NEW MULTIMEDIA QUALITY EVALUATION METHOD MSCQS

ABSTRACT

A new method MSCQS (Multi Stimuli using a Continuous Quality Scale) is proposed for the assessments of vision part of multimedia codecs or systems. Il is derived from the DSCQS method, but is efficient in the assessment of a large range of image quality as it provides reliable discrimination at both high and low quality levels. It makes use of both an explicit and a hidden reference and provides the observers with the adequate degree of liberty to permit stable ratings to be expressed. Practical experiments have shown both the stability of the test results and the suitability of the method to be applied to large range of multimedia image formats. In the context of wireless image transmission systems such as GPRS, UMTS, or STN and ADSL image transmission on PC screens, the MSCQS method was successfully tested the last 3 years. A work is in progress to propose the MSCQS method to standardization at EBU [1]and ITU-R [2].

1 INTRODUCTION

Today new audio-visual media are commonly used on Internet and Intranet Networks that offer more and more capabilities of communication exchanges. Among these new media, the audio video streaming is going to modify significantly individual and collective communication. It's important to be able to evaluate perceived quality of these audiovisual services. Now a large number of encoding systems that move continuously are in a hard competition. Most of them are not in accordance with MPEG4 standard. Except for television, there is no standardised method for evaluating image quality. Screens and viewing conditions are rather different than for the television world.

A point of great interest is to know whether the codec quality and performances claimed by manufacturers are actually met. The need is clearly identified for a standardized multimedia quality evaluation method in order to provide homogeneous results. The present document introduces the MSCQS method and gives examples of efficient application.

2 DIFFERENCES BETWEEN TV AND MULTIMEDIA DOMAINS

Main differences are summarized in table 1 of this paragraph. TV monitors are no longer adapted to display multimedia images. Then, it is necessary to play and evaluate quality of the multimedia images on a PC screen. Moreover, proprietary decoder-players are only PC compatible.

Unlike MPEG2 video sequences, the temporal update of data can vary in MPEG4 video sequences. Then in the multimedia field the fluidity perception axis is combined to the sharpness perception axis. The approach of quality evaluation becomes more complex.

Firstly in MPEG2 context, observers' judgement of images is based on only one axis of perception : the spatio-temporal sharpness of images. Concerning MPEG4 video sequences, the frequency of updated data is not constant in comparison to MPEG2 video. Thus, observers combine the sharpness axe perception to a second axis of perception: fluidity of images.

Secondly, the image format varies according to the types and the capabilities of networks. Moreover, the viewing distance is dependent on both image size (result of image and display formats) and each one's punctum proximum. Consequently, no viewing distance can be defined in the context of an individual multimedia display.

Thirdly, encoded multimedia images cannot be directly recorded on a video tape recorder, but only in data files.

A new specific approach is required to evaluate the quality of MPEG4 codecs.

	TV domain	Multimedia domain
Types of codecs	MPEG-2 codecs	MPEG-4 + proprietary codecs
Image format	fixed	variable
Update of image data	fixed	variable
Decoder type	standardized	various
Display	TV	PC, PDA,…
Viewing distance	standardized	undefined

Table 1. differences between TV and multimedia domains

3 MSCQS QUALITY EVALUATION METHOD

The MSCQS method (Multi Stimuli using a Continuous Quality Scale) that provides a global quality score for a short display duration (10s – 15s), is inspired by the standardize DSCQS (Double Stimulus using a Continuous Quality Scale : IUT-R BT.500) method. The commonly used methods were not designed to discriminate low qualities. The new method is able to discriminate low qualities as well as high qualities. With that in mind, it is necessary to combine quality evaluation capabilities and

the ability to discriminate near quality, using an implicit comparison process. The proposed solution is to use a random access to play sequence files. The individual observer can start or stop the evaluation, give, keep or change the current score when he wants. Additionally, he can replay sequences as much as he wants.
The MSCQS quality evaluation method developed in France Télécom R&D laboratories uses a continuous quality scale to provide a measurement of the intrinsic quality of video sequences. In fact each observer moves a slider on a continuous scale graded from 0 to 100 annotated by 5 quality items linearly arranged (Excellent, good, fair, poor, bad).
The quality evaluation is carried out scene after scene (cf. figure 1 of paragraph 3.3) including an explicit reference, a hidden reference and various algorithms.
To get a better understanding of the method, the following specific words are defined below:
Scene: audio-visual content
Sequence: Scene with combined processing or without processing
Algorithm: one or several image processing

3.1 Explicit, hidden references and algorithms

An evaluation method commonly includes quality anchors to stabilise results. Two high quality anchors are considered in the MSCQS method for the following reasons. Several tests were carried out that indicate minimised standard deviations of scores by using an explicit reference rather than a hidden or no reference. Particularly to evaluate codec performances, it is better to use an explicit reference to get the maximum reliability of results. A hidden reference is also added to evaluate intrinsic quality of the reference, instead of the explicit reference, because the presentation is anonymous as well as processed sequences. The explicit name "reference" have an influence on about 30% of observers. They give the highest possible score (100), which score is totally different from the corresponding score of the hidden reference. Exceptionally, when there is no available reference the test remains possible but the standard deviation is dramatically increased.
The MSCQS method is appropriated to multimedia context since it is possible to combine indifferently image processing such as codec type, image format, bit-rate, temporal updating, zooming, etc… A combination or a single processing is summarised by the name algorithm.

3.2 Test conditions

A 10 or 15 seconds maximum visualisation duration of a sequence is sufficient to get a stabilised and reliable quality score.
The images are separated from the proprietary skin to get an anonymous display in order to evaluate the reproduced quality performances of the players without being influenced by the knowledge of an environment.
The corresponding proprietary decoder-players have to be used to keep exactly the same viewing performances or by default a screen copy from the player must be carried out.
18 observers per a test are required to keep at least 15 of them after having applied rejection criteria.

3.3 Test organization

- The test is carried out scene after scene as it described in figure 1 of this paragraph.
- For the current scene, it is possible to play and score any sequence in any order. Each sequence can be played and scored again.
- From one scene to another, the sequence access is randomised and prevents the observers from attempting to vote in an identical way according to an established order. In fact, inside a test the algorithm order remains the same to simplify analysis and presentation of results. Only, the corresponding access from an identical button is randomised.
- For a first viewing, the current sequence must be totally played before being scored; otherwise it is possible to score and stop immediately.
- To test the next scene all sequences of the current scene must be scored.
- To finish the test all the sequences of all the scenes must be scored.

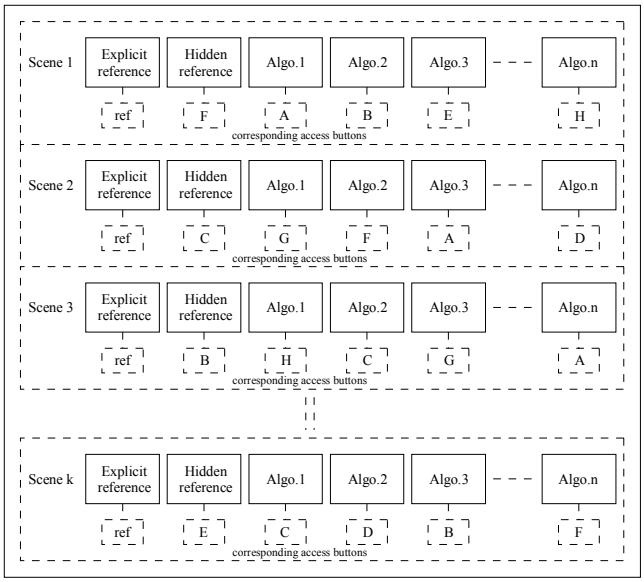

Figure 1. test organisation example

In conclusion, as for the other methods, a large quality range is required to stabilise observers'quality scores. In fact, when the quality range is reduced observers try to discriminate even if the difference of quality between sequences is not perceptible. Thus, the reliability of results decreases.

4 QUALITY OF EXPERIENCE

FTR&D has acquired a large experience in end user perceived quality in several domains such as digital TV and multimedia audio-video signals. FTR&D is involved in quality studies of GPRS, UMTS and ADSL services.

The last 3 years, a lot of tests have been carried out in the scope of these new services. The results provided by the MSCQS method are really stable and reliable whatever the observers, the codecs, the image formats and bit rates from 10 kps to 5000 kbps.

Three examples are listed below where the MSCQS method is an efficient way :

- Firstly, to evaluate and compare performances of multimedia codecs for a given image format. It can be also used to TV images by using a specific card that reads compatible TV format files and applies a real time transformation through a SDI output.
- Secondly, to optimise quality in order to determine the best encoding parameters, such as keyframe distance, buffer size and temporal update of image data.
- Thirdly, to determine what is the best compromise of format for a given bit-rate and a codec. It is often an example of service. In that test, the image format is variable.

5 EXAMPLES OF RESULTS

Image quality performances against bit-rate for two types of services are considered in figures 2 and 3: the first one is a streaming at a CIF format on a PC display and the second one is a streaming at a SmartPhone compatible format for the UMTS. Only, the average of quality scores are considered per codec. Taking account of the results for each service, a codec provides better results than the other ones and should be selected.

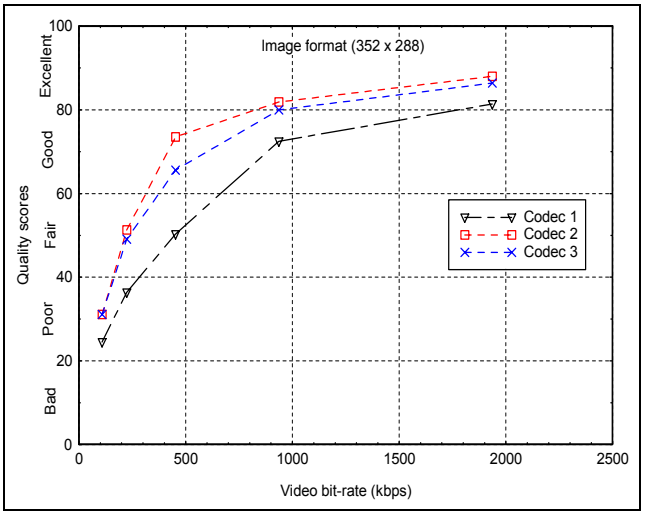

Figure 2 : Average quality scores over 5 scenes against bit-rate

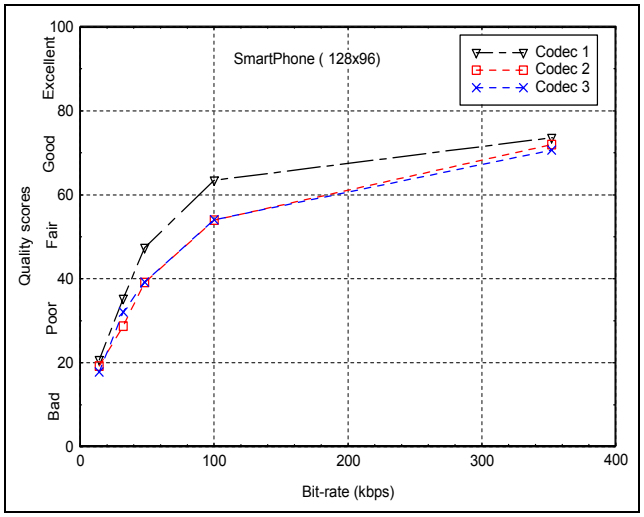

Figure 3 : Average quality scores over 6 scenes against bit-rate

6 REFERENCES

[1] EBU : European Broadcast Union, B/VIM (Video In Multimedia)group for subjective quality evaluation.

[2] ITU-R : International Telecommunication Union – Radiocommunication, BT.500 (Broadcast service Television), methodologies for subjective assessment of audio and video quality.