Signal Processing: *Image Communication* 42 (2016) 120–129

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

journal homepage: www.elsevier.com/locate/image

Using electroencephalography to analyze sleepiness due to low-quality audiovisual stimuli

Sebastian Arndt*, Jan-Niklas Antons, Robert Schleicher, Sebastian Möller

Quality and Usability Lab, Telekom Innovation Laboratories, TU Berlin

ARTICLE INFO

Article history:
Received 9 September 2014
Received in revised form
21 January 2016
Accepted 21 January 2016
Available online 12 February 2016

Keywords:
Quality of experience
Audiovisual
Physiology
Cognitive state
Alpha frequency band power

ABSTRACT

Standardized methods to assess the quality of experience (QoE) of multimedia focus typically on short sequences of approximately 10 s and a subjective judgment of the test participant. Two main problems occur when using this methodology: On the one hand these short sequences do not represent the typical media usage, and on the other hand it is still not completely understood how these subjective ratings are formed within the test participant. To overcome the second issue and to gain insight into the internal processing, electroencephalography (EEG) has been introduced successfully to the domain of QoE. As it is possible to use EEG to assess the quality perception continuously, we present two studies using standard documentary audiovisual clips with a length of 40 min and 60 min as stimuli. During the presentation of these quality-wise manipulated test sequences, we record an electroencephalogram and other physiological measures. We show that features of the EEG recordings indicate a change in the cognitive state of the test participant during the exposure to low-quality compared to high-quality sequences.

1. Introduction

The quality of multimedia services has been enhanced substantially in the past decade. At the beginning of the millennium standard definition (SD, resolution: 720×576) resolution was the standard for multimedia playback, today the standard is full HD (1920×1080), and soon to come 4k (3840×2160). An increase in resolution brings an increase in transmitted bit rate that is necessary to deliver the media to the user. Although the bit rate of broadband connections has been increased as well, there still is a mismatch, leading to the situation that bandwidth is a restricting variable. Thus, it is important for service providers to find the optimal tradeoff between transmitted bit rate and perceived quality at the customer side, in order to keep costs low and still deliver a satisfactory quality service. Especially, for popular live reporting events not the best possible video quality can be transmitted, as the required hardware setup would be too complex. In order to determine perceived quality, subjective tests have to be conducted. How these tests should be carried out is specified in several recommendations released by the International Telecommunication Unit (ITU).

For most subjective studies in the area of QoE, short sequences with a length of around 10 s are used as stimulus material, as recommended e.g. by ITU-T P.910 [1]. These rather short stimuli are not representing the typical use of media, therefore other recommendations have been released with a scope of longer videos, with up to 30 min. These are specified e.g. in ITU-R BT.500 [2]. Here, participants are asked to give a continuous quality judgment with the help of a slider with which they continuously can give a rating of the experienced quality, and therefore can react to spontaneous changes in quality. However, this setup has the disadvantage that test participants have to avert their gaze from time to time from the screen to the slider, as they need to check the adjustment of the slider.

* Corresponding author.
E-mail address: sebastian.arndt@mailbox.alumni.tu-berlin.de (S. Arndt).

http://dx.doi.org/10.1016/j.image.2016.01.011
0923-5965/

Thus, the flow experience might suffer. In order to overcome these problems, the use of different modified input devices has been investigated. Here, a steering wheel, joystick [3], or a glove [4] have been proposed. Still, the advantageous use could not been shown compared to the slider method. A different approach has been proposed by Borowiak et al. [5], which aims not at a direct quality evaluation, but the momentary quality can be adjusted by the test participant. This is done as the participant turns a knob until he or she perceives a perfect quality of the shown material.

However, using any of the above mentioned methods does not bring any information about underlying cognitive processes, and how these are affecting the mental state of the media consumer. In order to get some insights in what happens inside a consumer of multimedia content, physiological measures can be helpful.

Physiological measures, such as electroencephalography (EEG), have shown to bring additional results to subjective MOS (mean opinion score) values. This could be shown in studies using rather short stimuli, with a length of words or syllables. Using audio [6], visual [7], and audiovisual [8] stimuli, it was shown that the stronger the degradation is, the greater the cognitive response. Thus, a higher P300 peak amplitude was present when exposing test participants to low quality sequences compared to better quality. The P300, a positive deflection about 300 ms after stimulus onset, is part of the event-related potential (ERP) which is elicited when a participant is exposed to a stimulus. Commonly, the latency and amplitude of the P300 depends on several factors, such as e.g. the intensity or irregularity of a stimulus [9].

When analyzing long-term changes (i.e. several minutes) in the brain, an analysis of frequency bands is often performed. The band power can be divided into different sub-bands, delta (0.5–3.5 Hz), theta (3.5–7.5 Hz), alpha (7.5–12.5 Hz), beta (12.5–30 Hz), and gamma (30–60 Hz) [10]. The alpha band is most interesting when it comes to draw conclusions concerning cognitive states, such as fatigue,[1] as it is shown in [12] in the context of driver's fatigue, and in [13] for effects of mental fatigue. However, research has shown that identifying individual alpha bands per subject may be the more accurate way to analyze this [14], due to individual difference of the human brain.

First studies in the context of QoE using frequency band power analysis could already give some insights into the perception process during the presentation of multimedia content. In two studies, an audiobook was played to the test participants [15,16]. It could be shown that when confronting participants with a low-quality version of the audiobook, they tended to be more fatigued compared to the high-quality version. Alpha and theta portions were larger for low-quality conditions compared to high-quality conditions.

[1] The terms sleepiness and fatigue have been used interchangeably in the literature. In the literature, sleepiness is defined more detailed and is referred to changes in the neurobiological processes, whereas the term fatigue is indicating to not being able to sustain on a certain level of task performance [11]. Therefore, we will use the two terms, *sleepiness* and *fatigue*, interchangeable within this paper.

Using physiological measures assumptions about the cognitive state and its parameters, such as e.g. work load and fatigue, can be drawn.

Work load is defined as the needed processing resources to fulfill a task [17]. To perform a task a certain amount of resources is demanded and these resources are limited. The amount of available resources can be influenced by numerous factors e.g. stable influences such as individual differences between humans or time of day effects [18] or fluctuating influences such as time on task effects or monetarily cognitive/physical condition. The cognitive state describes the level of free resources which a user of a system has. In an increased/high cognitive state almost all of the available resources are used and in a decreased/low cognitive state almost all resources are available. A sustained period or high workload can lead to a lower cognitive state as fewer resources are available and lead to being more sleepy; users have less available resources for task processing available. This goes along with an increase in alpha band power [19,20]. Fatigue refers to the feeling of reduced alertness that impairs the capabilities and willingness to perform a task [21].

In this paper, we will present two studies using audiovisual material with quality degradations in both modalities. We recorded an EEG and analyzed the frequency band power concerning portions of alpha and theta band power, thus indicators of fatigue and reduced attention.

The remainder of this paper is organized as follows: In Section 2, we report on a first study using a stimulus impaired only in the visual modality, and in Section 3 on a second study where the material was degraded in both modalities. In Section 4, we discuss the two studies, before we conclude with the consequences for physiological recordings applied in QoE in Section 5.

2. Study 1

Motivated by [15], a study has been carried out using a commercially available documentary about sea life which consists in the audio track only of content related noise, such as e.g. movement of the waves, and almost no speech. As we have learned from [8], superimposition effects can appear when analyzing both modalities at the same time for an initial study, thus we decided to start with a simple stimulus setup. Initial results of this study were already reported in [22].

2.1. Stimulus

We took the documentary 'Ocean' as a stimulus. This video shows mainly sea life scenes, and is easy to follow. It was taken from a BluRay with full HD resolution (1920×1080), and the original video bit rate of 40,000 kB/s. The video shown was a 40 min excerpt. The cutting did not affect the story of the video. For the test, we had two conditions, (a) the original video without any quality reduction (named HQ from here on), and (b) a video with reduced bit rate (named LQ). For quality reduction, we used the H.264 codec, in the $\times 264$ implementation, and had a reduced bit rate of 2000 kB/s (constant bit rate settings).

The video was then divided into two parts, of which one part was shown in HQ and the other in LQ. Whether