Do I Smell Coffee? The Tale of a 360° Mulsemedia Experience

Ioan-Sorin Comşa
Brunel University London

Estêvão Bissoli Saleme
Federal University of Espírito Santo

Alexandra Covaci
University of Kent

Gebremariam Mesfin Assres
Brunel University London

Ramona Trestian
Middlesex University London

Celso A. S. Santos
Federal University of Espírito Santo

Gheorghiţă Ghinea
Brunel University London

***Abstract*—One of the main challenges in current multimedia networking environments is to find solutions to help accommodate the next generation of mobile application classes with stringent Quality of Service (QoS) requirements while enabling the Quality of Experience (QoE) provisioning for users. One such application class, featured in this article, is 360° mulsemedia—*multiple sensorial media*—which enriches 360° video by adding sensory effects that stimulate human senses beyond those of sight and hearing, such as the tactile and olfactory ones. In this article, we present a conceptual framework for 360° mulsemedia delivery and a 360° mulsemedia-based prototype that enables users to experience 360° mulsemedia content. User evaluations revealed that higher video resolutions do not necessarily lead to the highest QoE levels in our experimental setup. Therefore, bandwidth savings can be leveraged with no detrimental impact on QoE.**

Digital Object Identifier 10.1109/MMUL.2019.2954405
Date of publication 20 November 2019; date of current version 27 March 2020.

In recent years, innovations in electronics have led to a significant increase in demand for new services and use cases. Future wireless networks are expected to enable ubiquitous connectivity for users, machines, and devices and provide support for an unimaginable exchange of data and multimedia content. In such heterogeneous multiuser, multimachine, multidevice, multitechnology, and multiapplication environments, QoE is foreseen to become the main differentiator between network operators and content providers.[1,2]

Given current advances, the shift from content-delivery applications to deftness-delivery applications becomes more popular. According to this new paradigm, human–machine interaction meets a new technological dimension that enables real-time tactile control of remote objects. So-called tactile Internet applications find their usefulness in many domains such as health care, sports, entertainment, gaming, or smart grids.[3] Providing an enhanced QoE for mulsemedia applications constitutes an important challenge since ultrareliable and ultraresponsive connectivity is required. This puts unprecedented pressure on the underlying networks, where reaction times for haptic interactions should be swift.[4]

In addition to haptic communications, the tactile Internet also integrates audio and visual feedback to increase user-perceived quality.[5] In this multisensory, multiinformation environment, subjective QoE evaluations can be carried out by involving human subjects in order to test, for instance, the user-perceived effect of the speed of system responsiveness, video quality degradation, or masking of audio feedback in different application domains. While subjective QoE evaluations can be laborious and time consuming, QoS parameters related more to the communication system performance can be used to compute objective QoE metrics, which approximate the quality of the auditory, visual, and/or haptic feedback in the tactile Internet. While objective QoE metrics have been questioned with respect to their correlation with actual human judgments, it is foreseen that high performance of objective QoE for tactile Internet applications can be obtained if the following challenges are addressed: network slicing, reduced air-interface latency, intelligent radio resource management, and edge computing.[5]

Besides tactile Internet applications, virtual reality (VR), augmented reality (AR) and omnidirectional, or 360° video could also benefit from the availability of high-performance network infrastructures. However, these applications require massive computational capability, high bandwidth communications, and ultra-low end-to-end latency. In interactive 360° video streaming, 360° views of captured scenes can be explored according to the user viewport position. Here, the entire VR system must respond in less than 10 ms. Failing to meet this requirement due to overheads associated with factors such as encoding and compression, fast head movement or unfavorable network conditions, will result in users having a poor QoE. To alleviate this, some approaches adopt motion-based predictive models to improve both subjective and objective QoE evaluations of 360° video applications.[6,7]

Much akin to the case of the tactile Internet, 360° video applications could also benefit from the same principle to improve the overall QoE performance, regardless of other factors such as dynamic network conditions. In this context, we introduce a new application type entitled 360° mulsemedia.[8] In these applications, the 360° media is enriched with additional sensory effects such as haptic/wind, heat, or smell. Being equipped with VR headsets, users can experience different intensities of wind, heat, and scent from the captured scenes according to their viewport positions. In traditional mulsemedia systems, olfactory and wind capabilities are able to strongly improve the perceived sense of reality when experiencing conventional video content.[9,10] With this strong premise, it is foreseen that 360° mulsemedia could revolutionize the streaming technology and enhance the user-perceived quality even when the quality of the 360° video is degraded due to different factors.

In this article, we introduce a new conceptual framework for 360° mulsemedia delivery over networks. At the server-side, the 360° video content is enhanced with multisensory information like wind and scent and transmitted over the radio interface. At the user side, a prototype is built to playback the 360° mulsemedia content. The proposed prototype is validated through subjective tests. Furthermore, the impact of different 360° mulsemedia content is analyzed to find out whether 360° video quality variations, when presented with sensory effects, affect the user-perceived quality.

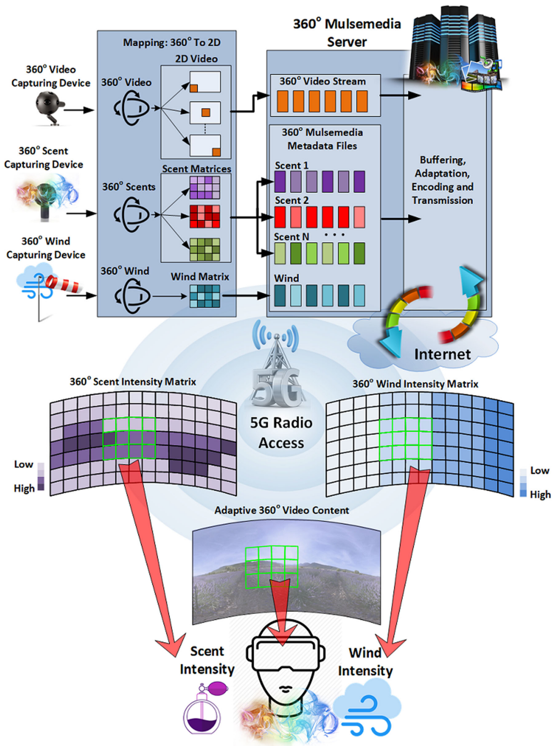

Figure 1. 360° Mulsemedia Concept.

360° MULSEMEDIA SYSTEM

The 360° mulsemedia delivery system consists of three main entities: 1) server-side, 2) network delivery side, and 3) the end-user side as depicted in Figure 1. At the server-side, the 360° mulsemedia content is captured, stored, and adapted according to the networking conditions and delivered to end-users. The 360° multisensory information is encoded by specialized devices and synchronized with 360° video scenes in both temporal and spatial dimensions. An alternative would be to derive the scent and wind effects based only on the omnidirectional video content, but this approach would involve extra image processing overhead.

Spatially, 360° video content must be decomposed in geometric layouts to be optimally processed by encoders. 360° mulsemedia can be implemented on any type of 360° video decomposition layouts such as equirectangular, pyramidal, cube map, or rhombic dodecahedron.[11] However, in this article, we opt for the equirectangular decomposition to help the reader to better understand the concept. According to this principle, 360° video frames are decomposed in larger equirectangular panoramic images. The proposed concept is flexible and works within a wired plus wireless environment. However, the user access side is assumed to be wireless in order to enable the VR connectivity. To reduce the bandwidth utilization of such large video content, for transmission over both wired and wireless networks, the tiling approach is mostly used in which case the panoramic images are spatially divided into small equirectangular tiles. Given the fact that most of the existing 2-D video encoders are operating on the equirectangular scales,[11] it is very efficient to use them for panoramic images. Since these encoders operate tile-by-tile, an increased scalability is foreseen to be obtained when varying the tiling granularity. The tiling granularity is supposed to be high enough to cover a wide range of user viewports. An optimal number of tiles must be determined for the video equirectangular representation in order to increase the encoding efficiency and minimize the storage overhead and bandwidth utilization. A study conducted by Graf *et al.*[12] investigated the use of different tiling patterns (columns x rows), such as 1×1, 3 x 2, 5 x 3, 6 x 4, and 8 x 5 for various video resolutions (e.g., 1920 x 960, 3840 x 1920, and 7680 x 3840). The authors conclude that the tiling pattern 6x4 provides the best view-port selection flexibility versus bitrate overhead versus bandwidth tradeoff. The associated 360° multisensory effects could follow the same layout and tiling procedure to enhance the user-perceived quality based on the experienced viewports.

360° Mulsemedia: Concept, Capturing, and Mapping

Alongside video tiles, 360° mulsemedia considers additional information regarding the types and intensities of the sensorial components of each video tile. An example is shown in Figure 1, where the wind and different scents are considered to be infused from the real world to improve the user experience. Conceptually, for each 360° video frame, a matrix of intensities for each individual sensory effect (wind and many scents, in this case) must be captured. The structures of matrices for wind and scent sensory effects (number of columns and rows) must be identical with the structure of the video matrix. The intensity values of scent and wind matrices represent the captured intensities of each sensory device which makes up the 360° sensory device.

Although in Figure 1 only wind and scents are considered as multisensory effects, the proposed system can be extended to incorporate other sensory effects such as vibration, heat, or lighting. Vibration can be correlated with the wind intensities, whereas heat and lighting can be derived based on the omnidirectional video content. For example, for a sudden increase of pixels intensity in a certain region of the omnidirectional video, the lighting effect can be activated according to the current user field of view (FoV).

By using 360° anemometers, the wind directions and intensities can be appropriately captured. As far as scents capture is concerned, one can utilize a specialized odor sensing system employing the principles of scentography. Here, a scent is captured by using a special type of sensor—a Quartz Crystal Microbalance (QCM)—coated with sensing films that are considered stationary phase materials for gas chromatography.[13] For a multiscent capture, several QCMs are needed for each odor and a machine learning-based system is needed to properly detect each scent intensity. The detection system is learned to give the right intensity of each captured scent based on *a priori* determined calibration curves. When recording 360° scents, an array of sensing systems is used to represent the intensity levels for each video tile.

The 360° intensity matrices for multisensory content must be harmonized in both time and spatial domains with the 360° video content. In the example, the capturing devices for both wind and scents sensory objects must work at the same sampling rate as and in sync with the sampling process of the 360° video and audio content. Conceptually, when capturing the multisensory effects, each device has its own sampling frequency. Sensory data can be multiplexed with the video content in different data streams, by using MPEG-V TS, for example.

360° Mulsemedia Server

For bandwidth management reasons, FoV is defined for each user device, comprising the tiles corresponding to the viewport together with its neighboring tiles, all of which are transmitted with higher quality than the remaining tiles. As shown in Figure 1, the server can have many representations of FoV-based panoramic video according to the viewport position of each end-user. Regarding the 360° multisensory effects, all matrices with intensities are delivered together with the FoV-based panoramic video. Due to fast head movements, the system may not react within 10 ms and the user viewport could experience scenes located outside of the FoV region. In this case, the video quality is poor, but the overall perceived quality could be improved when experiencing additional multisensory effects given by the matrices of sensorial intensities.

Different bandwidth adaptation solutions could be implemented at the server-side to match the 360° mulsemedia content to the momentary networking conditions. One possible solution would be to adopt the dynamic adaptive streaming over HTTP (DASH) protocol. At certain time periods, network conditions may require the server to decrease the FoV-based video quality in order to minimize the ratio of lost packets, and hence, to avoid a strong degradation of user-perceived quality. The information regarding the 360° multisensory effects can be sent periodically by using MPEG-V files. Additional specifications can be included here for each sensory effect, such as duration and intensity for each video tile.

360° Mulsemedia Content Rendering and Synchronization

Much like any other video content, 360° video is represented according to a standard coding format. To this end, the H.264 standard is able to encode/decode the panoramic 360° video frames tile-by-tile; thus, ensuring full compatibility with more traditional video formats. Since each tile is decoded independently, the tile synchronization for each frame must be performed properly before view rendering. An important aspect to be addressed here is the synchronization overhead. This can be reduced by setting an optimal number of tiles that depend on encoding efficiency, storage overhead, and bandwidth utilization.

On the end-user side, each user must be equipped with VR glasses connected to devices

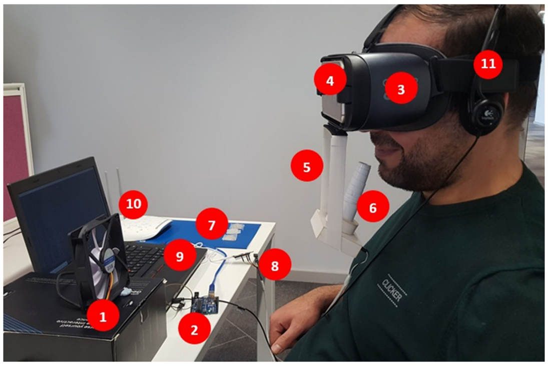

Figure 2. Setup of the experiment. (1) Wind blower fan. (2) Arduino Uno. (3) VR headset. (4) Smartphone. (5) Scent emitter. (6) Conic pipe. (7) Mesh-bags. (8) Arduino Nano. (9) Laptop. (10) Wi-Fi router. (11) Headphones.

capable of emitting sensory effects (e.g., olfaction, wind). An appropriate system of software and hardware is then needed to render the 360° mulsemedia content.

The aim of this work is to introduce a proof-of-concept prototype for the mulsemedia end-user side and analyze the impact of different parameters, such as sensory effects, video encoding qualities, and video content type on the perceived users' quality of experience through subjective tests. Consequently, for simplicity reasons, the tiling approach was not implemented in the practical video experience.

360° MULSEMEDIA: USER PROTOTYPE

With the aim of providing immersion, wearable devices such as VR goggles playing 360° media allow users to freely rotate their heads, to change the angle and look at interest points in the panoramic video. This is a step further compared to by now traditional headsets playing 2-D videos, some of which need an attached smartphone to play content whereas others are standalone and work mainly based on content streaming. Despite being successful, these commercial headsets do not support mulsemedia content yet, albeit there are some incipient proposals.[14] To overcome this hurdle, we have designed an end-user prototype as illustrated in Figure 2 that enriches the 360° video viewing by adding other sensory effects like scent and wind, which enables the users to immerse themselves more deeply into the experience.

We took into account the work of Saleme *et al.*,[14,15] who have explored both mulsemedia hardware and software solutions and have found gaps in the literature mainly related to reuse and adaptability of mulsemedia solutions. They proposed a framework named PlaySEM Sensory Effects Renderer (SER) to integrate different applications to varied devices in a straightforward fashion by reusing their mulsemedia renderer that supports MPEG-V, a standard to describe sensory effects metadata (SEM), and provides the necessary protocols to interact with our application. From that perspective and considering the fact that PlaySEM SER has been successfully applied to different scenarios of usage not only for its authors,[4] it catered for our prototype. Thus, the integration of our system reflects the end-user side of the conceptual approach introduced in Figure 1 and follows the tutorial proposed by Saleme *et al.*[14]. First, we annotated in the temporal domain only the videos with sensory effects, that is, 360° audiovisual content written in MPEG-V. The spatial annotation is currently under investigation and makes the objective of future work. Therefore, videos are encoded by using the tilling approach but only the temporal synchronization with multisensory components is deployed at this stage. The annotation files were not encoded with the videos, but instead, they were embedded into a Unity solution for Android capable of reproducing equirectangular videos and communicate via a wireless network with PlaySEM SER.

PlaySEM SER and the 360° application communicate via WebSocket. After opening the videos on the 360° application, their correspondent SEMs were sent to PlaySEM SER, which in turn, preprocessed the SEMs converting them to commands to handle the specific devices configured for the experiment, as described by Saleme *et al.* [15]. Upon the end of this process, the mulsemedia renderer sent a cue to the 360° mulsemedia application to start playing the video. Finally, PlaySEM SER, which was connected via Bluetooth to the Scent Device and via USB to the Wind Device, rendered the wind and scent effects synchronized with the 360° audiovisual content.

Table 1. Timeframe of sensory effects associated with the scenes in the video samples.

360° Video	Wind Effect			Scent Effect		
	Start	Stop	Intensity	Start	Stop	Intensity
Lavender Field	0 s	60 s	30%	0 s	60 s	100%
Coffee Shop	5 s	6 s	100%	37 s	60 s	100%
Roller Coaster	0 s	27 s	30%	27 s	37 s	100%
	27 s	37 s	100%			
	37 s	60 s	50%	43 s	48 s	100%

The prototype depicted in Figure 2 comprises: 1) Wind blower fan used to render the wind effect; 2) Arduino Uno which controls the wind blower fan; 3) Samsung Gear VR headset that provides a virtual reality experience; 4) Smartphone used to play the 360° videos on the VR headset; 5) Scent emitter prototype with a small size cooling fan inside and the scent cartridge; 6) Extensible spiral conic pipe to direct the scent from the fan towards the user's nose; 7) Mesh bags with scent crystals that are going to be placed in the scent cartridge; 8) DFRobot Bluno Nano is employed to control the scent emitter; 9) Laptop running a mulsemedia renderer; 10) WiFi router for wireless connectivity between the smartphone and the laptop; and 11) Headphone to enable the sound of the video and immerse the user fully into the experience. One of the most exciting novelties of the proposed prototype is the scent emitter and the spiral conic pipe shape, which were attached to the VR headset. The conic shape helps us to direct the scent toward the subject's nose and avoids the reduction in intensity and the spreading of the scent into the environment due to the wind blower fan, which creates the air-flow. Moreover, the spiral shape enables us to adjust the pipe length accordingly based on the subject.

SUBJECTIVE QUALITY ASSESSMENT

In order to evaluate the performance of the proposed prototype, subjective quality assessment was carried out. A total of 48 participants (27 male, 21 female) aged from 16 to 65 participated in the study. Most participants had prior VR experiences; 24 self-reported as unfamiliar with subjective video quality evaluation, 16 were familiar, while 8 had worked in the field before.

Three types of 360° mulsemedia content were used in our study. These comprised three 360° videos coupled with airflow (of varying intensities) and scent (diesel, coffee, and lavender) effects. They are further classified and described as follows: 1) Static: Lavender Field—the user is standing in a fixed position in the middle of a lavender field. There is no movement in the background and the user can only feel the wind and the lavender smell. 2) Semi-dynamic: Coffee Shop—the user is standing in a fixed position inside a coffee shop. However, this time there is movement in the background where the waiter is preparing the coffee. The user smells the coffee aroma and experiences a gust of air coming from the coffee machine signaling that the coffee preparation process had ended. 3) Dynamic: Roller Coaster—the user is located inside a fast-moving carriage. The background moves along with the carriage. The user experiences a whiff of diesel as well as the wind in the face as he or she is riding the roller coaster. Table 1 presents a timeframe when the sensory effects are started and stopped as well as their corresponding intensity. A value of 100% intensity for wind or for scent effect corresponds to the maximum power of the sensory device. The variations occur in function of the scenes in each video sample. Each 360° panoramic video was encoded at four different quality levels: HD, Full HD, 2.5K, and 4K. The framerate was 60 fps (Coffee Shop), 25 fps (Lavender Field), and 30 fps (Roller Coaster), except for 4K that was limited to 30 fps due to restriction on the Unity Application to reproduce the videos on the smartphone.

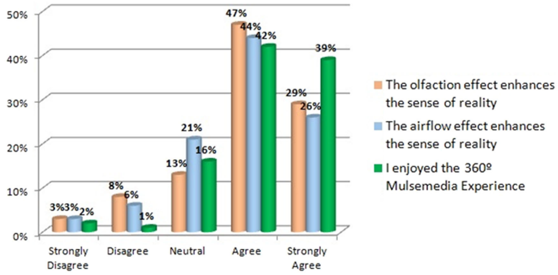

Figure 3. Results on the sensory effects and user enjoyment.

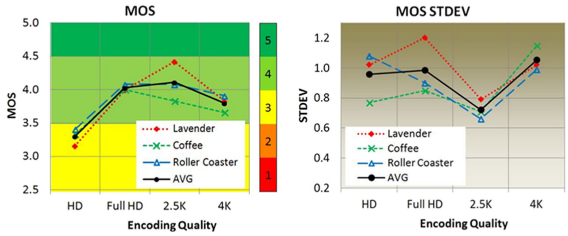

Figure 4. Subjective assessment results: On the left side MOS, on the right side MOS Standard Deviation.

Furthermore, the following codecs were used: H264-MPEG-4 AVC (part 10) (avc1); chroma location: left; and projection: equirectangular. The duration of each video sample was 60 s.

In Table 1, timeframes are set based on a consensus among a group of researchers after watching the videos and experiencing them on the developed prototype. According to Murray *et al.*[16] the timeframe setting for multisensory effects is critical from the perspective of user QoE enhancement in traditional mulsemedia applications and can be considered further as part of future work in 360° mulsemedia.

Each participant watched all three 360° mulsemedia content types at the same quality level in a randomized order. After watching each video, participants rated their overall quality level on a 5-point scale (1-Bad, 2-Poor, 3-Fair, 4-Good, 5-Excellent). Apart from assessing the user-perceived quality of the 360° content, subjects were also asked to rate the impact of the airflow and olfactory effects on the sense of reality and enjoyment on a 5-point scale ranging from 1-Strongly Disagree to 5-Strongly Agree.

Impact of Sensory Effects and User Enjoyment on the Experience

Participants' responses highlighted positive to very positive attitudes to the 360° mulsemedia experience—more than 80% of users agreed or strongly agreed that they enjoyed the experience, as seen in Figure 3. This is also reflected in a Mean Opinion Score (MOS) value of 3.9063, which a *one-sample t-test* highlights as being statistically significant ($t = 16.463$, $p < .001$). The same trend can be observed in terms of participants' agreement with the statements that olfaction and airflow effects heighten the sense of reality—76% of users agreeing or strongly agreeing with this in respect of olfaction and 70% as far as airflow is concerned. The MOS for the two types of sensorial effects are 3.9028 and 3.833, respectively, and again *one sample t-test* confirms the statistically significant nature of these results ($t = 10.847$, $p < .001$ for olfaction; $t = 10.254$, $p < .001$ for airflow). All underline the enhanced QoE that 360° mulsemedia affords users.

Impact of Video Encoding Quality on Human-Perceived Visual Quality

To study the impact of the user-perceived quality of the four encoding qualities of the 360° videos (Lavender, Coffee, and Roller Coaster), the participants were asked to rate the overall quality on a 1 to 5 scale (Bad to Excellent). For each encoding quality, the mean value represented by the MOS as well as the standard deviation (STDEV) of the statistical distribution of the assessment scores were computed and illustrated in Figure 4, respectively. The average values of the results for the four encoding qualities are also listed.

Analyzing the subjective assessment results, we give the following interpretations and remarks:

1. More data does not necessarily mean an increase in user-perceived quality. The results show that the Full HD, 2.5K, and 4K encoding qualities are all perceived as "Good" by the participants while the HD encoding quality is perceived as "Fair." This means that we can safely say that there is a threshold above which, even if we generate more data, it does not make a significant difference in the user-perceived quality. Thus, in our setup, this threshold is somewhere between Full HD and 2.5K as the MOS scores

for 4K start decreasing. The *Pearson correlation* further indicates that there is a weak decreasing relationship between MOS and STDEV values ($r = -0.31226$). Consequently, the scores across the participants tend to have a higher variation for the videos with lower perceived quality.

2. Transmitting high data rates traffic puts significant pressure on the underlying networks. Thus, the findings in the first interpretation could help the network operators in accommodating more users at lower encoding qualities while maintaining an acceptable user-perceived quality. However, to be able to meet the strict QoS requirements of the 360° mulsemedia traffic class and to maximize the overall QoS satisfaction, the integration of new mechanisms such as Machine Learning[8] needs to be considered. In this sense, two-stage reinforcement learning can be potentially used to improve the QoS provisioning when scheduling heterogeneous traffic in the 5G access networks.[17] In the first stage, the scheduling order of different traffic classes is decided at each time instant to avoid the overprovisioning of higher prioritized traffic classes such as 360° mulsemedia and the starvation of some lower prioritized services. In the second stage, the most convenient scheduler is selected in the frequency domain for each ordered traffic class to perform the radio resource allocation.
3. The decrease in user-perceived quality between the 2.5K and 4K video encoding rates as illustrated in Figure 4 highlights the fact that the use of 4K resolution in 360° mulsemedia might be excessive considering our setup, and therefore, this can be explored as proposed in the second interpretation. While this might have been influenced by hardware limitations of the HMD device, such as the FoV and screen-door effect,[18] it nonetheless reinforces (but in the mulsemedia arena) issues with which the community has been aware of.

Impact of Video Content on the 360° Mulsemedia Experience

Looking at the impact of the video content on the MOS obtained from the subjective tests, and illustrated in Figure 4, it can be noticed that regardless of the video content, the user-perceived quality follows nearly the same distribution with the HD video levels perceived as "Fair" and the remaining of Full HD, 2.5K, and 4K video levels perceived as "Good." According to the user profiles, 52% of the participants are mainly interested in watching dynamic video content, 23% are mainly watching semidynamic video content and 25% are mainly watching static video content on a regular basis on their mobile devices. However, the results indicated that on average the Lavender video was more enjoyable, followed by the Roller Coaster and then the Coffee video. The increase in MOS scores for the Lavender video can be related to the pleasant lavender smell and the static nature of the content. It was also noted that on average the participants found the intensity of the olfaction effect "Just Fine" toward "Strong" for Lavender and "Just Fine" for Coffee and Roller Coaster videos. For the airflow effect, the average intensity was reported as "Just Fine" across all contents. Moreover, the MOS results for each video content class still validate our first interpretation.

CONCLUSION

With the advent of next-generation wireless networks come also the promise of delivering innovative user experiences that will take advantage of the significant improvements in terms of throughput and latency that 5G has in stock. However, this pressurizes the network operators in trying to understand the QoS requirements for these new applications in order to provide an acceptable level of QoE to their customers. In this context, this article introduces a new type of application class referred to as 360° mulsemedia, which enriches the 360° video by adding other sensory effects like scent and wind to enable the end-users to immerse themselves more deeply into the 360° video experience. Indeed, we developed a prototype to enable the users to experience 360° mulsemedia content in an unprecedented fashion. User evaluations revealed that 360° mulsemedia delivered enhanced levels of QoE, but also that more data does not necessarily attract an increase in user-perceived quality. Indeed, 4K video quality used in a 360° mulsemedia context was found to lead to lower QoE levels than when a 2.5K video was employed to the same end taking into account our setup. This highlights and suggests that network operators could exploit significant bandwidth savings in the delivery of 360° mulsemedia with no consequent degradation in user QoE. The precise amount of network resources that can be saved will depend on the length and encoding parameters of the video that will be transmitted as well as the device characteristics on which it is experienced (e.g., VR Gear). This warrants further investigation. Also, as part of our future work, we are planning to find out the optimal number of tiles for the FoV and the optimal number of neighboring tiles that must be transmitted at a higher quality than the rest of the tiles from the panoramic video in order to get an acceptable QoE level. A worthwhile future study should also examine the impact of human factors in 360° mulsemedia QoE. Last but not least, future work will consider extending the prototype to an end-to-end solution integrating omnidirectional wind effects as well as other sensorial data. All our worthy future pursuits.

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. S. Sukhmani, M. Sadeghi, M. Erol-Kantarci, and A. El Saddik, "Edge caching and computing in 5G for mobile AR/VR and tactile internet," *IEEE MultiMedia*, vol. 26, no. 1, pp. 21–30, Mar. 2019.
2. R. Trestian, I.-S. Comşa, and M. F. Tuysuz, "Seamless multimedia delivery within a heterogeneous wireless networks environment: Are we there yet?," *IEEE Commun. Surv. Tuts.*, vol. 20, no. 2, pp. 945–977, Apr.–Jun. 2018.
3. G. P. Fettweis, "The tactile internet: Applications and challenges," *IEEE Veh. Technol. Mag.*, vol. 9, no. 1, pp. 64–70, Mar. 2014.
4. E. B. Saleme, C. A. S. Santos, and G. Ghinea, "Coping with the challenges of delivering multiple sensorial media," *IEEE MultiMedia*, vol. 26, no. 2, pp. 66–75, Apr.–Jun. 2019.
5. A. Aijaz, M. Dohler, A. H. Aghvami, V. Friderikos, and M. Frodigh, "Realizing the tactile internet: Haptic communications over next generation 5G cellular networks," *IEEE Wireless Commun.*, vol. 24, no. 2, pp. 82–89, Apr. 2017.
6. T. Zhang Bao, A. Pande, H. Wu, and X. Liu, "Motion-prediction-based multicast for 360-degree video transmissions," in *Proc. IEEE Conf . Sens . , Commun . Netw .*, 2017, pp. 1–9.
7. X. Hou, S. Dey, J. Zhang, and M. Budagavi, "Predictive view generation to enable mobile 360-degree and VR experiences," in *Proc. Morning Workshop Virtual Reality Augmented Reality Netw.*, 2018, pp. 20–26.
8. I.-S. Comşa, R. Trestian, and G. Ghinea, "360° mulsemedia experience over next generation wireless networks—A reinforcement learning approach," in *Proc. 10th Int. Conf. Quality Multimedia Experience*, 2018, pp. 1–6.
9. O. A. Ademoye, N. Murray, G.-M. Muntean, and G. Ghinea, "Audio masking effect on inter-component skews in olfaction-enhanced multimedia presentations," *ACM Trans. Multimedia Comput., Commun. Appl.*, vol. 12, no. 4, pp. 51–64, 2016.
10. Z. Yuan, G. Ghinea, and G.-M. Muntean, "Beyond multimedia adaptation: Quality of experience-aware multi-sensorial media delivery," *IEEE Trans. Multimedia*, vol. 17, no. 1, pp. 104–117, Jan. 2015.
11. X. Corbillon, G. Simon, A. Devlic, and J. Chakareski, "Viewport-adaptive navigable 360-degree video delivery," in *Proc. IEEE Int. Conf. Commun.*, May 2017, pp. 1–7.
12. M. Graf, C. Timmerer, and C. Mueller, "Towards bandwidth efficient adaptive streaming of omnidirectional video over HTTP: Design, implementation, and evaluation," in *Proc. 8th ACM Multimedia Syst. Conf.*, 2017, pp. 261–271.
13. B. Wyszynski, A. G. Galvez, and T. Nakamoto, "Improvement of ultrasonic atomizer method for deposition of gas-sensing film on QCM," *Elsevier Sensors Actuators B*, vol. 127, no. 1, pp. 253–259, 2007.
14. E. B. Saleme, A. Covaci, G. Mesfin, C. A. S. Santos, and G. Ghinea, "Mulsemedia DIY: A survey of devices and a tutorial for building your own mulsemedia environment," *ACM Comput. Surv.*, vol. 52, no. 3, pp. 58:1–58:29, 2019.

15. E. B. Saleme, C. A. S. Santos, and G. Ghinea, “A mulsemedia framework for delivering sensory effects to heterogeneous systems,” *Multimedia Syst.*, vol. 25, no. 4, pp 421–447, 2019.
16. N. Murray, O. A. Ademoye, G. Ghinea, and G. M. Muntean, “A tutorial for olfaction-based multisensorial media application design and evaluation,” *ACM Comput. Surv.*, vol. 50, no. 5, pp. 67:1–67:30, 2017.
17. I.-S. Comşa, A. De Domenico, and D. Ktenas, “Method for allocating transmission resources using reinforcement learning,” U.S. Patent 2019 / 0124667 A1, 2019, pp. 1–10.
18. J.-M. Cho, Y.-D. Kim, S. H. Jung, H. Shin, and T. Kim, “Screen door effect mitigation and its quantitative evaluation in VR display,” in *Proc. SID Symp. Dig. Tech. Papers*, 2017, pp. 78–81.

Ioan-Sorin Comșa is a research assistant in the Computer Science Department, Brunel University London, U.K. 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@brunel.ac.uk.

Estêvão Bissoli Saleme is currently working toward the Ph.D. degree in the Computer Science Department, Federal University of Espírito Santo, Brazil. His research interests include multimedia and human-computer interaction with a focus on digital multisensory systems, immersive technologies, and middleware and frameworks. Contact him at estevao.saleme@ieee.org.

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

Gebremariam Mesfin Assres is a research assistant in the Computer Science Department, Brunel University London, London, U.K. His research interests inlcude software engineering, usability, cross-platform smartphone applications, multimedia, mulsemedia, Web service, service-oriented architecture, and e-learning. Contact him at gebremariam.assres@brunel.ac.uk.

Ramona Trestian is a senior lecturer with the Design Engineering and Mathematics Department, Middlesex University, London, U.K. 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 is a professor with the Department of Informatics, Federal University of Espírito Santo, Brazil. His recent research interests focus on multimedia systems and applications, synchronization, digital video, and crowdsourcing systems. Contact him at saibel@inf.ufes.br.

Gheorghiță Ghinea is a professor of mulsemedia computing with the Computer Science Department, Brunel University London, U.K. 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.