An Evaluation of 360° Video and Audio Quality in an Artistic-Oriented Platform

Anderson Augusto Simiscuka[1], *Member, IEEE*, Mohammed Amine Togou[1], *Member, IEEE*, Rohit Verma[1], Mikel Zorrilla[2], Noel E. O'Connor[1] and Gabriel-Miro Muntean[1], *Senior Member, IEEE*

***Abstract*—Consuming immersive content over a browser is possible due to libraries such as WebGL, WebXR and Three.js. 360° videos, however, are very large in size, and high definition content can cause buffering events even in environments with high bandwidth. Immersiveness is further achieved in 360° content delivery with the addition of ambisonic audio, which indicates the source of sound, according to head tracking of users, in the 3-D space. In situations where video quality needs to be reduced due to network constraints, ambisonics can help increase user-perceived quality. This paper introduces a study of user experience with 24 participants located in various countries considering multiple quality levels of audio and video. An artistic-oriented platform for opera delivery is also described, in the context of the European Horizon 2020 TRACTION project.**

***Index Terms*—360 video, QoE, Ambisonics, Immersive Media**

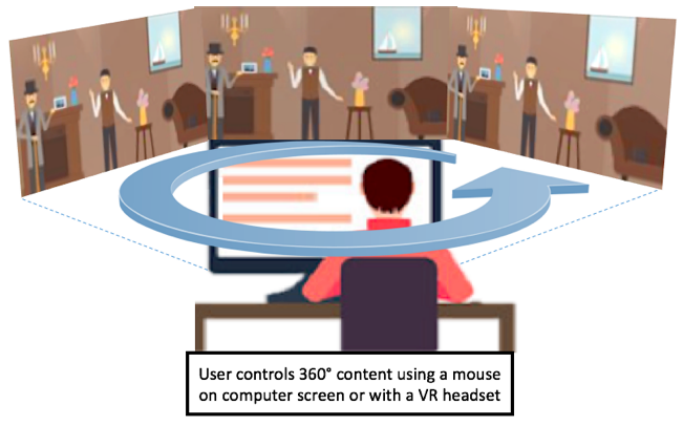

Fig. 1. 360° video user navigation

I. Introduction

CONTENT delivery of 360° videos to remote viewers brings many challenges to research and development. These challenges are mostly associated with the large data amounts to be exchanged and low streaming latency required to prevent user dizziness when consuming the content on various VR headsets [1].

The works presented in [2]–[4] indicate that rendering VR graphics in web browsers is a graphics processing unit (GPU) intensive job, and modest graphics and processing hardware may struggle to achieve a smooth user experience.

The proposal of adaptation schemes specifically designed for improving the delivery of immersive video aim to improve user experiences. Adaptation based on resolution and region of interest can improve the quality of the video. Region of interest-based adaptive schemes, for instance, perform adaptation at the level of regions within clip frames, based on user interest obtained from tracking eye movement [5], [6]. These schemes adjust the quality of less important regions from multimedia frames. Regions that the viewers are interested in, either do not change, or involve little adjustment, resulting in high overall end-user perceived quality.

360° VR videos and the underlying 3D geometry, as illustrated in Fig. 1, can also be divided into spatially partitioned segments/tiles in the 3D space, and can be adapted with various priorities, according to the regions the users are more likely to look at [7], [8].

It is important, however, to understand users' perception when adapting immersive video content. These experiences require high definition visuals, as users eyes can be very close to the screen when using headsets. One technique to improve user perception is improving the audio quality to create a masking effect of video imperfections [9], [10]. Higher order ambisonics greatly improves audio in immersive content delivery. It reproduces spatial audio building a 3D sound field with spherical harmonics. It provides high resolution and an approximation of the original sound field and can be executed with standard stereo speakers and headphones and also multichannel setups [11].

Authors in [12] acknowledge that few works have investigated the role of spatial audio in 360° videos and its impact in Quality of Experience (QoE). This is of extreme importance for web-based 360° players, as the audio quality can help offset video quality imperfections and maintain a higher-perceived user experience.

This paper is based on the research activities within the EU Horizon 2020 TRACTION project [13], which designs and builds a collaborative and participatory production toolset for the co-creation and co-design of operas. The toolset is envisaged to support community interaction, user-generated rich media capture, immersive audiovisual and 360° content, smart media editing, narrative engines and interactive adaptive media distribution [14]. TRACTION requires the development of collaborative and immersive multimedia players, therefore it is important that content can be processed and streamed without latency or delays.

[1] A. A. Simiscuka, M. A. Togou, R. Verma, N. E. O'Connor and G.-M. Muntean are with Dublin City University, Glasnevin, Dublin 9, Ireland. (e-mails: anderson.simiscuka2@mail.dcu.ie, {mohammedamine.togou, rohit.verma, noel.oconnor, gabriel.muntean}@dcu.ie).

[2] M. Zorrilla is with the Department of Digital Media, Vicomtech Foundation, Basque Research and Technology Alliance (BRTA), Donostia-San Sebastián, Spain (e-mail: mzorrilla@vicomtech.org).

This paper describes a research study to enable understanding of the effect of different audio and video qualities in immersive experiences in an artistic context. The impact of audio quality and the use of immersive audio in 360° videos with varying resolutions is also investigated. The study focuses on understanding users' perceptions of different audio and video qualities without adaptation, therefore each short video presented maintains the same audio and video quality from beginning to end. This study is highly important for the development of new adaptation algorithms for immersive content.

Twenty-four participants took part in the study, which follows the recommendations of the ITU-T P.913 standard [15] for video assessment. Participants accessed a web player remotely from various locations, including Spain, Ireland, United Kingdom, India and China.

The remaining sections of this paper are organised as follows. Section II presents related works while Section III discusses the web player design. Section IV describes the analysis of the user study. Conclusions and directions for future work end this paper in Section V.

II. Related Work

The articles presented in [16], [17] indicate that less than 1% of sold computers and 6.8% of sold smartphones are VR ready, a total of 13 million computers and just under 200 million smartphones. An estimate of 100 million VR-ready PCs were expected to be sold by 2020, a fraction of roughly 1.5 billion PCs being used worldwide [18]. Therefore, it is important to integrate VR experiences to existing hardware, employing, for instance, web browsers for the dissemination of immersive content.

The standard ISO/IEC MPEG Dynamic Adaptive Streaming over HTTP (MPEG-DASH) is a major video adaptation technique. It consists of the Media Presentation, segments that contain periods, adaptation sets, and representations that form a media, and the Media Presentation Description (MPD), an eXtensible Markup Language (XML) document that has the locations of all alternative segments [19], [20]. MPEG-DASH has been applied to regular 2-D and 360° video content in order to adapt video quality to network capabilities [21], [22].

WebVR and the newer WebXR, are specifications for VR applications for web browsers, both desktop and mobile web browsers. Content is rendered twice on both halves of a screen (one for each eye) and can be seen through smartphone-based VR headsets, such as the Google Cardboard. Applications made in WebVR and WebXR can also be executed in regular VR headsets such as the Oculus Rift [23].

Omnitone is an open-source JavaScript implementation for Ambisonics decoding. Web-based Virtual Reality applications can implement Omnitone to provide spatial audio capabilities, which means that the direction of the audio can be perceived by users as they move within the virtual environment. Omnitone also uses the Web Audio API and multi-channel Ambisonics files are streamed via the *AudioBufferSourceNode* interface. A rotation matrix monitors the user's head position, based on user interactions or sensor data. Binaural rendering is handled by Convolver and GainNode interfaces, both native to Web Audio, with the use of head related transfer functions (HRTFs) [24].

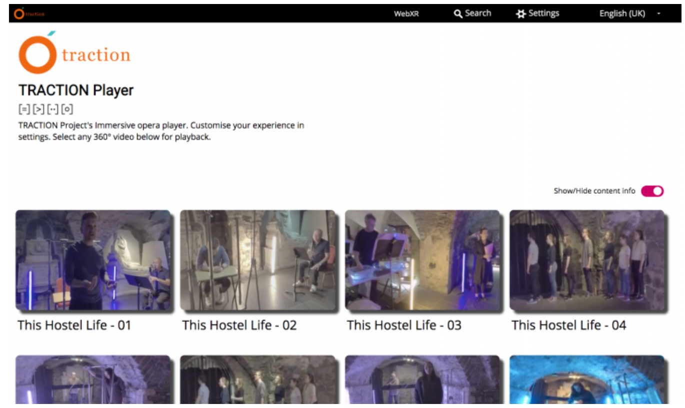

Fig. 2. 360° player with support to ambisonics - video selection page

Many modern video platforms support 360° content, such as YouTube and Vimeo. Other websites use libraries to customise their video players. These libraries include Three.js and Video.js. Three.js is a JavaScript library for the creation of 3D assets for web browsers. Video.js is a library for the creation of custom web-based video players through the use of available plugins. [25], [26].

The technologies described in this section provide the means for the creation of immersive and innovative user experiences, which however, can be affected by network instabilities. Some works review user perception of 360° experiences, and aim to understand how to improve the content delivery process.

Authors in [27] present a study that aims to understand how spatial audio influences visual attention in 360° videos. Authors suggest that users had a sense of integration with the spatial environment more quickly than the non-spatial environment.

The work presented in [28] investigates QoE in VR based on user profiles, considering age, gender, familiarity with 360 content, and interest in the content, in order to understand the impact of different factors that influence QoE (*e.g.,* encoding parameters, content type and device type) on users.

A tile-based 360° video adaptive streaming scheme is proposed in [29]. The approach contains multiple video quality levels based on bandwidth conditions, in order to improve QoE with high video bitrate while minimising stall time.

The works presented in the previous three paragraphs [27]–[29] do not consider the effect of different audio qualities in 360° content delivery, focusing on other metrics. Therefore, a study combining different video and audio qualities in 360° is useful for the understanding of the impact of maintaining and/or prioritising high quality audio when adapting video.

III. Web Player Design

The web-based immersive video player used in the tests, as seen in Fig. 2, can execute regular 2D and 360° content. Users can navigate in immersive videos by dragging the point of view with a mouse, touch screens or head movements when using virtual reality headsets. This player is based on the ImAc player developed by the EU Horizon 2020 ImAc project [30].

TABLE I
GROUPS OF THE 360° PLAYER PERCEPTUAL STUDY

User Groups	**Romeo and Juliet - Perceptual Audio Tests**			**This Hostel Life - Perceptual Video Tests**		
	2m50s; 2m49s; 2m48s (all 3840×2160)			**Low:** 1024×512 4m03s; **Medium:** 2048×1024 3m12s; **High:** 3840×1920 1m40s		
	Video 1	**Video 2**	**Video 3**	**Video 4**	**Video 5**	**Video 6**
Group A	Ambisonics	Mono	Stereo	High	Medium	Low
Group B	Mono	Stereo	Ambisonics	Medium	High	Low
User Groups	**This Hostel Life - Perceptual Video Tests**			**Romeo and Juliet - Perceptual Audio Tests**		
	Low: 1024×512 4m03s; **Medium:** 2048×1024 3m12s; **High:** 3840×1920 1m40s			2m50s; 2m49s; 2m48s (all 3840×2160)		
	Video 1	**Video 2**	**Video 3**	**Video 4**	**Video 5**	**Video 6**
Group C	Low	Medium	High	Stereo	Ambisonics	Mono
Group D	High	Low	Medium	Ambisonics	Stereo	Mono

The player contains a number of tools for accessibility, including subtitles, spatial audio description, sign language interpreting, voice control and large menus for visually impaired people. These can be accessed through the main player menu as shown in Fig. 3. The four icons [=] [>] [··] [o] represent text subtitles, sign language, audio subtitles and audio description, respectively. It also supports linking questionnaires after videos finish playing.

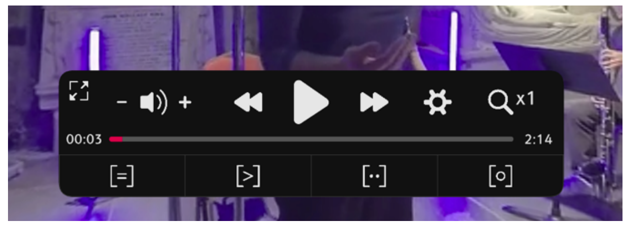

Fig. 3. 360° player menu

The player architecture, illustrated in Fig. 4, requires a webserver (*e.g.,* Tomcat), which hosts the web application and the required libraries, and a HTTP server (*e.g.,* Apache), for hosting the content. The player implements the Omnitone JavaScript library for ambisonic audio, dash.js for video reproduction, and Three.js for 360° content rendering. The player can run on desktop and mobile web browsers, and immersive audio can play on stereo speakers and headphones.

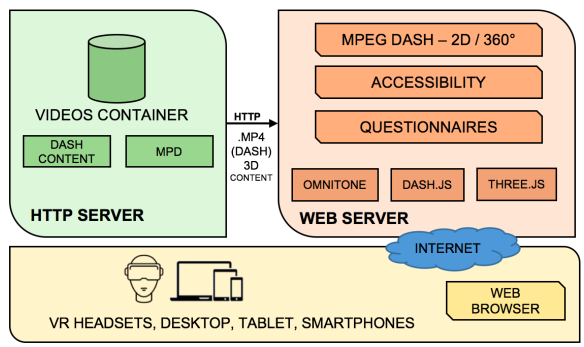

Fig. 4. Player architecture

IV. USER STUDY

The user study was informed by the ITU-T Recommendation P.913 [15], in relation to duration and number of participants. This section presents details about participants, assessment protocol, questionnaires, results and discussion.

A. Participants

The 24 participants, with 15 males and 9 females, accessed the web player remotely from various locations in Spain, Ireland, United Kingdom, India and China. They were allowed to use their preferred device (*i.e.,* smartphone, laptop, desktop, VR headset) to consume content.

Participants were recruited via e-mail and had to fill a consent form before participating. Participants also received a plain language statement and data management plan. These documents contain a detailed description of the testing scenario, research purpose, data processing and analysis, participant identity protection, etc. Ethical approval was obtained from Dublin City University Research Ethics Committee.

B. Assessment Protocol

Participants received a link to the web player. The home page of the player has an introduction to the experiment and a button to continue to the demographics form. Participants could fill forms at their own pace. After filling the demographics questionnaire, participants viewed 6 videos with durations between 1m40s and 4m03s.

The multimedia sequences, two opera plays, were provided by TRACTION project partners: "This Hostel Life" from the Irish National Opera and "Romeo and Juliet" from Barcelona's Gran Teatre del Liceu. After each of the six videos, participants answered a set of questions about the perceptual experience of video and audio quality. Finally, participants filled a usability questionnaire with questions about the player and the experience as a whole. The expected duration of the study was 36 minutes.

As seen in Table I, 4 groups of participants were created. Each group watched the videos in a different order to eliminate bias. The videos were presented in a combination of 3 levels of audio quality in *Romeo and Juliet* (*i.e.,* mono, stereo and ambisonics, as seen on Table II) and 3 levels of video quality in *This Hostel Life* (*i.e.,* low: 1024×512px, medium: 2048×1024px and high: 3840×1920px). The footage from Romeo and Juliet was focused on audio quality, so the video quality was fixed at 3840×1920px. The footage from This Hostel Life was focused on video quality, therefore audio was fixed at ambisonics (256kbps).

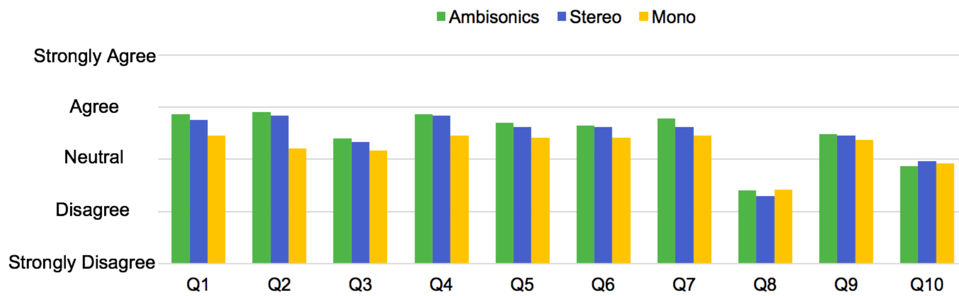

Fig. 5. Average scores of the content related to audio quality.

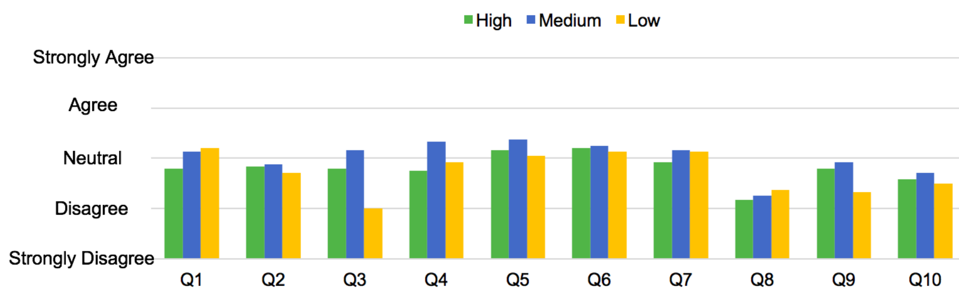

Fig. 6. Average scores of the content related to video quality.

TABLE II
AUDIO SETTINGS - 360° PLAYER USER STUDY

	Mono	Stereo	Ambisonics
Channels	1 channel	2 channels	4 channels
Sample Rate	48kHz	48kHz	48kHz
Bitrate	90kbps	90kbps	256kbps
Codec	AAC	AAC	AAC

C. Questionnaires

After each video clip, participants were asked to fill in a QoE questionnaire with 13 questions. These questionnaires employed a five-point Likert scale (*i.e.,* (1) Strongly disagree; (2) Disagree; (3) Neutral; (4) Agree; (5) Strongly agree). Some examples of questions asked after each video include: "The audio improved the immersiveness of the experience," "The immersive experience helped me to be more engaged in opera," "The 360°/VR effects were disturbing for me during the video," and "I believe that the immersive experience is comparable to a live opera."

A demographics questionnaire was provided to the participants before the start of the experiment, containing 26 questions, including user profiling, familiarity with the technology, questions about the network, equipment used and noise levels.

A usability questionnaire with 14 questions was provided at the end, in order to evaluate the player navigation, preferences and ease of use.

D. Results and Discussion

The results presented in this section are divided in two aspects investigated in the questionnaires presented to participants. The first is audio, which summarises the findings from the questionnaires presented after watching the footage of the opera *Romeo and Juliet*. These clips were all presented with a high resolution of 3840×2160px, and the audio was delivered in ambisonics, stereo and mono modes. The second aspect investigated is video quality, which was delivered in high, medium and low levels, and the opera footage was from *This Hostel Life*.

Participants were not informed about which content focused on audio or video, nor the quality levels being delivered. The footage sequence was also presented in four different orders. The same questions were asked for all clips of both operas. The answers were converted into a score ranging from 1 to 5 based on the Likert scale used (*i.e.,* from strongly disagree to strongly agree). The scores were then averaged per question in each of the quality levels of video and audio.

1) Audio: Fig. 5 depicts the average scores of the questions presented after the clips with varying audio qualities. More participants agreed or strongly agreed that the clip with ambisonic audio improved the immersiveness of the experience (Q1), especially in comparison to mono audio. The audio quality was perceived as good (Q2) more frequently in stereo and ambisonics modes. Even though the video quality was the same across the 3 clips, more participants felt the video quality was good (Q3) when ambisonics and stereo were used (*i.e.,* 14 participants answered agree or strongly agree for ambisonics while 9 participants answered the same for mono audio). Ambisonics and stereo audio also received more positive answers than mono when participants were asked if they enjoyed the experience presented (Q4), if the immersive experience helped assimilating the performance (Q5), if the immersive experience helped engaging with opera (Q6), and if it was enjoyable to watch the opera piece as an immersive experience (Q7). In all three scenarios, participants found that the 360°/VR effects were not disturbing during the videos (Q8). The perception of the colours of the footage being clear/vivid (Q9) did not seem to be affected by audio quality. Finally, participants were more neutral regarding the immersive experience being comparable to a live opera (Q10).

2) Video: The average scores of the questions presented after the clips with varying video qualities are shown in Fig. 6. The changes in video quality impacted user perception, especially in the scenario with low resolution (*i.e.,* 1024×512px), with participants disagreeing that the video quality was good (Q3). Interestingly, participants seemed to have a similar and at times better perception of the medium quality footage, indicating that videos with a resolution of 2048×1024px cause a similar impression to videos with a resolution of 3840×1920px. The perception of the colours of the footage being clear/vivid (Q9) was also more affected in videos with low resolution.

V. Conclusions and Future Research Directions

This paper presented a study of the impact of audio and video quality in 360° videos in a web-based player. The experiment was performed in the context of the EU Horizon 2020 TRACTION project, which promotes co-creation and dissemination of opera via novel technologies. 24 participants viewed 6 clips from two operas in different video and audio qualities, including ambisonic audio. The answers to the questionnaires provided to participants indicated that higher quality audio and ambisonics positively affects the perception of the 360° experience.

The study aims to inform the design of future algorithms for streaming 360° content through the web. For future work, we aim to compare this study with a new experiment with an adaptive version of the player.

Acknowledgements

This work was supported by the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement no. 870610 for the TRACTION project. The support of the Science Foundation Ireland (SFI) Research Centres Programme via grants SFI/12/RC/2289_P2 (Insight) and SFI/16/SP/3804 (ENABLE), co-funded by the European Regional Development Fund, is also gratefully acknowledged.

References

[1] M. Abdelrahman, M. Elbamby, and V. Räisänen, "Proactive scheduling and caching for wireless vr viewport streaming," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, 2021, pp. 1–6.

[2] V. Parthasarathy, A. A. Simiscuka, N. O'Connor, and G.-M. Muntean, "Performance Evaluation of a Multi-User Virtual Reality Platform," in *International Wireless Communications and Mobile Computing (IWCMC)*, 2020, pp. 934–939.

[3] R. Pathak, A. A. Simiscuka, and G.-M. Muntean, "An adaptive resolution scheme for performance enhancement of a web-based multi-user vr application," in *Proc. IEEE Int. Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, 2021, pp. 1–6.

[4] S. Abbas, A. A. Simiscuka, and G. M. Muntean, "A Platform Agnostic Solution for Inter-Communication between Virtual Reality Devices," in *Proc. IEEE World Forum on Internet of Things (WF-IoT)*, 2019, pp. 189–194.

[5] S. B. Lee, A. F. Smeaton, and G. M. Muntean, "Quality-oriented multiple-source multimedia delivery over heterogeneous wireless networks," *IEEE Transactions on Broadcasting*, vol. 57, no. 2 PART 1, pp. 216–230, jun 2011.

[6] B. Ciubotaru, C. H. Muntean, and G. Muntean, "Mobile multi-source high quality multimedia delivery scheme," *IEEE Transactions on Broadcasting*, vol. 63, no. 2, pp. 391–403, 2017.

[7] M. Hosseini, "View-aware tile-based adaptations in 360 virtual reality video streaming," in *Proc. IEEE Virtual Reality Conference (VR)*, 2017, pp. 423–424.

[8] A. Yaqoob and G.-M. Muntean, "A combined field-of-view prediction-assisted viewport adaptive delivery scheme for 360° videos," *IEEE Transactions on Broadcasting*, vol. 67, no. 3, pp. 746–760, 2021.

[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, aug 2016.

[10] Y. Pan and L. Xie, "The influence of video quality on auditory masking effect experiment," in *Proc. Int. Conf. on Signal Processing Systems*, vol. 2, 2010, pp. V2–50–V2–52.

[11] S. R. Quackenbush and J. Herre, "Mpeg standards for compressed representation of immersive audio," *Proceedings of the IEEE*, vol. 109, no. 9, pp. 1578–1589, 2021.

[12] A. Hirway, Y. Qiao, and N. Murray, "A QoE and Visual Attention Evaluation on the Influence of Audio in 360° Videos," in *Proc. IEEE Int. Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM)*, 2020, pp. 191–193.

[13] "Traction Project – Opera co-creation for social transformation." [Online]. Available: https://www.traction-project.eu/

[14] A. A. Simiscuka, M. Zorrilla, P. Cesar, N. O'Connor, and G.-M. Muntean, "Performance Analysis of Multi-Source Wireless Multimedia Content Delivery," *NEM Summit*, 2020.

[15] "P.913: Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment." [Online]. Available: https://www.itu.int/rec/T-REC-P.913-202106-I/en

[16] "Less than 1% of PCs can run virtual reality, BBC News." [Online]. Available: https://www.bbc.com/news/technology-35220974

[17] J. B. E. 19th April 2017, "6.8% of smartphones are ready for vr." [Online]. Available: https://www.gamesindustry.biz/articles/2017-04-19-6-8-percent-of-smartphones-are-ready-for-vr

[18] "PCs Installed Base Worldwide 2013-2019." [Online]. Available: https://www.statista.com/statistics/610271/worldwide-personal-computers-installed-base

[19] L. Zou, T. Bi, and G. Muntean, "A dash-based adaptive multiple sensorial content delivery solution for improved user quality of experience," *IEEE Access*, vol. 7, pp. 89 172–89 187, 2019.

[20] A. Polakovic, R. Vargic, and G. Rozinaj, "Adaptive Multimedia Content Delivery in 5G Networks using DASH and Saliency Information," in *Proc. Int. Conf. on Systems, Signals, and Image Processing*, 2018.

[21] D. You, E. Jeong, and D. H. Kim, "Mpeg-dash mpd for tile-based hybrid stereoscopic 360-degree video streaming," in *Proc. Int. Conf. on Ubiquitous and Future Networks (ICUFN)*, 2018, pp. 682–684.

[22] D. Podborski, E. Thomas, M. M. Hannuksela, S. Oh, T. Stockhammer, and S. Pham, "360-degree video streaming with mpeg-dash," *SMPTE Motion Imaging Journal*, vol. 127, no. 7, pp. 20–27, 2018.

[23] B. Maclntyre and T. F. Smith, "Thoughts on the future of webxr and the immersive web," in *Proc. IEEE Int. Symposium on Mixed and Augmented Reality Adjunct (ISMAR)*, 2018, pp. 338–342.

[24] "Omnitone Spatial Audio Rendering on the Web." [Online]. Available: https://googlechrome.github.io/omnitone/home

[25] S. A. Sarker, M. Rahman, N. Muslim, and S. Islam, "Performance analysis of video streaming at the edge and core cloud," in *Proc. Int. Conf. on Networking, Systems and Security (NSysS)*, 2018, pp. 1–7.

[26] M. Stanton, T. Hartley, F. Loizides, and A. Worrallo, "Dual-mode user interfaces for web based interactive 3d virtual environments using three.js," in *Human-Computer Interaction – INTERACT 2017*, R. Bernhaupt, G. Dalvi, A. Joshi, D. K. Balkrishan, J. O'Neill, and M. Winckler, Eds. Springer International Publishing, 2017, pp. 441–444.

[27] A. Hirway, Y. Qiao, and N. Murray, "A qoe and visual attention evaluation on the influence of spatial audio in 360 videos," in *Proc. IEEE Int. Conf. on Artificial Intelligence and Virtual Reality (AIVR)*, 2020, pp. 345–350.

[28] M. S. Anwar, J. Wang, A. Ullah, W. Khan, Z. Li, and S. Ahmad, "User profile analysis for enhancing qoe of 360 panoramic video in virtual reality environment," in *2018 Int. Conf. on Virtual Reality and Visualization (ICVRV)*, 2018, pp. 106–111.

[29] J. Zhang, Y. Zhong, Y. Han, D. Li, C. Yu, and J. Mo, "A 360° video adaptive streaming scheme based on multiple video qualities," in *Proc. IEEE/ACM Int. Conf. on Utility and Cloud Computing (UCC)*, 2020, pp. 402–407.

[30] M. Montagud, I. Fraile, J. A. Nuñez, and S. Fernández, "Imac: Enabling immersive, accessible and personalized media experiences," in *Proc. ACM Int. Conf. on Interactive Experiences for TV and Online Video (TVX)*, 2018, p. 245–250.