Sense of Presence Multimodal Dataset

Anne-Flore Perrin, He Xu, Eleni Kroupi, Martin Rerabek, Touradj Ebrahimi
Multimedia Signal Processing Group (MMSPG)
Ecole Polytechnique Federale de Lausanne (EPFL)
EPFL/STI/IEL/GR-EB, Station 11, CH-1015 Lausanne, Switzerland
{anne-flore.perrin, he.xu, eleni.kroupi, martin.rerabek, touradj.ebrahimi}@epfl.ch

ABSTRACT

Measuring the quality of a set of videos with various properties, as the content, quality, resolution and sound system, is not an easy task because of the subjectivity of the human perception. A way to express the quality of a video is to assess the Quality of experience (QoE), especially the Sense of Presence (SoP). The latter, not so explored, is the principle we would like to assess. To be independent of the human way to express the feelings and experiences, the measure is built thanks to biological signals (electroencephalography (EEG), electrocardiography (ECG) and respiration).
The project is to create a data set containing all the biological signals enabling the SoP study, process those data and demonstrate the data set is well constructed and functional for the assessment of the sense of presence for instance.

1. INTRODUCTION

[PARAGRAPH ON QoE] As digital television technologies aim to provide higher quality multimedia experience, possibly with various Quality of service (QoS), the Sense of Presence (SoP) should be investigated to understand its impact on the Quality of experience (QoE).

According to Sadowski & Stanney (2002), the SoP also called immersiveness level, whether physical or psychological in nature, allows the sense of belief that the user as left the real world and is now "present" in a virtual environment. The aim to create a database on the SoP is to study the impact of an high immersiveness experience on the QoE, especially during the visualization of a multimedia content.

Traditionally, the perceived quality of a multimedia content is measured thanks to subjective quality assessment, where perceived quality of selected visual stimuli is obtained from a number of subjects. The subjects have to explicitly rate the quality of each stimulus in a pre-defines rating scale. This procedure is replicated by directing the questionnaire on the immersivennes level experienced during a stimulus.

The expression way of each people is specific and for instance cultural and educational dependent. Thus the subjective ratings contains a subjective bias. A valuable survey of the SoP can not only rely on subjects subjective rates. Based on the Kroupi results on the assessment of 2D vs 3D quality, subjects'biological signals such as brain activity (electroencephalography (EEG)), heart activity (electrocardiography (ECG)) and respiration are objective data adequate to assess the SoP complementarily to the subjective rates.

This paper assesses the creation of a database representing the differences in users'experiences during low, middle and high immersivness level stimuli visualization through EEG, and peripheral physiological signals including ECG and respiration as well as subjective rates.

We conduct subjective experiments, in which various immersivness stimuli multimedia content were presented to users, and both explicit subjective ratings and implicit EEG and physiological response were recorded.
Then an investigation of the felt experience transcribed in the subjective ratings allows to associate some QoS properties to the SoP. Finally the construction of a subject-independent classification system distinguishing the various immersiveness levels of stimuli based on EEG and/or peripheral physiological signals.

The remainder of this paper is organized as follows. The next section describes how we conducted experiments to collect subjective ratings and physiological responses. Section 3 presents the results of subjective rating analysis and user-independent physiological classification. Finally, conclusion is given in Section 4.

2. DATA COLLECTION

2.1 Participants

The height females and the twelve males participants had from 18 to 30 years old (23 median and average years of age). The 20 subject were screened for correct visual acuity (no errors on 20/30 lines) and color vision using Snellen and Ishiara charts respectively. They all provided written consents forms. Before each experiment, oral instruction were provided to the participants to explain their tasks. Additionally, a training session was organized to allow participants to familiarize with the assessment procedure. The content shown in the training session was selected by experts viewers in order to include examples of all evaluated aspects.

2.2 Audio-visual stimuli

Video stimuli are coming from nine video sequences ex-

ACM MM 2015 , Brisbane, Australia

tracted from four open source movies published by the Blender Foundation (Big buck bunny, Elephant dream, Sintel and Tears of Steel)[1]. A supplementary sequence content was chosen for the training session.
The one minutes selected video contents have the highest audio, spatial end temporal energy and are related to the scene cuts. The twenty seven video stimuli shown during an experiment are the combination of the nine video content with the three levels of immersiveness described below.
Low, middle and high immersiveness level were defined respectively thanks to the audio sound system, the video quality\level of compression and the resolution. The table 1 illustrates the previous comment.

Immersive senario	Low	Middle	High
Audio	No Audio	Stereo	Surround
Quality (QP)	36	20	20
Resolution	SD	HD	UHD

Table 1: immersiveness levels

The video stimuli order could impact the data and so the results. Thus the sessions are built to allow this study : The first session will display the video stimuli from the the lowest immersiveness level stimuli to ones with the highest. The second session order is middle immersiveness level stimuli, followed by the low immersiveness level stimuli and then the high immersiveness level stimuli. The last session will display the video stimuli from the video with the highest immersiveness level to ones with the lowest. Besides these constraints, the video order is different for each volunteer, thanks to a pseudo-random function.

2.3 Monitor, sound system and environment

Professional high-performance 4K/QFHD LCD reference 56-inch monitor Sony Trimaster SRM-L560[2] was used to display video stimuli. As recommended in [TO SET comparsing upscaling algorithms from HD to UHD...] the viewing distance was set at 1.6H (H - Height of the screen). [SOUND SYSTEM]

2.4 Physiological signal acquisition

The nets used for the recording of the EEG signals contains 256 electrodes placed at the standard position on the scalp. An EGI's Geodesic EEG System (GES) 300 was used to record, amplify, and digitized the EEG signals while the participants were watching the stimuli. Additionally, two standard ECG leads were placed on the lower left ribcage and on the upper right clavicle, as well as two respiratory inductive plethysmography belts (thoracic and abdomen). All signals were recorded at 250 Hz.

2.5 Experimental protocol

The experiments were conducted in three sessions. A ten-minutes break was given between two sessions in order to avoid subject fatigue and lack of attention. Nine video sequences were presented in each session leading to a total of 27 video sequences, and thus, to a total of 27 trials. Each trial consisted of a ten-second baseline period and a stimulus period. The biosignals recorded during the baseline period were used to remove stimulus-unrelated variations from the signals obtained during the stimulus period.

During the baseline periods, the subjects were instructed to remain calm and focus on a 2D white cross on a black background presented on the screen in front of them. Once this baseline period was over, a video sequence was pseudo-randomly selected and presented. After the video sequence was over, the subjects were asked to provide their self-assessed ratings for the particular video sequence without any restriction in time, following the Absolute Category Rating (ACR) evaluation methodology [TO SET]

Regarding the self-assessed ratings, subjects were asked to evaluate the video sequences in terms of five different aspects, namely interest in the video content, perceived video quality, interest in audio content, immersiveness level and surrounding awareness. A 9-point rating scale was used that ranged from 1 to 9, with 1 representing the lowest value, and 9 the highest value of each aspect. In particular, the two extremes (1 and 9) correspond to "low" and "high" for interest in video and audio content as well as the perceived video quality, "no immersion" and "full immersion" for the immersion level and "no conscience of my environment" and "full conscience of my environment" for the surrounding awareness.

Once a trial was over, the next baseline period was recorded and the next video sequence was pseudo-randomly selected and presented. The procedure was repeated until all 27 video stimuli were presented and rated. Although the experiments lasted for almost two hours, including the training and the set up, the subjects did not report fatigue.

An illustration of a session is presented figure 1.

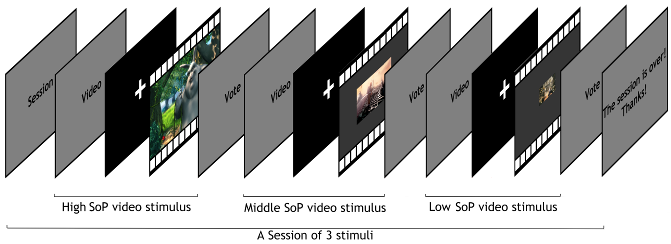

Figure 1: Example of a 3 video stimuli session progress

3. ANALYSIS

3.1 Subjective ratings analysis

3.2 Physiological signal analysis

3.2.1 Pre-processing

3.2.2 Correlation

4. CONCLUSION

[1] *http://media.xiph.org/*
[2] *http://pro.sony.com/bbsccms/assets/files/cat/mondisp/brochures/di0195_srm1560.pdf*