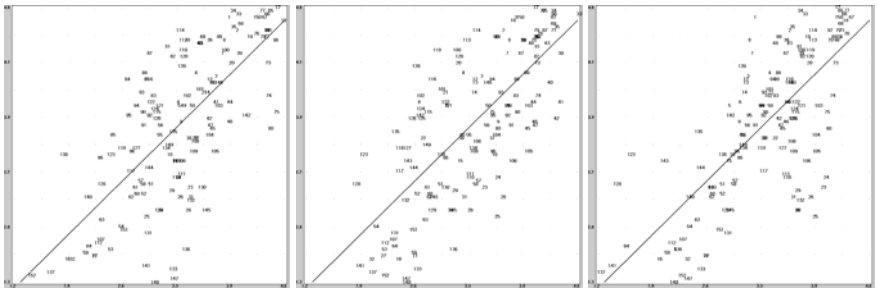
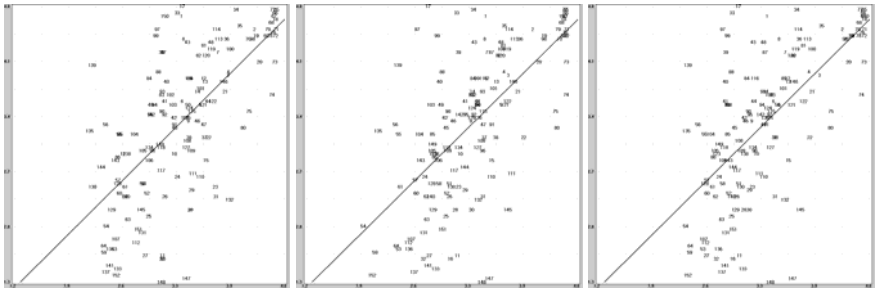


Fig. V.37 QCIF10

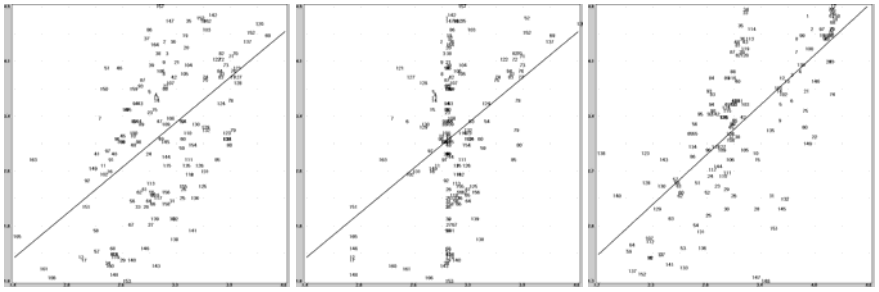
(a) NTT (b) OPTICOM (c) Psytechnics(FR)



(d) Yonsei_FR (e) Yonsei_RR1k (f) Yonsei_RR10k



(g) Psytechnics(NR) (h) SwissQual(NR) (i) PSNR(NTT)



(j) PSNR(NTIA) (k) PSNR(Yonsei)

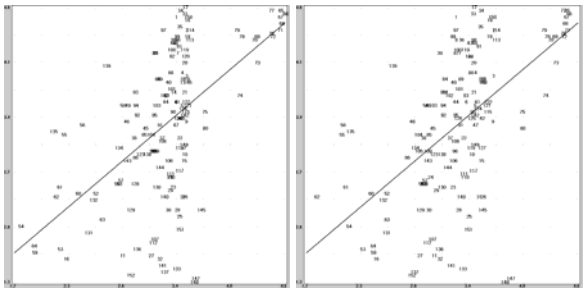
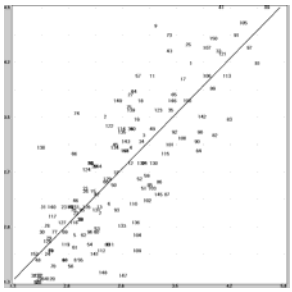
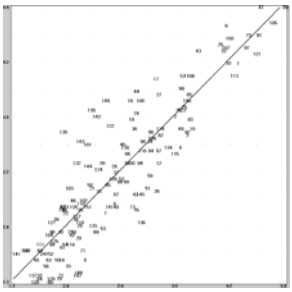
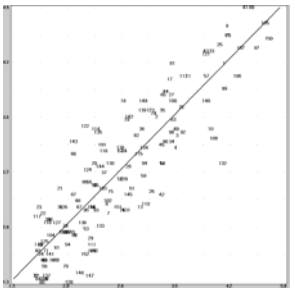
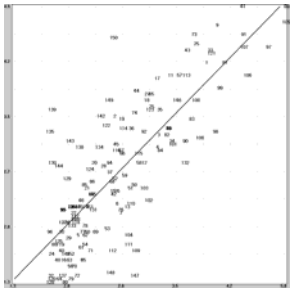
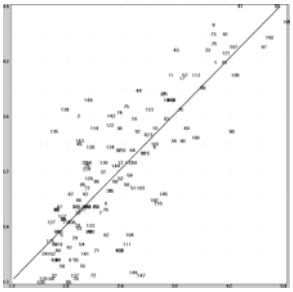
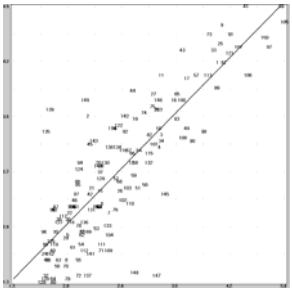


Fig. V.38 QCIF11

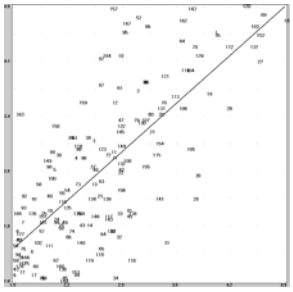
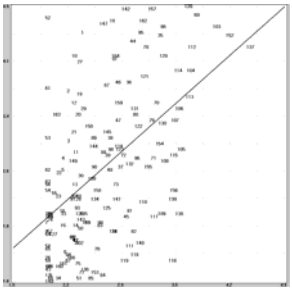
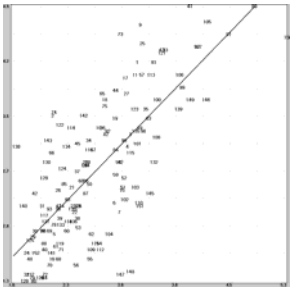
(a) NTT (b) OPTICOM (c) Psytechnics(FR)

(d) Yonsei_FR (e) Yonsei_RR1k (f) Yonsei_RR10k

(g) Psytechnics(NR) (h) SwissQual(NR) (i) PSNR(NTT)

(j) PSNR(NTIA) (k) PSNR(Yonsei)

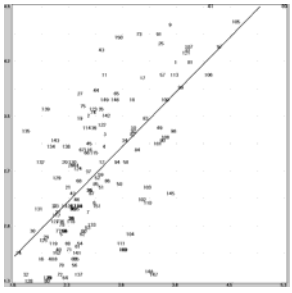
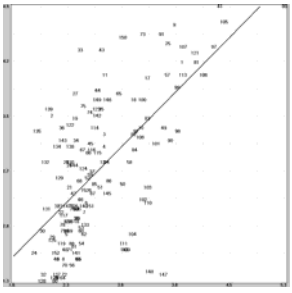
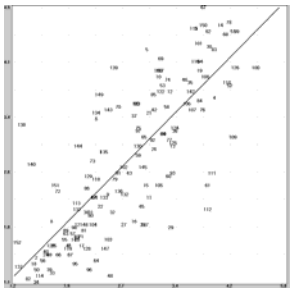
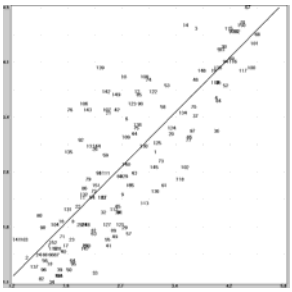
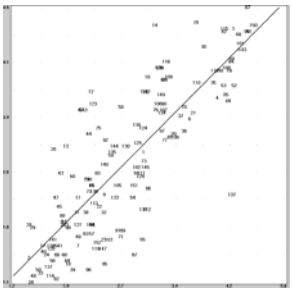
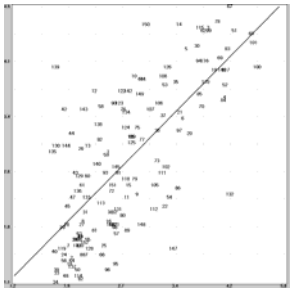
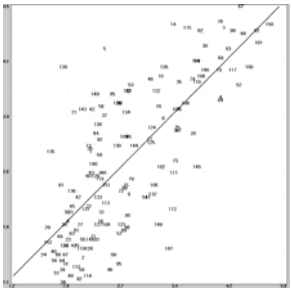
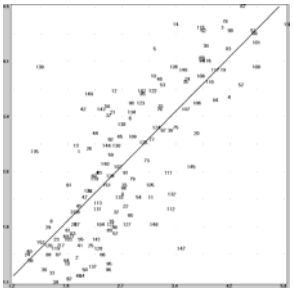



Fig. V.39 QCIF12

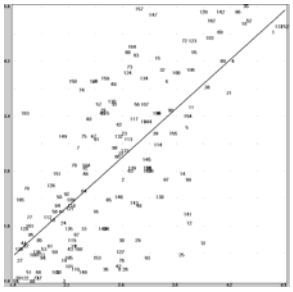
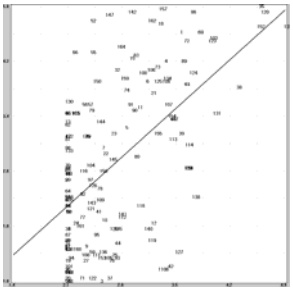
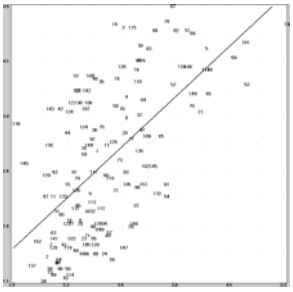
(a) NTT (b) OPTICOM (c) Psytechnics(FR)

(d) Yonsei_FR (e) Yonsei_RR1k (f) Yonsei_RR10k

(g) Psytechnics(NR) (h) SwissQual(NR) (i) PSNR(NTT)

(j) PSNR(NTIA) (k) PSNR(Yonsei)

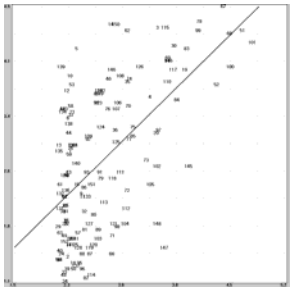
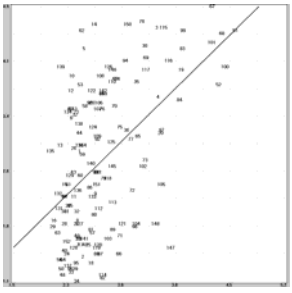
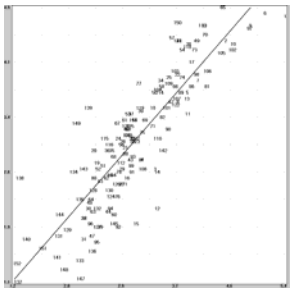
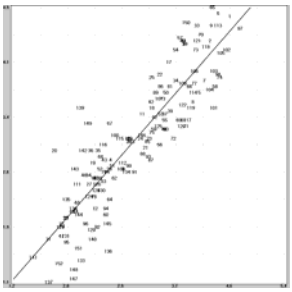
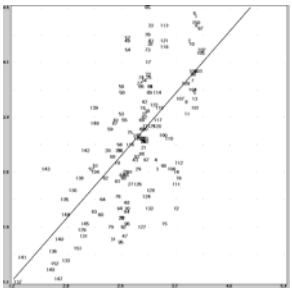
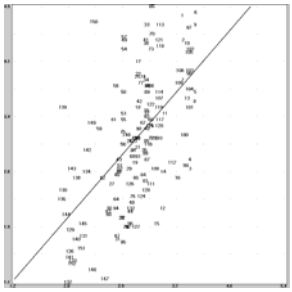
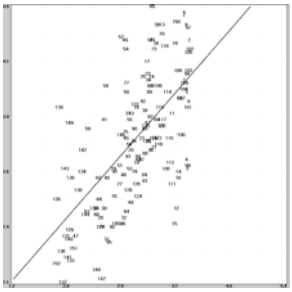
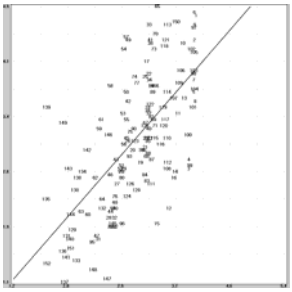



Fig. V.40 QCIF13

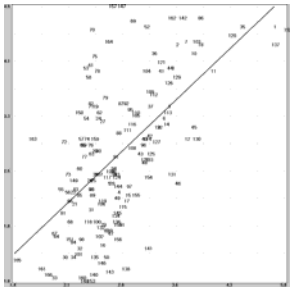
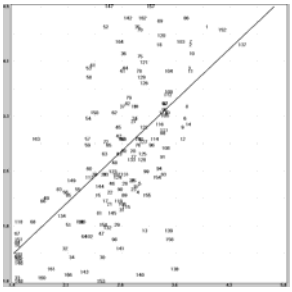
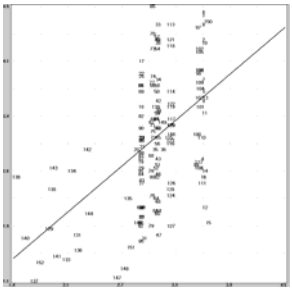
(a) NTT (b) OPTICOM (c) Psytechnics(FR)

(d) Yonsei_FR (e) Yonsei_RR1k (f) Yonsei_RR10k

(g) Psytechnics(NR) (h) SwissQual(NR) (i) PSNR(NTT)

(j) PSNR(NTIA) (k) PSNR(Yonsei)

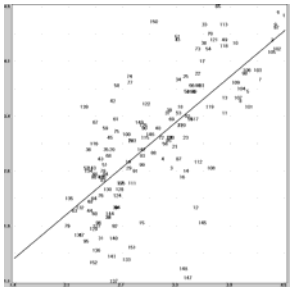
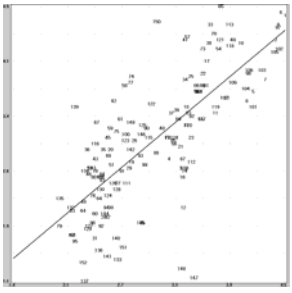



Fig. V.41 QCIF14

Appendix VI Proponent Comments

Note: The proponent comments are not endorsed by VQEG. They are presented in this Appendix to give the Proponents a chance to discuss their results and should not be quoted out of this context.

Appendix VI.1 NTT

Proponent Comment (NTT)

Needs for two supplementary analyses:

per-sample analysis without common video clips and per-condition analysis

1 Background

In the final report, the performance of objective video quality estimation models was primarily evaluated on per-sample basis, i.e., objective video quality for each video clip was compared with subjective quality to investigate estimation accuracy. This is one of the essential analysis for performance verification of these models. However, this approach has two drawbacks.

The first one is the effects of repetitive use of common video clips. Objective models that show better performance for these PVSs are evaluated too highly because these specific PVSs were evaluated more than 10 times in the analysis. Therefore, per-sample analysis without common video clips is recommended for fair evaluation of models.

The other one is the lack of investigation on the estimation of average quality over various contents. For the optimization and/or characterization of a codec or system, which is one of the most important applications for FR, one usually does not optimize the codec or system from the viewpoint of specific video content. Rather, he/she tries to tune the system to maximize the average quality of several video contents. Therefore, estimating the average quality over various types of content, per-condition analysis, is of great interest as well.

2 Result of two supplementary analyses

12.1 Per-sample analysis without common video clips

12.1.1 VGA

	NTT FR			OPTICOM FR			Psytechnics FR			Yonsei FR			NTIA PSNR search		
	Correl	RMSE	Outlier	Correl	RMSE	Outlier	Correl	RMSE	Outlier	Correl	RMSE	Outlier	Correl	RMSE	Outlier
V01	0.8850	0.4947	0.5313	0.7834	0.6605	0.6016	0.8877	0.4893	0.5078	0.8241	0.6020	0.5625	0.8640	0.5359	0.5859
V02	0.7809	0.5799	0.5391	0.9134	0.3756	0.3516	0.5491	0.7727	0.7031	0.7743	0.5850	0.5625	0.6457	0.7057	0.6328
V03	0.7902	0.5849	0.5156	0.7332	0.6431	0.5469	0.7155	0.6640	0.5703	0.4647	0.8351	0.7266	0.7098	0.6665	0.6016
V04	0.8033	0.5743	0.4375	0.7713	0.6060	0.5156	0.7600	0.6257	0.4688	0.7870	0.5889	0.3750	0.7159	0.6656	0.4844
V05	0.9363	0.3539	0.2500	0.9457	0.3302	0.2656	0.8810	0.4664	0.3281	0.9202	0.3863	0.3594	0.8268	0.5591	0.4922
V06	0.8614	0.5186	0.4766	0.8793	0.4820	0.3672	0.8972	0.4472	0.3203	0.8539	0.5219	0.4375	0.7384	0.6776	0.5859
V07	0.7909	0.6242	0.4766	0.8907	0.4641	0.3906	0.8389	0.5569	0.5000	0.8546	0.5341	0.4844	0.7906	0.6249	0.5391
V08	0.8687	0.4987	0.4038	0.6282	0.7797	0.5481	0.8224	0.5549	0.4231	0.8463	0.5192	0.3365	0.7834	0.6114	0.5000
V09	0.6748	0.6822	0.6328	0.7525	0.6064	0.6406	0.7781	0.5782	0.6406	0.6853	0.6709	0.6328	0.6085	0.7304	0.7344
V10	0.6126	0.8365	0.6406	0.8307	0.5882	0.5313	0.8928	0.4756	0.4453	0.8661	0.5278	0.5625	0.8024	0.6302	0.6406
V11	0.5240	0.9419	0.6016	0.8448	0.5875	0.4922	0.8609	0.5596	0.4922	0.7604	0.7150	0.6328	0.7748	0.6944	0.5703
V12	0.7342	0.6570	0.5313	0.7567	0.6332	0.5469	0.8256	0.5460	0.5625	0.6422	0.7425	0.6641	0.4994	0.8380	0.7188
V13	0.8882	0.5227	0.5703	0.8900	0.5180	0.4766	0.9229	0.4393	0.4297	0.8940	0.5131	0.5469	0.6413	0.8777	0.6719
average	0.7808	0.6053	0.5082	0.8169	0.5596	0.4827	0.8178	0.5520	0.4917	0.7825	0.5955	0.5295	0.7231	0.6783	0.5968

12.1.2 CIF

	NTT FR			OPTICOM FR			Psytechnics FR			Yonsei FR			NTIA PSNR search		
	Correl	RMSE	Outlier	Correl	RMSE	Outlier	Correl	RMSE	Outlier	Correl	RMSE	Outlier	Correl	RMSE	Outlier
C01	0.7225	0.7037	0.5625	0.8131	0.5931	0.5938	0.8353	0.5594	0.5234	0.8412	0.5592	0.5156	0.7424	0.6851	0.6484
C02	0.8891	0.4044	0.3438	0.8616	0.4447	0.4141	0.8871	0.4047	0.3359	0.7419	0.5870	0.4922	0.7463	0.5815	0.5234
C03	0.8114	0.5485	0.5313	0.7001	0.6727	0.6484	0.8335	0.5196	0.4844	0.7208	0.6506	0.5859	0.7390	0.6322	0.6016
C04	0.8312	0.4761	0.3906	0.8343	0.4719	0.3750	0.7914	0.5222	0.4375	0.7292	0.5845	0.4688	0.7443	0.5710	0.5391
C05	0.8738	0.5311	0.4844	0.8798	0.5160	0.4531	0.9074	0.4546	0.4219	0.8973	0.4860	0.3750	0.7667	0.6967	0.6094
C06	0.8977	0.4930	0.4375	0.9136	0.4574	0.4219	0.9094	0.4637	0.4219	0.9160	0.4562	0.4219	0.8540	0.5848	0.5469
C07	0.8340	0.4797	0.4297	0.7602	0.5619	0.4688	0.8150	0.5006	0.4141	0.7561	0.5644	0.4844	0.4438	0.7757	0.6250
C08	0.7822	0.5408	0.4609	0.8862	0.3992	0.2734	0.8468	0.4553	0.4375	0.8935	0.3876	0.3516	0.7218	0.6014	0.5781
C09	0.7448	0.5340	0.4844	0.8473	0.4227	0.3750	0.8559	0.4104	0.3906	0.8386	0.4337	0.4063	0.5826	0.6481	0.5703
C10	0.7526	0.6862	0.5859	0.7337	0.6997	0.5781	0.7621	0.6647	0.6016	0.8403	0.5654	0.4922	0.6551	0.7789	0.6563
C11	0.7426	0.6872	0.6484	0.7540	0.6639	0.5625	0.8024	0.6039	0.5625	0.7497	0.6691	0.5313	0.4788	0.8875	0.6641
C12	0.8363	0.4800	0.4844	0.7578	0.5684	0.5859	0.7643	0.5619	0.5234	0.7260	0.5979	0.5234	0.6512	0.6652	0.5156
C13	0.8221	0.6253	0.6484	0.8817	0.5207	0.5859	0.9174	0.4327	0.5859	0.7275	0.7450	0.7422	0.7482	0.7244	0.7969
C14	0.8870	0.4626	0.5313	0.9529	0.3103	0.3047	0.9018	0.4314	0.4609	0.9026	0.4465	0.4453	0.8332	0.5605	0.5313
average	0.8162	0.5466	0.5017	0.8269	0.5216	0.4743	0.8450	0.4989	0.4715	0.8058	0.5524	0.4883	0.6934	0.6709	0.6004

12.1.3 QCIF

	NTT FR			OPTICOM FR			Psytechnics FR			Yonsei FR			NTIA PSNR search		
	Correl	RMSE	Outlier	Correl	RMSE	Outlier	Correl	RMSE	Outlier	Correl	RMSE	Outlier	Correl	RMSE	Outlier
Q01	0.8808	0.5590	0.4688	0.7955	0.7121	0.5781	0.8946	0.5274	0.4609	0.8963	0.5350	0.4844	0.7169	0.8225	0.6484
Q02	0.8688	0.4488	0.4766	0.7909	0.5502	0.4844	0.9013	0.3928	0.3906	0.8333	0.5118	0.4766	0.7611	0.5870	0.5703
Q03	0.7949	0.6169	0.5000	0.7627	0.6530	0.4609	0.6754	0.7477	0.5938	0.6343	0.7782	0.6328	0.7984	0.6133	0.5313
Q04	0.8606	0.4406	0.4141	0.8346	0.4764	0.4219	0.7535	0.5681	0.5625	0.7515	0.5718	0.5156	0.7401	0.5853	0.5234
Q05	0.8955	0.4417	0.3906	0.9359	0.3536	0.3203	0.9230	0.3814	0.3359	0.9325	0.3726	0.3281	0.8729	0.4915	0.5000
Q06	0.8859	0.4547	0.4219	0.9320	0.3580	0.3516	0.9546	0.2912	0.2031	0.9519	0.3181	0.2266	0.8356	0.5461	0.5234
Q07	0.8615	0.5257	0.5000	0.9008	0.4483	0.3828	0.9228	0.3975	0.3438	0.8994	0.4577	0.3672	0.7349	0.7011	0.5703
Q08	0.8861	0.4425	0.4063	0.8811	0.4496	0.4375	0.8815	0.4491	0.3828	0.8997	0.4188	0.3984	0.6147	0.7536	0.6719
Q09	0.7948	0.5938	0.5859	0.9064	0.4121	0.2891	0.8663	0.4871	0.4141	0.8482	0.5204	0.5000	0.6416	0.7535	0.6797
Q10	0.7970	0.4829	0.4297	0.8972	0.3633	0.3516	0.7115	0.5592	0.4766	0.6165	0.6277	0.6328	0.8094	0.4784	0.4844
Q11	0.7445	0.5981	0.6094	0.6900	0.6421	0.6719	0.6714	0.6552	0.5625	0.6701	0.6563	0.5156	0.6498	0.6824	0.5469
Q12	0.8405	0.5242	0.5469	0.9292	0.3595	0.3359	0.9055	0.4107	0.4609	0.8840	0.4636	0.5313	0.7887	0.6042	0.5469
Q13	0.8477	0.5822	0.5703	0.8575	0.5564	0.5625	0.8528	0.5619	0.5078	0.8125	0.6394	0.7031	0.6739	0.8018	0.7500
Q14	0.8717	0.4001	0.4063	0.9017	0.3542	0.3516	0.5313	0.6892	0.5703	0.4900	0.7093	0.5938	0.7782	0.5175	0.4531
average	0.8450	0.5080	0.4805	0.8583	0.4778	0.4286	0.8175	0.5085	0.4475	0.7943	0.5415	0.4933	0.7440	0.6384	0.5714

12.2 Per-condition analysis

12.2.1 VGA

	NTT FR		OPTICOM FR		Psytechnics FR		Yonsei FR		NTIA PSNR search	
	Correl	RMSE	Correl	RMSE	Correl	RMSE	Correl	RMSE	Correl	RMSE
V01	0.9760	0.2534	0.9345	0.4671	0.9584	0.3133	0.8667	0.5434	0.9329	0.4371
V02	0.9088	0.3909	0.9596	0.2608	0.6757	0.6770	0.9192	0.4232	0.6981	0.6600
V03	0.9719	0.2163	0.9578	0.2700	0.9072	0.3678	0.5909	0.6680	0.9098	0.3563
V04	0.9552	0.3003	0.9372	0.3473	0.9757	0.3459	0.8862	0.4417	0.8631	0.5097
V05	0.9824	0.1965	0.9927	0.1477	0.9685	0.2779	0.9843	0.2431	0.9478	0.4134
V06	0.9499	0.3146	0.9705	0.2520	0.9918	0.1594	0.9815	0.2405	0.9382	0.4438
V07	0.8196	0.4023	0.9342	0.2576	0.7424	0.4486	0.8675	0.3581	0.7231	0.4623
V08	0.9206	0.3447	0.5474	0.7080	0.9075	0.3867	0.8825	0.4443	0.7426	0.5546
V09	0.6631	0.5715	0.8665	0.3776	0.8634	0.3815	0.8074	0.4493	0.7203	0.5539
V10	0.8259	0.5091	0.9603	0.3070	0.9317	0.2707	0.8684	0.3653	0.7398	0.4905
V11	0.9210	0.4676	0.9646	0.2402	0.9750	0.2125	0.8769	0.4418	0.8757	0.4362
V12	0.7606	0.6567	0.9071	0.5022	0.9069	0.4480	0.7804	0.6747	0.7093	0.8123
V13	0.9339	0.3791	0.9552	0.3659	0.9291	0.3470	0.9240	0.3973	0.7180	0.8253
average	0.8914	0.3848	0.9144	0.3464	0.9026	0.3566	0.8643	0.4378	0.8091	0.5350

12.2.2 CIF

	NTT FR		OPTICOM FR		Psytechnics FR		Yonsei FR		NTIA PSNR search	
	Correl	RMSE	Correl	RMSE	Correl	RMSE	Correl	RMSE	Correl	RMSE
C01	0.9402	0.4634	0.9209	0.4570	0.8716	0.5024	0.9333	0.5090	0.8253	0.6541
C02	0.9369	0.2905	0.9386	0.3145	0.9777	0.2269	0.8839	0.4478	0.8066	0.5196
C03	0.8789	0.3518	0.8210	0.4150	0.9119	0.3177	0.9037	0.3818	0.7761	0.4864
C04	0.9471	0.2885	0.9501	0.2867	0.9251	0.3290	0.8136	0.4682	0.8248	0.4594
C05	0.9511	0.3497	0.9766	0.3231	0.9410	0.3653	0.9691	0.3710	0.8723	0.6026
C06	0.9686	0.3254	0.9645	0.3534	0.9602	0.3576	0.9777	0.3637	0.9309	0.5312
C07	0.9184	0.2734	0.9220	0.3211	0.9112	0.2989	0.9097	0.2997	0.9090	0.5506
C08	0.9330	0.2957	0.9403	0.2864	0.8839	0.3382	0.9413	0.2572	0.8958	0.5137
C09	0.8926	0.3346	0.9192	0.2887	0.9240	0.2655	0.9356	0.2763	0.8184	0.5509
C10	0.7826	0.5557	0.7783	0.5825	0.7189	0.6018	0.9205	0.4021	0.6581	0.6666
C11	0.7972	0.6377	0.8794	0.5074	0.9205	0.4004	0.8549	0.5445	0.6375	0.8438
C12	0.9560	0.2713	0.9097	0.3328	0.9047	0.3293	0.8626	0.4233	0.8703	0.5204
C13	0.9351	0.5136	0.9605	0.3627	0.9525	0.3104	0.6072	0.7562	0.6770	0.7174
C14	0.9688	0.2875	0.9903	0.1713	0.9820	0.2532	0.9807	0.3061	0.9395	0.4945
average	0.9148	0.3742	0.9194	0.3573	0.9132	0.3498	0.8924	0.4148	0.8173	0.5794

12.2.3 QCIF

	NTT FR		OPTICOM FR		Psytechnics FR		Yonsei FR		NTIA PSNR search	
	Correl	RMSE	Correl	RMSE	Correl	RMSE	Correl	RMSE	Correl	RMSE
Q01	0.9411	0.3424	0.9643	0.2709	0.9262	0.3844	0.9616	0.3029	0.7711	0.6681
Q02	0.9225	0.3363	0.9275	0.3821	0.9510	0.2797	0.9114	0.4326	0.8055	0.5661
Q03	0.9063	0.4157	0.8782	0.4739	0.7533	0.6229	0.7087	0.6538	0.9322	0.4746
Q04	0.9632	0.2610	0.9380	0.3141	0.9097	0.4033	0.8626	0.4606	0.8435	0.4990
Q05	0.9512	0.3055	0.9829	0.2093	0.9698	0.2743	0.9885	0.2799	0.9501	0.4265
Q06	0.9564	0.2950	0.9860	0.2162	0.9849	0.1770	0.9904	0.2404	0.9127	0.5084
Q07	0.9563	0.2661	0.9612	0.3017	0.9667	0.2584	0.9820	0.3142	0.9685	0.6422
Q08	0.9564	0.3091	0.9648	0.3123	0.9530	0.3306	0.9831	0.3384	0.9211	0.7758
Q09	0.9353	0.3356	0.9535	0.2814	0.9517	0.2978	0.9777	0.2960	0.9309	0.6848
Q10	0.9327	0.3303	0.9507	0.2637	0.9273	0.4726	0.7846	0.6229	0.8909	0.4099
Q11	0.9416	0.3211	0.7614	0.4153	0.8530	0.3555	0.7811	0.4561	0.8064	0.5160
Q12	0.9681	0.3726	0.9952	0.1376	0.9762	0.2518	0.9945	0.3022	0.9437	0.5113
Q13	0.9019	0.5155	0.9212	0.4581	0.9010	0.5018	0.8401	0.6389	0.7659	0.7929
Q14	0.9518	0.3113	0.9388	0.2967	0.8504	0.6811	0.7413	0.7257	0.9042	0.4346
average	0.9418	0.3370	0.9374	0.3095	0.9196	0.3780	0.8934	0.4332	0.8819	0.5650

3 Discussion

12.3 Per-sample analysis without common video clips

Some observations from the above results are shown below.

i. The performances of all proposed models are significantly better than that of PSNR.

ii. The model which achieves best performance for all subjective tests doesn't exist.

iii. The model which achieves best performance for all resolutions doesn't exist.

iv. The ranking of performance from this analysis is slightly different from that of "primary analysis" from the viewpoint of average correlation coefficient for all subjective tests.

For VGA the FR models from OPTICOM and Psytechnics perform slightly better than the two others. However, every tested model performs poorly in some experience, implying that there is not an absolutely best model. For CIF the performance of all FR models are very close. For QCIF the FR models from OPTICOM and NTT perform slightly better than the two others.

12.4 Per-condition analysis

Per-condition analysis shows in principle similar characteristics as per-sample analysis. However, the correlation coefficients generally increase about 0.1 for all subjective tests. For VGA the FR models from OPTICOM and Psytechnics perform slightly better than the two others. However, every tested model performs poorly in some experience, implying that there is not an absolutely best model. For CIF the performance of all FR models are very close. For QCIF, where the FR models show the best performance, the model from NTT shows the best prediction accuracy. The NTT model shows no disadvantages for any experiment (all correlation coefficients above .90)

4 Proposal

From these analysis, there are no critical differences in estimation accuracy among proposed FR models. Therefore, we propose these four models to be recommended in the new Recommendation.

Appendix VI.2 OPTICOM

Data Analysis Performed by OPTICOM

5 General Remarks on the Data Analysis

OPTICOM believes that the entire test has been performed in a fair and professional manner. It proved to be wise that most decisions related to the evaluation of the test were taken before the models were submitted. OPTICOM is convinced that changing some of these decisions after the model submission would be an unfair bias of the test. One such decision was to include the common data set in all experiments and to evaluate it for all experiments and models. Certainly this may panellize a model if it has difficulties with one sequence from the common set, but the same risk exists for all models. Also, one must consider that the same data were also included in all subjective tests. Other decisions that fall into this category would be to compare the FR and RR models to the MOS instead of the DMOS. It was decided to train the models against DMOS and if a model by chance predicts the MOS values with higher accuracy, this should be disregarded.

6 Alternative Data Aggregation Based on Ranking Calculation

The VQEG Multimedia testplan specifies three metrics for the statistical analysis of the benchmark results, namely the Pearson Correlation, the RMSE and the Outlier Ratio. For all three metrics the 95% confidence intervals as well as significance tests are specified. The testplan also specifies that priority is given to the correlation and not to the RMSE, the outlier ratio is not mentioned in this context (MM Testplan V1.19, chapter 8.3.2) and the fitting process as described in the testplan does not take it into account at all. When it comes to aggregating the data from the different experiments, the testplan only mentions the average values of the correlations, RMSE and OR values across all experiments. While this is a simple procedure, it has the drawback that the confidence intervals and significance tests are not taken into account. The alternative aggregation method described here is based on the above metrics and uses significance tests to calculate the ranking between the models for individual experiments. A method to estimate the ranking across all experiments is proposed as well. The following chapters describe the method and the results obtained by applying it to the VQEG MM test results.

12.5 Limitations of the Alternative Aggregation Method

We do not see any limitations as far as calculating the top rank for each individual experiment is concerned, since the procedure is strictly based on statistically sound metrics described in the VQEG MM testplan and uses the priority between the metrics as defined by VQEG (that the OR should have the least priority was implied since it is not mentioned in the testplan). The distinction between ranks two and below should however take the multiple comparisons involved into account, which is not the case here. Since ranks below two are rare for the tested models, this simplification seems acceptable.

Nevertheless, the aggregation of the ranks by summing them up should not be seen as the ultimate truth for the following reasons:

- Similar as for the averaged correlations etc., there is no confidence interval known for the rank sum. In contrast to the averages of the plain metrics however, the proposed method takes the confidence intervals of the underlying metrics into account when calculating the ranks for the individual experiments.

- If model A and B differ in only one experiment, this should not be over weighted since it might be by chance and if more or slightly different experiments were conducted, the situation could be vice versa.
- If model A occupies rank three in one experiment and B is twice on rank two, and both models occupy the same rank otherwise, their rank sum would be the same and we don't know of any method to decide which model is better in this case.
- Due to the involved "Fisher's z transformation" and its non-linearity, the significance test for the correlations is very tolerant if the correlations are low and very strict if the correlations are high. This may lead to false impressions for experiments where a model has correlations below 0.8. Nevertheless, the decision is statistically correct.
- Due to the large confidence intervals we consider the method of limited use if the correlations of the two compared models are low (<0.75)
- Due to the statistically small number of samples (152 for the FR models) each individual outlier contributes 0.0065 to the OR. This is a fairly coarse quantisation.
- If all models in question have a rank sum which is noticeable higher than the optimum rank sum would be, the meaning of the ranking becomes less significant. This is an indication that all models fail from time to time, or that they simply swap ranks between different experiments.
- The tests involve comparisons to hard thresholds. This may lead to a different ranking between two models due to round off errors.

Due to these uncertainties we propose to see two models as performing equally good if their rank sum does not differ by more than three. If this is sufficiently large can be discussed, but smaller values make certainly no sense.

We do not claim that the rank sum represents the optimum procedure to identify the overall ranking, but it can give valuable additional evidence for a certain ranking. In any case it should not be seen isolated. Furthermore additional aggregated parameters like average correlations etc. should be taken into account as well.

12.6 Results from the Ranking Procedure

This analysis has been performed for the FR models only. The results are shown in Table 1 to Table 3.

	PSNR	Psytechnics_FR	OPTICOM_FR	Yonsei_FR	NTT
Sum	33	18	20	21	25
Top Rank Count	*0*	*9*	*7*	*6*	*4*

Table 1, Ranking of the FR models for all VGA experiments

	PSNR	Psytechnics_FR	OPTICOM_FR	Yonsei_FR	NTT
Sum	36	15	18	21	24
Top Rank Count	*0*	*13*	*10*	*9*	*6*

Table 2, Ranking of the FR models for all CIF experiments

	PSNR	Psytechnics_FR	OPTICOM_FR	Yonsei_FR	NTT
Sum	39	17	19	32	23
Top Rank Count	*1*	*11*	*9*	*1*	*6*

Table 3, Ranking of the FR models for all QCIF experiments

12.7 Discussion of the Ranking Results

The best models according to this method would be:

- VGA: OPTICOM plus two other models
- CIF: OPTICOM plus one other model
- QCIF: OPTICOM plus one other model

These results are very similar to those based on analysing the average correlations by human reason. The overall ranking remains the same independent of whether the rank sum is calculated or whether it is counted how often a model occupies the top rank.

7 Special Remarks to the OPTICOM Model

The OPTICOM model showed excellent performance and very few outliers. Due to the preparation of this report and the ongoing data analysis very little time remained for a detailed investigation of individual outliers. Nevertheless, many could be fixed already by simple modifications. The fixed model performs better than 0.8 correlation for all individual VGA experiments, although the degree of freedom for this improved version is lower than it was for the submitted version since one more or less unused internal indicator has been removed. The processing requirements of this improved version are also lower.

Appendix VI.3 Psytechnics

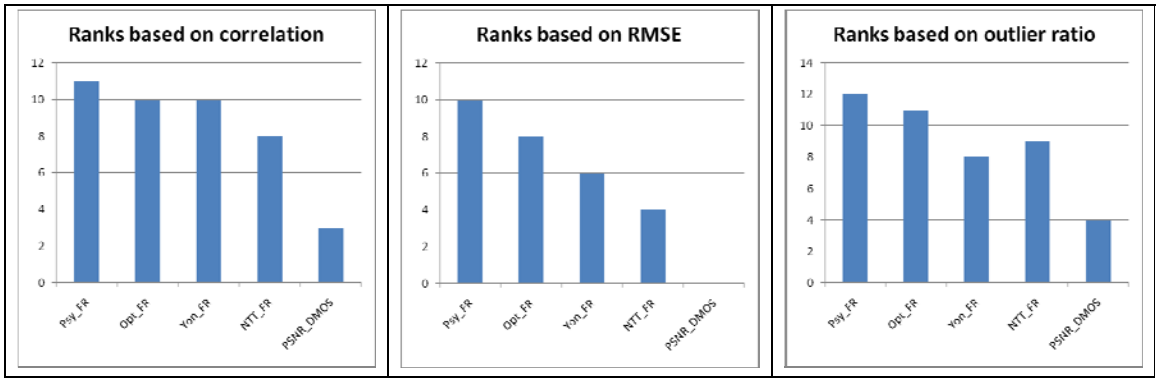
8 Comments on the performance of the Psytechnics FR model

VQEG agreed on 3 performance evaluation metrics (correlation, RMSE and outlier ratio) and on the corresponding statistical significance tests to discriminate the difference in performance between the objective models. The significance tests were applied per experiment using each of the metrics to check if the difference of performance between models was significant or not on that experiment. A number of times a model is at the top (rank 1) can therefore be calculated for each image resolution.

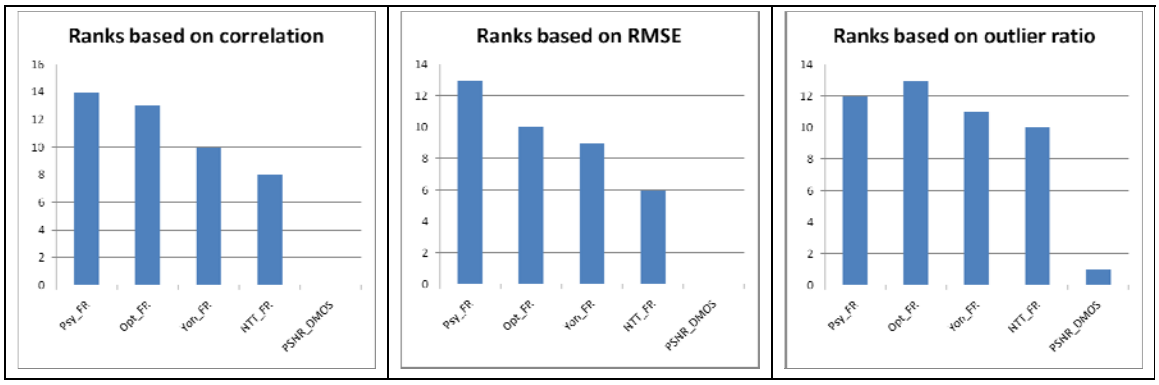
Based on the data analysis provided by the Independent Lab Group (ILG), the Psytechnics FR model was always ranked top at each of the 3 resolutions (QCIF, CIF and VGA) and based on any of the 3 metrics (See Psy_FR in following graphs):

- Based on correlation, the Psytechnics model had the highest number of occurrences of being at rank 1 (top performing) for all resolutions.
- Based on RMSE, the Psytechnics model had the highest number of occurrences of being at rank 1 (top performing) for all resolutions.
- Based on outlier ratio, the Psytechnics model had the highest number of occurrences of being at rank 1 (top performing) for QCIF and VGA. For CIF, the absolute value of the number of occurrences is not the highest but is statistically equivalent to the highest.
- For QCIF, the Psytechnics model had the highest number of occurrences at rank 1 for all metrics, i.e. top if ranking is based on correlation and top if ranking is based on RMSE and top if ranking is based on outlier ratio.
- For CIF, the Psytechnics model had the highest number of occurrences at rank 1 for correlation and RMSE, whereas for outlier ratio the number is statistically similar to the highest value.
- For VGA, the Psytechnics model had the highest number of occurrences at rank 1 for all metrics, i.e. top if ranking is based on correlation and top if ranking is based on RMSE and top if ranking is based on outlier ratio.

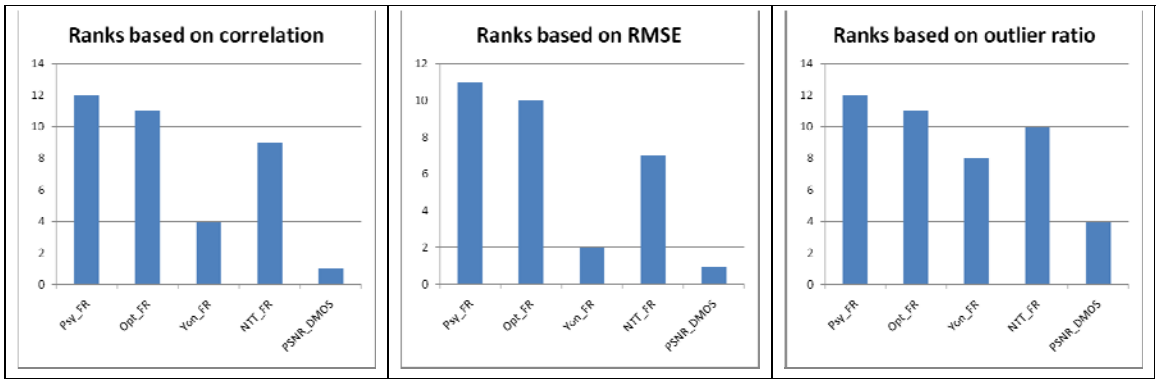
For VGA:



For CIF:



For QCIF:

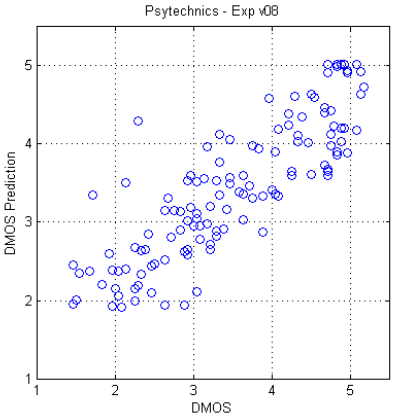
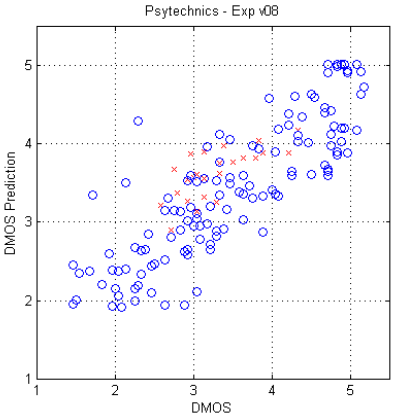
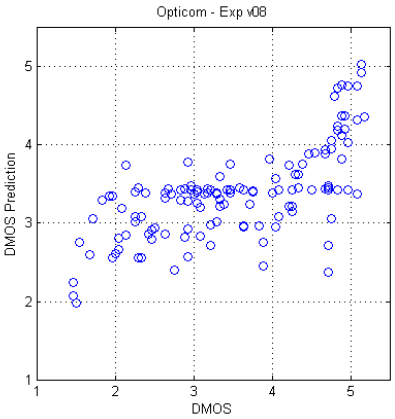
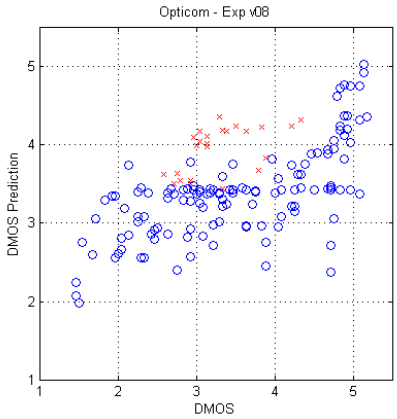
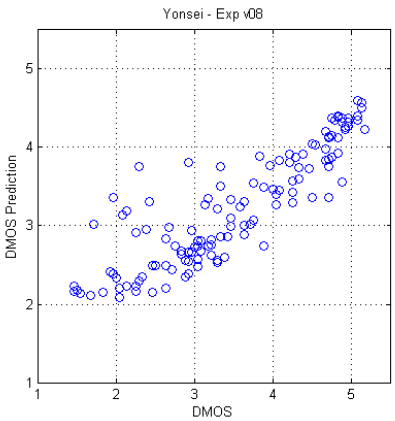
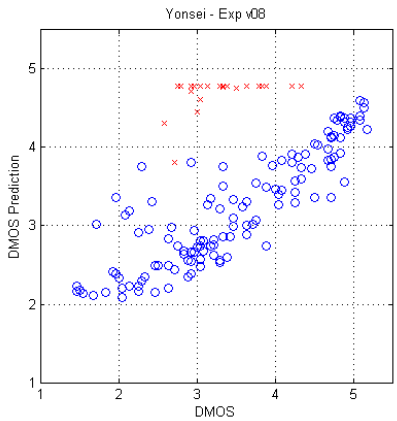


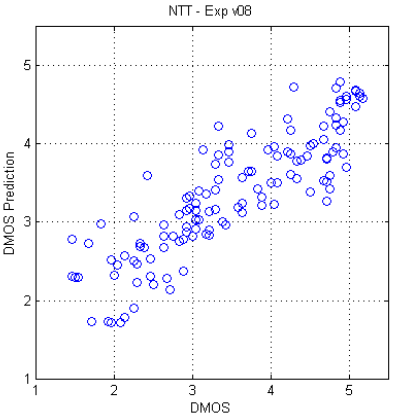
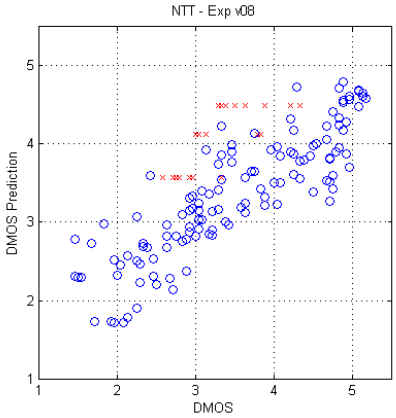
9 Exclusion of some data points

For experiment v08, VQEG decided to remove 3 test conditions - HRC 7, 8 and 9 - in the official data analysis because these test conditions exhibited only temporal degradations (i.e. frame freezing due to transmission errors) without any spatial degradation (lossless coding). This represents 24 data points in experiment v08.

The scatter plots of the candidate models are shown below respectively when (a) excluding and when (b) including these test conditions in the performance evaluation. In plots (b), the 24 files corresponding to the 3 test conditions are marked by 'x'.

We observe that the Psytechnics model can handle well these conditions that were removed from data analysis.

Psytechnics		
OPTICOM		
Yonsei		

NTT		
	(a)	(b)

For all models: (a) Scatter plots excluding HRC 7/8/9; (b) Scatter plots including HRC 7/8/9

10 Test files corresponding to quality enhancement condition and low-quality reference video

Some reference videos received a very low subjective quality with MOS < 4. In total, there were 2 reference videos in QCIF, 13 reference videos in CIF and 10 reference videos in VGA with MOS<4. For a reference (SRC) with low MOS, it is possible to have a degraded video (PVS) of higher quality than the reference (i.e. DMOS>5) corresponding to a test condition corresponding to a quality enhancement.

This case scenario was not part of the scope of the MM Phase I test and the Psytechnics model was not designed to address quality measurement for cases of quality enhancement where the PVS is of higher quality than the reference.

Furthermore, the model expects a reference of high quality (with MOS>4) and therefore might have been less accurate to evaluate the quality of a processed video for which the corresponding reference video received a low MOS. The ILG however decided to keep all these data points in the analysis.

When removing all data points for which the corresponding reference video received MOS<4 (101 files for VGA, 85 files for CIF and 21 files for QCIF) and all data points corresponding to DMOS>5 (60 files for VGA, 18 files for CIF and 14 files for QCIF), improvement in performance of the Psytechnics model is observed for the following experiments:

	All data			Data excluding cases with DMOS>5 and cases for which reference MOS<4		
	Correlation	RMSE	Outl ratio	Correlation	RMSE	Outl ratio
v01	0.884	0.505	0.566	0.887	0.489	0.560
v03	0.749	0.669	0.572	0.750	0.627	0.555
v04	0.735	0.652	0.507	0.803	0.575	0.478

v05	0.892	0.486	0.368	0.894	0.471	0.350
v07	0.843	0.556	0.487	0.849	0.525	0.444
c01	0.823	0.587	0.546	0.831	0.574	0.541
c03	0.823	0.550	0.513	0.828	0.533	0.500
c04	0.796	0.525	0.480	0.800	0.514	0.458
c07	0.804	0.535	0.454	0.808	0.524	0.439
c08	0.826	0.503	0.487	0.834	0.487	0.476
c09	0.852	0.432	0.434	0.857	0.425	0.426
c10	0.769	0.663	0.605	0.764	0.658	0.593
c13	0.897	0.472	0.625	0.895	0.468	0.620

11 Data fitting

As described in the VQEG Multimedia Test Plan, the metrics (correlation coefficient, RMSE and outlier ratio) were obtained after fitting of the raw objective data (i.e. raw model output) to the subjective data per experiment using a 3rd-order monotonic polynomial fitting function. This data fitting is done per experiment. Data fitting is performed as it is not reasonable to expect that objective models of video quality can replicate the limitations of subjective testing, e.g., subjective ratings compressed at the ends of the rating scale, difference in culture and language.

A comparison between the correlation obtained when using the fitted objective data on the one hand and using the raw objective data on the other hand provides an indication of the robustness and applicability of the model in the real world as fitting functions are not usually applied on the model's prediction in a real-world application. If there is little difference in correlation when using the fitted objective data or the raw objective data, this indicates that the model will be robust in the real-world. On the other hand if there is substantial difference in correlation when using fitted or raw data, this indicates that the model's performance is artificially enhanced by the fitting of the data.

The Psytechnics model presents little difference in correlation when using the fitted data or raw data to evaluate its performance. The fitting of the data increases the average correlation by only 1.2%, 0.07% and 0.06% respectively for VGA, CIF and QCIF. This shows that the raw output of the model (without data fitting) has already a good linear relationship with subjective data.

12 Comments on the performance of the Psytechnics NR model

No-reference models are primarily used in applications where measurements can be repeated over a large number of samples. Analysing large data sets mitigates the effects of the measurement noise inherent in no-reference model predictions and can be used to identify systematic trends and problems.

Primary analysis by VQEG uses a per-file analysis for computing all performance metrics. However, for NR models, the secondary analysis as agreed by VQEG is highly relevant. A NR model that provides good per-condition performance has a use, which is to identify systematic problems through statistical analysis of multiple measurements (as opposed to alarming on single events). There are many areas where systematic problems can occur, e.g., sub-optimal configuration of a codec.

13 Comments on the VQEG Multimedia Phase I tests

The 41 MM subjective experiments covered a very wide range of test condition parameters in terms of image resolution, codecs, bit rates, frame rates, transmission errors, and additional processing (such as colour space conversions). These experiments therefore included a very wide range of visual distortions and represented a very difficult challenge for candidate objective models.

Due to this very wide range of distortions and the very high number of test video files (more than 5000 test files), it would not be expected that a particular objective model would perform very well on all 41 subjective experiments. The VQEG Multimedia Phase I Validation represents until today the only independent evaluation and most critical benchmarking of video objective models. For comparison, VQEG FRTV Phase 2 evaluated the objective models included in ITU-T J.144 using only 2 subjective experiments, with a total of 128 test files (less than the number of files in one single experiment in this MM Phase I).

Appendix VI.4 SwissQual

Proponent Analysis of Results: SwissQual

Introduction

SwissQual has submitted a no-reference MOS prediction model to VQEG for an independent performance evaluation. The model is part of the VMon analysis suite and targets the QCIF, CIF and VGA resolution groups as well as provides a predicted overall video MOS.

A no-reference model only analyzes the video sequence that is received during a test. As a result, this model has a lower prediction accuracy than a full-reference model, which also analyzes the reference signal.

Content dependency of perceived quality and prediction problems

A no-reference model can detect typical compression and transmission distortions, but cannot separate distinguish between these artifacts and content areas. For example, naturally occurring content with soft edges, such as a cloudy sky or a meadow, is scored as blurry, a graphical object is scored as a compression artifact, and a cartoon containing only a few different colors in wide areas is scored as unnatural. However, if the content has a natural spatial complexity and a minimum of movement, a no-reference model can deliver worthwhile results.

Application of no-reference models

Unlike a full-reference model where a user has full control over the video sequences, pure codec evaluation and tuning is not the focus of a no-reference model. Instead, a no-reference model is typically applied in a situation where a user does not have access to the source video, for example, in-service monitoring of networks, streaming applications from unknown sources, and live TV applications. In these cases, a user is determined to find the best compromise between codec settings and the current network behavior.

Although a no-reference model is optimized for this purpose, usage guidelines and the interpretation of results must also be considered. To demonstrate the performance of the SwissQual no-reference MOS prediction of VMon, the following typical use cases are considered:

1. Quality evaluation of a specific transmission chunk or a specific location while requesting video streams from a live TV server. This evaluation is used for service optimization or benchmarking.
2. Network monitoring by an in-service observation to find severe quality problems.

In **use case 1)**, the aim is to analyze the general behavior of a transmission channel from a user perspective by using the service over a period of time. For this type of analysis, the user behavior is determined by analyzing a series of typical video examples and not by analyzing a short individual video sequence. This series can consist of several samples that are taken from a longer sequence or of several samples that are taken from typical content categories during a longer observation period.

For simplification, the model uses a combination of compression ratios, frame-rates, and specific error patterns to target a specific codec type. By averaging across the different contents in a transmission condition (known as HRC in this document), the model can create a general view of a channel.

Furthermore, averaging across the individual contents for each condition dramatically minimizes the content dependency of the perceived quality as well as the content dependency of the model.

The following procedures can be used for content averaging:

HRC 1 is the method that is used for secondary analysis in this report. Each predicted MOS value is transformed by a third order mapping function that is derived from the entire set of samples in an experiment. After the transformation, the predicted and the subjective MOS are averaged over the different contents. The correlation coefficient and RMSE are then calculated (excluding common set). The average values over all experiments for each resolution are shown in Table 1.

HRC 2 is the method that is usually applied in ITU-T for speech quality measures. In this method, the predicted MOS and the subjective MOS are averaged over the contents and then the third order mapping is applied to all 'per-condition' values (excluding common set).

Table 1: Mean correlation coefficient over all experiments for each format.

Format	mean cor (PVS)	mean cor (HRC 1)	mean cor (HRC 2)
QCIF	0.661	0.864	0.903
CIF	0.543	0.800	0.836
VGA	0.476	0.789	0.835

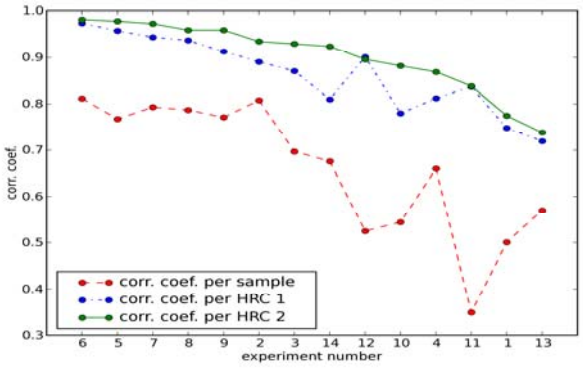
Format	mean RMSE (PVS)	mean RMSE (HRC 1)	mean RMSE (HRC 2)
QCIF	0.717	0.549	0.362
CIF	0.820	0.630	0.446
VGA	0.885	0.681	0.443

Table 1 shows that the performance for both kinds of averaging procedures significantly increases, i.e. the correlation coefficient is larger.

The principal behavior for both methods is similar. Upon closer examination of the design of the experiments, it can be stated that the methods perform well for experiment 5 to 9 for all resolutions. This performance is the result of the straight design that applies most test conditions, such as compression ratios and error conditions, to one codec type only. Since the type of distortion remains similar but the amount varies, this approach leads to very consistent experiments in the subjective domain and especially in objective prediction.

Experiment 13, which is a combination of compression and transmission errors for 7 different codecs, yields the poorest performance.

Figure 1: Correlation coefficients for different evaluation methods, QCIF format, sorted with respect to second averaging method.



In **use case 2)**, the behavior of a transmission channel in a live scenario should be observed and critical quality issues should be signalized accordingly. This signaling can be seen as a threshold-based trigger. For simplification, the threshold is only applied to the pure predicted MOS value of each sample. In a real world application, all the partial results can be used to produce more confident results.

The following rules are applied to the data:

Threshold signalizing bad quality: *< 2.5*
Uncertainty of subjective test results: *0.2 MOS*
Criteria A 'False Rejection': *MOS > 2.7 & MOSpred < 2.5*
Criteria B 'False Acceptance': *MOS < 2.3 & MOSpred > 2.5*

Table 2: False Acceptance and false rejection ratio over all experiments for each format.

Format	mean fA (PVS)	mean fR (PVS)	mean fA (HRC 1)	mean fR (HRC 1)	mean fA (HRC 2)	mean fR (HRC 2)
QCIF	0.119	0.080	0.080	0.025	0.034	0.042
CIF	0.164	0.114	0.143	0.042	0.059	0.071
VGA	0.176	0.085	0.142	0.050	0.060	0.069

The results in Table 2 show that an alarm is incorrectly raised in approximately 10% of the cases based on a per-sample evaluation and that this percentage decreases significantly after HRC averaging. However, no-spotted quality problems remain within a range of 15%.

In a real world application, such decisions are not exclusively based on an MOS. Instead, these decisions also take partial results of the analysis into account, which leads to even more confident results.

No-reference models can be used in certain applications which cannot be addressed by full-reference approaches and can deliver worthwhile results.

Appendix VI.5 Yonsei University

14 Un-proportional representation of the common sets

In each format (QCIF, CIF and VGA), a test consists of 152 video clips which include 24 common clips. Since the common sets are included in every test, they are un-proportionally weighted. Tables 1-3 show the performance comparison of the three metrics (correlation, RMSE, outlier ratios) before and after the common sets are excluded. Significant improvements were observed for Yonsei FR and RR models for QCIF.

Table 1. Averages of the three metrics for VGA (with/without the common set)

VGA	NTT FR	OP FR	Psy FR	Yonsei FR	Yonsei RR10k	Yonsei RR64k	Yonsei RR128k	PSNR/NTIA
Cor	0.786 /0.781	0.825 /0.818	0.822 /0.818	0.805 /0.784	0.803 /0.790	0.803 /0.791	0.803 /0.791	0.713 /0.724
RMSE	0.621 /0.599	0.571 /0.554	0.566 /0.547	0.593 /0.591	0.599 /0.589	0.599 /0.590	0.598 /0.589	0.714 /0.674
OR	0.523 /0.516	0.502 /0.486	0.523 /0.499	0.542 /0.529	0.556 /0.541	0.553 /0.537	0.552 /0.535	0.615 /0.600

Table 2. Averages of the three metrics for CIF (with/without the common set)

CIF	NTT FR	OP FR	Psy FR	Yonsei FR	Yonsei RR10k	Yonsei RR64k	PSNR/NTIA
Cor	0.777 / 0.818	0.808 / 0.828	0.836 / 0.845	0.785 / 0.807	0.780 / 0.802	0.782 / 0.802	0.656 / 0.699
RMSE	0.604 / 0.539	0.562 / 0.517	0.526 / 0.497	0.594 / 0.546	0.593 / 0.549	0.590 / 0.548	0.720 / 0.663
OR	0.538 / 0.487	0.513 / 0.473	0.506 / 0.465	0.521 / 0.481	0.518 / 0.486	0.510 / 0.483	0.632 / 0.580

Table 3. Averages of the three metrics for QCIF (with/without the common set)

QCIF	NTT FR	OP FR	Psy FR	Yonsei FR	Yonsei RR1k	Yonsei RR10k	PSNR/NTIA
Cor	0.819 / 0.851	0.841 / 0.858	0.830 / 0.818	0.756 / 0.799	0.771 / 0.797	0.791 / 0.832	0.662 / 0.749
RMSE	0.551 / 0.496	0.516 / 0.475	0.517 / 0.506	0.617 / 0.527	0.604 / 0.542	0.578 / 0.491	0.721 / 0.627
OR	0.497 / 0.458	0.461 / 0.423	0.457 / 0.447	0.523 / 0.463	0.505 / 0.479	0.486 / 0.450	0.596 / 0.557

Tables 4-6 show the significant test results of the three metrics for VGA, CIF and QCIF FR models before and after the common sets are excluded. The tables show the occurrences in the top group (models which are statistically identical with the best performing model). Noticeable improvements were observed for Yonsei FR models for QCIF.

Table 4. Number of occurrences in the top group for VGA FR models only (with/without the common set).

VGA	NTT FR	OP FR	Psy FR	Yonsei FR	PSNR/NTIA
Cor	8 / 9	10 / 10	11 / 11	10 / 9	3 / 3
RMSE	4 / 5	8 / 8	10 / 9	6 / 3	0 / 1
OR	9 / 9	11 / 11	12 / 11	8 / 8	4 / 5

Table 5. Number of occurrences in the top group for CIF FR models only (with/without the common set)

CIF	NTT FR	OP FR	Psy FR	Yonsei FR	PSNR/NTIA
COR	8 / 8	13 / 12	14 / 13	10 / 8	0 / 1
RMSE	6 / 7	10 / 9	13 / 10	9 / 7	0 / 0
OR	11 / 12	13 / 13	12 / 11	11 / 11	1 / 4

Table 6. Number of occurrences in the top group for QCIF FR models only (with/without the common set)

QCIF	NTT FR	OP FR	Psy FR	Yonsei FR	PSNR/NTIA
COR	9 / 9	11 / 12	12 / 10	4 /9	1 / 2
RMSE	7 / 8	10 / 11	11 / 8	2 / 7	1 / 1
OR	10 / 9	11 / 11	12 / 10	8 / 8	4 / 3

Tables 7-9 show the significant test results of the three metrics for the FR/RR models before and after the common sets are excluded. It is noted that the significant tests for the RR models were applied to the combined pool of the FR and RR models.

Table 7. Number of occurrences in the top group for VGA FR/RR models (with/without the common set). The significant tests were applied to the combined pool of the FR and RR models.

VGA	NTT FR	OP FR	Psy FR	Yonsei FR	Yonsei RR10k	Yonsei RR64k	Yonsei RR128k	PSNR/NTIA
Cor	8 / 9	10 / 10	11 / 11	10 / 9	8 / 8	8 / 8	8 / 8	3 / 3
RMSE	4 / 5	8 / 8	9 / 8	6 / 3	7 / 5	7 / 5	7 / 5	0 / 1
OR	9 / 9	11 / 10	12 / 11	8 / 8	7 / 7	7 / 8	7 / 8	4 / 5

Table 8. Number of occurrences in the top group for CIF FR/RR models (with/without the common set). The significant tests were applied to the combined pool of the FR and RR models.

CIF	NTT FR	OP FR	Psy FR	Yonsei FR	Yonsei RR10k	Yonsei RR64k	PSNR/NTIA
COR	8 / 8	13 / 12	14 / 13	10 / 8	10 / 7	10 / 8	0 / 1
RMSE	5 / 6	10 / 8	13 / 10	9 / 7	6 / 7	6 / 7	0 / 0
OR	10 / 10	12 / 12	12 / 11	10 / 11	12 / 10	11 / 9	1 / 3

Table 9. Number of occurrences in the top group for QCIF FR/RR models (with/without the common set). The significant tests were applied to the combined pool of the FR and RR models.

QCIF	NTT FR	OP FR	Psy FR	Yonsei FR	Yonsei RR1k	Yonsei RR10k	PSNR/NTIA
COR	9 / 8	11 / 11	11 / 10	4 /9	6 / 8	7 / 11	1 / 2
RMSE	7 / 8	10 / 9	10 / 6	2 / 7	2 / 4	5 / 10	0 / 1
OR	9 / 7	11 / 11	12 / 10	8 / 8	8 / 10	9 / 10	3 / 3

15 Registration error

In the Multimedia testplan (Ver. 1.19), it is stated (2. List of Definitions):

"Pausing without skipping (formerly frame freeze) is defined as any event where the video pauses for some period of time and then restarts without losing any video information."

Then, in section 6.3.4, it is also stated that:

“Pausing without skipping events will not be included in the current testing.”

However, if there is one-bit information loss, anything would be allowed, including “pausing without skipping.” Due to this ambiguity and misunderstanding, substantial changes had to be made to the registration routines just before model submission. After Yonsei models were submitted, some minor errors were found. Once the errors are corrected, the performance was noticeably improved. Figures 1-6 show performance improvement after the error correction with the common sets included. Tables 10-12 show the three metrics after error correction. Tables 13-15 show the significant test results for the FR models after error correction. It is noted that the significant tests for the FR models were applied to the FR models only. Tables 16-18 show the significant test results of the three metrics for the FR/RR models. It is noted that the significant tests for the RR models were applied to the combined pool of the FR and RR models. With the error correction, Yonsei FR and RR models show noticeable improvement.

Figure 1.FR Correlation & RMSE (per-clip) after error correction – VGA (common set included)

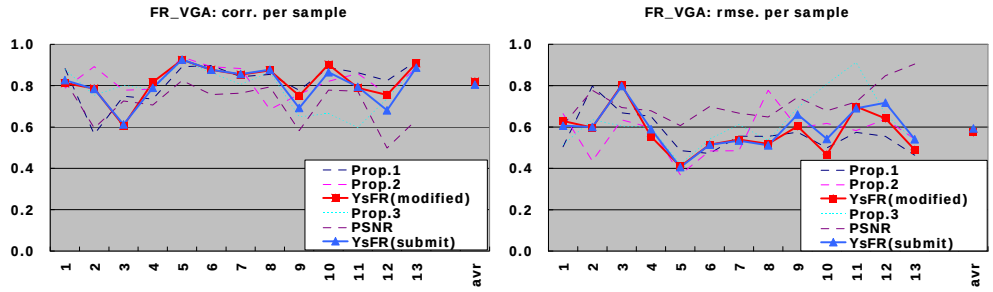


Figure 2.RR Correlation & RMSE (per-clip) after error correction – VGA (common set included)

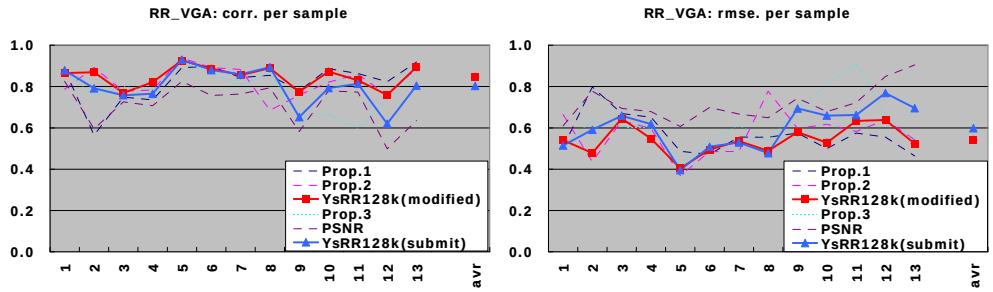


Figure 3. FR Correlation & RMSE (per-clip) after error correction – CIF (common set included)

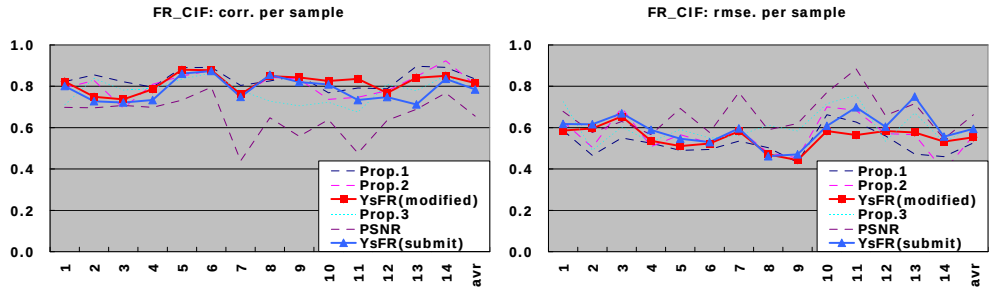


Figure 4. RR Correlation & RMSE (per-clip) after error correction - CIF (common set included)

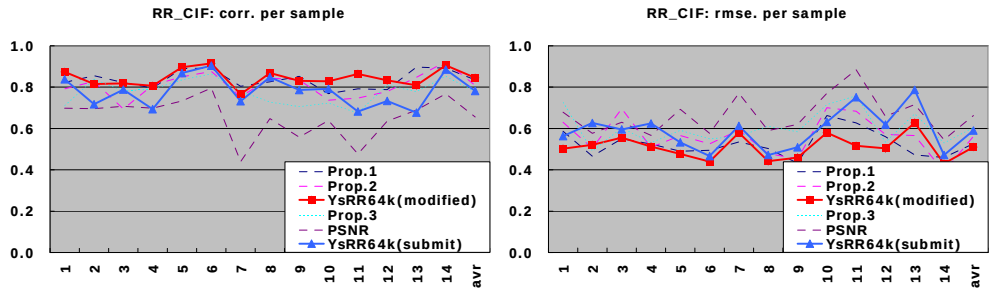


Figure 5. FR Correlation & RMSE (per-clip) after error correction - QCIF (common set included)

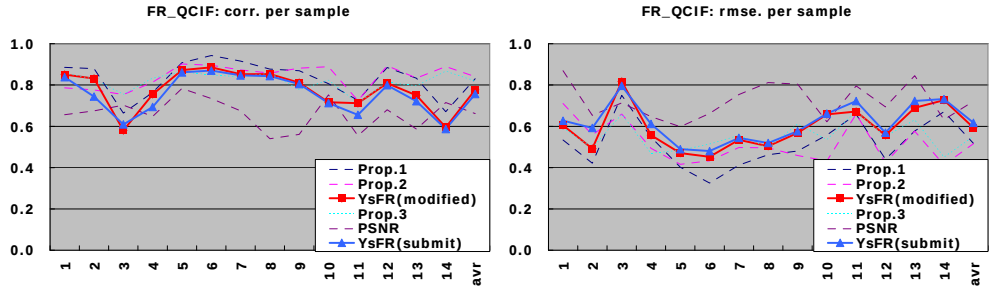


Figure 6. RR Correlation & RMSE (per-clip) after error correction - QCIF (common set included)

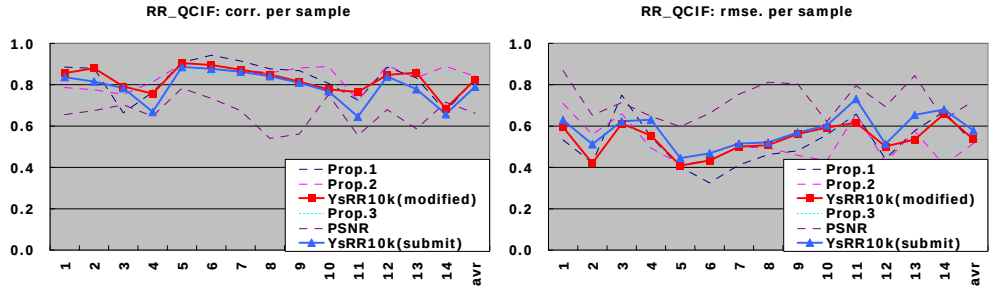


Table 10. Averages of the three metrics for VGA after error correction (with/without the common set)

VGA	NTT FR	OP FR	Psy FR	Yonsei FR	Yonsei RR10k	Yonsei RR64k	Yonsei RR128k	PSNR/NTIA
Cor	0.786 /0.781	0.825 /0.818	0.822 /0.818	0.820 /0.796	0.837 /0.829	0.846 /0.837	0.847 /0.837	0.713 /0.724
RMSE	0.621 /0.599	0.571 /0.554	0.566 /0.547	0.575 /0.577	0.557 /0.547	0.544 /0.535	0.542 /0.534	0.714 /0.674
OR	0.523 /0.516	0.502 /0.486	0.523 /0.499	0.533 /0.519	0.531 /0.512	0.531 /0.509	0.527 /0.503	0.615 /0.600

Table 11. Averages of the three metrics for CIF after error correction (with/without the common set)

CIF	NTT FR	OP FR	Psy FR	Yonsei FR	Yonsei RR10k	Yonsei RR64k	PSNR/NTIA
Cor	0.777 / 0.818	0.808 / 0.828	0.836 / 0.845	0.816 / 0.836	0.829 / 0.844	0.845 / 0.853	0.656 / 0.699

RMSE	0.604 / 0.539	0.562 / 0.517	0.526 / 0.497	0.553 / 0.507	0.535 / 0.502	0.511 / 0.487	0.720 / 0.663
OR	0.538 / 0.487	0.513 / 0.473	0.506 / 0.465	0.493 / 0.443	0.489 / 0.451	0.467 / 0.435	0.632 / 0.580

Table 12. Averages of the three metrics for QCIF after error correction (with/without the common set)

QCIF	NTT FR	OP FR	Psy FR	Yonsei FR	Yonsei RR1k	Yonsei RR10k	PSNR/NTIA
Cor	0.819 / 0.851	0.841 / 0.858	0.830 / 0.818	0.777 / 0.812	0.771 / 0.797	0.826 / 0.844	0.662 / 0.749
RMSE	0.551 / 0.496	0.516 / 0.475	0.517 / 0.506	0.593 / 0.508	0.604 / 0.542	0.536 / 0.480	0.721 / 0.627
OR	0.497 / 0.458	0.461 / 0.423	0.457 / 0.447	0.500 / 0.445	0.505 / 0.479	0.459 / 0.434	0.596 / 0.557

Table 13. Number of occurrences in the top group for VGA FR after error correction (with/without the common set).

VGA	NTT FR	OP FR	Psy FR	Yonsei FR	PSNR/NTIA
Cor	8 / 9	10 / 10	11 / 11	9 / 9	2 / 3
RMSE	4 / 5	8 / 8	9 / 9	8 / 5	0 / 1
OR	9 / 9	12 / 11	12 / 11	8 / 9	4 / 5

Table 14. Number of occurrences in the top group for CIF FR after error correction (with/without the common set)

CIF	NTT FR	OP FR	Psy FR	Yonsei FR	PSNR/NTIA
Cor	7 / 7	11 / 11	14 / 13	11 / 10	0 / 1
RMSE	6 / 6	9 / 8	13 / 9	10 / 8	0 / 0
OR	10 / 11	11 / 11	11 / 11	12 / 13	1 / 3

Table 15. Number of occurrences in the top group for QCIF FR after error correction (with/without the common set)

QCIF	NTT FR	OP FR	Psy FR	Yonsei FR	PSNR/NTIA
Cor	9 / 8	11 / 12	12 / 10	7 / 10	1 / 2
RMSE	7 / 7	10 / 11	11 / 7	3 / 8	1 / 1
OR	10 / 8	11 / 10	12 / 9	8 / 9	4 / 3

Table 16. Number of occurrences in the top group for VGA FR/RR after error correction (with/without the common set). The significant tests were applied to the combined pool of the FR and RR models.

VGA	NTT FR	OP FR	Psy FR	Yonsei FR	Yonsei RR10k	Yonsei RR64k	Yonsei RR128k	PSNR/NTIA
Cor	8 / 9	10 / 10	11 / 11	9 / 9	13 / 12	13 / 13	13 / 13	2 / 2
RMSE	4 / 5	8 / 8	9 / 8	8 / 5	9 / 4	11 / 5	12 / 5	0 / 1
OR	9 / 9	11 / 11	12 / 11	8 / 9	9 / 9	9 / 9	11 / 9	3 / 3

Table 17. Number of occurrences in the top group for CIF FR/RR after error correction (with/without the common set). The significant tests were applied to the combined pool of the FR and RR models.

CIF	NTT FR	OP FR	Psy FR	Yonsei FR	Yonsei RR10k	Yonsei RR64k	PSNR/NTIA
Cor	5 / 7	10 / 11	13 / 12	11 / 10	13 / 11	13 / 11	0 / 1
RMSE	5 / 5	7 / 7	11 / 8	7 / 8	12 / 10	13 / 10	0 / 0
OR	8 / 9	10 / 10	11 / 11	12 / 13	13 / 11	13 / 13	1 / 1

Table 18. Number of occurrences in the top group for QCIF FR/RR after error correction (with/without the common set). The significant tests were applied to the combined pool of the FR and RR models.

QCIF	NTT FR	OP FR	Psy FR	Yonsei FR	Yonsei RR1k	Yonsei RR10k	PSNR/NTIA
Cor	9 / 7	11 / 12	11 / 9	6 / 10	4 / 6	10 / 10	1 / 1
RMSE	6 / 6	10 / 8	10 / 5	3 / 7	1 / 3	7 / 9	0 / 1
OR	8 / 7	11 / 10	12 / 9	8 / 8	8 / 9	10 / 9	3 / 3