Multisensory 360° Videos Under Varying Resolution Levels Enhance Presence

A. Covaci, E. B. Saleme, *Member, IEEE*, G. Mesfin, I-S. Comsa, R. Trestian, C. A. S. Santos, and G. Ghinea, *Member, IEEE*

Abstract—Omnidirectional videos have become a leading multimedia format for Virtual Reality applications. While live 360° videos offer a unique immersive experience, streaming of omnidirectional content at high resolutions is not always feasible in bandwidth-limited networks. While in the case of flat videos, scaling to lower resolutions works well, 360° video quality is seriously degraded because of the viewing distances involved in head-mounted displays. Hence, in this paper, we investigate first how quality degradation impacts the sense of presence in immersive Virtual Reality applications. Then, we are pushing the boundaries of 360° technology through the enhancement with multisensory stimuli. 48 participants experimented both 360° scenarios (with and without multisensory content), while they were divided randomly between four conditions characterised by different encoding qualities (HD, FullHD, 2.5K, 4K). The results showed that presence is not mediated by streaming at a higher bitrate. The trend we identified revealed however that presence is positively and significantly impacted by the enhancement with multisensory content. This shows that multisensory technology is crucial in creating more immersive experiences.

Index Terms—multisensory, 360-degree videos, encoding quality, presence, mulsemedia.

1 Introduction

VR (Virtual Reality) is envisaged to be one of the killer-apps of the future enabled by the advent of 5G networks, revolutionising the way we perceive and interact with media. Through artificial sensory stimulation, both computer generated VR and CVR (Cinematic Virtual Reality) - where content is captured with 360° cameras - have the ability to immerse users in engaging experiences. However, to achieve true engagement, the quality of the visual feedback [20] has to be maintained during streaming, while end-to-end latency has to be kept below 15-20 milliseconds [29]. With the surge of 360° videos, which have become catalysts for new forms of journalism and social experiences [21], [33], [45], this vision is hard to fulfil.

360° videos present many technical challenges and are extremely bandwidth intensive, as they require a complete high-quality 360° frame for interactivity. However, only a fraction of what is downloaded is displayed on the device, resulting in bandwidth waste. The absence of continuous optimal bandwidth conditions and the encoding/decoding quality degradation lead to video deterioration that was traditionally measured through QoS (Quality of Service). However, during the recent decade, one important paradigm change in communication networking research has been the evolution trend from QoS to QoE (Quality of Experience) and QoL (Quality of Life) [13], [41]. While QoS describes the performance of the system based on low-level network metrics, QoE is its user-centric counterpart defined as "the degree of delight or annoyance of the user of an application or service. It results from the fulfillment of his or her expectations with respect to the utility and/or enjoyment of the application or service in the light of the user's personality and current state" [22]. The main purpose of QoE is to characterise the human side of the multimedia experience, however it is not trivial to provision a high QoE in the context of persistent delivery of 360° videos. To prevent QoE from being affected, and manage any potential scarcity of resources, adaptive streaming solutions have been proposed where video qualities are adjusted according to users' viewports [18].

Also with a view towards resource savings, a different approach explores what happens to QoE when we go beyond audiovisual interfaces and focus on multisensory setups. In multisensory systems, information from various sensory modalities (such as sight, sound, touch, smell, and taste) are integrated to achieve unique experience. Many emerging forms of content can provide users with a multisensory experience, thus eliciting a wide range of emotions or knowledge [19]. This requires the annotation of mulsemedia (multiple sensorial media) [15] with additional metadata that allows the control of corresponding rendering multisensory interfaces [10], [15]. Recent studies have shown that for multisensory CVR applications, the augmentation of audiovisual content with media targeting extra sensorial channels leads to an improvement of subjective QoE evaluation and masks quality degradation. Since metadata driven approaches (e.g., MPEG V) are negligible in size compared to voluminous 360° videos, these results show that mulsemedia can be employed in a resource-saving process [9], [17]. This enables us to remark that the benefits of multisensory enhanced CVR greatly outweigh its costs,

- *A. Covaci is with the University of Kent, UK.* *E-mail: a.covaci@kent.ac.uk*
- *E. B. Saleme is with the Federal Institute of Espírito Santo, Brazil.*
- *C. A. S. Santos is with the Federal University of Espírito Santo, Brazil.*
- *I.-S. Comsa is with the Swiss Distance University of Applied Sciences, Brig, CH-3900, Switzerland*
- *G. Ghinea is with Brunel University London, UK.*
- *G. Mesfin is with Kristiania University College, UK.*
- *R. Trestian is with Middlesex University, UK.*

thus further research should look into adjacent areas of QoE, such as UX (User Experience). This will allow researchers to understand and quantify the overall user experience.

The convergence of new enabling technologies (e.g., new multimedia experiences: augmented reality, VR, CVR; multisensory interfaces; affordable wearables: smart watches, health bands - that can capture physiological signals; data mining technologies) can improve key services in our society. If these applications fail to meet the quality requirements, their impact will be limited. To prevent this, we should move beyond concepts like service or product quality to more broad approaches that consider and assess the combination of technologies to QoE. As a result, new work needs to be undertaken to address the challenges of these new services and to capture the influence factors that might differ between applications. For instance, for CVR applications, the evaluation of QoE has to consider other key aspects (e.g., sense of presence) besides the perceptual quality. The sense of presence is the central goal for UX in alternate reality experiences [42], thus its evaluation should become sine qua non when measuring user experiences in CVR.

Our Contribution

The relationship between video degradation in 360° videos and the impact of the overall experience is insufficiently studied, notwithstanding the potential resource saving impact that such a study might have on bandwidth intensive applications such as CVR. To this end, we propose to mitigate the impact of video degradation by leveraging multisensory stimulation to improve the overall user experience in CVR. Specifically, we focus on presence as one of the influencing factors when experiencing 360° videos.

For the remaining of the paper we will refer to traditional 360° videos as *360°multimedia* and to their enhancement with multisensory content as *360°mulsemedia*.

2 Related Work

CVR streaming challenges

Several major broadcasters (e.g., BBC, Sky) have started to explore CVR and Augmented Reality (AR) technologies for enhancing and expanding their current services, with a focus on music or sport events. Their offer does not only include pre-recorded highlights but also 360° video coverage, where viewers can select from a variety of vantage points. In this context, specialists predict that by 2022, CVR and AR traffic will increase 12-fold compared to 2017 [7]. However, despite the increased availability and popularity of CVR content, the infrastructure required for its distribution is still lacking and users trying 360° videos often report their grainy, pixelated appearance [6].

360° videos are about 5 times larger than regular videos and involve complex projections [55]. Despite this, the streaming approach of major CVR providers (e.g., YouTube, Facebook) is to display the entire panoramic view, causing a significant growth in data requirements. In the absence of continuous optimal bandwidth conditions and bad encoding/decoding quality for the given bitrate, various errors might occur that cause the deterioration of video experience. Moreover, this streaming approach leads to significant bandwidth waste [5].

In this context, QoE is important for multimedia services, where its main purpose is to manage the scarcity of resources. Most of the studies in this area, look into the evaluation and improvement of the perceptual quality component and propose tile-based methods based on QoE-driven adaptive streaming systems [37], [53]. 360° video frames are cropped into tiles that are then encoded into multi-bitrate segments and pre-fetched by the client based on the predicted viewport. However, despite the apparent flexibility of these solutions, obtaining the optimal tiles to provide a high QoE steaming service is not trivial [49].

Mulsemedia - a solution?

QoE enhancement was also investigated from a different perspective - in setups that stimulate multiple sensory dimensions. In the context of mulsemedia [10], studies show that engaging more senses (e.g., smell, taste, and touch) can improve the overall QoE when viewing audiovisual content. For example, olfactory-enhanced multimedia was shown to mask audio degradation [3] and to enrich the user experience [1], [2], [12], [14], [26], [51].

In [9], the authors explored further the potential of mulsemedia and look into increasing the QoE of CVR experiences with a minimal impact on the underlying networking resources. For this, they investigated how multiple sensory cues mask quality degradation and if they can be employed in a resource-saving process for streaming omnidirectional videos. Their results demonstrated that mulsemedia is promising for enhancing QoE in 360° videos even when the encoding quality is reduced.

With the evolution of VR-related technologies, there has been a push forward in the current state of the art for multisensory systems where several senses are stimulated at the same time, and users are presented with 'real experiences' designed in virtual worlds. Unlike traditional audiovisual setups, multisensory environments provide more sensory information to the user and this should enable a more immersive, coherent, and credible experience, thus possibly raising the level of presence.

Presence in VR and CVR

The concept of QoE has been mainly applied to multimedia systems and services, but research on this focuses on managing the scarcity of resources. However, interactive services (e.g., online gaming, video conferencing, VR) are much more complex to deal with and user behaviour, actions and context impact the perceived quality [13]. Given that in VR setups, dimensions such as presence and immersion are strongly related to the user experience, they should be equally considered in QoE evaluations. Some efforts have been undertaken in this direction [34], [36], however they do not consider multisensory setups.

Consciousness of our immediate surroundings emerges when incoming sensory information (vision, sound, touch, force, taste, and smell) is broadcast globally to multiple cognitive systems that process the incoming data. Bottom-up processing of the sensory inputs is actively combined with top-down processing based on our previously existing model of the world. Thus, our perceptual system can infer the full spatial models of a place even when we see a small proportion of it - think about how fast we get to ”know” a new room. These processes are behind the fact that VR works even when the rendering is poor and the environments simplistic [40]. The cues offered by VR allow our perceptual system to hypothesise about the nature of a place and to use a fill-in mechanism based on existing internal models of that type of place. VR aims to replace real sensory perceptions with computer generates ones - if this substitution is effective, the brain is ”tricked” into believing that the virtual world is in fact the surrounding physical world. However, to achieve effective substitution, we need to consider the sensory systems we want to include - vision and auditory are typical, touch, smell and taste are rare. This subjective illusion of ”being there” in a virtual environment – in spite of the fact that you know for sure that you are not actually there - is referred to in the literature as presence or ”place illusion”.

Presence is a concept describing a core subjective sensation, thus the most commonly used method for measuring this illusion of ”being there” in a virtual environment is via questionnaires developed and validated over more than two decades [39]. One of the most frequently used and cited presence questionnaire is developed by Witmer and Singer [48]. This has 32 items grouped into four core groups (control, sensory, distraction and realism factors) using three subscales: involvement/control, naturalness, and interface quality. However, despite its popularity, this questionnaire was critisised because it does not give a measure of presence that is constructed independently from the factors that might influence it [38]. Another approach for measuring presence was proposed by Slater et al. [43]. The Slater-Usoh-Steed (referred to as SUS) questionnaire has six questions focusing on three factors: 1) the subjects’ sense of “being” in the virtual environment; 2) the extent to which the virtual environment becomes the dominant reality and 3) the extent to which the virtual environment is remembered as a place rather than just a visual stimulus (referred to as locality).

Aside from the use of questionnaires, a limited amount of existing research explores the reliability of behavioral and physiological data for evaluating presence. Promising results about various physiological measurements (including heart rate, skin temperature and skin conductance) during exposure to virtual environments were reported in [24], [25] and [46]. These indicate that both skin conductance and heart rate data are correlated with the sense of presence reported by answers to subjective questionnaires. Nevertheless, objective measurements come also with limitations - physiological measures require a baseline comparison for each user and additional equipment which might lead to breaks in presence [35].

Adding sensory cues (that require little computation) to virtual environments was shown to increase the sense of presence and the memory of the environment without lowering system responsiveness [11]. In [30], the authors take this research a step further and build a multisensory head mounted display (HMD) to explore different seasons reporting the enhancement of the sense of presence in the multisensory setup. A novel haptic display (exploring light touch, texture and temperature as actuation channels) based on a robot arm attached on an HMD is introduced in [47] indicating improvements in the multisensory environment. In [16], the authors focus on credible VR scenarios enhanced with multisensory stimuli and show that in these setups it is more difficult to raise presence. Moreover, this also depends on the combinations of the considered multisensory stimuli.

The viewing of 360° videos on VR headsets can provide novel immersive user experiences and, by extension, enhanced levels of QoE [23], [42], [55]. However, the impact of combining 360° multimedia or mulsemedia on presence has been largely uncharted. In [42], the authors claim to evaluate presence as one of the key QoE aspects for 360° videos. However, their approach is not based on a validated and published bespoke questionnaire for measuring presence. Presence in 360° videos is also the focus of [36], however the authors focus on the evaluation of long vs. short version of the Witmer and Signer presence questionnaire.

Consequently, the study detailed in this paper explored the following research questions:

RQ1: *What is the impact of mulsemedia on the sense of presence in 360° videos?*

RQ2: *What is the impact of video degradation on the sense of presence in 360° mulsemedia?*

Moreover, given that high motion levels in omnidirectional videos viewed through an HMD might be a significant contributor to QoE [56] and presence [44], this was a further issue which we explored in the study, via the following research question:

RQ3: *What is the impact of video motion levels on the sense of presence in 360° mulsemedia?*

3 Methods and Materials

3.1 Participants

We recruited a total of 48 participants (27 male, 21 female) from campuses of three universities (University of Kent, Brunel University, and Middlesex University). Recruitment was by email advertising. The final sample consisted of 12 participants in each of the four encoding quality conditions (HD, FullHD, 2.5K, 4K). Their age was between 16 and 65 years old (33% between 16 - 25; 31% between 26 - 35; 36% over 35 years). All participants had normal or corrected-to-normal vision and were screened for contraindications for VR (e.g., epilepsy, psychoactive drugs treatment). Overall sample size and participants per condition are in line with similar studies reported elsewhere [27], [31]. 69% of the participants reported having prior experience with VR.

3.2 Instruments

We used Unity to develop a VR application that reproduces equirectangular videos annotated with Sensory Effects Metadata (SEM) of the MPEG-V standard. This application communicates with a mulsemedia renderer to send the associated sensory effects metadata via a wireless local network provided by a WiFi router. For our experiment, we selected as audiovisual stimuli three 360°videos downloaded from YouTube, based on the following criteria: 1) the audiovisual content had to afford the association of semantically-congruent scents and matching airflow effects; 2) the three videos had to present varying degrees

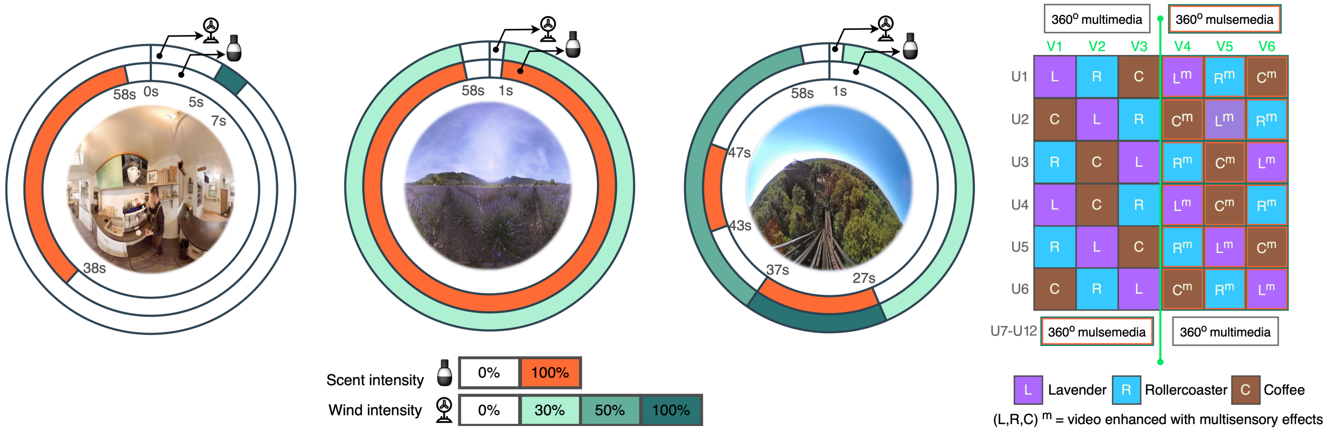


Fig. 1. a) Timeline of multisensory effects activation and intensity (left); b) Viewing order for the first 12 users (U1 - U12). Video resolution: HD (right)

of dynamism/content motion (static, semi-dynamic, and dynamic). Accordingly, the videos, their features and the associated sensory effects are described below.

Coffee shop. Description: a barista preparing a coffee. Scent: coffee. Airflow effect: puff of air made when the barista steams the milk for the cappuccino. Camera position: fixed. Content: semi-dynamic.

Lavender field. Description: a meander through a field of lavender. Scent: lavender. Airflow effect: breeze. Camera position: fixed. Content: static.

Rollercoaster. Description: a ride with the rollercoaster. Scent: diesel (because of the mechanical association between this particular scent and burnt rubber). Airflow effect: strong wind in the user's face when going downhill. Camera position: moving. Content: dynamic.

The duration of each video was 60s and multisensory effects were applied following the timeline of activation and intensity presented in Figure 1 (a). Each video was encoded with four levels of quality (HD, Full HD, 2.5K, 4K) using H.264/MPEG-4 Part 10 and had the chroma location: left; and projection: equirectangular.

3.3 Apparatus

To enable users to experience 360° multisensory videos, we designed and built a 360° mulsemedia head-mounted prototype (Figure 2). This consisted of: 1) a Samsung Galaxy S6 smartphone with a Super AMOLED capacitive touchscreen and 16M colors, 5.1 inches (71.5 cm^2) screen size, and 1440 x 2560 pixels (and 577 PPI density) used for rendering 360° videos; 2) a scent device; 3) a wind emitter device. The scent and wind-emitter device were controlled by a DFRobot Bluno Nano. The scent device emitted smell through a re-sizeable pipe connected to the headset. The scent was activated by a micro fan located at the basis of the pipe, blowing air through mesh bags with scent crystals. The power supply of the wind device was modified so that it can be used with an AC power source. An Arduino Uno microcontroller was used to control both the power supply and the wind blower fan.

A laptop running a mulsemedia effects renderer called PlaySEM SER [32] was also used to logically integrate the 360° video applications to the wind and smell devices. The laptop was a quad-core Intel Core i7-6700 HQ running at

Fig. 2. User with 360° mulsemedia prototype

2.6GHz, 16 GB RAM, 260 GB SSD, and GTX960M 4 GB GPU. We employed a TP-LINK WiFi router[1] to wirelessly connect the laptop and the smartphone.

Last but not least, mention must be made that participants sat on a swivel-chair which enabled them to spin around and experience the 360°videos. All participants wore i-shine[2] headphones during the experiment.

3.4 Measures

To measure the user's subjective sensation of presence, we used the original version of the SUS questionnaire. This assesses presence considering three themes: 1) Being there: the sense of 'being there' when experiencing VR; 2) Dominant reality: the extent to which participants perceive a virtual environment as the dominant reality; 3) Images or places: the extent to which the virtual environment makes the participants perceive they are visiting a place as opposed to viewing images. Each item is rated on a 7-point scale where a high rating is indicative of presence. Based on the

1. https://www.tp-link.com/eg/home-networking/access-point/tl-wa901nd/
2. https://www.ishine-trade.com/Headphones-Earphones

individual responses to the six questions associated with each experimental condition, we followed similar previous studies [4], [43] and we calculated two measures: SUS Count and SUS Mean. SUS Count indicates the number of responses of 6 and 7 among the six questions of the SUS questionnaire. SUS Mean is the mean score across the six questions.

3.5 Experimental Design

Our study had three independent variables (ENCODING QUALITY, MOTION DYNAMISM, and SENSORY EFFECTS) and one dependent variable (SENSE OF PRESENCE). A mixed factorial design was employed, whereby SENSORY EFFECTS and MOTION DYNAMISM were within subject variables, whereas ENCODING QUALITY was a between subject variable.

4 Results and Discussion

In this section, we investigate how multisensory effects, video degradation and motion dynamism influence the sense of presence in a 360° video experience. Accordingly, we are going to structure this section in keeping with the three research questions identified in Section 2, the answers and implications of which we proceed to discuss.

We used Cronbach's alpha to ascertain the internal consistency of the questionnaire. Results indicated a high level of consistency ($\alpha = 0.879$).

4.1 RQ1: What is the impact of mulsemedia on the sense of presence in 360° videos?

Results of the SUS Presence questionnaire are presented in Table 1. SUS Mean was computed by averaging the 7-point scores of the SUS questionnaire. Results show that 360° mulsemedia induces a higher level of presence (SUS Mean = 4.93, SD = 1.23) than 360° multimedia (SUS Mean = 4.17, SD = 1.30). The values of SUS Count (the mean of the test count of scores of 6 or 7 for the six questions) are consistent with the mean values for presence. The 360° mulsemedia condition has a higher count with a mean count of 2.53 (SD = 1.41). To test whether the differences in presence evaluation between the two conditions are statistically significant, we further analysed the data with an independent sample t-test. This indicated that values for SUS Mean and SUS Count are significantly higher when multisensory effects are present, t(286) = 5.074, $p < 0.001$; t(286) = 4.512, $p < 0.001$.

Condition	SUS Count	SUS Mean
360° mulsemedia	2.53 ± 1.41	4.93 ± 1.23
360° multimedia	1.36 ± 1.99	4.17 ± 1.30

TABLE 1
Means and Standard Deviations of Questionnaire Scores

We explored this further by looking into how 360° mulsemedia influences different themes of the SUS Presence questionnaire. Figure 3 illustrates the distribution of ratings across the different questionnaire items and themes. All the three presence themes (*Being there*, *Dominant reality* and *Images or place*) have higher scores (more 6s and 7s) when multisensory effects are employed, with *Being there* being there showed the greatest difference between multimedia and mulsemedia (131 scores of 6 and 7 in 360° mulsemedia vs 65 scores of 6 and 7 in 360° multimedia).

Accordingly, *mulsemedia enhances the sense of presence in CVR*. Our results clearly highlighted that, irrespective of encoding quality employed, the use of mulsemedia in 360° immersive environments leads to an enhanced sense of presence. Accordingly, on average, the sense of presence is enhanced by 18% when multisensory effects are employed. This finding underlines the usefulness of enriching CVR environments with multisensory content and confirms similar work [52] undertaken in the context of traditional desktop computing based settings.

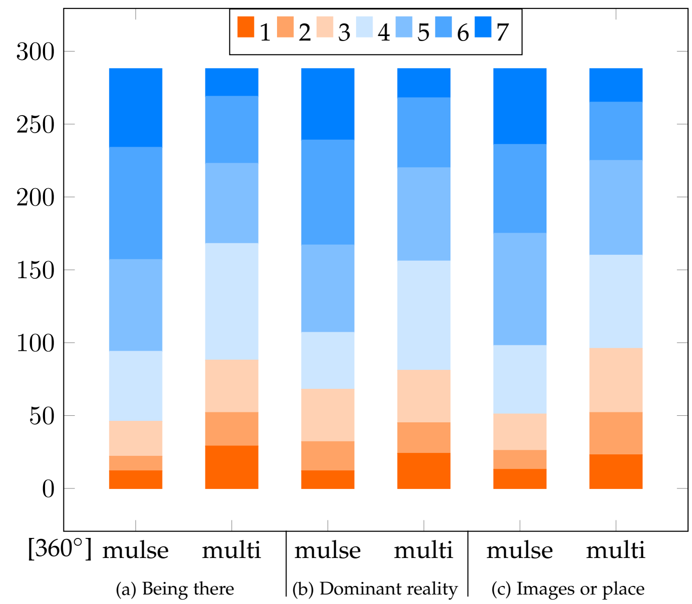


Fig. 3. Visualization of the distribution of ratings (1-7) across the three SUS questionnaire themes for 360° mulsemedia and 360° multimedia.

4.2 RQ2: What is the impact of video degradation on the sense of presence in 360° mulsemedia?

To investigate how video degradation in 360° videos impacts the sense of presence, we computed the SUS Mean values for the four encoding qualities (see Figure 4). Results showed the sense of presence is stronger in the 360° mulsemedia condition, irrespective of the underlying encoding quality.

To assess statistical significance, we analyzed presence ratings with a 2 (SENSORY EFFECTS: 360° multimedia, 360° mulsemedia) x 3 (MOTION DYNAMISM: slow, medium, fast) x 4 (ENCODING QUALITY: HD, Full HD, 2.5K, 4K) between-subjects ANOVA followed by post-hoc Bonferroni adjusted pair-wise comparisons. We found that the main effect of ENCODING QUALITY on presence evaluation was significant, F(3,264) = 6.58, $p < 0.001$. The main effect of SENSORY EFFECTS on presence evaluation was also significant, F(1,264) = 26.93, $p < 0.001$. There was no significant interaction between ENCODING QUALITY and SENSORY EFFECTS F(3, 2.64) = 1.65, p = 0.18. Post-hoc Bonferroni-adjusted comparisons indicated that in the mulsemedia condition, SUS Means for HD (SUS Mean = 4.67, SD = 1.27), Full HD (SUS Mean = 4.92, SD = 1.01) and 4K (SUS Mean = 4.99, SD = 1.56) were

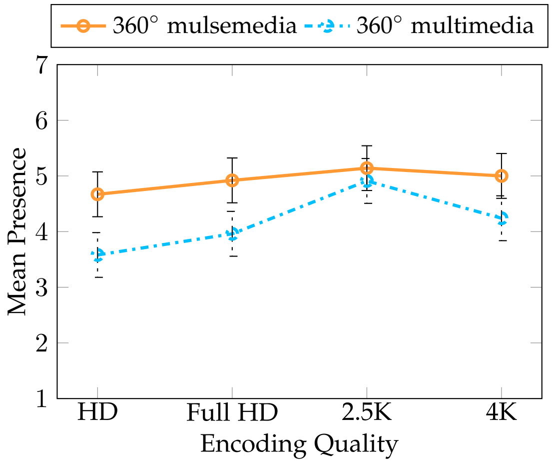


Fig. 4. Mean Presence values for different encoding qualities of 360° multimedia and 360° mulsemedia

significantly different (HD: $p < 0.001$, Full HD: $p = 0.01$ and 4K: $p = 0.09$) from the presence evaluation in the multimedia condition (HD: SUS Mean = 3.58, SD = 1.16; Full HD: SUS Mean = 3.96, SD = 1.18; 4K: SUS Mean = 4.23, SD = 1.49).

Increasing the video bitrate level does not necessarily lead to an an improvement in the perceived sense of presence. As our results have underlined, the assumption that higher encoding quality - and, by extension, data rates - will be automatically followed by an enhanced sense of presence is not borne out. Whilst, perhaps counter-intuitive to some degree, this might well be explained by the fact that encoding quality was a between-subjects independent variable. The reasons for doing so were precisely so that subjects were blind to other possible qualities and shows that one does not need to employ a high encoding quality in 360° videos to achieve an enhanced sense of presence for the users. This leads us to our next point.

Transmitting high data rates traffic puts significant pressure on the underlying networks. As our results have shown, this might not be necessary in the context of CVR. Indeed, transmitting at lower data rates (and lower encoding qualities) did not lead to a detrimental impact on the user sense of presence. The improvement in sense of presence from Full HD to 2.5K is marginal (and not warranted if there are insufficient network resources to accommodate this) and the extent to which participants felt present in our CVR actually decreased when going from 2.5k to 4K (which will be elaborated upon further in our next point). Consequently, what our findings also showcase is the need for adaptive bandwidth-allocation protocols incorporating user-centric measures (such as the sense of presence), which lead to more efficient resource allocation strategies in practice. This is reinforced by results from previous work of ours [8], which have highlighted the same qualitative aspect of the impact of video encoding rates of mulsemedia QoE in 360 VR° and it is important that presence, a significant experience-related dimension in VR is likewise affected.

Our study suggests that a good level of perceived presence while experiencing CVR can be achieved at a resolution of 2.5K, therefore going higher might not be necessary. Whilst the hardware characteristics of the equipment used in our study could be an influencing factor (e.g., CPU throttling or resolution degradation may have happened but has not been tested), the fact remains that the typical VR kit used by an average user is - due to cost constraints - not likely to be top-spec (and could well be mid-range, as in our study). Therefore, overloading the network as well as the processing capacity of the end user device with what is considered to be, from a technical viewpoint, the best possible encoding quality will not lead to an enhanced sense of presence. These findings underpin two observations: firstly, the importance of considering the actual user device in the context of presence in CVR and, secondly, the need not to go beyond 2.5K if sense of presence is important for the CVR application. We recognise that this is a somewhat counter-intuitive result and more research must be undertaken to explore its possible causes, with a more diverse set of experimental configurations.

QoE has the same pattern. Elsewhere [9] we have reported the results of a study exploring the impact on user quality of experience of employing mulsemedia in 360° videos. It is noteworthy that, just like in this study exploring the impact of mulsemedia on the sense of presence, qualitative results are strikingly similar. Accordingly, perceived quality of omnidirectional videos was higher when multisensory effects were employed. Airflow and olfactory effects boost the overall perceived quality of 360° videos by 12% and enjoyment by 13% across the four encoding qualities. This is similar to results obtained in the study currently reported. Moreover, in terms of user-perceived quality, the improvement from HD to Full HD is evident and significant. The gain for transition from Full HD to 2.5K is marginal, and there is a decrease in user-perceived quality when going from 2.5K to 4K. A similar profile, is again observed here, and, it is safe to say that encoding quality can be reduced in 360° videos mulsemedia to Full HD without any significant detrimental impact on either QoE or the sense of presence.

4.3 RQ3: What is the impact of video motion levels on the sense of presence in 360° mulsemedia?

The analysis of the influence played by motion dynamism on the sense of presence perceived by the participants - in both multimedia and mulsemedia conditions - revealed that the interaction between these variables was not statistically significant F(6, 264) = 0.487, p = 0.82. Simple main effects analysis showed that motion dynamism is close to being significant F(2, 264) = 2.90, p = 0.057. Mean values for presence in 360° multimedia and mulsemedia show that multisensory effects mitigate the effects of both video degradation and motion level on the sense of presence (see Figures 5 and 6).

Whilst there is substantial evidence that content is king in multimedia QoE (i.e. the particular dynamism - or lack thereof - of multimedia content influences QoE) [28], [54], here we show a tendency for mulsemedia in a 360° context to mitigate the influence of content dynamism in as far as overall presence is concerned. So, although content dynamism/motion level impacts QoE, *when multisensory effects are employed, content does not seem to impact the sense of presence*. Whilst highlighting that QoE and sense of presence are different concepts and impacted upon differently by

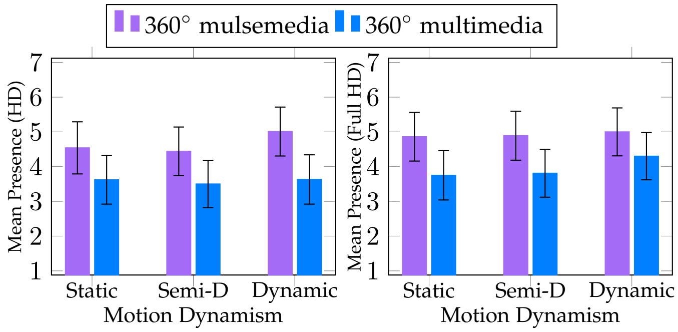


Fig. 5. Impact of the motion level on the presence experienced by participants in both multimedia and mulsemedia conditions for HD (left) and Full HD (right) encoding qualities.

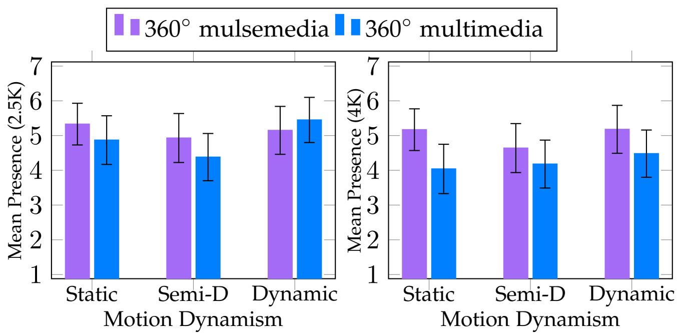


Fig. 6. Impact of the motion level on the presence experienced by participants in both multimedia and mulsemedia conditions for 2.5K (left) and 4K (right) encoding qualities.

mulsemedia, the finding nonetheless also emphasises the benefit of using mulsemedia in 360° VR as - even in the narrow context of the three clips employed in our study (chosen solely based on their varied dynamism/motion level) - it enhances the sense of presence, irrespective of the type of content being experienced. However, further work with an even more diverse and representative content set needs to be undertaken to validate the results of our exploratory study.

5 Conclusion

Advances in compression technologies have significantly increased access to digital video content. These technologies make possible the delivery of high-quality audio-visual streaming, usually by employing a process of perceptual encoding to reduce the size of audio-visual files while keeping both visual and audio quality degradation to a minimum. The perceptual (also called psychovisual and psychoacoustic) coding techniques exploit the limitations of human auditory and visual systems to discard the data captured during digitization which are not needed or cannot be used by users.

In traditional videos, the user has a view of the content on a flat screen and pixel-based scaling to lower resolutions works well. On the other hand, the user experiences a high degree of presence and interactivity with 360° video, using a head-mounted display that enables spherical viewing direction. However, the visual quality requirements are more stringent compared to traditional video because of the viewing distances involved in head-mounted displays and therefore the more perceptible quality degradation is presented at lower resolutions in 360°.

Streaming of omnidirectional content at high resolutions is yet a challenge in bandwidth-limited networks. In this paper, we proposed the enrichment of 360° video presentations with multisensory effects as a solution to mitigate visual degradation caused by the reduction of video encoding quality and motion level while preserving an overall acceptable user experience with the omnidirectional content. Thereby, we have studied how quality degradation impacts the sense of presence in VR mulsemedia, whereby scent and airflow effects have been synchronized with three 360° equirectangular videos encoded with different qualities (HD, FullHD, 2.5K, 4K).

As a limitation of our work, we acknowledge that the propagation of the airflow and scents on the physical environment may decrease the intensity of the sensory effects, which might affect the way users perceive them. To mitigate this, we placed the participants on a swivel-chair close enough to the wind fans. As for the smell, we used an adjustable pipe for directing the scent to the users' nostrils. As there was no walk-through in the VR environment, we constrained the participants to spin around and to watch the videos from the swivel-chair at a distance where we figured out they could feel the sensory effects properly. Thus, we did not consider the depth of sensory effects in the videos presented to the users, which was out of the scope of our work and can be the subject of further studies in the area. Nevertheless, we took into account different intensities of sensory effects in relation to the scenes in each video sample to avoid underkill/overkill effects. Moreover, with evolving - and improving - hardware comes the possibility that results reported therein could change. This is yet another possibility for future work.

In this paper, a study of 360° mulsemedia from the perspective of presence - one of the dimensions of QoE - is provided, in which, the 360° video content is locally stored. Assessing QoE is very important in the context of CVR as it allows for an evaluation of media through the user attitude and expectation concerning the employed prototype. As stated by Xu et al. [50], "since omnidirectional video ultimately outputs to human eyes, subjective visual quality assessment is more rational than an objective one in assessing visual quality." Alongside QoE, QoS indicators measure the network performance when delivering the 360° video with remotely stored content. Although measuring QoS could bring value to the proposed prototype, implementing and emulating such end-to-end system is costly and difficult. A first challenge in this direction is to ensure the physical infrastructure, such as multiple 360° mulsemedia equipment and a local radio base station. The second challenge is to enable the operational processes, such as scheduling multiple 360° mulsemedia users, ensuring proper wireless conditions, equipment synchronization, etc. To this end, further study can consider QoS measures to evaluate the system performance when serving simultaneously multiple 360° mulsemedia users.

Whilst we have acknowledged limitations of our work, this should not detract from noteworthy results obtained therein. Accordingly, in our experiments the gathered subjective data showed that 360° mulsemedia induces a higher level of presence in comparison to 360° multimedia in all scenarios. Thus, bearing in mind that the encoding quality of 360° videos can increase the sense of presence, what becomes evident is the capacity of enhancement provided by those sensory effects when integrated into such immersive videos.

Acknowledgments

This work has been performed in the framework of the Horizon 2020 project NEWTON (ICT-688503). It was also financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES) —Finance Codes 88881.187844/2018-01 and 88882.317673/2019-01. E. B. Saleme also acknowledges support from the Federal Institute of Espírito Santo.

References

[1] Oluwakemi A Ademoye and Gheorghita Ghinea. Synchronization of olfaction-enhanced multimedia. *IEEE Transactions on Multimedia*, 11(3):561–565, 2009.

[2] Oluwakemi A. Ademoye and Gheorghita Ghinea. Information recall task impact in olfaction-enhanced multimedia. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 9(3):17, 2013.

[3] Oluwakemi A. Ademoye, Niall Murray, Gabriel-Miro Muntean, and Gheorghita Ghinea. Audio masking effect on inter-component skews in olfaction-enhanced multimedia presentations. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 12(4):51, 2016.

[4] Mariano Alcañiz, Beatriz Rey, Jose Tembl, and Vera Parkhutik. A neuroscience approach to virtual reality experience using transcranial doppler monitoring. *Presence: Teleoperators and Virtual Environments*, 18(2):97–111, 2009.

[5] Yanan Bao, Huasen Wu, Tianxiao Zhang, Albara Ah Ramli, and Xin Liu. Shooting a moving target: Motion-prediction-based transmission for 360-degree videos. In *2016 IEEE International Conference on Big Data (Big Data)*, pages 1161–1170. IEEE, 2016.

[6] Marc Van den Broeck, Fahim Kawsar, and Johannes Schöning. It's all around you: Exploring 360 video viewing experiences on mobile devices. In *Proceedings of the 25th ACM International Conference on Multimedia*, MM '17, page 762–768, New York, NY, USA, 2017. Association for Computing Machinery.

[7] VNI Cisco. Cisco visual networking index: Forecast and trends, 2017–2022, 2018.

[8] I. Comşa, E. B. Saleme, A. Covaci, G. M. Assres, R. Trestian, C. A. S. Santos, and G. Ghinea. Do i smell coffee? the tale of a 360 mulsemedia experience. *IEEE MultiMedia*, 27(1):27–36, 2020.

[9] Alexandra Covaci, Ramona Trestian, Estêvão Bissoli Saleme, Ioan-Sorin Comsa, Gebremariam Assres, Celso AS Santos, and Gheorghita Ghinea. 360 mulsemedia: a way to improve subjective qoe in 360 videos. In *Proceedings of the 27th ACM International Conference on Multimedia*, pages 2378–2386, 2019.

[10] Alexandra Covaci, Longhao Zou, Irina Tal, Gabriel-Miro Muntean, and Gheorghita Ghinea. Is multimedia multisensorial? - a review of mulsemedia systems. *ACM Comput. Surv.*, 51(5):91:1–91:35, September 2018.

[11] Huong Q Dinh, Neff Walker, Larry F Hodges, Chang Song, and Akira Kobayashi. Evaluating the importance of multi-sensory input on memory and the sense of presence in virtual environments. In *Proceedings IEEE Virtual Reality (Cat. No. 99CB36316)*, pages 222–228. IEEE, 1999.

[12] Darragh Egan, Conor Keighrey, John Barrett, Yuansong Qiao, Sean Brennan, Christian Timmerer, and Niall Murray. Subjective evaluation of an olfaction enhanced immersive virtual reality environment. In *Proceedings of the 2nd International Workshop on Multimedia Alternate Realities*, pages 15–18. ACM, 2017.

[13] Markus Fiedler, Sebastian Möller, Peter Reichl, and Min Xie. A glance at the dagstuhl manifesto 'qoe vadis?'. In *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–3. IEEE, 2018.

[14] Gheorghita Ghinea and Oluwakemi Ademoye. The sweet smell of success: Enhancing multimedia applications with olfaction. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 8(1):2, 2012.

[15] Gheorghita Ghinea, Christian Timmerer, Weisi Lin, and Stephen R Gulliver. Mulsemedia: State of the art, perspectives, and challenges. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 11(1s):1–23, 2014.

[16] Guilherme Gonçalves, Miguel Melo, José Vasconcelos-Raposo, and Maximino Esteves Bessa. Impact of different sensory stimuli on presence in credible virtual environments. *IEEE transactions on visualization and computer graphics*, 2019.

[17] Álan LV Guedes, G de A Roberto, Pascal Frossard, Sérgio Colcher, and Simone Diniz Junqueira Barbosa. Subjective evaluation of 360-degree sensory experiences. In *2019 IEEE 21st International Workshop on Multimedia Signal Processing (MMSP)*, pages 1–6. IEEE, 2019.

[18] Mohammad Hosseini and Viswanathan Swaminathan. Adaptive 360 vr video streaming: Divide and conquer. In *2016 IEEE International Symposium on Multimedia (ISM)*, pages 107–110. IEEE, 2016.

[19] Y. Ikei, K. Abe, K. Hirota, and T. Amemiya. A multisensory vr system exploring the ultra-reality. In *2012 18th International Conference on Virtual Systems and Multimedia*, pages 71–78, Sep. 2012.

[20] Jason Jerald. *The VR book: Human-centered design for virtual reality*. Morgan & Claypool, 2015.

[21] Sarah Jones. Disrupting the narrative: immersive journalism in virtual reality. *Journal of Media Practice*, 18(2-3):171–185, 2017.

[22] Patrick Le Callet, Sebastian Möller, Andrew Perkis, et al. Qualinet white paper on definitions of quality of experience. *European network on quality of experience in multimedia systems and services (COST Action IC 1003)*, 3(2012), 2012.

[23] Xing Liu, Qingyang Xiao, Vijay Gopalakrishnan, Bo Han, Feng Qian, and Matteo Varvello. 360 innovations for panoramic video streaming. In *Proceedings of the 16th ACM Workshop on Hot Topics in Networks*, pages 50–56. ACM, 2017.

[24] Michael Meehan, Brent Insko, Mary Whitton, and Frederick P Brooks Jr. Physiological measures of presence in stressful virtual environments. *Acm transactions on graphics (tog)*, 21(3):645–652, 2002.

[25] Michael Meehan, Sharif Razzaque, Brent Insko, Mary Whitton, and Frederick P Brooks. Review of four studies on the use of physiological reaction as a measure of presence in stressfulvirtual environments. *Applied psychophysiology and biofeedback*, 30(3):239–258, 2005.

[26] Niall Murray, Oluwakemi A Ademoye, George Ghinea, Yuansong Qiao, Gabriel-Miro Muntean, and Brian Lee. Olfactory-enhanced multimedia video clips datasets. In *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–5. IEEE, 2017.

[27] Niall Murray, Yuansong Qiao, Brian Lee, Gabriel-Miro Muntean, and AK Karunakar. Age and gender influence on perceived olfactory & visual media synchronization. In *2013 IEEE international conference on multimedia and expo (ICME)*, pages 1–6. IEEE, 2013.

[28] Pradip Paudyal, Federica Battisti, and Marco Carli. Impact of video content and transmission impairments on quality of experience. *Multimedia Tools and Applications*, 75(23):16461–16485, 2016.

[29] Cristina Perfecto, Mohammed S Elbamby, Javier Del Ser, and Mehdi Bennis. Taming the latency in multi-user vr 360: A qoe-aware deep learning-aided multicast framework. *IEEE Transactions on Communications*, 2020.

[30] Nimesha Ranasinghe, Pravar Jain, Nguyen Thi Ngoc Tram, Koon Chuan Raymond Koh, David Tolley, Shienny Karwita, Lin Lien-Ya, Yan Liangkun, Kala Shamaiah, Chow Eason Wai Tung, et al. Season traveller: Multisensory narration for enhancing the virtual reality experience. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2018.

[31] Nimesha Ranasinghe, Pravar Jain, David Tolley, Shienny Karwita Tailan, Ching Chiuan Yen, and Ellen Yi-Luen Do. Exploring the use of olfactory stimuli towards reducing visually induced motion sickness in virtual reality. In *Symposium on Spatial User Interaction*, pages 1–9, 2020.

[32] Estêvão Bissoli Saleme, Celso A. S. Santos, and Gheorghita Ghinea. A mulsemedia framework for delivering sensory effects to heterogeneous systems. *Multimedia Systems*, 25(4):421–447, Aug 2019.

[33] Marcio Carneiro dos Santos. Narrativas e desenvolvimento de conteúdo imersivo: aplicação de sistemas bi e multissensoriais de realidade virtual no jornalismo. *Intercom: Revista Brasileira de Ciências da Comunicação*, 42(3):133–150, 2019.

[34] Raimund Schatz, Andreas Sackl, Christian Timmerer, and Bruno Gardlo. Towards subjective quality of experience assessment for omnidirectional video streaming. In *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–6. IEEE, 2017.

[35] Valentin Schwind, Pascal Knierim, Nico Haas, and Niels Henze. Using presence questionnaires in virtual reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2019.

[36] Ashutosh Singla, Rakesh Rao Ramachandra Rao, Steve Göring, and Alexander Raake. Assessing media qoe, simulator sickness and presence for omnidirectional videos with different test protocols. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 1163–1164. IEEE, 2019.

[37] Robert Skupin, Yago Sanchez, Cornelius Hellge, and Thomas Schierl. Tile based hevc video for head mounted displays. In *2016 IEEE International Symposium on Multimedia (ISM)*, pages 399–400. IEEE, 2016.

[38] Mel Slater. Measuring presence: A response to the witmer and singer presence questionnaire. *Presence*, 8(5):560–565, 1999.

[39] Mel Slater. How colorful was your day? why questionnaires cannot assess presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 13(4):484–493, 2004.

[40] Mel Slater and Maria V Sanchez-Vives. Enhancing our lives with immersive virtual reality. *Frontiers in Robotics and AI*, 3:74, 2016.

[41] C. Timmerer, T. Ebrahimi, and F. Pereira. Toward a new assessment of quality. *Computer*, 48(3):108–110, Mar 2015.

[42] Huyen TT Tran, Nam Pham Ngoc, Cuong T Pham, Yong Ju Jung, and Truong Cong Thang. A subjective study on qoe of 360 video for vr communication. In *2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP)*, pages 1–6. IEEE, 2017.

[43] Martin Usoh, Ernest Catena, Sima Arman, and Mel Slater. Using presence questionnaires in reality. *Presence: Teleoperators & Virtual Environments*, 9(5):497–503, 2000.

[44] Aleksander Väljamäe. *Self-motion and presence in the perceptual optimization of a multisensory virtual reality environment*. PhD thesis, Chalmers tekniska högskola, 2005.

[45] Kristin Van Damme, Anissa All, Lieven De Marez, and Sarah Van Leuven. 360 video journalism: Experimental study on the effect of immersion on news experience and distant suffering. *Journalism Studies*, 20(14):2053–2076, 2019.

[46] Brenda K Wiederhold, Dong P Jang, Mayumi Kaneda, Irene Cabral, Yair Lurie, T May, IY Kim, Mark D Wiederhold, and SI Kim. An investigation into physiological responses in virtual environments: an objective measurement of presence. *Towards cyberpsychology: Mind, cognitions and society in the internet age*, 2001.

[47] A. Wilberz, D. Leschtschow, C Trepkowski, J. Maiero, E. Kruijff, and B. Riecke. Facehaptics: Robot arm based versatile facial haptics for immersive environments. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 2020.

[48] Bob G Witmer and Michael J Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence*, 7(3):225–240, 1998.

[49] Lan Xie, Zhimin Xu, Yixuan Ban, Xinggong Zhang, and Zongming Guo. 360probdash: Improving qoe of 360 video streaming using tile-based http adaptive streaming. In *Proceedings of the 25th ACM international conference on Multimedia*, pages 315–323, 2017.

[50] M. Xu, C. Li, Z. Chen, Z. Wang, and Z. Guan. Assessing visual quality of omnidirectional videos. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(12):3516–3530, 2019.

[51] Zhenhui Yuan, Shengyang Chen, Gheorghita Ghinea, and Gabriel-Miro Muntean. User quality of experience of mulsemedia applications. *ACM Trans. Multimedia Comput. Commun. Appl.*, 11(1s):15:1–15:19, October 2014.

[52] Zhenhui Yuan, Gheorghita Ghinea, and Gabriel-Miro Muntean. Quality of experience study for multiple sensorial media delivery. In *2014 International Wireless Communications and Mobile Computing Conference (IWCMC)*, pages 1142–1146. IEEE, 2014.

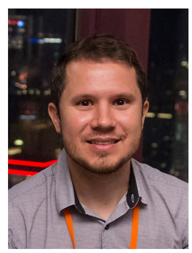
[53] Alireza Zare, Alireza Aminlou, Miska M Hannuksela, and Moncef Gabbouj. Hevc-compliant tile-based streaming of panoramic video for virtual reality applications. In *Proceedings of the 24th ACM international conference on Multimedia*, pages 601–605, 2016.

[54] Tiesong Zhao, Qian Liu, and Chang Wen Chen. Qoe in video transmission: A user experience-driven strategy. *IEEE Communications Surveys & Tutorials*, 19(1):285–302, 2016.

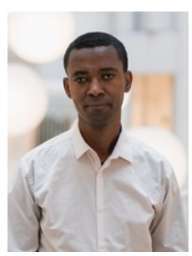
[55] Chao Zhou, Zhenhua Li, and Yao Liu. A measurement study of oculus 360 degree video streaming. In *Proceedings of the 8th ACM on Multimedia Systems Conference*, pages 27–37. ACM, 2017.

[56] M. Zink, R. Sitaraman, and K. Nahrstedt. Scalable 360video stream delivery: Challenges, solutions, and opportunities. *Proceedings of the IEEE*, pages 1–12, 2019.

Alexandra Covaci She is a researcher in the field of virtual reality, currently Lecturer in Digital Arts and Technology at the University of Kent, UK. Her research activities lie at the confluence of virtual reality, multisensory media, human-computer interaction, and psychology. Contact her at a.covaci@kent.ac.uk.

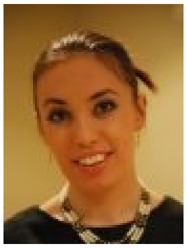


Estêvão B. Saleme He holds a PhD degree in Computer Science from UFES, Brazil, 2019. During his PhD, he was an Academic Visitor at Brunel University London, UK. His research interests include multimedia and human-computer interaction with a focus on digital multisensory systems, quality of experience of users, immersive technologies, signal processing, middleware, and frameworks. Contact him at estevao.saleme@ieee.org.



Gebremariam Mesfin He is an Associate Professor at Kristiania University College, Norway. Prior to his current position, he worked as a Research Assistant in the Computer Science Department at Brunel University London. His research interests span software engineering, usability, cross-platform smartphone applications, multimedia, mulsemedia, Web service, service-oriented architecture, and e-learning. Contact him at gebremariam.assres@kristiania.no.

Ioan-Sorin Comsa He is a Data Scientist with the Institute for Research in Open-, Distance- and eLearning, Swiss Distance University of Applied Sciences, Brig, Switzerland. His research interests include intelligent radio resource and QoS management, reinforcement learning, data mining, distributed and parallel computing, adaptive multimedia/mulsemedia delivery. Contact him at ioan-sorin.comsa@ffhs.ch.



Ramona Trestian She is a Senior Lecturer with the Design Engineering and Mathematics Dept., Middlesex University, London, UK. Her research interests include mobile and wireless communications, user-perceived quality of experience, multimedia streaming, handover, and network selection strategies. Contact her at r.trestian@mdx.ac.uk.

Celso A. S. Santos He is a Professor in the Department of Informatics at Federal University of Espírito Santo, Brazil. His current research interests focus on multimedia systems and applications, multisensorial experiences, and crowdsourcing systems. Contact him at saibel@inf.ufes.br.

George Ghinea He is a Professor of Mulsemedia Computing in the Computer Science Department at Brunel University London, UK. His work focuses on building adaptable cross-layer end-to-end communication systems incorporating user multisensorial and perceptual requirements. Contact him at george.ghinea@brunel.ac.uk.