DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-128

Towards prediction of Sense of Presence in immersive audiovisual communications

Anne-Flore Perrin, Martin Řeřábek and Touradj Ebrahimi
Multimedia Signal Processing Group (MMSPG)
École Polytechnique Fédérale de Lausanne (EPFL)
Lausanne, Switzerland

Abstract

Tremendous progress has been made in audiovisual communication technologies in the last decades offering a large variety of multimedia experiences. Consequently, Quality of Experience (QoE), which brings a user-centric assessment of the multimedia experience has become popular. Assessment of QoE is challenging because it depends not only on the content and users but also on the context in which the latter consume the former. QoE has become essential in order to allow creators, engineers, designers as well as product and services developers to offer increasingly richer multimedia experiences to users. This paper focuses on QoE in immersive multimedia communications. More specifically, the Sense of Presence (SoP) is explored as an important factor influencing the QoE. To reach this objective, a series of experiments have been conducted in typical situations, where users consume audiovisual content in various contexts, defined by the type of the devices used. Such experiments consisted in presenting one-minute video stimuli to twenty subjects, on three different devices: iPhone, iPad, and Ultra High Definition (UHD)TV. Subjective evaluation scores were recorded together with physiological signals of users. More particulary, Electroencephalography (EEG), Electrocardiography (ECG), and respiration signals were acquired during consumption of audiovisual stimuli tailored to each device. Furthermore, a publicly available multimodal dataset containing all acquired physiological signals together with corresponding subjective ratings was created. The resulted dataset can help in design, implementation and validation of metrics to predict SoP experienced by users in typical use cases.

Introduction

Audiovisual communication technologies have made tremendous progress in the last few decades offering a large variety of multimedia experiences. As an example, television broadcast has evolved from standard resolution at 25 interlaced frames per second, in black and white (gray scaled), to colour, and more recently higher resolutions up to 8K at frame rates above 120 in progressive sampling. Capture and rendition of High Dynamic Range (HDR) content with Wide Colour Gamut (WCG) have become reality thanks to cameras with advanced sensors and optics, and monitors featuring brightness levels up to 10K nits, with increased frame rates by several fold, when compared to the past. Attempts have been made to capture and display three dimensional (stereoscopic as well as multi view) content with some success, and omnidirectional and immersive video technologies seem to be around the corner, as witnessed by recent prototypes and early products in form of Head Mounted Display (HMD). Similar progress has been made in audio technologies, where mono has been replaced by stereo, surround, and 3D sound in the last decades. Furthermore, the modes of consumption of audiovisual contents are changing with a wide range of services offering new experiences. Moreover, mobile and immersive audiovisual communications are enabled due to powerful and feature rich devices and infrastructures as well as massive innovations in how information is stored and delivered to consumers and professionals. In parallel with this progress, attempts have been made to assess such new experiences in a reliable, accurate, and reproducible manner, not only to quantify them better, but also to optimize the entire processing chain of media from its creation to its consumption, for a targeted level of experience.

This paper positions itself in this latter category and aims at defining assessment methodologies that can measure an important dimension in emerging immersive media, namely, to implicitly evaluate the SoP experienced by subjects consuming immersive audiovisual content in different typical situations. This is achieved by building upon a past similar attempt by the authors of this paper [12], where the SoP in a variety of audiovisual experiences was measured as a function of the modality and the quality of audiovisual content, from lower quality standard definition video with no audio, to higher quality ultra high definition video with surround audio. The SoP was evaluated both explicitly, by means of a questionnaire, as well as implicitly, by recording physiological signals, namely, EEG, ECG, and respiration. This work showed that there is a correlation between SoP as experienced by human subjects and their physiological signals.

In this paper, the previous attempt mentioned above is extended, while still relying on the same fundamental approach of finding the correlation between SoP as experienced and explicitly assessed by subjects, and implicit prediction of the latter by analyzing physiological signals recorded from the same subjects during the same experiences. The type of immersive audiovisual experiences under consideration was changed to include those of typical situations, such as when watching a movie on a mobile phone or a tablet, or watching it on a UHDTV monitor. This change was motivated by the desire of assessing QoE and SoP in an as realistic as possible context of consumption of multimedia content. In addition to the analysis of the conducted experiment based on the explicit ratings, this paper offers a publicly available dataset containing physiological signals and subjective scores acquired within an assessment of SoP in immersive com-

DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-128

(a) C1 (b) C2 (c) C3 (d) C4 (e) C5

(f) C6 (g) C7 (h) C8 (i) C9 (j) C10

Figure 1: Typical frame of each test sequence used in experiments. Sequences C1 - C9 ((a)-(i)) are used for testing and sequence C10 (j) is used for training.

munication[1].

Conducted experiments were based on the assumption that each device will induce a different SoP, referred to as Immersiveness Levels (ILs) in the remaining of this paper. Analysis of subjective ratings obtained after conducting experiments have confirmed that three distinct levels of immersiveness were observed by subjects. Nevertheless, the differences between the ILs induced by each device largely depend on the content.

The reminder of the paper is structured as follows. In the next section, we derive the test from related work. Then, we describe design and implementation of the conducted experiments. The following section presents the results of the subjective tests and their analysis. In the last section, a conclusion is provided.

Related work

Multimedia providers tend to use QoE rather than Quality of Service (QoS) in order to improve their services to be more user-centric. An overview of the strong relationship between SoP and QoE is given in [3]. This paper focuses on assessment of SoP as an important factor influencing QoE.

Subjective assessment of QoE is influenced by emotional, cultural, educational, and environmental differences across subjects [2, 4]. In order to be as independent as possible from the bias induced by different subjects, as well as to build the right contextual information for prediction of QoE, researchers increasingly consider to record and use physiological signals of subjects. Such signals can be used to evaluate emotional state of subjects as well as to implicitly assess the QoE without a need for its explicit expression given by a subject.

For instance, in [10], the frustration created by a human-computer interface was evaluated via ECG, skin conductance, blood pressure and respiration. The authors of [1] investigated the emotional states of users interacting with mobile phone applications in studying the frontal alpha asymmetry of EEG signals. The experience provided by tone-mapped HDR content was evaluated both explicitly (questionnaire) and implicitly (EEG, skin conductance, blood pressure, respiration and skin temperature) in [11]. The authors, also provide an overview of the physiological studies involving EEG and peripheral signals for various media type such as speech, video, 3D, and sense of reality.

All previously mentioned work assess technology-centric scenarios where the conducted experiments assess various technologies, such as HDR, 3D, and human machine interfaces. In this paper we propose to study the SoP during consumption of audiovisual content on three typical devices and situations. Hence, design of the experiment considers typical multimedia end-user consumptions. At the end of each media consumption, explicit ratings are collected by means of a questionnaire. Additionally, physiological signals (EEG, ECG and respiration) are recorded during the media consumption.

Experiment

In this section, the video stimuli creation process is first described in details. Then, the test methodology as well as test equipment and environment, together with physiological signals acquisition are described. Finally, additional information about participants in experiments is provided.

Test stimuli

A test dataset was created as a collection of one-minute test sequences extracted from four Blender open source movies[2]. The uncompressed original movies were in YUV 4:2:0 8-bit format at 24 fps. The original movies were in either HD or UHD. Audio signals were available in stereo and 5.1 surround and represented in FLAC format.

A total of ten test sequences were produced for the dataset. Nine test sequences, C1-C9, were used in actual assessments. The last test sequence, C10, was used for training purposes. Test sequences were selected based on a careful analysis of the entire original movies. More specifically, the energy level in surround audio channels, as well as temporal and spatial properties of luminance component (TI and SI [9]) of each original movie were considered. Selected test sequences generally correspond to the highest values of the above mentioned properties (TI, SI, and audio energy), whereas the scene cuts of the original movie were also taken into account. In particular, the selection was made to prevent abrupt changes at the beginning and at the end of each test sequence. An example frame of each test sequence used in experiments is illustrated in Figure 1. Additional information for each test sequence including its original movie, spatial resolution, and

[1] *http://mmspg.epfl.ch/SoPMD*

[2] *http://media.xiph.org/*

DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-128

Sequence	Original movie	Original resolution	Start frame
C1	Big Buck Bunny	1920x1080	5388
C2	Big Buck Bunny	1920x1080	9379
C3	Elephant Dream	1920x1080	2160
C4	Elephant Dream	1920x1080	6516
C5	Sintel	4096x1714	6900
C6	Sintel	4096x1714	12607
C7	Tears of Steal	4096x1744	4152
C8	Tears of Steal	4096x1744	10392
C9	Elephant Dream	1920x1080	11280
C10	Sintel	4096x1714	360

Table 1: Original movie, original resolution and start frames of test sequences

the position of the first frame within its original movie is reported in Table 1

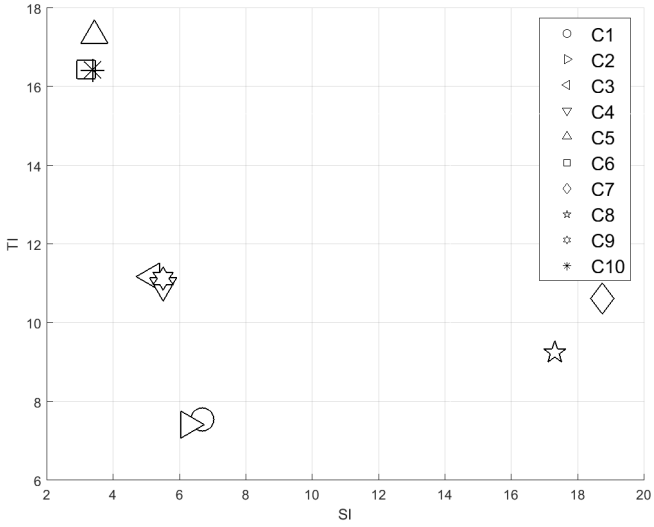


Figure 2: Values of SI and TI of test sequences

An analysis of the TI, SI, and audio energy was subsequently conducted on selected test sequences to better understand their properties, namely, spatial complexity, amount of motion, and impact of the audio rear signals. The distribution of test sequences along their SI and TI values are presented in Figure 2. It is observed that test sequences extracted from *Sintel* movie have large TI and small SI values. Thus the C5, C6 and C10 test sequences contain a lot of motion and a low level of details. The test sequences originating from *Elephant Dream* and *Big Buck Bunny* movies have small TI and SI values. This means that there is few motion and a low level of details in C1, C2, C3, C4, and C9 test sequences. The test sequences extracted from *Tears of Steal* movie present high SI and fair TI values. Hence, C7 and C8 test sequences contain few motion with detailed spatial information. This can be explained by the fact that the three first movies are computer-generated, whereas the last movie contains a real world environment with additional computer-generated information. The results of audio channels analysis are presented in Table 2. Values E_L and E_R express the level of audio energy in percentage as a ratio between audio energy of left and right channels (both front and rear) and total audio energy (sum of all channels). Moreover, values E_{SL} and E_{SR} represent the ratio between

Sequence	E_L[%]	E_R[%]	E_{SL}[%]	E_{SR}[%]
C1	38.12	48.16	1.06	0.91
C2	34.87	34.88	4.97	4.19
C3	45.51	42.48	15.29	15.06
C4	40.21	39.45	4.42	4.38
C5	37.84	38.11	23.01	24.17
C6	28.15	26.89	11.67	14.78
C7	47.36	44.99	12.94	13.49
C8	45.87	51.49	20.52	19.79
C9	36.35	32.61	6.22	6.35
C10	29.92	30.56	26.11	23.41

Table 2: Test sequences audio ratios in percentage: E_L and E_R are ratios between audio energy of both front and rear channels and total audio energy. E_{SL} and E_{SR} are ratios between rear and side channels of each side

the level of audio energy of rear channel and side channels (both front and rear) in percentage of each side. General conclusion of what we can see from these numbers is that, in average, a relative balance is observed between left and right audio signals. The test sequences having the highest amount of information in surround channels are C8, C5, and C10. The test sequences presenting the lowest amount of information in rear left and right channels are C1, C2, and C4.

Three different devices were used in this study. Details describing the properties of each device are given further in this paper. The audiovisual stimuli were generated from the test sequences to match rendering properties of each device. All audiovisual stimuli were compressed with an AVC/H.264 encoder, to achieve the best quality in terms of their corresponding device rendering capabilities. The audiovisual stimuli generated from C3 test sequence were encoded with a QP of 25, whereas the remaining audiovisual stimuli were encoded with a QP of 20. Expert viewing sessions were conducted to confirm the transparent quality of all audiovisual stimuli. To recapitulate, 27 audiovisual test stimuli were generated from 9 test sequences, C1-C9, and 3 audiovisual training stimuli from C10 training sequence.

Experimental protocol

The experiment consisted of three sessions. During each session, 9 stimuli were visualised on the same device. A minimum of one day separated every two sessions to avoid any statistical bias between the evaluation of two different devices and fatigue of subjects. Each session lasted approximatively one hour, including the training phase and set up of physiological signals acquisition devices. A training session was inserted at the beginning of each test session to recapitulate the test procedure and to remind subjects of each IL. After the training, subjects were asked to calm down in order to record baseline signals in the following order,

- eyes closed during 5-7 seconds,
- eyes open for 5-7 seconds,
- looking up/down/left/right and moving eyes back to the center, 5 times each,
- blinking, 5 times,
- gritting their teeth (clamping jaws shut), 5 times,
- making a snoring sound, twice,
- if possible, moving one's ears, 3 times.

DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-128

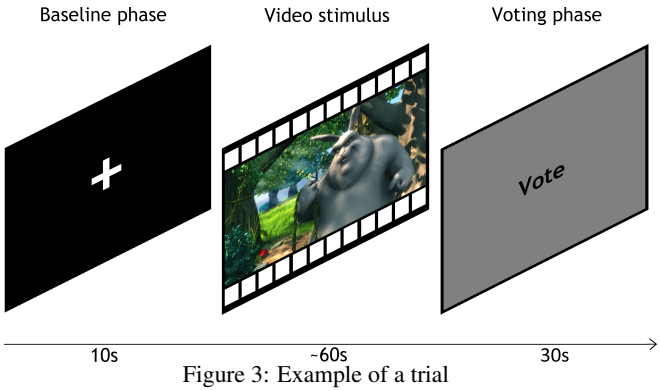


Figure 3: Example of a trial

Nine audiovisual stimuli were presented in each session leading to a total of 27 audiovisual stimuli forming 27 trials per subject. Each trial consisted of a ten-second baseline phase, a stimulus period and voting phase. During the baseline phase, subjects were instructed to remain calm and focus on a white cross on a black background presented on the screen in front of them. The physiological signals recorded during the baseline period were used to remove stimulus-unrelated variations from the signals acquired during the stimulus period. Once the baseline period was over, an audiovisual stimulus was presented. After the audiovisual stimulus was over, subjects were asked to provide their corresponding ratings in 30 seconds (about 5 seconds per question). The next trial followed after the voting phase until the completion of the entire session. Figure 3 illustrates an example of one test trial including baseline, audiovisual stimuli, and voting. The order of trials were randomised in every session for each subject.

Regarding the ratings, subjects were asked to evaluate the audiovisual stimuli according to six different criteria, namely, Interest in Audio content (IA), perceived Audio Quality (AQ), Interest in Video content (IV), IL, perceived Overall Quality (OQ), and Awareness of their Surrounding (SA). A 9-point rating scale, following the Absolute Category Rating (ACR) evaluation methodology [9], was used ranging from 1 to 9, with 1 representing the lowest and 9 the highest value of each criteria. In particular, the two extremes (1 and 9) correspond to "low" and "high" for IV and IA as well as the OQ, "no immersion" and "full immersion" for IL, and "not aware" and "fully aware" for SA. The 9-point rating scale was presented with clear separation lines between low (1-3), middle (4-6), and high (7-9) ratings creating three classes of ratings. Subjects were instructed to evaluate the stimuli in the 3-classes ratings (low, middle, and high) and then further refine their assessment.

Test equipment and environment

Low, middle, and high ILs were expected to be induced to the subjects by the devices used for rendering of the audiovisual stimuli. The iPhone5[3] and iPad4[4] were used to render audiovisual stimuli corresponding to low and middle ILs, respectively. The professional high-performance 4K/QFHD LCD reference 56-inch monitor Sony Trimaster SRM-L560 [5] was used to render high IL stimuli. The Table 3 illustrates characteristics of the three devices. As recommended in [6], the viewing distance of the UHDTV was set to 1.6H. To prevent additional noise in recorded EEG signals and to reduce influence caused by movement of portable devices, subjects did not hold iPhone nor iPad. Instead, they were fixed in front of them at a viewing distance of 6H (30cm) and 4H (60cm) respectively. H corresponds to the height of the display area of each device. The estimated foveal area of the video stimuli on each device was therefore around 15%, 10% and 3% of the total video stimuli, on iPhone, iPad, and UHDTV, respectively.

Parameters Setting	IL		
	Low	Middle	High
Device	iPhone5	iPad4	UHDTV
Audio	Stereo	Stereo	5.1 Surround
Native Device Resolution	1136x640	2048x1536	3840x2160
Video Resolution	1280x720	1920x1080	3840x2160
Foveal area[px]	171	202	121
Foveal area[%]	15	10	3
Viewing distance	6H	4H	1.6H
Viewing distance[cm]	30	60	110

Table 3: Immersiveness Level (IL) settings based on devices characteristics

Two sound systems were used during tests. Stereo audio signals were used for iPad and iPhone, and 5.1 surround audio signals were used for UHDTV. The stereo sound was provided by a professional headset Sennheiser HD 280 Pro[6] (accurate for linear sound reproduction in critical monitoring applications and attenuating up to 32 dB of ambient noise). The Altec Lansing 5.1 THX speaker system super subwoofer was used as 5.1 surround sound system.

The laboratory setup provided a quiet environment and the ambient light was set in order to ensure subjects comfort during bright and dark scenes, as well as during voting and resting periods. The luminosity of the iPhone and iPad was set at 75% of their maximum brightness.

Physiological signal acquisition

To record brain activity, a 256-electrodes net was placed at the standard positions on the scalp. An EGI's Geodesic EEG System (GES) 300 was used to record, amplify, and digitize the EEG signals while the participants were watching the stimuli. The heart activity was recorded from two standard ECG electrodes placed on the lower right rib cage and the upper left clavicle. Two respiratory inductive plethysmography belts (thoracic and abdomen) were used to acquire the respiration. All signals were recorded at 250 Hz.

Participants

Twenty subjects participated in this study (ten females and ten males). They were from 18 to 25 years old (21 in average with a 2.17 standard deviation in number of years). Subjects were screened for correct visual acuity (no errors on 20/30 lines) and color vision using Snellen and Ishiara charts respectively[5, 13].

A prior informative text containing the detailed description of the subject assessment was provided to subjects. Moreover,

[3] *https://support.apple.com/kb/SP655?*
[4] *https://support.apple.com/kb/SP662?*
[5] *http://pro.sony.com/bbsccms/assets/files/cat/mondisp/brochures/di0195_srm1560.pdf*
[6] *http://en-us.sennheiser.com/professional-dj-headphones-noise-cancelling-hd-280-pro*

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-128

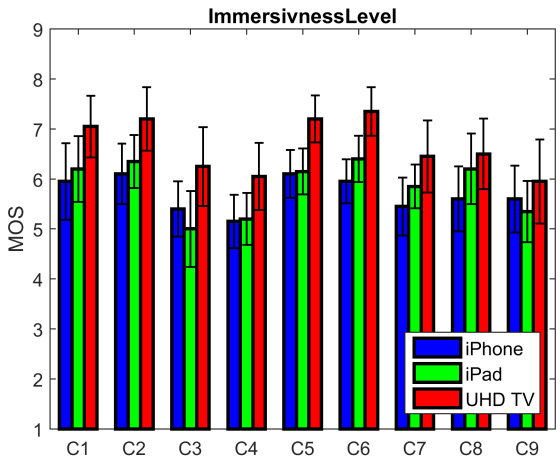


Figure 4: MOSs and CIs for experienced ILs per test sequence

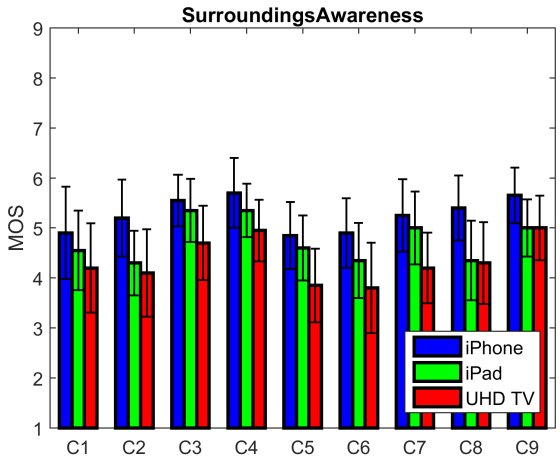


Figure 5: MOSs and CIs for SAs per test sequence

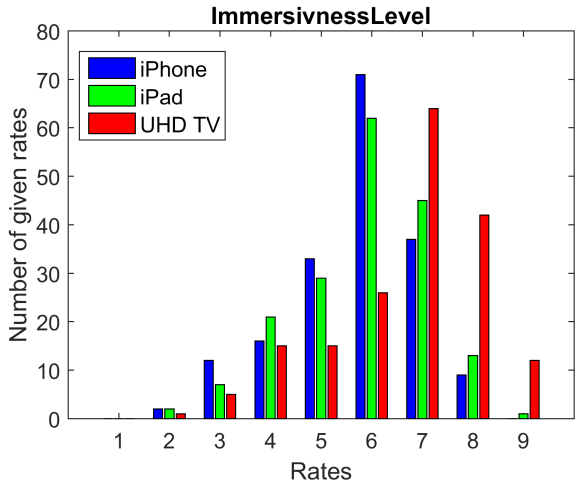


Figure 6: Rate distribution histograms regarding Immersiveness Level (IL) for each device

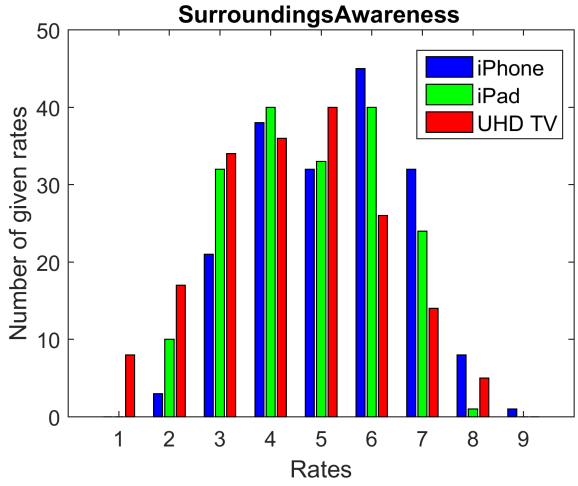


Figure 7: Rate distribution histograms regarding Awareness of their Surrounding (SA) for each device

oral instructions were given to participants before they signed a consent form. Additionally, a training session was organized to illustrate low, middle, and high ILs in order to guide subjects to bound their own perceived overall ratings. The training also allowed subjects to get familiar with the assessment procedure.

Analysis of subjective ratings

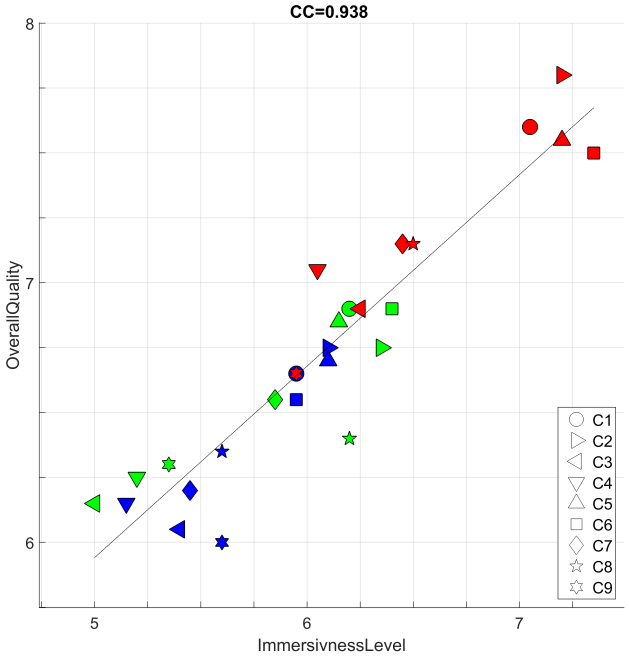
This section describes the analysis carried out on the subjective ratings to investigate how IL is perceived in an explicit way, as well as to explore how various evaluation criteria are correlated. The conducted analysis on subjective ratings assumed a Student's t-distribution of the subjective rates. Outliers detection, resulting distribution histograms, MOSs and associated 95% CIs, as well as Pearson's correlations are presented in the remaining of this section.

A detection and elimination of outliers was performed according to the guidelines described in Section 2.3.1 of Annex 2 of [7]. Based on the scale of the ILs ratings, no outliers were detected. Thus, the non-significant deviation across subjects ratings is ensured.

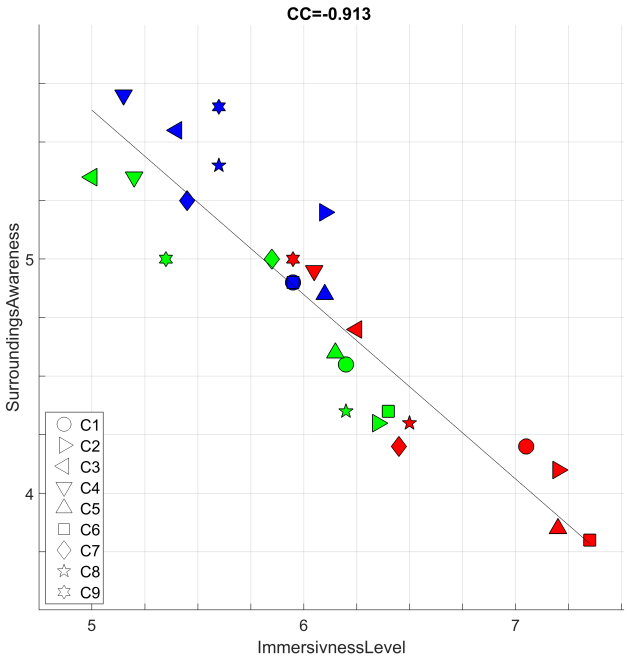
The ratings analysis comprises MOSs and its associated 95% CIs as recommended in [8]. Figures 4 and 5 depict MOS and CI values for each content and device, corresponding to all ILs or SAs, respectively.

Figure 4 confirms that each device brings different ILs to subjects. More particularly, iPhone induces the lowest IL, whereas UHDTV induces the highest IL. However, differences between ratings remain small as respective CIs overlap. Moreover, the MOS values are condensed in a range from 5 to 7.5, which means that the full range of immersive experiences is not covered by different stimuli and different devices. This could be explained by the fact that the audiovisual stimuli rendered on the used devices in the laboratory environment did not induce neither very low IL nor very high IL on subjects. Further, it can be observed that for $C3$ and $C9$ test sequences the ILs experienced by subjects are in average lower for iPad when compared to iPhone. $C3$ and $C9$ test sequences can be compared to the test sequence $C4$, which originates from the same movie and exhibits very similar audiovisual characteristics (SI, TI, foveal area, aspect ratio, and audio energy). Therefore, the observed behaviour of $C3$ and $C9$ test sequences is most likely due to statistical differences between subjective ratings. On the other hand, Figure 5 illustrates

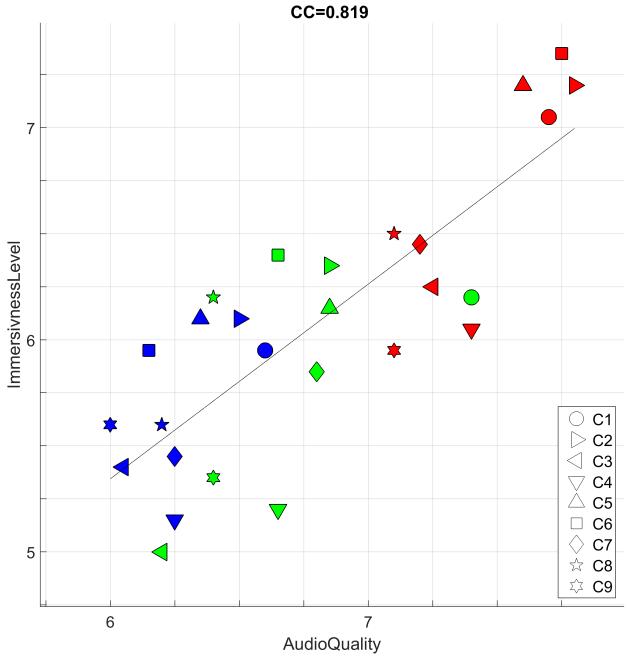
DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-128



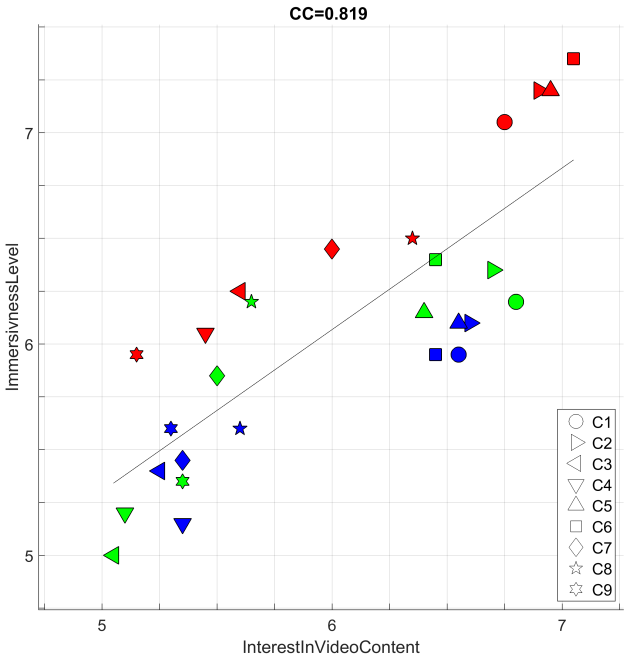
(a) IL and OQ correlation



(b) IL and SA correlation



(c) IL and AQ correlation



(d) IL and IV correlation

Figure 8: Correlation between the experienced IL and (a) perceived Overall Quality (OQ), (b) Awareness of their Surrounding (SA), (c) perceived Audio Quality (AQ), and (d) Interest in Video content (IV). Blue, green and red markers are iPhone, iPad and UHDTV stimuli respectively

how the average perceived SA is distributed among different sequences and devices. These results are in line with those reported in Figure 4. Interestingly the anomalies observed for $C3$ and $C9$ test sequences are not present, although differences between SAs between iPad and UHDTV are negligible for $C9$ and $C10$.

Figure 6 represents the distribution histograms of subjective scores related to ILs for each device. As it can be observed, these histograms form a left-skewed distribution. It should be noted that the lowest ILs are mostly experienced with the iPhone, whereas the highest ILs are largely experienced with UHDTV. The observed rate distribution histograms show that the subjective ratings are centered on the score 6 for iPhone and iPad, and on the

DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-128

	Surrounding Awareness	Overall Quality	Video Content Interest	Audio Content Interest	Audio Quality
Immersiveness Level	**-0.913**	**0.938**	**0.819**	**0.723**	**0.819**
Surrounding awareness	-	-0.869	-0.756	-0.705	-0.781
Overall quality	-	-	0.764	0.712	0.914
Video Content Interest	-	-	-	0.889	0.557
Audio Content Interest	-	-	-	-	0.609

Table 4: Pearson's correlation ρ (also called CC) coefficients between the ratings of different evaluation criteria.

score 7 for the UHDTV. Figure 7, representing the rate distribution histograms for the SA criterion shows a plateau-like distribution, which contradicts our a priori assumptions that all levels of SA will be experienced and SA will be negatively correlated to IL.

To understand the impact of all evaluation criteria on each other, Pearson's correlation was applied between MOS values of each two criteria. Pearson's ρ measures the degree of linear dependence between two variables. A ρ value of 1 is a total positive correlation, 0 is no correlation, and -1 is a total negative correlation. Table 4 summarizes obtained correlation results. High correlation between QoE and OQ has already been observed in the state of the art. Results of our tests show that a similar high correlation also exists between IL and OQ ($\rho > 0.938$) confirming that overall quality of audiovisual stimuli has an important impact on the immersive experience. Similarly, results show a strong negative correlation between SA and IL ($\rho < -0.913$) which confirms the previous conclusion. Moreover, OQ and AQ are strongly correlated ($\rho = 0.914$) which hints to the fact that high quality audio plays an important role in the overall quality experienced by subjects. There is also good correlation between IL and IV ($\rho > 0.819$) indicating that the level of interest in the video content plays a role in the immersive experience. Likewise, IL and AQ exhibit good correlation ($\rho > 0.819$) indicating audio quality has an impact on the immersive experience. The inter-relation between IL and IA is less strong ($\rho < 0.723$) therefore less impacting the IL than the previous criteria. Figure 8 shows the correlations between IL and other evaluation criteria for each device and content in more details. The color and shape of each marker indicate which device and test sequence were used for the multimedia consumption, respectively. In correlation graphs, the distinction between the three ILs is not always straightforward. Referring to results illustrated in Fig. 8(a), UHDTV audiovisual stimuli form a quite separate cluster, whereas iPhone and iPad stimuli are more interleaved. It can be observed that IL of each stimuli is often improved when an iPad is used instead of an iPhone, and when a UHDTV is used instead of an iPad or an iPhone.

Conclusion

In this paper, the results of subjective evaluation experiments assessing sense of presence in typical multimedia consumption scenarios are presented. More precisely, typical end-user consumption of multimedia contents rendered on three different devices, namely, iPhone, iPad, and UHDTV was evaluated by means of analysis of subjective ratings. Twenty subjects participated in experiments evaluating audiovisual stimuli in terms of interest in audio and video content, audio quality, perceived overall quality, surrounding awareness, and sense of presence. The analysis of subjective ratings shows that each device induces different level of sense of presence. In average, the sense of presence brought to subjects by UHDTV is higher when compared to iPad and iPhone. Moreover, the analysis revealed a strong correlation between the sense of presence and the perceived overall quality. Additionally, physiological signals (EEG, ECG, and respiration) were recorded for each test condition, i.e. combination of device and content. The recorded physiological signals together with subjective ratings form a publicly available dataset, which can help future studies to further investigate the quality of experience in immersive multimedia environment.

Acknowledgements

The work behind this paper was possible thanks to European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 643072 - Network QoE-Net, and Innovation Action immersiaTV. Authors especially acknowledge funding received for these projects from Swiss State Secretariat for Education, Research and Innovation SERI.

References

[1] J. Chai, Y. Ge, Y. Liu, W. Li, L. Zhou, L. Yao, and X. Sun. *Engineering Psychology and Cognitive Ergonomics: 11th International Conference, EPCE 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014. Proceedings*, chapter Application of Frontal EEG Asymmetry to User Experience Research, pages 234–243. Springer International Publishing, Cham, 2014.

[2] J. P. Forgas. On feeling good and being rude: Affective influences on language use and request formulations. *Journal of Personality and Social Psychology*, 76(6):928, 1999.

[3] I. Galloso, C. Feijóo, and A. Santamaría. Novel approaches to immersive media: From enlarged field-of-view to multi-sensorial experiences. In *Novel 3D Media Technologies*, pages 9–24. Springer, 2015.

[4] X. Geng. Cultural differences influence on language. *Review of European Studies*, 2(2):p219, 2010.

[5] S. Ishihara. Test for colour-blindness. *Tokyo: Hongo Harukicho*, 1917.

[6] ITU-R BT.2022. General viewing conditions for subjective assessment of quality of SDTV and HDTV television pictures on flat panel displays. *International Telecommunication Union*, August 2012.

[7] ITU-R BT.500-13. Methodology for the subjective assessment of the quality of television pictures. *International Telecommunication Union*, January 2012.

[8] ITU-T, P.800. Methods for subjective determination of transmission quality. *International Telecommunication Union*, August 2012.

[9] ITU-T P.910. Subjective video quality assessment methods for multimedia applications. *International Telecommunication Union*, April 2008.

DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-128

[10] A. Kathol and E. Shriberg. The SRI biofrustration corpus: Audio, video, and physiological signals for continuous user modeling. In *2015 AAAI Spring Symposium Series*, 2015.

[11] S.-E. Moon and J.-S. Lee. Perceptual experience analysis for tone-mapped HDR videos based on EEG and peripheral physiological signals. *Autonomous Mental Development, IEEE Transactions on*, 7(3):236–247, 2015.

[12] A.-F. Perrin, H. Xu, E. Kroupi, M. Řeřábek, and T. Ebrahimi. Multimodal dataset for assessment of quality of experience in immersive multimedia. In *Proceedings of the 23rd Annual ACM Conference on Multimedia Conference*, pages 1007–1010. ACM, 2015.

[13] S. Songden and E. Ike. Colour vision performance test. *Journal of Natural Sciences Research*, 3(11):19–24, 2013.

Author Biography

Anne-Flore Perrin *received the Dipl.-Ing. degree in Computer Science, with specialisation in Image Processing and Computer Graphics, from the engineering school of Rennes (ESIR), Rennes, France, in 2014. She is currently working as PhD student in the Multimedia Signal Processing Group (MMSPG) wherein the Department of Electrical Engineering, at the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland. Her research interests are in the fields of image and video processing and encoding to measure, improve and provide QoE in immersive multimedia technologies.*

Martin Řeřábek *received his PhD degree in 2013 from the Faculty of Electrical Engineering (FEE), Czech Technical University in Prague (CTU), Prague, Czech Republic. The main topic of his thesis was modeling of space variant optical systems and their impact to precision of scientific (astronomical) data processing. Since September 2010 he worked first as a doctoral assistant, and then since 2013 as a postdoc, in Multimedia Signal Processing Group, at École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland. His main research interest is focused on image and video processing in immersive multimedia (UHD TV, 3D, HDR), evaluation and improvement of quality of experience, astronomical image processing, and biomedical signal processing.*

Touradj Ebrahimi *received his M.Sc. and Ph.D., both in Electrical Engineering, from the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, in 1989 and 1992 respectively. In 1993, he was a research engineer at the Corporate Research Laboratories of Sony Corporation in Tokyo, where he conducted research on advanced video compression techniques for storage applications. In 1994, he served as a research consultant at AT&T Bell Laboratories working on very low bitrate video coding. He is currently Professor at EPFL heading its Multimedia Signal Processing Group. He was also adjunct Professor with the Center of Quantifiable Quality of Service at Norwegian University of Science and Technology (NTNU)between 2008 and 2012. Prof. Ebrahimi became a Fellow of the international society for optical engineering (SPIE) in 2003. He is also the head of the Swiss delegation to MPEG, JPEG and SC29, and acts as the Chairman of Advisory Group on Management in SC29. He is a co-founder of Genista SA, a high-tech start-up company in the field of multimedia quality metrics. His research interests include still, moving, and 3D image processing and coding, visual information security (rights protection, watermarking, authentication, data integrity, steganography), new media, and human computer interfaces (smart vision, brain computer interface). Prof. Ebrahimi is a member of IEEE, SPIE, ACM and IS&T.*