Progress of the Monitoring of Audio Visual Quality by Key Indicators (MOAVI) Project

Mikołaj Leszczuk, Silvio Borer, and Emmanuel Wyckens

The VQEG MOAVI project is an open collaborative for developing No-Reference models for monitoring audio-visual service quality. The goal is to develop a set of key indicators that describe service quality in general and to select subsets for each potential application. MOAVI models predict the presence or absence of these key indicators, not the overall quality.

The MOAVI project has accomplished the following tasks from inception through 2013.

- Implementation of 7 metrics for the following artifacts:
  - Blockiness – the probability of correct classification: 98.48%
  - Blur – the probability of correct classification: 80.52%
  - Exposure time distortion
  - Noise
  - Framing
  - Freeze
  - Blackout
- Initial values of thresholds for particular metrics were settled
- Development of metrics for audio artifacts (mute and clipping) in a MATLAB® environment
- Development of metrics for block loss and interlace artifacts in a MATLAB environment
- Preliminary tests of subjective opinion with the purpose of improving the approach to thresholds
- Design and construction of the website where the metrics are publicly available (vq.kt.agh.edu.pl)
- Writing a paper about the MOAVI project for the SIGCOMM conference in Hong-Kong and VPQM conference in Arizona
- SIGCOMM and VPQM conferences reviewers have provided some feedback comments that should be analyzed and taken into account for future steps of the MOAVI project. The most important weakness detected is the lack of any presentation of

actual results in the articles, although there is a set of metrics of artifacts ready.

- Therefore, a set of video and audio files has been created to test the metrics developed in previous months (Mute, Clipping and the Voice Activity Detector). These results of the metrics on those videos are ready to be compared with some ground truth determined by the researchers or eventually the results of subjective tests.
- In the case of the Voice Activity Detector in particular, its accuracy in detecting the voice activity in the audio clips extracted from the database has been measured by comparing the results obtained from the detector with the ground truth determined by both the observation of the waveforms and listening to the sound.
- The metric to detect Lip Activity in the videos has been enhanced during this month and the results of the temporal activity in the region of the mouth for the videos in the database have been stored for future analysis. The main goal of the latter is the establishment of a threshold for considering the video frame as "lip active" or not.
- A set of test videos has been created with the following characteristics:
  - Frontal view of talking faces.
  - Duration around 20 s.
- Real delay introduced to make the tests compared with the delay detected by the metric:
  - Average deviation = 130 ms.
  - The metric discriminates positive and negative delays.
- For the supercomputing cluster calculations we had to move the Temporal Activity and Spatial Activity metrics to C++, which we think may also contribute to the small progress in the MOAVI project.
- Also just creating all the databases with the results of the MOAVI project metrics required the use of the project applications, which can be considered as a solid test (for a total of more than 7500 videos).

Below, the results of key indicators verification tests are presented. For each metric the test consists of two parts: one is setting of the threshold of distortion visibility; the second is key indicators checking process. Before the test the results of subjective experiments are randomly split into two independent sets for each part of the test. These two sets are training set and the verification set respectively.

Setting metric threshold values

For each metric the procedure of determining the visibility threshold includes the following steps:

1. For all video sequences from the appropriate subjective experiment the value of the metric is calculated.
2. We assume each successive value of the metric as candidate thresholds th_{TEMP}. For values less than th_{TEMP} we set the key indicator to 0 and for values the same or above we set it to 1.
3. For each th_{TEMP} we calculate the accuracy rate of the resulting assignments. It is the fraction of key indicators which match with indications given by humans from the training set.

$$accuracy(th_{TEMP}) = \frac{number\ of\ matching\ results}{number\ of\ results} \quad (1)$$

4. We set the threshold of the metric to the candidate th_{TEMP} with the best (maximum) accuracy. In the case of several th_{TEMP} values with the same accuracy, we select the lowest value.

Figure 1 illustrates the procedure of determining the threshold for the blur key indicator. The threshold values are shown in Table 1.

***Mikołaj Leszczuk**, PhD, is an assistant professor at the Department of Telecommunications, AGH University of Science and Technology (AGH-UST), Krakow, Poland). His current activities are focused on e-Health, multimedia for security purposes, P2P, image/video processing (for general purposes as well as for medicine), the development of digital video libraries, particularly video summarization, indexing, compression and streaming subsystems.*

***Silvio Borer** received his Master degree in mathematics from the University of Zurich and the Ph.D. degree in science from EPF Lausanne, Switzerland, in 2004. He is currently with SwissQual AG, where he is a Senior Research Engineer. His research interests include video and audio quality estimation algorithms.*

***Emmanuel Wyckens,** research engineer for operational audiovisual services, graduated in Electronic Engineering at Valenciennes University (France), specializing in the design of audiovisual broadcasting solutions.*

In 2000 he joined Orange labs, being involved in the improvement of MPEG-2/4 codec video quality and the optimization of audiovisual chain tools. Mr Wyckens' current activities are in the field of subjective, objective audiovisual quality in multimedia applications, and for high definition television domains. He participated in the work of EBU/ITU projects for standardizing subjective methodologies.

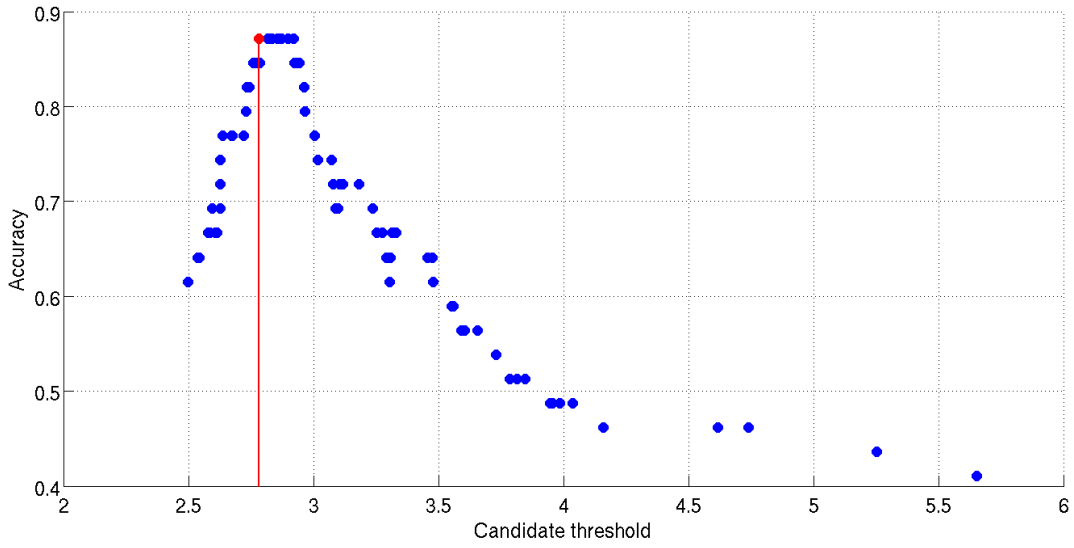

Figure 1. Blur metric threshold determination. Points represent the relation between candidate thresholds and accuracy. The line is drawn at the best candidate, which is chosen to be the metric threshold.

Key indicators verification

In the second part of the test, the correctness of the key indicator is checked. Accuracy of the indicator is calculated according to (1) and compared with indications from the verification set. Table 1 presents the verification results.

Table 1. Key Indicators verification – probability of distortion detection.

Metric	Probability of distortion detection	Value of threshold
Blur	0.86	2.78
Exposure Time Distortions	0.81	78 and 178
Noise	0.85	3.70
Block loss	0.84	5.3
Blockiness	0.94	0.85
Freezing	0.80	0
Slicing	0.85	7