Subjective video quality assessment

The problems of subjective video quality assessment

By nature, perceived quality is a subjective notion. It is not something that can be exactly computed and, considering video quality, each observer uses its own internal scale to construct his judgement. That's why quality judgements vary from one person to another and therefore obtaining precise measures of the perceived quality (also called "perceptual quality") of a video is very complicated.

But complicated doesn't mean impossible and since quality is a subjective notion and since it varies a lot from one observer to another, the best thing to do to measure it is to realize subjective quality assessment tests. By asking several observers to judge the quality (on a scale, ranging from 1 to 5 or from 0 to 100 for example) and by computing the mean value of the collected votes, it is possible to know the mean opinion score (MOS) which is the average judgement of the observers.

During such tests, human observers are asked to judge the quality of video they are presented to them. To be efficient, these tests have to be conducted with many precautions. Indeed, the worst thing would be to realize such tests without any guideline since votes would be collected, a mean quality score could be computed but which confidence could we have in these scores? Would we be sure that the observers really answered the good question? If no explanations were given, did they really understand what they had to do? And if they voted when they wanted, wouldn't have they missed important distortions simply because they were not watching? And what if they had been too far from the screen to see them? Or maybe they were too close and therefore were too much sensitive to small artefacts? And what about the screen's calibration? Were the colours normally reproduced? And maybe the tests were too long and the observers got bored and voted mechanically? And what if some observers had answered randomly? OK, now you should begin to see that designing and realizing subjective tests is a very complicated task since you have to deal with humans, reproduce the same experience for all of them, minimize the factors that could disturbing them, detect the incoherent votes and measure the precision of the votes you collect.

So how can these tests be performed ?

Several ITU recommendations indicate how to realize subjective quality assessment tests. Beginners should start reading the ITU-R BT.500-11 recommendation.

These documents give recommendations about:

- The viewing distance;
- The room illumination;
- The normalized protocols;
- The tests duration;
- The observers recruitment;
- The methods to detect incoherent observers (to reject their votes);
- The methods to compute the precision of the mean opinion score.

Let's just say a word about the protocols, a protocol rules the way that videos are presented to the observers: When do they vote? How much time do they have to vote? Do they see the reference (not distorted) video before judging the distorted one? Can they replay the videos several times to precise their judgements? etc...

For video quality assessment, they are two famous protocols that are used:

- SAMVIQ : Subjective Assessment Methodology. for Video Quality;
- ACR : Absolute Category Rating.

The main difference between these two protocols is that, with SAMVIQ, the observers can replay the videos to precise their judgements whereas with ACR they see the video only once and have to judge its quality. SAMVIQ gives a more precise quality on each vote but requires a longer time to perform the tests than ACR. ACR enables to collect more subjective judgements in a limited time and therefore permits to have more votes to compute the mean opinion score (and the more votes, the better precision). In the end, these two protocols have quite similar performances in terms of precision.

All these precautions that have to be taken lead the subjective video quality assessment to be:

- complex;
- time consuming;
- risky (when someone designs a test campaign for the first time, it is frequent that the results can't be exploited because there were errors during the tests design);
- expensive (due to the need for a dedicated room and for human resources).

The main problem with these tests is they are very much time consuming. Often, the laboratories that realize subjective video quality assessment tests have to recruit between 25 and 40 observers (depending on the test complexity) to get an acceptable precision on the mean opinion score.

And many company (video equipment manufacturers) need to know the video quality quickly, for example to take the good decisions when improving an encoding algorithm or to validate a new product. Therefore they can't rely on subjective video quality assessment since this method is very slow (designing and realizing tests to measure the quality of about 30 to 50 videos takes between 10 days and a month).

Therefore, many people avoid the subjective video quality assessment tests and prefer using a video quality metric which is an algorithm able to compute video quality scores. These algorithms use display device modelling and visual perception models to compute quality scores. These video quality metrics are useful if they respect one condition: the quality scores they compute must be well correlated with human judgements.

For More Information: http://www.acceptv.com
info@acceptv.com
Tel : +33 (0)6 10 71 11 64