An Integrated Object-Based System for Performing Audiovisual Subjective Assessments

Ulrich Reiter

Q2S - Centre for Quantifiable Quality of Service in Communication Systems, NTNU Trondheim, Norway, reiter@q2s.ntnu.no

Introduction

This paper presents an overview of an MPEG-4 based system created for performing bimodal subjective assessments in the auditory and visual domain. The assessment system has been created by the author during his stay at the Institute for Media Technology, Technische Universität Ilmenau, Germany, as there were no other tools readily available offering the functionality necessary for performing interactive audiovisual assessments in a fast and reliable way.

System Structure

The system described here consists of three main elements: 1) the I3D MPEG-4 player that serves as a rendering platform for the reproduction of interactive audiovisual content, 2) the Input Device that is used by the test subject to give feedback on the perceived quality, and 3) the logging tool SALT that allows to write a protocol file of all relevant events taking place during an evaluation session. Communication between the three elements is organized using a standard MIDI (Musical Instruments Digital Interface) protocol easy to implement in soft- and hardware. Fig. 1 shows the three elements and the corresponding communication structure. As the I3D is the central node, it is possible to add arbitrary data, e.g. a time stamp or information of the current state of the I3D, to every message coming from the Input Device, and save it to the log file.

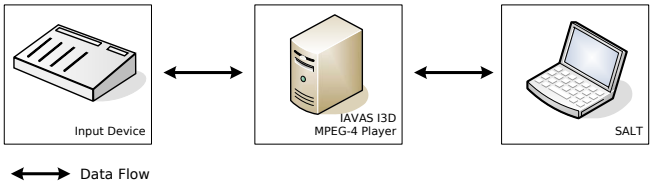

Figure 1: The three elements of the system and the necessary communication structure.

I3D MPEG-4 Player

The main parts of the I3D MPEG-4 player date back to the IAVAS project funded by the Thuringian Ministry of Science, Research and the Arts between 2002 and 2004 [1]. The I3D is based on the IM1 (Implementation Model One) reference implementation of the MPEG, the so-called core module [2]. Large parts of the code have been re-designed for better performance. It has been enhanced significantly to include more codecs as e.g. MP3, H.263, and H.264. Other formats can be integrated easily because the *libavcodec* part of *FFmpeg*, a collection of software libraries that can record, convert and stream digital audio and video in numerous formats, is included in the I3D [3]. In the field of audio, I3D provides a modular real-time rendering engine called TANGA which can be used to perform room acoustic simulations. The parameters describing the acoustic properties of the room are provided by the MPEG-4 scene description language (Advanced)AudioBIFS.

In areas in which the MPEG-4 standard does not provide sufficient means of external control, the I3D has been enhanced beyond the standard (e.g. the MIDI functionality implemented). The MIDI extension allows to modify the scene description via (nearly) arbitrary input devices in real-time. Also, the exchange of scene control and status data is possible. This for the first time allows to use an MPEG-4 player as a rendering device for assessments evaluating the perceived quality of interactive audiovisual scenes. This is especially interesting since the MPEG-4 paradigm of object-based scene description is widely believed to be one of the main innovations to the way how audiovisual content will be produced, distributed and consumed in the future.

Input Device

The traditional methods for performing audiovisual assessments are the so-called introspective methods that encompass perceptual and affective measurements. In order to determine the perceived quality, the test subjects need to reflect on the percepts and rate them according to their expectations or in relation to a reference stimulus. Rating is usually done by dragging a computer mouse to move virtual sliders on a computer screen or by using pen and paper, which is a process that inevitably cuts in the flow of visual percepts coming from the system under assessment. It is therefore necessary for the test subjects to "switch" between different perceptual situations (the one under assessment and the act of rating the assessed system itself). This is a process which diminishes the continuity of percepts and which therefore inevitably influences the subjects' perception.

This makes it necessary to find a method with which test subjects can deliver their ratings in such a way that the cuts in the flow of percepts are as small as possible. One important idea is to reduce the amount of distraction originating from the assessment system itself as much as possible by using a different, unused modality for the feedback channel. Whereas in the subjective quality assessments of audiovisual systems auditive and visual percepts are evaluated, haptics is a modality which is still "free" to be used in a different context. Therefore the

use of the haptic channel is proposed for receiving feedback from the test subjects: they have to manually move hardware sliders on the Input Device into a position that corresponds to their rating on a scale from 0 to 100. Of course, the scale is not necessarily a numbered one but can also consist of semantic designators, e.g. *bad* to *excellent*, as described in ITU-R BS.1284 [4].

The Input Device is based on a micro-controller running the open-source operating system MIOS [5]. Different front plates can be used to customize the Input Device according to the type of assessment to be performed (e.g. multi-stimulus, absolute category rating (ACR), etc.). When exchanging front plates, the underlying micro-controller hardware remains the same, see fig. 2.

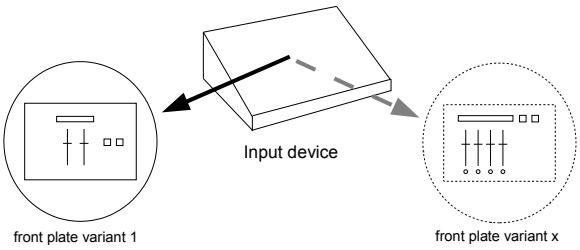

Figure 2: Front plates for the Input Device can be exchanged according to the type of assessment to be performed.

SALT Logging Tool

SALT is an acronym for *Subjective Assessment Logging Tool*. It is a cross-platform JAVA application which mainly serves three purposes: the recording of personalized test subject data (e.g. age, gender, ...), the complete logging of input provided by the test subject via the Input Device during a test session, and the automation of the Input Device's motorized sliders to start from a neutral slider position. Furthermore, SALT provides also functionality to export logged data into formats readable by statistical analysis software like SPSS or R.

SALT provides a graphical user interface and needs to be configured according to the current front plate of the Input Device. On the other hand, SALT must know the structure of the subjective assessment to organize the recorded data in a sensible way. Therefore, two main paradigms exist in SALT: the *design* and the *session*.

A *design* specifies the general settings regarding the Input Device (e.g. number of buttons and sliders present on the front plate) as well as the subjective assessment itself. The structure of a subjective assessment is often organized in so-called *trials* and *items*. An item is an instance of the attribute to be rated by the test subject. A trial groups the items according to the type of assessment.

Each test subject is related to a *session*. The session holds parameters that identify and describe a test subject (age, gender, acuity of vision, listening experience, ...). Also, additional information about the presentation order of items are recorded with the session. This is important because the presentation order is not necessarily defined before the assessment, but can also be created at random by incorporating an appropriate script into the MPEG-4 scene description itself.

Conclusions

The flexibility of the MPEG-4 scene description language allowed to create complex, self-contained test scenarios. These could be reproduced, (randomly) varying a number of parameters relevant to the perceived audiovisual quality, by including the corresponding instructions into the scene description itself. In principle, each and every parameter of the scene description could be changed during an assessment, if necessary.

The haptic Input Device as a means of collecting quality feedback from test subjects has proven to work efficiently and flawlessly. Test subjects were, according to their own estimation, able to concentrate fully on the audiovisual percepts presented in the assessments. Especially for sequences of single-stimulus items (ACR) this was observed. The usage of the Input Device was regarded as straight-forward and intuitive.

SALT has been used extensively to record the test subjects' ratings and all other types of events and data (high scores in interactive game-like assessments, presentation order of items and trials, subject related data, ...) during different audiovisual assessments. The fact that complete assessment setups can be saved and restored at a later time has considerably helped in performing assessments in a lab that was frequently used by other experimenters in between the test sessions.

The MIDI-based communication structure proved to be simple to employ and stable. The MIDI protocol is easy to implement and fast enough for the purpose.

Finally, the system is not limited to the field of audiovisual assessments, but can also be applied in more traditional unimodal auditory or visual quality tests.

References

[1] Drumm, H.; Kühhirt, U.; Rittermann, M.; Reiter, U.: *Application Systems for MPEG-4*, IEEE/ISCE'02 International Symposium on Consumer Electronics, Erfurt, Germany, September 23-26, 2002.

[2] Int. Std. (IS) ISO/IEC 14496-5:2001, *Information technology - Coding of audio-visual objects - Part 5: Reference software*, Geneva, Switzerland, 2001.

[3] Fast Forward MPEG - The FFmpeg Homepage, http://ffmpeg.mplayerhq.hu/

[4] Recommendation ITU-R BS.1284, *General methods for the subjective assessment of sound quality*, International Telecommunication Union, Geneva, Switzerland, 2003.

[5] Midibox Hardware Platform and MIOS Midi Operating System for Microchip PIC micro-controller devices, http://www.ucapps.de/