NEW PSYCHO-PHYSICAL METHOD OF TELEVISION PICTURE QUALITY EVALUATION (EBU-II)

Hoffmann H., Wood D.

EBU Technical Department, Switzerland

Itagaki T.

Brunel University, School of Engineering and Design, United Kingdom

НОВИЙ ПСИХОФІЗИЧНИЙ МЕТОД ОЦІНКИ ЯКОСТІ ТЕЛЕВІЗІЙНИХ ЗОБРАЖЕНЬ (EBU-II)

Гофман Г., Вуд Д.

Технічний департамент EBU, Швейцарія

Ітагакі Т.

Школа техніки і дизайну Брюнельського Університету, Велика Британія

A new method for subjective video quality assessments, particular suited to HDTV, is proposed, the Triple Stimulus Continuous Evaluation Scale (TSCES) method. The method allows the evaluation and ranking of the performance of different television formats, by rating the format under test with respect to two reference pictures: a defined lower quality anchor and a higher quality anchor.

Introduction: Measuring television picture quality is essential for the development and selection of video systems. The current methodologies for subjective assessment of the quality of television pictures are given in ITU-R Recommendation BT.500-11 [1] and in [2]. Current methods, however, have shortcomings: the scales have to be translated into the languages in which the tests are being done; there are variable intervals between the meanings of the descriptors in the scale in the same language, and variable intervals across different languages; furthermore the reference pictures are displayed on the same screen as the test condition. The task for the research was to define a new method following the requirements below:

- provides reproducible results, with a standard deviation determined by the spread of opinions
- independent of language adjectives describing the perceived image quality
- can cope with a wide range of picture qualities, SDTV and HDTV formats
- can be used to accurately measure a video system's quality and failure characteristics
- can be used with large and medium-sized flat panel displays and allows a direct comparison of HDTV and SDTV formats.

Method: Assessors are presented with three monitors one above the other, as shown in Figure 1. The identical displays are adjusted at an angle in such a way that a reference viewer at an eye-height of 1.2 m has always a constant eye-distance of three times picture height (3h) to all three displays. The 'design viewing distance' for HDTV is 3h, and this is used here. Having the monitors mounted above one another, the observers quickly graspwhat needs to be done, and the arrangements suit well wide screen displays. ITU-R BT.500-11 viewing conditions are recommended. We define the top display as an upper reference, providing a high-quality anchor and the bottom display as the low anchor with a defined impairment. The type of impairment used for the bottom anchor should be of a similar one to

those in the scenes under test, and which are shown on the middle display. All three displays show the same scene content at the same time.

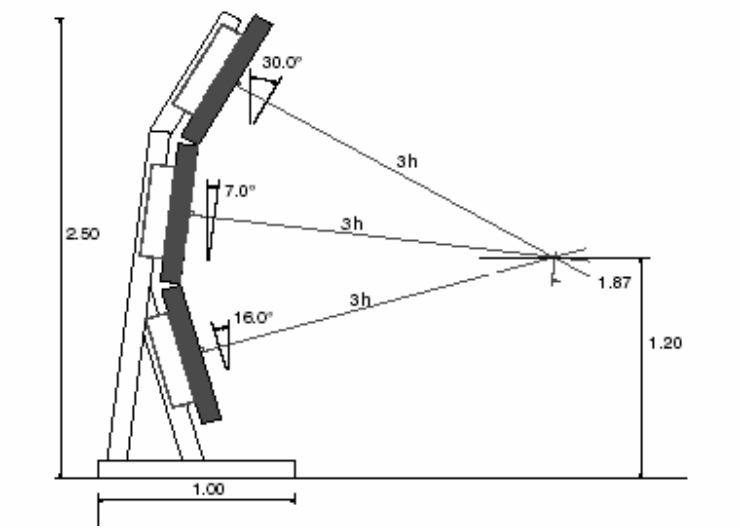

Figure 1 – Display mechanical setup for 50 in flat panel displays (developed by E. Wilson, EBU)

Presentation and voting: First a training session and explanation is given. Each test sequence has a length of 10 s and is repeated four times before voting. The middle display test sequences are shown in randomised order, and the test sequences include the upper and lower anchors to verify the consistency of the assessors.Assessors are given a continuous 100 mm line (Figure 2) to make the assessment. The top of the line represents the quality of the top monitor and the bottom of the line represents the quality of the bottom monitor. The assessors mark where the overall quality of the central monitor falls between the top and the bottom.

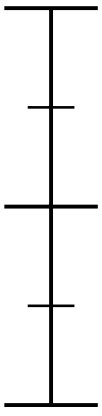

Figure 2 – Scale for voting on paper – original 100 mm length

Reporting of results: The viewing conditions are specified and the precise instructions and voting procedure are documented. The statistical analysis of the results including assessors' screening can follow ITU-R BT.500-11. The display conditions, upper anchor signal, the impaired signal, and how the lower anchor signal was generated in particular need to be clearly documented.

Results of series of tests: The method was tested with 178 student assessors and performed very well. HDTV formats and their failure characteristics were assessed according to the following conditions:

– Upper display high image quality anchor: uncompressed 1080p = 50 HDTV signal.

– Middle display images under test: 1080p50, 1080i = 25, 720p50 HDTV and 576i = 25 SDTV formats under test at various bit-rates compressed with MPEG-4 H.264 AVC Part 10.

– Bottom display low image quality anchor: 576i = 25 SDTV format down-converted from a 1080i = 25 HDTV source, compressed with the MPEG-4 AVC (i.e. the H.264 JM11 reference encoder available via http://iphome.hhi.de/suehring/tml/download/) reference encoder at 3 Mbit/s, providing a Grade 1 quality ('bad') impairment.

It was thus quite easy for assessors to evaluate where in quality space the test condition lies. The tests with the new method using non-expert assessors showed standard deviations of about 12 % (12 % equals about half a grade of the ITU-R BT.500-11 continuous quality scale) calculated over all votes.

Conclusions: The method has proven to be accurate and language independent, and the results are reproducible. The method should lead to easier agreement on video formats and systems. It can be used for many of the video studies being undertaken in the world today, and should contribute eventually to higher picture qualities for viewers throughout the world. Acknowledgments: E. Wilson from the EBU technical department for the engineering of the display rack; and A. Eichmueller, R. Schreiner, Dr. Schnoell and Prof. Hedkte supporting the tests at the University of Applied Sciences, Wiesbaden.

References

1 ITU-R BT.500-11: 'Methodology for the subjective assessment of the quality of television pictures', International Telecommunication Union, 2003, Geneva, Switzerland

2 Bernath, K., Kretz, F., and Wood, D.: 'The EBU method for organizing subjective tests of television picture quality', EBU Tech. Rev., 1981, 186, pp. 66–75.