360° Mulsemedia: A Way to Improve Subjective QoE in 360° Videos

Alexandra Covaci
University of Kent
Canterbury, United Kingdom
a.covaci@kent.ac.uk

Ramona Trestian
Middlesex University
Hendon, United Kingdom
r.trestian@mdx.ac.uk

Estêvão Bissoli Saleme
Federal University of Espírito Santo
Vitória-ES, Brazil
estevao.saleme@aluno.ufes.br

Ioan-Sorin Comsa
Brunel University London
Uxbridge, United Kingdom
ioan-sorin.comsa@brunel.ac.uk

Gebremariam Assres
Brunel University London
Uxbridge, United Kingdom
gebremariam.assres@brunel.ac.uk

Celso A. S. Santos
Federal University of Espírito Santo
Vitória-ES, Brazil
saibel@inf.ufes.br

Gheorghita Ghinea
Brunel University London
Uxbridge, United Kingdom
george.ghinea@brunel.ac.uk

ABSTRACT

Previous research has shown that adding multisensory media—mulsemedia—to traditional audiovisual content has a positive effect on user Quality of Experience (QoE). However, the QoE impact of employing mulsemedia in 360° videos has remained unexplored. Accordingly, in this paper, a QoE study for watching a 360° video—with and without multisensory effects—in a full free-viewpoint VR setting is presented. The parametric space we considered to influence the QoE consists of the encoding quality and the motion level of the transmitted media. To achieve our research aim, we propose a wearable VR system that provides multisensory enhancement of 360° videos. Then, we utilise its capabilities to systematically evaluate the effects of multisensory stimulation on perceived quality degradation for videos with different motion levels and encoding qualities. Our results make a strong case for the inclusion of multisensory effects in 360° videos, as they reveal that both user-perceived quality, as well as enjoyment, are significantly higher when mulsemedia (as opposed to traditional multimedia) is employed in this context. Moreover, these observations hold true independent of the underlying 360° video encoding quality—thus QoE can be significantly enhanced with a minimal impact on networking resources.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**; **Virtual reality**;

MM '19, October 21–25, 2019, Nice, France

ACM ISBN 978-1-4503-6889-6/19/10…$15.00
https://doi.org/10.1145/3343031.3350954

KEYWORDS

Mulsemedia, 360° videos, Multisensory virtual reality, QoE, Subjective testing, Immersive media

ACM Reference Format:
Alexandra Covaci, Ramona Trestian, Estêvão Bissoli Saleme, Ioan-Sorin Comsa, Gebremariam Assres, Celso A. S. Santos, and Gheorghita Ghinea. 2019. 360° Mulsemedia: A Way to Improve Subjective QoE in 360° Videos. In *Proceedings of the 27th ACM International Conference on Multimedia (MM '19), October 21–25, 2019, Nice, France.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3343031.3350954

1 INTRODUCTION

The evolution of network capabilities and the development of hardware has led to a rapid advance in video services. Applications like multimedia mobile streaming gained popularity and providers are constantly making available new generations of video streaming technologies. In this context, omnidirectional videos (also known as 360° videos) are emerging as a new type of applications that promise an immersive visual experience, where users have the freedom to control the view from a full spherical panorama. 360° videos can be seen in browsers or via Head-Mounted Displays (HMDs) - where the viewport is controlled through standard peripheral devices such as keyboard and mouse or by moving the device in sync with one's head movements. Omnidirectional videos play a key role in enabling Virtual and Augmented Reality (VR/AR) applications that are pushed by industry and by faster mobile networks (5G). VR broadcasting, focused on music or sport events, is gaining popularity as a new business model. In this context, it is envisaged that by 2022, VR and AR traffic will increase 12-fold compared to 2017 [5].

However, this immersive experience that allows for 3D view and user interaction does not come without cost. Compared with 4K traditional content, the processes behind VR are more complex—from content creation to viewing. 360° videos are large (about 5 times larger than regular videos) and involve complex projections (e.g., Oculus encodes each omnidirectional video into 88 offset cubic projections [39]). The streaming approach of major 360° video providers (e.g., YouTube, Facebook) is monolithic—the entire panoramic view is displayed—thus, leading to 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.

Indeed, the user experience is important in any multimedia context, and that of 360° VR/AR is no exception. If for no other reason, a better understanding of the user experience could lead to novel insights about perceptual redundancies, which in turn could be potentially exploited to reduce the aforementioned processing and bandwidth costs of 360° VR/AR applications. Whilst efforts in this direction generally lead to the possibility of accommodating a greater number of 360° VR/AR applications with no deterioration in user Quality of Experience (QoE), the question that begs itself is: *are we able to increase QoE further with a minimal impact on the underlying networking resources?*

One potential avenue for accomplishing this is given by mulsemedia—*multiple sensorial media* [7, 10]. Noting that 360° VR/AR applications are very much like traditional multimedia, which overwhelmingly engage but two of the human senses (vision and audition), the augmentation of audiovisual content with media targeting extra sensorial channels (i.e. mulsemedia) delivers not only a novel user experience but, bearing in mind that we experience the world in the plenitude of our five Aristotelian senses, a potentially more realistic and enhanced one as such. Moreover, given the fact that the majority of mulsemedia experiences are rendered at the client by a variety of extra-sensorial devices and initiated remotely through meta-data driven approaches (e.g. MPEG V), which are negligible in size compared to voluminous video, the comparative costs associated with 360° mulsemedia VR/AR are far outweighed by the potential benefits.

To explore this potential further, this paper reports the results of a user evaluation study which investigated the effects of 360° mulsemedia of varying encoding quality and dynamism/content motion on QoE. Accordingly, the structure of this paper is as follows. Section 2 details research related to our work, whilst Section 3 describes the concept of 360° mulsemedia as well as the design and implementation of a wearable 360° mulsemedia prototype, used in our empirical study. The methodology of the study is detailed in Section 4 and its results presented and analyzed in Section 5. Finally, concluding observations are drawn and opportunities for future work identified in Section 6.

2 BACKGROUND AND RELATED WORK

2.1 Why is QoE Important for QoS?

Improving the Quality of Service (QoS) of 360° video streaming is one of the main concerns of several studies that focus on metrics such as throughput, bandwidth, jitter, etc. [6, 12, 19]. However, it has become clear that for systems delivering audiovisual content, objective metrics like these fail to describe the perceptual QoS from the users' perspective (i.e. QoE). For instance, a fluctuating bitrate, even within allowable thresholds, can lead to changes in quality levels that annoy viewers.

QoS is concerned with interaction management between applications running in end-user terminals and networking conditions, whereas QoE characterises "the human side of the service provision and consumption" [16]. While controlling QoS parameters is important, there are a number of caveats which concern the sole focus on QoS assessment: (1) although QoE depends on the QoS delivered by the underlying networks, it also takes further aspects into account; (2) enhancing QoS can lead to increased operating costs for service providers and capital expenditures for network operators.

The advent of new consumer video technologies calls for putting the user in the centre of the investigation through a paradigm shift from QoS to QoE [35]. This user-centricity is also one of the main expectations from 5G networks that should understand the end-user's and the service's needs and consider them in performing personalised network management [18]. We believe that in the current context, where user expectations of video quality are steadily rising, QoE is a fundamental concept that balances the trade-off between user expectations, QoS and expenditures. Thus, it is important to understand how viewers perceive the viewing of 360° videos and what contributes to the QoE in these new interactive setups.

Little is known about QoE in the area of 360° video delivery, thus there is a need for new subjective assessment experiments where additional metrics should be considered. For instance, high motion levels in omnidirectional videos viewed through an HMD might be a significant contributor to QoE [40].

2.2 Mulsemedia QoE

Previous research investigated different ways of enhancing QoE for 360° videos and identified significant factors that influence its assessment: stalling patterns [33], average quality [14].

In [11], the authors propose a set of QoE metrics for adaptive 360° videos, however, they do not assess their validity through subjective evaluation, whilst [34] focuses on the integral quality experienced with different commercial HMDs and reports better results for HTC Vive. In [36], the authors propose a testbed for the evaluation of the perceived quality of 360° pictures.

However, we argue that not enough emphasis was put on the immersive dimension of omnidirectional videos. 360° videos provide a different type of media (not tied to traditional audiovisual formats), where users go beyond watching—they experience the content in an immersive way. The affordances given by a system play an important role in determining to what extent how users perceive matches what they perceive. Thus, stimulating multiple sensory dimensions might have an effect on the users' perception of quality.

Mulsemedia enables multisensory stimulation in digital environments [7] and previous studies showed its positive effects on QoE. Accordingly, in the sphere of olfactory-enhanced multimedia, research has shown that it leads to good user experience [2, 9, 38] and has benefits in terms of masking potential degradation in audio [3]. The influence of varying QoS parameters (e.g., network delay, jitter, buffer control) on QoE has also been explored in a mulsemedia context. Proposed solutions incorporated on top of audiovisual content, haptic [8, 15], olfactory [1, 20, 21] as well as combined haptic-olfactory stimuli [13, 22]. Stimulation of taste sensations in interactive systems has also known an active growth in recent years [24, 25, 28] and innovative applications have been developed for stimulating multisensory flavor experiences such as a digital lollipop [26] or a virtual cocktail [29].

While there is a large body of work on the impact of multisensory effects on the QoE of traditional audiovisual content, what happens when we stimulate various senses in 360° videos is much less understood. In work most closely related to ours, [27] presents a VR experience enhanced with olfactory and haptic effects and evaluate the impact of multisensory stimulation on the sense of presence. However, the influence of varying QoS factors (e.g. encoding quality) on QoE in 360° mulsemedia has, to the best of our knowledge, been unelucidated—this is precisely what we focus on in this paper. To this end, we have built an innovative prototype within the 360° mulsemedia arena and we now proceed to detail the two concepts.

3 360° MULSEMEDIA

In this section, we present the concept and implementation of 360° mulsemedia—our proposed solution to quality degradation as a result of bandwidth limitations. This is based on viewport-adaptive streaming for multisensory content.

3.1 Concept

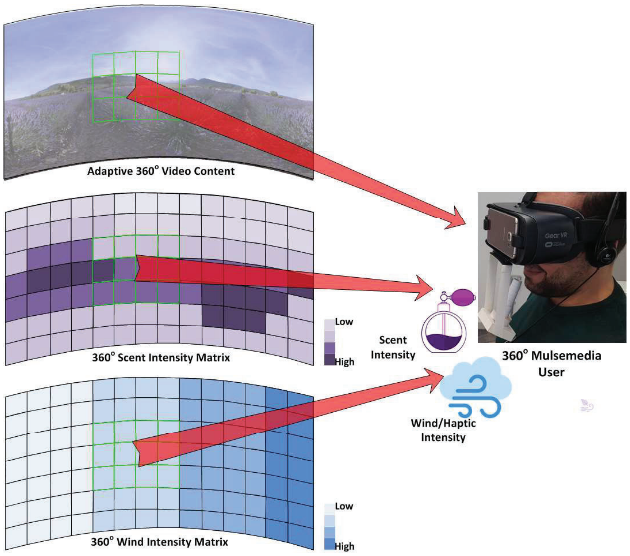

Figure 1: Conceptual 360° mulsemedia.

When delivering 360° video content over communications networks, the service providers are facing the problem of bandwidth waste. The HMD device displays only a fraction from the entire downloaded content which represents the user Field-of-View (FoV) or user viewport. Thus, it is bandwidth inefficient to transmit the entire 360° video content at the same quality since most of the downloaded scenes remain unused. On top of that, the entire system should react to head movements as fast as the HMD refresh rate (120Hz). Failing to achieve this requirement will lead to significant user QoE depreciation.

To make the streaming process possible for these bandwidth hungry and delay sensitive applications, the data format of 360° videos should be compatible with existing processing and transmission systems. In this sense, 360° videos are decomposed in specific geometric layouts where the spherical video is projected for processing by existing video encoding standards. Regardless of the chosen geometric layout (i.e. equirectangular panorama, cube map, pyramid, dodecahedron), the mapped video content must be adapted to the networking conditions in order to enhance user QoE and save bandwidth.

The main idea of the bandwidth reduction principle is to transmit at high quality the decomposed video inside FoV while the rest of the content is encoded at a much lower quality. Then, bandwidth adaptation techniques can be used to adapt the quality levels according to the momentary networking conditions. However, the data load associated with FoV-based video region can require significant bandwidth especially when network conditions are very favorable. Then, most of the techniques aim to reduce at minimum the area for which the video is encoded at higher quality. In this sense, the Quality Emphasized Region (QER) is defined being concentrated around the central axis of user viewport.

For better spatial separation between QER and the rest of the video content, the tiling approach is used in this sense. By considering the equirectangular decomposition as a study case, the rectangular panoramic video images are divided into smaller equirectangular tiles. An optimal number of tiles can be determined based on the trade-off between encoding efficiency, storage overhead and bandwidth utilization. Both encoders and decoders operate per frame basis in time domain and tile-by-tile in the spatial domain ensuring this way a full compatibility with more traditional video formats. At each frame, the tiles located inside of QER are transmitted at much higher quality than the rest. However, due to very fast head movements and poor networking conditions, the system may not react in the specified refreshing rate of HMD devices. As a result, the user QoE can be strongly affected while experiencing the displayed video content situated outwards of QER region.

One way to improve the user QoE under these circumstances with minimum bandwidth requirement is to use additional 360° sensory effects transmitted and synchronised with traditional 360° video content. However, the implementation of 360° mulsemedia concept requires two main aspects. First, the captured 360° multisensory information must follow the same geometric representation of 360° videos. Instead of video tiles, matrices of sensory intensities are considered for 360° multisensory representation. These matrices can be captured by using specialised sensory capturing devices for live streaming and annotation files for archived streaming. Second, the decomposed video tiles and matrices of intensities must be perfectly synchronized in both time and spatial domains. These involve the same sampling rates for 360° multisensory and video capturing devices. Additionally, at each video frame rate, each intensity level from each sensory matrix must be perfectly mapped with the corresponding video tile from the panoramic rectangular representation to deliver realistic sensory effects according to the user view-port position.

Figure 1 presents the concept of 360° mulsemedia where the user experiences 360° video in a free viewport VR system enhanced with olfactory and wind effects that are rendered by using specialised devices. For 360° video, only the tiles associated with QER are transmitted at higher quality, while the rest are encoded at much lower rates. As mentioned, due to fast head movements, users can experience low level quality of video tiles being located outwards of FoV region. We propose to overcome this fundamental drawback by multisensory stimulation. Thus, the matrices of olfactory and wind intensities are coded together with QER-based video content. Since the video tiles are synchronised with the intensity matrices, users experience 360° mulsemedia at the intensity corresponding to the viewport.

3.2 360° Mulsemedia Prototype Design

From the aforementioned 360° mulsemedia conceptual approach to a prototype, some issues should be taken into account. Systems' components such as audiovisual gears, a mulsemedia renderer, and olfactory and wind devices are physically and logically integrated.

However, whilst the concept is technologically feasible, for practical purposes the physical realisation of the prototype involves making an arrangement of components in such a way that they are well-connected to deliver the experience in immersive environments. To this end, we have adapted a VR goggle to support the attachment of an olfactory device and created an intensity-based fan to deliver granular wind level. Figure 2 displays the appliances and their integration. An Arduino (b) microcontroller was used to handle the wind fan (a) to blow air towards the user. A DFRobot Bluno Nano (f) microcontroller was employed to drive a scent emitter (d) that also blows wind from a Mini Dupont Brushless Cooling Fan placed within a small box further loaded with mesh bags filled with crystal scents (e) to deliver odor. The difference between Arduino and DFRobot Bluno Nano is that the latter is more compact and supports Bluetooth Low Energy which is useful to compound a wearable device. Both microcontrollers were managed by a mulsemedia renderer (i) without, at this stage, mapping sensory effects to tiles. The user was wearing a headphone (g) and an HMD Samsung Gear VR (c) to accommodate a smartphone (h). The smartphone was running a 360° VR application that reproduced equirectangular videos annotated with Sensory Effects Metadata (SEM) of the MPEG-V standard and communicated with the mulsemedia renderer to send the associated sensory effects metadata using a wireless local network provided by a WiFi router (j).

For the scent emitter, an innovative spiral conic pipe was exclusively developed. It conveys the scent directly towards the user's nose and steers clear of issues related to lingering effects which might spread the odor throughout the environment, and as a result, spoil the user's experience.

The mulsemedia renderer was used to logically integrate the 360° VR application to the devices. For this purpose, we used PlaySEM SER 2 [32] following the tutorial for building immersive 360° mulsemedia environments presented in [30]. Its responsibilities include SEM processing (conversion of abstract data representation to specific commands for sensory effects devices), device communication (taking into account emerging versatile computing equipment), and end-user applications integration (to be reused in different scenarios meeting different requirements). In other words, PlaySEM SER seamlessly interweaves applications and devices enabling a myriad of protocols and standards to be integrated without the need for supplementary coding as shown in the case studies described in [31, 32]. Therefore, whilst PlaySEM SER saved us time, our focus was directed on developing the 360° VR application in Unity to reproduce audiovisual content annotated with MPEG-V to be sent to the mulsemedia renderer, as well as on the development of devices.

4 METHODOLOGY

4.1 360° Dataset

For this experiment we used a dataset downloaded from YouTube. Since one of our aims is to investigate the impact of varying content motion on QoE, our dataset choice was determined by the availability of 360° videos with different degrees of dynamism as well as the videos lending themselves to be associated with olfactory and airflow effects. Accordingly, three video clips were chosen, each boasting a different level of camera and content motion, as detailed below:

- **Lavender field (*Static*)** - Camera position: fixed. Content: static—a meander through a field of lavender.
- **Coffee shop (*Semi-Dynamic*)** - Camera position: fixed. Content: semi-dynamic—a barista preparing a cappuccino.
- **Rollercoaster (*Dynamic*)** - Camera position: moving. Content: dynamic—background that moves with the camera located in the carriage of a rollercoaster.

Each of the videos had four encoding settings (HD, Full HD, 2.5K, 4K)—see Table 1. The three video streams of the sample dataset were encoded with H.264/MPEG-4 Part 10 and had the chroma location: left; and projection: equirectangular. The duration of each video sample was 60 seconds.

Table 1: Features of the sample videos in our dataset.

Video	Encoding Quality	Resolution	Frame Rate (fps)	Bitrate (mbps)
Lavender field  Static	HD Full HD 2.5K 4K	1280x720 1920x1024 2560x1400 3840x2048	25 25 25 25	2.2 5.2 8.4 15
Coffee shop  Semi-Dynamic	HD Full HD 2.5K 4K	1280x720 1920x1024 2560x1400 3840x2048	60 60 60 30	0.7 1.6 2.7 6.3
Rollercoaster  Dynamic	HD Full HD 2.5K 4K	1280x720 1920x1024 2560x1400 3840x2048	30 30 30 30	5.6 10.5 15.8 15.2

4.2 Mulsemedia Effects

As stated above, one of the considerations behind our choice of the three videos employed in our study was their suitability to be associated with olfactory and wind effects in order to create a 360° mulsemedia environment. Semantically-congruent scents were thus associated with the content of each of the three videos—lavender,

Figure 2: Prototype used in the 360° mulsemedia experiment. On the left side, its components: (a) Wind blower fan, (b) Arduino Uno, (c) Samsung Gear VR headset, (d) Scent emitter, (e) Mesh bags with scent crystals, (f) DFRobot Bluno Nano, (g) Headphone, (h) Samsung Galaxy S6 smartphone, (i) Laptop running a mulsemedia renderer, and (j) WiFi router. On the right side, a user in the environment with integrated components.

coffee, and diesel, respectively used with the *Lavender field*, *Coffee shop* and *Rollercoaster* clips. Whilst in the case of lavender and coffee scents their use is self-evident, in that of the diesel scent, justification lies in the mechanical association between this particular scent (reminiscent of lubricant aromas coupled with mildly pungent accents associated with burnt rubber) and the video content experienced. Additionally, wind effects were employed in the clips to simulate the wind in a user's face when going downhill in the rollercoaster ride, the wind breeze as one wanders through the lavender field, and the puff of air made by the coffee machine when the barista steamed the milk in order to prepare the cappuccino. The timings of the release of these effects in the context of the three videos are given in Figure 3.

Figure 3: Videos' annotation in seconds. W - Wind, S - Scent.

4.3 Participants

In total, 48 participants (27 male, 21 female) aged from 16 to 65 participated in our experiment. They were recruited through bulletin boards and email advertising from three UK universities: *Brunel*, *Middlesex*, and *University of Kent*. Participants did not receive any monetary compensation. The study had obtained ethical clearance from all three universities.

4.4 Experimental Design

Our study had three independent variables: mulsemedia (with two levels—with and without), video clip encoding quality (four levels—HD, Full HD, 2.5K, and 4K), and clip dynamism (three levels—static, semi-dynamic, and dynamic). The dependent variable was the self-reported QoE, as shall be described in Section 4.5. Accordingly, we employed a mixed factorial design in our study, in which mulsemedia and clip dynamism were within-subjects variables (each user experienced all clips both with and without mulsemedia), whilst encoding quality was a between-subject variable (each user experienced the video clips with only one particular video clip encoding quality).

Two measures were taken in the experimental design so as to avoid order effects. The first targeted the order in which users experienced wind and olfactory effects. Accordingly, the first 6 users in our study experienced each of the 360° videos with audiovisual stimuli only (360° multimedia), after which they experienced the same videos with olfactory and wind stimuli (360° mulsemedia). The next batch of 6 users then experienced the 360° videos in the opposite order—360° mulsemedia followed by 360° multimedia. This pattern was then repeated for all 48 users of our study, and Table 2 details the clips' viewing order for the first 6 participants. The second measure taken to counteract any potential order effects was that the order of presentation of videos was cyclically varied, ensuring that the number of times each video was seen first, second or third was exactly the same for all videos.

A final measure was taken in order to avoid recency effects. To this end, within each condition (multimedia/mulsemedia) participants watched the videos in the same respective order. This ensured that between watching a particular video clip in one condition (e.g. 360° *Lavender field* multimedia) and its equivalent in the other condition (i.e. 360° *Lavender field* mulsemedia) exactly two other 360° video clips would always have been experienced by all participants.

4.5 QoE Questionnaire

The QoE questionnaire comprised a series of questions targeting the user experience. The response to each question was expressed

Table 2: Viewing order for the first 6 participants.

	360° multimedia			360° mulsemedia		
User 1	Lavender field	Coffee Shop	Rollercoaster	Lavender field	Coffee Shop	Rollercoaster
User 2	Rollercoaster	Lavender field	Coffee Shop	Rollercoaster	Lavender field	Coffee Shop
User 3	Coffee Shop	Rollercoaster	Lavender field	Coffee Shop	Rollercoaster	Lavender field
User 4	Lavender field	Rollercoaster	Coffee Shop	Lavender field	Rollercoaster	Coffee Shop
User 5	Coffee Shop	Lavender field	Rollercoaster	Coffee Shop	Lavender field	Rollercoaster
User 6	Rollercoaster	Coffee Shop	Lavender field	Rollercoaster	Coffee Shop	Lavender field

on a 5 point Likert scale, as detailed below. Accordingly, for all video clips participants answered the following three questions:

(1) *Please rate the overall quality of the video clip.*
{Bad, Poor, Fair, Good, Excellent}.
(2) *The quality of the visual display is appropriate:*
{Strongly Agree, Agree, Neutral, Disagree, Strongly Disagree}.
(3) *I enjoyed the 360° experience:*
{Strongly Agree, Agree, Neutral, Disagree, Strongly Disagree}.

Additionally, in the *with mulsemedia* condition, participants also answered a set of 8 more questions targeting olfactory and wind effects. For brevity, only the olfactory-related questions are detailed below, as the questions targeting airflow are analogous. All were answered in a {Strongly Agree, Agree, Neutral, Disagree, Strongly Disagree} scale.

(1) *The olfactory effects enhance the sense of reality.*
(2) *The olfactory effects are distracting.*
(3) *The olfactory effects are annoying.*
(4) *I enjoy watching the video with olfactory effects.*

4.6 Procedure

The study was carried out in dedicated laboratories at each of the three sites and lasted for approximately 30 minutes per participant. At the outset, participants were explained the procedure and tasks involved in the experiment. Moreover, they were also reminded that their participation was voluntary and that they could withdraw at any stage without giving any reasons for doing so. They were then asked if they had any questions and, once such questions (if any) had been clarified, they proceeded to fill in an online demographic questionnaire. Once this had been done, participants were then given the 360° mulsemedia prototype to put on; when they confirmed that they were comfortable and satisfied with the whole setup, they then continued to view the 360° video clips.

After watching each video, participants filled in the QoE questionnaire for that clip. If a participant was watching the clip in the *with mulsemedia* condition, whilst they were completing the QoE questionnaire the scents were changed in the 360° mulsemedia prototype by the experimenters. Moreover, this elapsed time interval also allowed for any lingering scents to dissipate. Once a participant had finished their run, they were thanked for their efforts. The minimum time between two different participant runs was 10 minutes, during which the laboratory windows were opened—a final measure to ensure the elimination of any lingering scents.

5 RESULTS

In this section, we investigate the impact of providing multisensory effects in a 360° video experience. More specifically, we look into how 360° mulsemedia impacts the perceived quality, the realism, annoyance, and distraction factors as well as the overall QoE.

5.1 Do Multisensory Effects Impact User-Perceived Quality of 360° Videos?

In order to answer this question, we asked the participants to rate the overall quality of the 360° experience on a 1 to 5 scale (from Bad to Excellent). We computed the Mean Opinion Score (MOS) for each encoding quality. Results are illustrated in Figure 4. It is remarkable that the user-perceived quality for 360° mulsemedia scored better (overall MOS = 3.81) than in 360° multimedia (overall MOS = 3.41), irrespective of the underlying encoding quality.

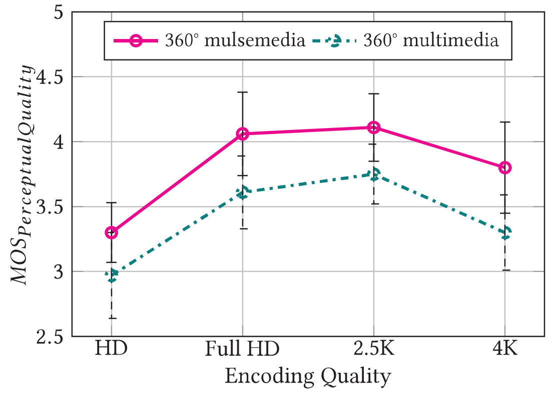

Figure 4: Impact of the different encoding qualities on the user-perceived quality of 360° videos in the presence (360° mulsemedia) or absence (360° multimedia) of multisensory content (5=Excellent, 1=Bad; Error bars represent 95% CI).

For a deeper analysis of the influence of different effects on the quality rating, a two-way independent-measures ANOVA (content-type: two levels—360° multimedia and 360° mulsemedia; encoding quality: 4 levels—HD, Full HD, 2.5K, 4K) was performed on these data. This showed a significant main effect of the encoding quality ($F(3, 280) = 114.6$, $p < 0.001$). Post-hoc Bonferroni tests revealed that the perceived overall quality was similar for Full HD and 2.5K, $Mean_d = 0.092$, p = 1. However, the perceived overall quality for HD is significantly lower compared to Full HD ($Mean_d = 0.69$), 2.5K ($Mean_d = 0.79$), $p < 0.001$. The statistical analysis indicates also that multisensory content has a significant main effect on the MOS ($F(1, 280) = 148.4$, $p < 0.001$): regardless of the encoding quality, the perceived quality was higher for 360° mulsemedia than for 360° multimedia content. The ANOVA does not reveal any significant interaction between encoding quality and the presence of multisensory content ($F(3, 280) = 0.1$, *ns*). This is indicated also in Figure 4, where across all four encoding qualities, the effects of experiencing 360° mulsemedia and multimedia are similarly marked.

What is remarkable is the unexpected decrease in MOS for 360° 4K videos observed in Figure 4 compared to its 2.5K counterparts. This alludes to the fact that playing 4K videos might be redundant, since there is no improvement in user-perceived quality, but on the contrary a deterioration. As remarked above, this characteristic is also similar for the 360° multimedia results. Whilst the underlying reasons might have to do with hardware display limitations, our results nonetheless indicate that overall quality is evaluated approximately 12% higher in the case of 360° mulsemedia compared to 360° multimedia for all the considered resolutions.

5.2 Does Video Dynamism Impact User-Perceived Quality and Enjoyment?

When experiencing free-viewpoint 360° videos, motion magnitude can have negative effects such as general discomfort, headaches, fatigue, and disorientation. These can be the result of rendering distortions, the vergence-accommodation effect of delays between the user's movements and the updating of the virtual scene. Previous research has shown that observers tend to experience simulator sickens when the exposure to the VR scenario induces significant amounts of vection [4, 17].

The motion level of the VR video content is one of the most important factors in determining the overall degree of physical discomfort [17]. This VR sickness experienced by the user can trigger degradation of the perceptual quality and enjoyment. In order to assess the impact of different motion levels on the quality assessment and overall experience in the 360° video setup, we performed a two-way independent-measures ANOVA (motion level and clip dynamism) on the data. As indicated by the Figure 5, the presence of multisensory content has a significant main effect on both perceived quality and enjoyment ($F(2, 282) = 78.23$, $p < 0.05$; ($F(2, 282) = 236.26$, $p < 0.005$)). The subjective assessment of these factors was significantly higher for the 360° mulsemedia condition. Clip motion level had a significant effect on the enjoyment factor ($F(2, 282) = 28.89$, $p < 0.05$), but not on perceived quality ($F(2, 282) = 2.12$, $p > 0.05$). Tests did not reveal any significant interaction effect.

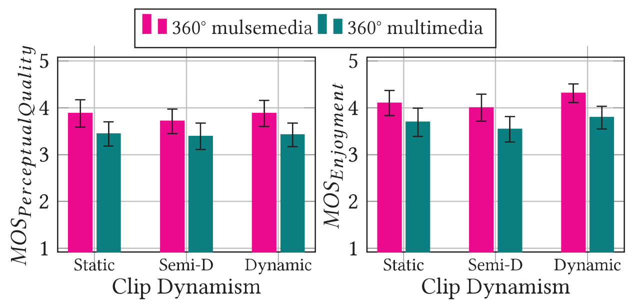

Figure 5: Impact of the motion level on the perceived quality (left) and enjoyment (right) in the absence and presence of multisensory effects (5 = Excellent; 1 = Bad).

5.3 Do Multisensory Effects Enhance the Sense of Reality in 360° Mulsemedia?

Figure 6 depicts the impact of different encoding qualities (left) and clip dynamism degrees (right) on the subjective realism assessment. For both types of multisensory effects (olfactory and airflow), there are no significant differences between the ratings of realism under different encoding qualities and users were generally enthusiastic about the 360 ° mulsemedia experience. This suggests that even if there is a significant difference in the perceived quality between HD and the other encoding qualities, the degree of realism is maintained by the additional stimuli that become part of the experience. These results are in line with investigations on traditional mulsemedia, where users cannot control the viewport and the direction of sensory effects [9, 38].

The impact of the content-type turns out to be less relevant than expected (see Figure 6). MOS results do not differ significantly between the three omnidirectional videos that feature different levels of camera and content motion. The video dynamism did not influence the realism added by the airflow ($F(2, 141) = 2.90$, $p = 0.058$) or olfaction effects ($F(2, 141) = 1.03$, $p = 0.36$).

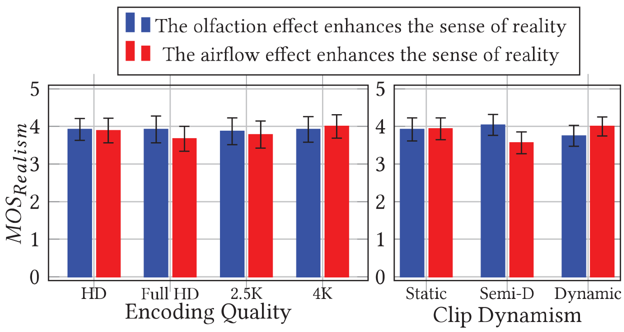

Figure 6: Impact of encoding quality (left) and video type (right) on the realism score in the presence of olfaction and airflow effects (5 = Excellent; 1 = Bad).

5.4 Do Multisensory Effects Annoy Users in 360° Mulsemedia?

Figure 7 depicts how different encoding qualities and motion levels impact the user annoyance in the 360° mulsemedia condition. Whilst some users did remark on the synthetic nature of smells, MOS values are roughly similar across conditions (4, corresponding to Disagree) indicating that multisensory effects do not cause annoyance independently of resolution and clip dynamism.

5.5 Do Multisensory Effects Distract Users in 360° Mulsemedia?

When it comes to distraction, we observe a similar behaviour as in the case of annoyance. MOS values are presented in Figure 8 and show that users disagree with finding the multisensory effects as sources of distraction. A one-way independent-measures ANOVA showed no significant difference between the level of distraction in different encoding conditions or clip dynamism conditions. However, it seems that when the multisensory content contributes more to the added realism (e.g., the airflow effect in the Rollercoaster video), users reported less distraction (MOS = 4.21).

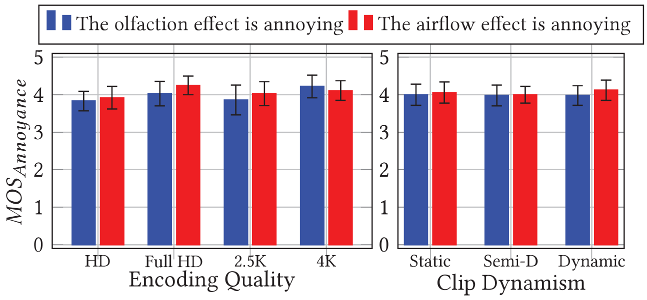

Figure 7: Impact of encoding quality (left) and video type (right) on subjective assessment ratings of annoyance (5 = Strongly Disagree; 1 = Strongly Agree).

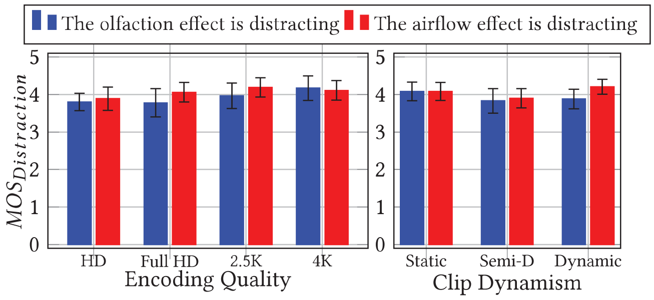

Figure 8: Impact of encoding quality (left) and video type (right) on subjective assessment ratings of distraction (5 = Strongly Disagree; 1 = Strongly Agree).

5.6 Do Multisensory Effects Enhance User Enjoyment of 360° Videos?

A two-way independent-measures ANOVA (encoding quality and content-type) was performed on answers assessing the enjoyment during the 360° experience. Results showed a significant increase of enjoyment in the presence of multisensory effects ($F(3, 280) = 17.84$, $p < 0.05$). This is visible also in Figure 9, where the MOS reflecting participants' enjoyment of the 360° mulsemedia are significantly higher than for 360° multimedia (overall MOS = 4.14 vs 3.67), irrespective of the underlying encoding quality of the video.

The same test indicated no significant difference in the enjoyment assessment depending on encoding qualities. Despite this, MOS values across qualities show a difference in how participants enjoyed the experience at different resolutions: for HD—MOS = 3.57, while for 2.5K—MOS = 4.18. These values suggest that multisensory content is masking the degradation of quality and contributes to an enhancement enjoyment.

6 CONCLUDING DISCUSSION

In this paper, we have presented the results of a study which explored the impact of mulsemedia in 360° videos. Our findings make a compelling case for the inclusion of multisensory effects in 360° environments, since results show that there there is a statistically significant improvement in both user-perceived quality as well as enjoyment when mulsemedia (as opposed to traditional multimedia) is employed. Additionally, we make the following observations:

Observation 1: Our results show that irrespective of the video quality, the user-perceived overall quality—as well as self-reported

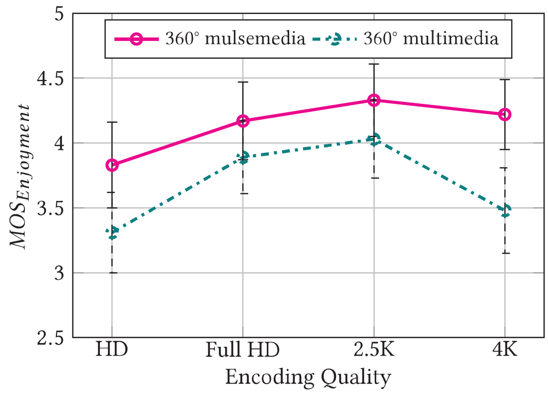

Figure 9: Impact of the different encoding qualities on the user enjoyment of 360° videos in the presence (360° mulsemedia) or absence (360° multimedia) of multisensory content (5=Excellent, 1=Bad; Error bars represent 95% CI).

enjoyment—of the omnidirectional videos was higher when multisensory effects were employed (360° mulsemedia). This indicates that multiple sensory cues mask quality degradation and can be employed in a resource-saving process. Airflow and olfactory effects boost the overall perceived quality of 360° videos by 12% and enjoyment by 13% across the four encoding qualities we tested.

Observation 2: Whilst the improvement in user-perceived quality from HD to Full HD is evident and significant, the gain in the case of the transition from Full HD to 2.5K is marginal and downright deteriorates in the case of 4K video. This shows that 2.5K video is the quality level to use for 360° mulsemedia (and multimedia) applications, should there be enough bandwidth on the network. However, should the network experience congestion one can reduce video quality to Full HD without negatively affecting the user experience. This has important implications in network resource allocation for it highlights the fact that an important QoS parameter—encoding quality—can be reduced in 360° mulsemedia to Full HD without any detrimental impact on the QoE.

Observation 3: Clip motion level significantly affects user enjoyment of 360° videos in the case of mulsemedia, but not so in respect of user-perceived quality. This highlights the importance of the purpose for which one experiences 360° mulsemedia in QoE. Accordingly, if the 360° viewing experience is primarily for entertainment purposes (e.g. VR games) then designers have to carefully reflect on the dynamism of the created content; this is especially important given the verifiable link in VR between user enjoyment and motion sickness [17].

Summarising—mulsemedia has been shown to have a great potential in enhancing QoE in 360° videos compared with traditional multimedia. Whilst this might be, to some extent, expected, what is totally surprising and has been borne out by the results of our study is that the best 360° video user-perceived quality and enjoyment are experienced at the relatively low encoding rates of Full HD and 2.5K—4K being shown to be overkill. Reasons for this might be because of FoV limitations of the HMD used [23], or indeed the VR screen-door effect [37]; all are worthy future pursuits, as is the investigation of head movement and adaptive QER (Quality Emphasized Regions) on 360° video QoE.

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. 2009. Synchronization of olfaction-enhanced multimedia. *IEEE Transactions on Multimedia* 11, 3 (2009), 561–565.

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

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

[4] Manuela Chessa, Guido Maiello, Alessia Borsari, and Peter J. Bex. 2019. The perceptual quality of the oculus rift for immersive virtual reality. *Human–computer interaction* 34, 1 (2019), 51–82.

[5] VNI Cisco. 2018. Cisco Visual Networking Index: Forecast and Trends, 2017–2022.

[6] Ioan-Sorin Comsa, Ramona Trestian, and Gheorghita Ghinea. 2018. 360° mulsemedia experience over next generation wireless networks-a reinforcement learning approach. In *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 1–6.

[7] Alexandra Covaci, Longhao Zou, Irina Tal, Gabriel-Miro Muntean, and Gheorghita Ghinea. 2018. Is Multimedia Multisensorial? - A Review of Mulsemedia Systems. *ACM Comput. Surv.* 51, 5, Article 91 (Sept. 2018), 35 pages. https://doi.org/10.1145/3233774

[8] Fabien Danieau, Julien Fleureau, Audrey Cabec, Paul Kerbiriou, Philippe Guillotel, Nicolas Mollet, Marc Christie, and Anatole Lécuyer. 2012. Framework for enhancing video viewing experience with haptic effects of motion. In *2012 IEEE Haptics Symposium (HAPTICS)*. IEEE, 541–546.

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

[10] Gheorghita Ghinea, Christian Timmerer, Weisi Lin, and Stephen R Gulliver. 2014. Mulsemedia: State of the art, perspectives, and challenges. *ACM Trans. Multimedia Comput. Commun. Appl.* 11, 1s (2014), 17.

[11] Arnob Ghosh, Vaneet Aggarwal, and Feng Qian. 2017. A Rate Adaptation Algorithm for Tile-based 360-degree Video Streaming. *CoRR* abs/1704.08215 (2017).

[12] Mario Graf, Christian Timmerer, and Christopher Mueller. 2017. Towards Bandwidth Efficient Adaptive Streaming of Omnidirectional Video over HTTP: Design, Implementation, and Evaluation. In *Proceedings of the 8th ACM on Multimedia Systems Conference (MMSys'17)*. ACM, New York, NY, USA, 261–271. https://doi.org/10.1145/3083187.3084016

[13] Sosuke Hoshino, Yutaka Ishibashi, Norishige Fukushima, and Shinji Sugawara. 2011. QoE assessment in olfactory and haptic media transmission: Influence of inter-stream synchronization error. In *2011 IEEE International Workshop Technical Committee on Communications Quality and Reliability (CQR)*. IEEE, 1–6.

[14] T. Hoßfeld, M. Seufert, C. Sieber, and T. Zinner. 2014. Assessing effect sizes of influence factors towards a QoE model for HTTP adaptive streaming. In *2014 Sixth International Workshop on Quality of Multimedia Experience (QoMEX)*. 111–116. https://doi.org/10.1109/QoMEX.2014.6982305

[15] Kai Iwata, Yutaka Ishibashi, Norishige Fukushima, and Shinji Sugawara. 2010. QoE assessment in haptic media, sound, and video transmission: Effect of playout buffering control. *Computers in Entertainment (CIE)* 8, 2 (2010), 12.

[16] Kalevi Kilkki. 2008. Quality of experience in communications ecosystem. *J. UCS* 14, 5 (2008), 615–624.

[17] Hak Gu Kim, Wissam J Baddar, Heoun-taek Lim, Hyunwook Jeong, and Yong Man Ro. 2017. Measurement of exceptional motion in VR video contents for VR sickness assessment using deep convolutional autoencoder. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*. ACM, 36.

[18] Eirini Liotou, Hisham Elshaer, Raimund Schatz, Ralf Irmer, Mischa Dohler, Nikos Passas, and Lazaros Merakos. 2015. Shaping QoE in the 5G ecosystem. In *Quality of Multimedia Experience (QoMEX), 2015 Seventh International Workshop on*. IEEE, 1–6.

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

[20] Niall Murray, Brian Lee, Yuansong Qiao, and Gabriel-Miro Muntean. 2014. Multiple-scent enhanced multimedia synchronization. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 11, 1s (2014), 12.

[21] Niall Murray, Yuansong Qiao, Brian Lee, A. K. Karunakar, and Gabriel-Miro Muntean. 2013. Subjective evaluation of olfactory and visual media synchronization. In *Proceedings of the 4th ACM Multimedia Systems Conference*. ACM, 162–171.

[22] Seiya Nakano, Yoshihiro Maeda, Yutaka Ishibashi, Norishige Fukushima, Pingguo Huang, and K Psannis. 2015. Influence of network delay on fairness between players in networked game with olfactory and haptic senses. *IEICE Technical Report, CQ2014-94* (2015).

[23] NVIDIA and GEFORCE. Accessed: February 27, 2019. [Online]. Available: https://forums.geforce.com/default/topic/831444/don-t-do-4k-on-a-tv-field-of-view-affects-video-games-/. Technical Support for GeForce Technologies.

[24] Nimesha Ranasinghe, Adrian Cheok, Ryohei Nakatsu, and Ellen Yi-Luen Do. 2013. Simulating the sensation of taste for immersive experiences. In *Proceedings of the 2013 ACM international workshop on Immersive media experiences*. ACM, 29–34.

[25] Nimesha Ranasinghe, Adrian David Cheok, and Ryohei Nakatsu. 2012. Taste/IP: the sensation of taste for digital communication. In *Proceedings of the 14th ACM international conference on Multimodal interaction*. ACM, 409–416.

[26] Nimesha Ranasinghe and Ellen Yi-Luen Do. 2017. Digital lollipop: Studying electrical stimulation on the human tongue to simulate taste sensations. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 13, 1 (2017), 5.

[27] 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, Ching Chiuan Yen, and Ellen Yi-Luen Do. 2018. Season Traveller: Multisensory Narration for Enhancing the Virtual Reality Experience. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 577, 13 pages. https://doi.org/10.1145/3173574.3174151

[28] Nimesha Ranasinghe, Kasun Karunanayaka, Adrian David Cheok, Owen Noel Newton Fernando, Hideaki Nii, and Ponnampalam Gopalakrishnakone. 2011. Digital taste and smell communication. In *Proceedings of the 6th international conference on body area networks*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, 78–84.

[29] Nimesha Ranasinghe, Thi Ngoc Tram Nguyen, Yan Liangkun, Lien-Ya Lin, David Tolley, and Ellen Yi-Luen Do. 2017. Vocktail: A virtual cocktail for pairing digital taste, smell, and color sensations. In *Proceedings of the 25th ACM international conference on Multimedia*. ACM, 1139–1147.

[30] Estêvão B. Saleme, Alexandra Covaci, Gebremariam Mesfin, Celso A. S. Santos, and Gheorghita Ghinea. 2019. Mulsemedia DIY: A Survey of Devices and a Tutorial for Building Your Own Mulsemedia Environment. *ACM Comput. Surv.* 52, 3, Article 58 (June 2019), 29 pages. https://doi.org/10.1145/3319853

[31] E. B. Saleme, C. A. S. Santos, and G. Ghinea. 2019. Coping With the Challenges of Delivering Multiple Sensorial Media. *IEEE MultiMedia* 26, 2 (April 2019), 66–75. https://doi.org/10.1109/MMUL.2018.2873565

[32] Estêvão Bissoli Saleme, Celso A. S. Santos, and Gheorghita Ghinea. 2019. A mulsemedia framework for delivering sensory effects to heterogeneous systems. *Multimedia Systems* 25, 4 (01 Aug 2019), 421–447. https://doi.org/10.1007/s00530-019-00618-8

[33] R. Schatz, A. Sackl, C. Timmerer, and B. Gardlo. 2017. Towards subjective quality of experience assessment for omnidirectional video streaming. In *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*. 1–6. https://doi.org/10.1109/QoMEX.2017.7965657

[34] Ashutosh Singla, Stephan Fremerey, Werner Robitza, and Alexander Raake. 2017. Measuring and comparing QoE and simulator sickness of omnidirectional videos in different head mounted displays. In *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 1–6.

[35] Ramona Trestian, Ioan-Sorin Comsa, and Mehmet Fatih Tuysuz. 2018. Seamless multimedia delivery within a heterogeneous wireless networks environment: Are we there yet? *IEEE Communications Surveys & Tutorials* 20, 2 (2018), 945–977.

[36] E. Upenik, M. Rerabek, and T. Ebrahimi. 2016. Testbed for subjective evaluation of omnidirectional visual content. In *2016 Picture Coding Symposium (PCS)*. 1–5. https://doi.org/10.1109/PCS.2016.7906378

[37] VRheads. Accessed: March 22, 2019. [Online]. Available: https://www.vrheads.com/what-screen-door-effect-and-why-does-it-happen. What is the "screen-door effect" and why does it happen?

[38] Zhenhui Yuan, Shengyang Chen, Gheorghita Ghinea, and Gabriel-Miro Muntean. 2014. User Quality of Experience of Mulsemedia Applications. *ACM Trans. Multimedia Comput. Commun. Appl.* 11, 1s, Article 15 (Oct. 2014), 19 pages. https://doi.org/10.1145/2661329

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

[40] M. Zink, R. Sitaraman, and K. Nahrstedt. 2019. Scalable 360° Video Stream Delivery: Challenges, Solutions, and Opportunities. *Proc. IEEE* (2019), 1–12. https://doi.org/10.1109/JPROC.2019.2894817