

# 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 technologies in the last decades. Consequently, new technologies quality measures evolve and trend to be more user-centric. This is the reason why the Quality of Experience (QoE) assessment is presently meaningful and challenging, especially for users typical experiences during multimedia content consumption. Such an evaluation is the aim of this paper. More specifically, the Sense of Presence (SoP) was explored in place of QoE as it is a factor influencing the QoE. This paper presents the conducted subjective test investigating typical and practical user experiences. This latter consists of presenting one-minute video stimuli to twenty subjects, on three different devices (iPhone, iPad and UHD screen). Annotated subjective scores were collected and physiological signals (EEG, ECG, and Respiration) were recorded during the conducted subjective test. The resulting multimodal dataset, aiming an alternative evaluation of human experience while consuming multimedia, is publicly available.*

## Introduction

Audiovisual communication technologies have made tremendous progress in the last few decades. 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. Capture and rendition of High Dynamic Range content with Wide Colour Gamut have become reality thanks to cameras with advanced sensors and optics, and monitors featuring brightness levels up to 4K 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.

Similar progress has been made in audio technologies, where mono has been replaced by stereo, surround and 3D in the last decades. Furthermore, the modes of consumptions of audiovisual content are changing, with a wide range of services offering new experiences thanks to innovations in how information is stored, and delivered to consumers and professionals, thanks to powerful and feature rich devices and infrastructures that enable mobile and immersive audiovisual communications.

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 Sense of Presence (SoP) felt by subjects experiencing immersive audiovisual content in different situations.

This is achieved by building upon a past similar attempt by the authors of this paper [13], where the sense of presence in a variety of audiovisual experiences was measured as a function of the modality and the quality of audiovisual content, from lower quality and no audio, standard definition video, to higher quality surround audio and ultra high definition video. The SoP was evaluated both explicitly, by means of a questionnaire, as well as implicitly, by recording selected physiological signals such as EEG, ECG, and respiration.

This prior work showed that there is a correlation between sense of presence as experienced by human subjects and their physiological signals. In this paper, the previous earlier attempt mentioned above is extended, while still relying on the same fundamental approach of finding the correlation between SoP as experienced and explicitly assessed from 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 extended in order to include those possible in typical mobile devices (such as when watching a movie on a phone or a tablet) in addition to Ultra High Definition (UHD) TV viewing experiences. A special attention was to scope out the typical and practical nature of users multimedia consumption.

In addition to the analysis of the conducted experiment based on the explicit ratings, this paper offers a publicly available database in assessment of SoP in immersive communication<sup>1</sup>.

Conducted experiments were based on the assumption that each device will induce to users different SoPs, referred to as Immersiveness Levels (ILs) in the remaining of this paper. Subjective ratings analysis showed that three immersiveness levels are observed. Nevertheless, the differences between the ILs produced by each device within used content is rather small.

The reminder of the paper is structured as follows. In first Section, we derive the test from related work. We then describe the experiment design and run. The Section Subjective ratings analysis presents the results of the subjective tests regarding explicit assessments. In the last Section, a conclusion is given.

---

<sup>1</sup><http://mmspg.epfl.ch/SoPMD>



Figure 1: Frame example of each sequence: testing dataset ((a)-(i)) and testing dataset (j).

## Related work

Multimedia providers tend to use Quality of Experience (QoE) to improve their services rather than Quality of Service (QoS) in order to be more user-centric. An overview of the evidences of the relation between SoP and QoE are given in [4]. Therefore we decided to assess the SoP instead of the QoE in this paper.

Human explicit assessment are influenced by emotional, cultural, educational and environmental differences across subjects [3, 5]. In order to be as independent as possible from the bias induce by subjects way of expression, the research community considered the study of physiological signals to assess the emotional states of users during their experience. Hence these implicit methodologies reveal the human body states.

For instance, in [11], the frustration created by a human-computer interface was evaluated via Electrocardiography (ECG), skin conductance, blood pressure and respiration. The authors of [2] investigate the emotional states of users interacting with mobile phone applications in studying the frontal alpha asymmetry of Electroencephalography (EEG) signals. The experience provided by tone-mapped HDR content is evaluated both explicitly (questionnaire) and implicitly (EEG, skin conductance, blood pressure, respiration and skin temperature) in [12]. The authors, also provide an overview of the physiological studies involving EEG and peripheral signals for various media type such as speech, video, 3D video, and sense of reality.

All the previously mentioned related work have assessed technology-centric scenario as the conducted experiments assess various technologies, such as HDR, 3D, and a developed interfaces. We propose here to study the SoP during the multimedia visualisation on three different devices. The design of the experiment considers typical multimedia end-user consumptions. At the end of each media consumption, explicit assessment was collected thanks to a subjective questionnaire. Additionally, physiological signals (EEG, ECG and respiration) were recorded during the media consumption.

## Experiment

In this section the video stimuli creation process is firstly described in detail. Then test methodology, used equipment and environment, and physiological signals acquisition are described. Finally, the information about the participants are provided.

## Audiovisual Stimuli

The one-minute audiovisual sequences presented during the subjective tests are extracted from four Blender open source movies<sup>2</sup>. The uncompressed movies were in format raw yuv 4:2:0, 8 bits, and 24 fps. The selected movies are originally HD or UHD contents. Audio signals associated to used movies were available in flac, stereo and 5.1 surround.

Ten sequences were extracted from the four movies. Nine sequences, C1-C9, were used to generate test stimuli. The last sequence, C10, will be dedicated to the three training stimuli creation. The sequences to extract must create as immersion as possible. Select sequences with high spatial and temporal complexities guarantee the subjects interest in the content while high surround audio energy ensure an audio 3D world which favour immersion. The selection of sequences was based on a conducted analysis on the audio energy level in surround channels, as well as on the temporal and spatial information on the luminance component (TI and SI [10]) of the movies. Parts of the contents with the highest values of the above mentioned properties (TI, SI and audio energy) were cut out, while the original movie scene editing was taken into account. The special attention paid to the scene cuts prevents abrupt changes at the beginning and the end of sequences.

A frame of each content is illustrated in Figure 1. The sequence properties, such as the movie they are extracted from, their resolution, and the number of the first frame of each sequence in the original movie are mentioned in Table 1

An analysis of the TI, SI and audio energy was then conducted on the selected sequences in order to have more knowledge about their properties such as spatial complexity, amount of movement and impact of the audio rear signals. The distribution of the sequences regarding their SI and TI are presented in Figure 2. It is observed that the sequences extracted from *Sintel* movie, have large TI values and small SI values. Thus the C5, C6 and C10 sequences contain a lot of movement and a low level of details representation. The sequences originating in *ElephantDream* movie and *BigBuckBunny* movie have small TI and SI values. This means there is few movements and a low level of details in C1, C2, C3, C4, and C9 sequences. The sequences extracted from *TearsofSteal* presents high SI values and fair TI values. Hence, C7 and C8 sequences are scenes with few movement which contain high amount of spatial information. This is

<sup>2</sup><http://media.xiph.org/>

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 Steel	4096x1744	4152
C8	Tears of Steel	4096x1744	10392
C9	Elephant Dream	1920x1080	11280
C10	Sintel	4096x1714	360

Table 1: Original movie, Original resolution and Start frames of sequences

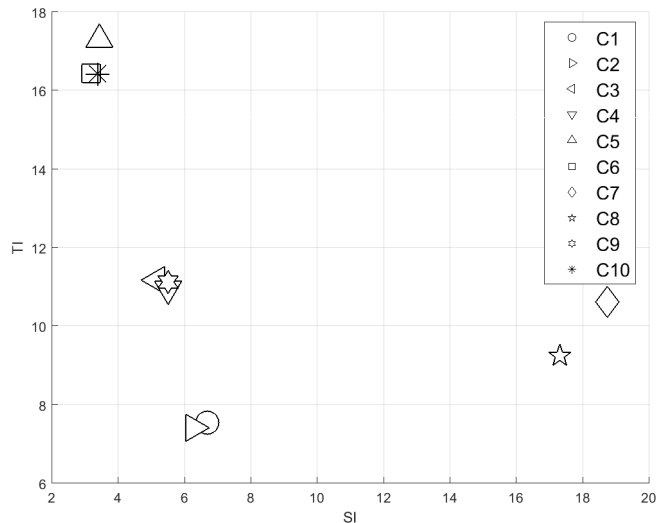


Figure 2: Values of SI and TI of sequences

explained by the fact that the three first movies are computer generated while 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 front and rear channels of each side in percentage. General conclusion of what we can see from these numbers is the audio energy distribution between left and right channels in comparison to central channel during the duration of the sequences.

Three different devices were used in this study. Details describing the devices are given in Section "Test equipment and environment". The audiovisual stimuli were generated from the sequences to match rendering properties of used devices. All stimuli were encoded, with h264 codec, to achieve the best quality in terms of each device rendering capabilities. The stimuli generated from the C3 sequence were encoded with a QP of 25, whereas the remaining stimuli were encoded with a QP of 20. Expert viewing session was conducted to confirm the transparent quality of all test material. To recapitulate, 27 test stimuli were generated from 9 sequences, C1-C9, and 3 training stimuli from the C10 sequence.

Sequence	$E_L$ [%]	$E_R$ [%]	$E_{SL}$ [%]	$E_{SR}$ [%]
C1	52.1	39.5	0.3	0.3
C2	61.7	26.9	0.9	2.6
C3	50.3	49.2	4.1	3.0
C4	8.8	7.9	15.3	19.6
C5	41.3	54.9	7.3	7.1
C6	23.8	11.3	67.8	12.8
C7	60.0	39.5	13.1	9.9
C8	53.2	46.8	34.9	34.4
C9	23.8	15.6	0.6	0.7
C10	37.6	43.3	40.6	27.9

Table 2: Sequences Audio Ratios in percentages:  $E_L$ ,  $E_R$ ,  $E_{SL}$ ,  $E_{SR}$

### Experimental protocol

The experiment consisted of three sessions. During each sessions, 9 stimuli were visualised on one device. Minimum one day break is done between two sessions to avoid any statistical bias. These breaks prevent lack of attention and fatigue of subjects and insure their comfort. Each session lasted approximatively one hour, including the training and the set up of physiological signals acquisition devices.

A training session is repeated at the beginning of each session to recapitulate the test procedure and to remind subjects of each IL. After the training, subjects were asked to calm down and noise signals were recorded. In coming 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
- Moving their jaw (biting), 5 times
- Making a snoring sound, twice
- If possible, moving ears, 3 times

Nine audiovisual stimuli were presented in each session leading to the total of 27 audiovisual stimuli forming 27 trials.

Each trial consisted of a ten-seconds baseline phase, a stimulus period and voting phase. During the baseline periods, 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. 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, a video stimulus was presented. After the video stimulus was over, subjects were asked to provide their ratings for the particular video stimulus in 30 seconds (about 5 seconds per question). After the voting phase start the next trial.

The Figure 3 illustrates an example of one test trial including baseline, stimuli, and vote period.

The order of sessions and trials were randomised.

Regarding the ratings, subjects were asked to evaluate the video stimuli according to six different criteria, namely, Interest in Audio content (IA), perceived Audio Quality (AQ), Interest in Video content (IV), Immersiveness Level (IL), perceived Overall Quality (OQ), and Awareness of their Surrounding (SA). A 9-point rating scale, following the Absolute Category Rating (ACR)

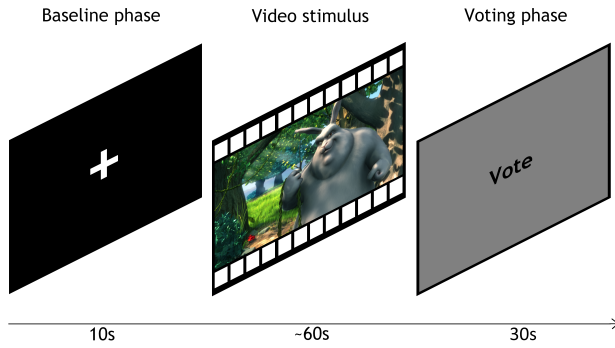


Figure 3: Example of a trial

evaluation methodology [10], 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 interest in video and audio content as well as the perceived overall quality, "no immersion" and "full immersion" for level of immersiveness and "not aware" and "fully aware" for awareness of the surrounding. The 9-points rating scale was presented with a clear separation line 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 to refine their assessment. The 3 defined classes were intended to be used for classification purposes.

### Test equipment and environment

Low, middle and high ILs were expected to be induced to the subjects by the devices used for the visualisation of the audiovisual stimuli. The iPhone5<sup>3</sup> and iPad4<sup>4</sup> were used to display low and middle ILs stimuli respectively. The professional high-performance 4K/QFHD LCD reference 56-inch monitor Sony Trimaster SRM-L560<sup>5</sup> was used to display high IL stimuli.

The Table 3 illustrates the characteristic settings of the three devices. As recommended in [7], the viewing distance of the UHD screen is 1.6H. Due to laboratory constraints, more specifically the fact the the subject can't hold the iPhone and the iPad to prevent EEG noise, their viewing distance was set at 6H (30cm) instead of 4.8H (24.50cm) and 4H (60cm) instead of 3.2H (47cm) respectively. H is the height of the screen of each devices. Visual attention is the amount of information in pixel or in % of the content which is perceived instantly by the fovea. The computed visual attention results shows that 15% (respectively 10% and 3%) of the video is perceived on the iPhone (r. iPad and UHD screen) by the fovea.

Two sound systems were used during the test. Stereo audio signals were used for iPad and iPhone stimuli, 5.1 surround audio signals were used for UHD screen stimuli. The stereo sound was provided by a professional headset Sennheiser HD 280 Pro<sup>6</sup>

<sup>3</sup><https://support.apple.com/kb/SP655?>

<sup>4</sup><https://support.apple.com/kb/SP662?>

<sup>5</sup>[http://pro.sony.com/bbsccms/assets/files/cat/mondisp/brochures/di0195\\_srm1560.pdf](http://pro.sony.com/bbsccms/assets/files/cat/mondisp/brochures/di0195_srm1560.pdf)

<sup>6</sup><http://en-us.sennheiser.com/professional-dj-headphones-noise-cancelling-hd-280-pro>

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

Table 3: Immersiveness levels settings based on devices characteristics

(accurate for linear sound reproduction in critical monitoring applications and attenuates 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 position 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 years). The 20 subjects were screened for correct visual acuity (no errors on 20/30 lines) and color vision using Snellen and Ishiara charts respectively[6, 14].

A prior informative file containing the detailed description of the subject assessment was provided to the subjects. Moreover, oral instructions were provided to the participant before they filled the consent form. Additionally, a training session was used to illustrate low, middle, and high ILs to guide subjects to bound their own perceived overall ratings. The training also allows subjects to get familiar with the assessment procedure.

### Subjective ratings analysis

This chapter 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 evaluated criteria are interrelated. The conducted analysis on subjective ratings assumed a Student's *t*-distribution of the subjective rates. Outlier detection, resulting distribution histograms, Mean Opinion Scores (MOS) and associated 95% Confidence Interval (CI), as well as Pearson's correlations are presented in the remaining of this section.

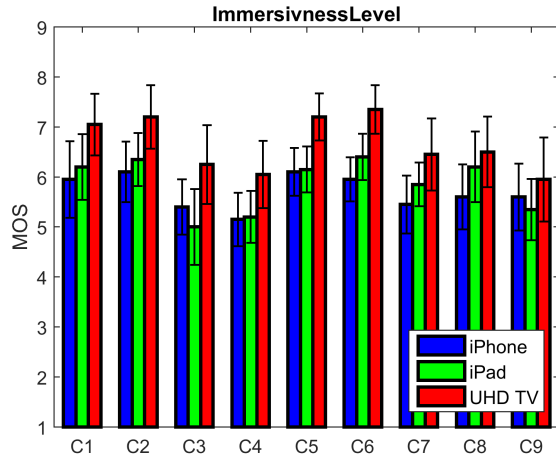


Figure 4: MOS and CIs for experienced ILs

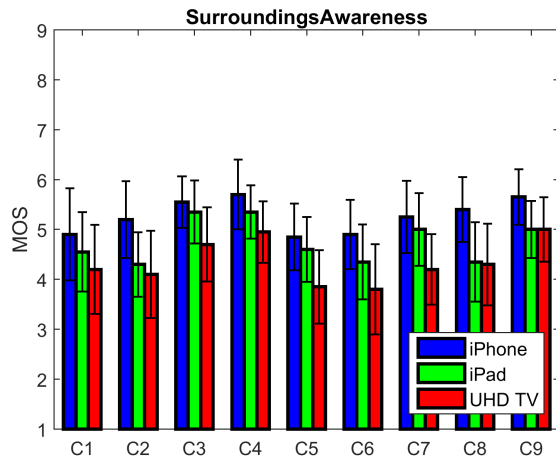


Figure 5: MOS and CIs for surrounding awareness

A detection and elimination of outliers was performed accordingly to the guidelines described in Section 2.3.1 of Annex 2 of [8]. 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 [9]. The Figures 4 and 5, depict MOS and CI for each content, regarding each ILs or SAs, induced by the used device.

The Figure 4 shows that used devices create to subjects different ILs. The iPhone induces the lowest IL while the UHD screen induces the highest IL. However the differences between these ILs are small as CIs considerably overlap for each content, for each IL. Moreover the MOSs are condensed in the 5 to 7.5 interval of values. The experiences are then assessed as high-middle or low-high SoP. Results could be explained by the fact that it is hard to produce low IL with a content of high quality in laboratory conditions as it is hard to produces a high IL with a one-minute stimulus. MOS of middle IL of C3 and C9 sequences are lowest than their MOSs of low IL. C3 and C9 can be compared to sequence C4 as they have similar characteristics such as original sequence, aspect ratio, SI, TI and

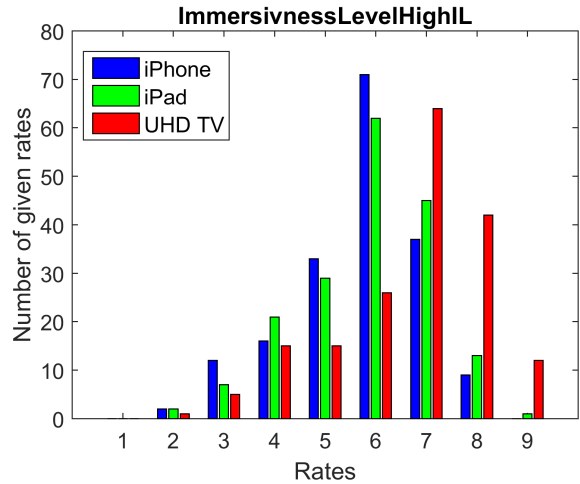


Figure 6: Distribution histogram regarding immersivness levels

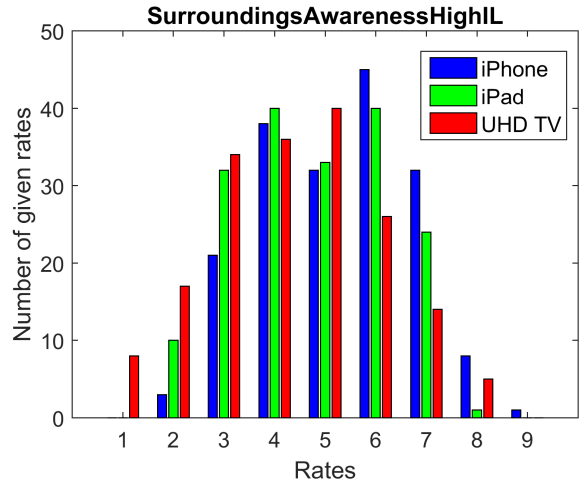
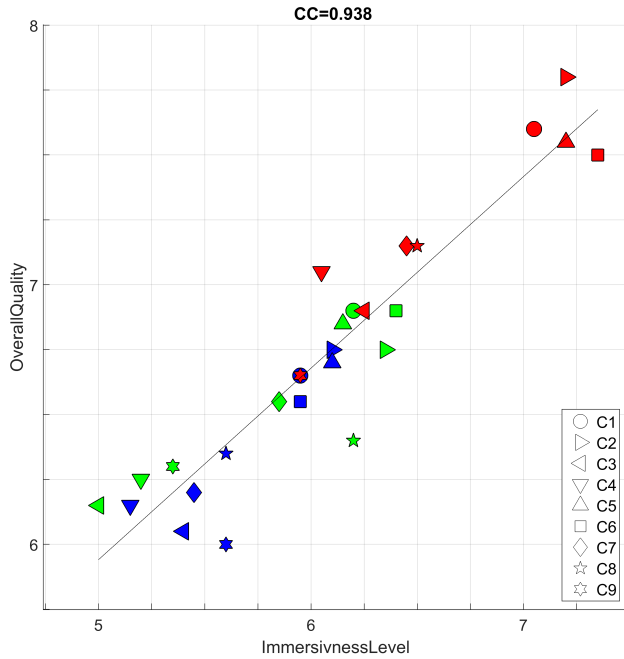


Figure 7: Distribution histogram regarding surrounding awareness

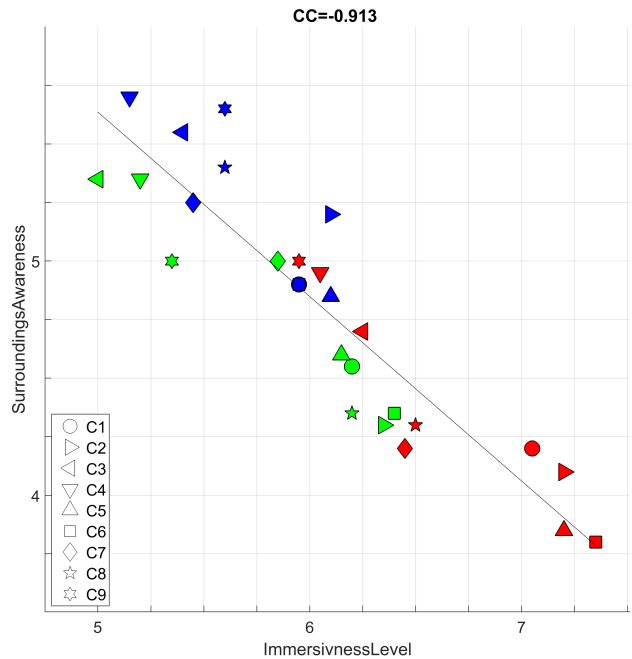
fovea attention. The differences found between the characteristics of these sequences lies in their audio signal energy. However no trend are observed between the differences of audio energy ratios of the sequences C3, C9 and C4. This leads to the conclusion that the fact that the IL of the iPad stimuli are assessed lower than the IL of iPhone stimuli is due to the small difference between ILs. The figure 5 fully illustrates the presence of three different ILs. The tiny difference between three ILs are observed. As mentioned previously, this behaviour is explained by the experiment design.

Rate distribution histograms shows the percentage of given rate regarding one of the six criterion.

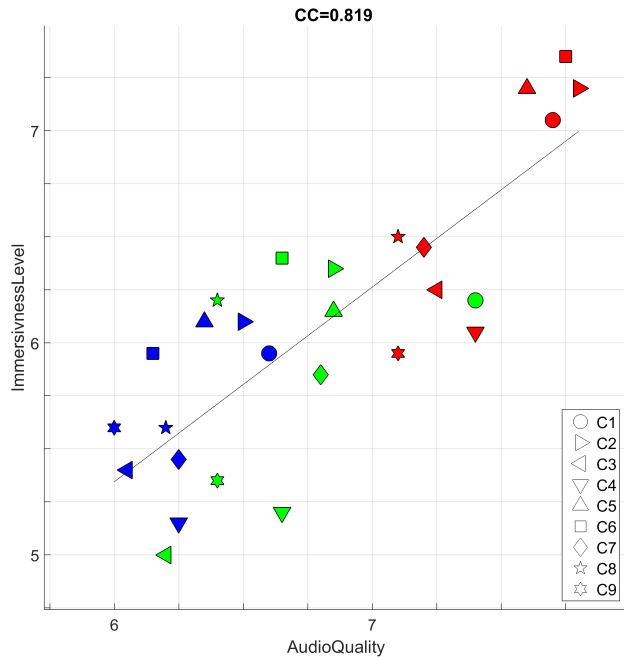
The Figure 6 represents the distribution histogram of the criteria Immersiveness Level (IL). The IL criteria follows a left-skewed distribution. It should be noted that the lowest ILs are mostly experiences with the iPhone while the best ILs are largely experienced thanks to the UHD screen. The observed rate distribution histograms regarding IL criteria and more specifically regarding the device used for the stimuli visualisation, shows that the iPhone's and iPad's distributions are centered on the score 6, and the UHD screen's distribution is centered on the score 7. The



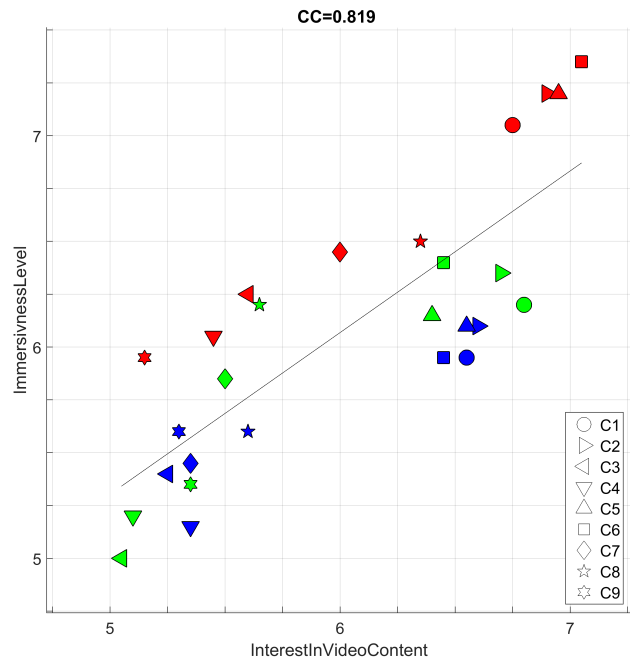
(a) IL and OQ correlation



(b) IL and SA correlation



(c) IL and AQ correlation



(d) IL and IV correlation

Figure 8: Correlation between the experienced IL and (a) overall quality, (b) surrounding awareness, (c) audio quality, and (d) interest in video content

Blue, green and red markers are iPhone, iPad and UHD screen stimuli respectively

Figure 7, representing the distribution histogram for the SA criterion, shows a plateau like distribution. The expectations for this criteria was to follow such a distribution, showing that all awareness were experienced. The SA should be negatively correlated to IL. The presented result questioned this hypothesis as ILs and SAs do not follow opposite distributions.

To understand the impact of the evaluation criteria, such as IA, AQ, IV, OQ, IL, and SA, on each other, Pearson's correlation was applied between MOSs of two criteria. Pearson's  $\rho$  measures the degree of linear dependence between two variables. A  $\rho$  value of 1 is a total positive correlation, 0 is no correlation, -1 is a total negative correlation. The Table 4 summarizes obtained correla-

	Surrounding Awareness	Overall Quality	Video Content Interest	Audio Content Interest	Audio Quality
<b>Immersiveness Level</b>	<b>-0.913</b>	<b>0.938</b>	<b>0.819</b>	<b>0.723</b>	<b>0.819</b>
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  $\rho$  (also called CC) coefficients between the ratings of different evaluation criteria.

tion results. Knowing that QoE is highly dependant of QoS and OQs can be seen as the assessment of the QoS perception, the resulting strong correlation between ILs and OQs ( $\rho > 0.93$ ) was expected. Then the almost total and negative correlation between SAs and ILs ( $\rho < 0.91$ ) is verified. Obviously, the OQs and the AQs are strongly correlated ( $\rho = 0.914$ ). There are also strong correlations between IL and IV and AQ ( $\rho > 0.81$ ) criteria. However the IVs is more correlated with the IAs than with the ILs. The inter-relation between IL and IA criteria is strong ( $\rho < 0.73$ ) but is less impacting the IL than the previous modalities. The Figure 8 details the correlations between ILs and OQs, SAs, AQs, and IVs. The markers’ colors indicate which device was used for the media consumption, markers’ shapes point out the stimuli visualised. In all correlation graphs, the distinction between the three IL is not straightforward. However, in the graphs (c) and (d), the UHD stimuli shape a separate class, whereas iPhone and iPad stimuli are blended. It can be observed that IL of each stimuli is mostly improved when an iPad is used instead of an iPhone and when a UHD screen is used in place of an iPad. However, the C3 and C9 stimuli middle IL is lowest than the low IL instead of being the highest (or the opposite) in the subfigures (a), (b) and (c).

## Conclusion

In this paper we analysed the results of our experiments in which stimuli visualized on different devices, with various resolutions and viewing distances for instance, were presented to human subjects. Each trial was assessed thanks to subjective assessments about various aspects and recorded EEG, ECG, and respiration signals.

The experiment was user-centric as it assesses the typical content consumption of multimedia services end-users on three devices, namely an iPhone, an iPad, and a UHD screen. Regarding analysis of subjective ratings, three classes of SoP were observed. The small differences between the three levels of SoP is explained by the experiment design. The presented analysis demonstrates the coherence of the ratings and their high quality.

The resulting database is available to research community and especially those interested in assessment of SoP in immersive communication, so that research in this field can progress further.

## Acknowledgements

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 643072 - Network QoE-Net.

## References

- [1] M. M. Bradley and P. J. Lang. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 25(1):49–59, 1994.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] X. Geng. Cultural differences influence on language. *Review of European Studies*, 2(2):p219, 2010.
- [6] S. Ishihara. Test for colour-blindness. *Tokyo: Hongo Harukicho*, 1917.
- [7] 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.
- [8] ITU-R BT.500-13. Methodology for the subjective assessment of the quality of television pictures. *International Telecommunication Union*, January 2012.
- [9] ITU-T, P.800. Methods for subjective determination of transmission quality. *International Telecommunication Union*, August 2012.
- [10] ITU-T P.910. Subjective video quality assessment methods for multimedia applications. *International Telecommunication Union*, April 2008.
- [11] 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.
- [12] 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.
- [13] 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.
- [14] S. Songden and E. Ike. Colour vision performance test. *Journal of Natural Sciences Research*, 3(11):19–24, 2013.

## Author Biography

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.

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.

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.