"A Crowdsourcing-based QoE Evaluation of an Immersive VR Autonomous Driving Experience"

Guilherme Daniel Gomes
Software Research Institute
Athlone Institute of Technology
Athlone, Ireland
g.dgomes@research.ait.ie

Dr. Ronan Flynn
Software Research Institute
Athlone Institute of Technology
Athlone, Ireland
rflynn@ait.ie

Dr. Niall Murray
Software Research Institute
Athlone Institute of Technology
Athlone, Ireland
nmurray@research.ait.ie

Abstract— **Due to COVID-19, crowdsourcing has gained momentum as an alternative methodology for continuing research and for Quality of Experience (QoE) assessment. Employing this approach, we remotely evaluated the user perceived QoE of two different visual rendering formats as part of an Autonomous Vehicles (AVs) simulation. The aim was to investigate the participant's QoE when testing AV technology in distinct visual rendering qualities (*low-poly vs high-poly*) of an online streamed 360º car riding experience. In addition, a scoring model based on the expected reliability of each level of the remote assessment was designed. Findings suggest that the consumer's preferences towards the adoption of AV technology is highly determined by the system and human effects on Influence Factors (IFs). Moreover, the adequacy of reliability into a mathematical model is highlighted as a potential turning point for QoE assessment, by carrying out the evaluation tasks from the laboratory environment into the internet, particularly relevant in pandemic times.**

Keywords—Quality of Experience, Crowdsourcing, Autonomous Vehicles, Virtual Reality, COVID-19, Visual quality, Photogrammetry

I. INTRODUCTION

Information and Communications Technology (ICT) has been chosen as common ground on different fields because of the pandemic. As pointed out by [1], online-based products and computer services have found new users during this period. For instance, the usage of Virtual Reality (VR) as a replacement of real-life experiences in tourism was researched by [2], [3] and [4], besides the familiar gaming and training purposes in this area [5] [6] [7]. In addition, the digital infrastructure was studied to minimize the overall impact on education during the transition from traditional face-to-face classes to online setups in this time [8] [9].

Research community was also affected by the pandemic as concerns have been raised regarding ethical issues while conducting research with humans. The closure of academic institutions as a result of the lockdown measures have directly affected ongoing research at the same time that laboratory protocols have been reviewed to ensure safe participant assessments [10] [11] [12]. For that reason, remote research has been adopted as a substitute to the laboratory environment. To achieve that, many studies have welcomed online assessment formats as mainstream methodology to remotely collect participant's data, as seen in [13] [14] [15].

In this context, the obtainment of data from a multitude can be sourced via the Internet. That is previewed by crowdsourcing, which is defined by [16] as "the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, large group of people, generally in the form of an open call". It has been successfully applied to different areas, especially in on-demand workforce focused on performing tasks virtually (e.g., Amazon Mechanical Turk [17]). In addition, crowdsourcing has also been used to collect data regarding people's opinion about technologies, including those used in AVs (also known as self-driving cars), the background of this research. A survey conducted by [18] with respondents from 112 countries investigated people's opinion about automated driving. In [19], people's preferences on auditory, visual and vibrotactile Take-Over Requests (TORs) when experiencing automated driving were investigated using the same remotely-addressed assessment configuration.

QoE evaluation has also been employed in the crowdsourcing format [20] [21]. The human experience is a stream of perceptions constantly weighted by the judgment of the quality characteristics of each of the independent entities involved [22]. With the new challenges imposed by the pandemic (e.g., limiting the access of subjects and conventional evaluation metrics to conduct on site research tests) an opportunity for an in-depth understanding of the crowdsourcing aspects on remotely-delivered QoE experiments in AV field has arisen. To be more specific, the multimedia existent in AV simulations can be remotely assessed from a QoE perspective and crowdsourced.

This paper addresses this type of assessment on a VR based simulation focused on investigating the impact of visual quality on overall participant's perception of AV technology. After a related work discussion, the methodology presents the 360º video formats that were adopted, allowing participants to interact with the experience using their own computer displays at home – no Head Mounted Display (HMD) required. To validate the crowdsourced data, reliability checks were designed to keep research integrity. Results indicate statistical relevant findings on the score system designed for that purpose, besides substantiating the effects of Influence Factors (IFs) (e.g., human, system and context [23]) on AV perception according to the visual quality format adopted.

II. RELATED WORK

This research is based on two pillars: (A) the crowdsourcing format employed to QoE assessment and (B) the use of VR to address distinct visual qualities applied to AV simulations.

A. Crowdsourcing in QoE

The use of online methods to solicit responses from users has gained significant attention from the research community in recent years: it allows researchers to possibly access large numbers of users (over 3.4 billion people connected to the internet [24]); allows to capture larger and more representative models of user perception [16]; and enables new possibilities for evaluation tasks far beyond what is possible in a laboratory environment. By all means, real-life conditions and a diverse population were made possible to take part in research campaigns with the access of a global pool of subjects [25]. A foundation about crowdsourcing was defined by [26] with recommendations proposed by [27]. Based on a literature review towards integrating the terms used by different authors, the elements of a crowdsourcing research (Fig. 1) are:

1) *Process:* it is the outsourcing operation in the direction of solving a specific problem. It is inferred from [16] as the outcome of an activity directed to a multitude.
2) *Crowd:* composed by a group of people, which number and quality will depend on the evaluation task [28]. The crowd will undertake the test towards solving the problem defined by the crowdsourcing business. It can be voluntary contribution or paid.
3) *Initiator:* the person, company or organization that requests the crowdsourcing process. Also called "crowdsourcer".

There are various approaches to solicit engagement in crowdsourcing such as: financial and entertainment compensation, besides social recognition [26]. The investigations conducted among researchers are usually returned in a payback format (e.g., people that are entitled to participate as crowd in one project might also be conducting a crowdsourcing study as an initiator in another one). From the entertainment perspective, the characteristics inherent to the process might influence the human factor curiosity, as studied by [29] within the gaming background.

In that context, "perceived quality" is a central aspect when evaluating an application using a QoE approach [22]. The quality-relevant standards, according to the type of the application, depends on the perceptual and cognitive processes that underlies the quality formation on the person that experiences an application [30]. From this perspective, tasks delivered remotely needs to be evaluated considering the properties and characteristics inherent to the application, besides the context that the person is included. IFs can be directly addressed to the process and the crowd in the crowdsourcing research.

Crowdsourcing has brought to QoE research the possibility of evaluating tasks outside laboratories (i.e., "in the wild"). The impact of the everyday life experience is transferred to the application, making check mechanisms necessary to guarantee the trustworthiness of the crowdsourcing research. The design of the experiment should never incentivise participants to cheat. That includes the split of the application into smaller tests, avoiding too difficult questions. Moreover, application layer monitoring should be included in the test to detect contradictions and possible outliers. The framework that contains the set of procedures that includes these safety checks are defined by [25] as Reliability Mechanisms.

B. Virtual Reality and Autonomous Vehicles simulations

Virtual Reality (VR) is often used in scientific contexts to provide simulations and get replicable results. The user of this type of application has a sense of presence with possibilities to explore the space and interact with elements of the environment [7]. The use of VR to create simulated environments stands as an alternative to real-world data collection [31]. It has been used to both artificial intelligence training models and user assessment, as it is capable to replicate real world conditions and collect measures repeated times [32].

Immersion is a key component of VR environments. It is stated by [33] that users feel and behave in a different manner in immersive applications. A successful example is the VR phobia therapy. It works by exposing the participant to an uneasy situation (e.g., a patient in a VR conference to treat the fear of speaking in public, or in a VR mountain to treat the fear of height). Immersive VR applications have also been used in military training [34], entertainment and medical fields [35]. One of the successful components of immersion is realism. High-fidelity stimuli is directly related to a realistic experience in the simulated environment [33]. In transportation research, however, the realism component is not often included and has not been evaluated from an immersion perspective or delivered in a standardized setup, as can be observed in [36] [37] [38] [39].

Simulations in VR have been applied in the development of AVs [42] and used for AV virtual testing [43] [44] [45]. A common problem related to VR is the range of symptoms that might be perceived during and after experiencing the application, like disorientation, headaches, eye strain, nausea, sweating and others. The difference between the movement speed perceived in the virtual environment (e.g., vehicle steering and acceleration) and the actual body steadiness (e.g., users in front of computer undertaking the VR experience) might result in cybersickness. By gradually introducing participants to virtual environments, adaptation programs can be used as an approach to reduce this effect [40].

Although previous studies focused on the impact of graphics quality comparing how different types of content presentation influence user perceptual quality and immersion [46] [47], the differences in the user's explicit responses while experiencing AV technology in VR simulations were not assessed from a QoE perspective. In addition, crowdsourcing applied to evaluate AV simulations is still an unexplored field. From the system perspective, it has been applied in transportation network systems to gather data from multiple users and build datasets used in autonomous systems [48]. However, crowdsourcing has mostly been used to outsource data to promote and understand other entities but not so frequently from a user-centered perspective to assemble the shape of human behaviour when evaluating novel technologies applied to simulations.

III. Methodology

The initial phase of this project has focused on designing a VR autonomous driving experience based on a real street of Athlone, Ireland. In order to establish a link between

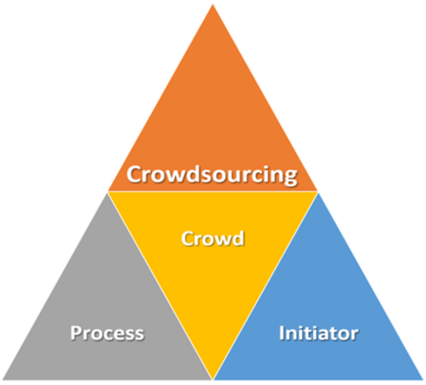

Fig. 1- Pyramids representing the elements of Crowdsourcing.

the quality formation process that involves the participant's consciousness when experiencing AV technology with VR simulations, a between-subject design comparison of distinct video renderings was made. By using a conventional (*low-poly*, game-based) [49] and a realistic (*high-poly*, photogrammetry-based) [50] concepts to create 3D visual content, the impact of realism was addressed in the AV ride (Fig. 2). The focus was to investigate the influence of visual quality on user QoE perception of AV-based technology in a VR simulation.

The research methodology adopted was experimental and designed to happen virtually, by virtue of current restrictions to assess subjects in the laboratory environment due to COVID-19. The *recruitment phase* consisted of a *crowd* formed by a diverse population (N=135; 62 male and 73 female, ranging from age 16 to 60, from 24 countries) recruited from crowdsourcing groups on social networks and through an online survey platform [51]. In the *information phase*, participants were informed how the tests would be conducted, the duration time (approximately 15 minutes, including questionnaire responses) and how they would interact with the VR content. They could only proceed if they agreed to take part in the study. In the *screening phase*, they were requested to undertake an online screening tool from Zeiss [52] in order to check their visual acuity, video contrast and colour blindness.

With the objective to reduce cybersickness symptoms by delivering an "adaptation process to VR" described by [40] a training framework was developed based on the simulation characteristics (e.g., street shapes, vehicle movement) and addressed in the *training phase.* This stage was also used to perform a system check and participant's performance, described in Section B. Finally, participants were invited to undertake the simulation in the *testing phase.* Post-experience questionnaires were addressed as subjective metrics that consisted of Likert-scale responses based on immersion [53], user experience [54], cybersickness [55], emotions [56] and memory-related statements. Additional resources were required in order to deliver the experiment in the crowdsourcing format and will be described in the next sections.

A. *Online resources*

Google Forms [57] was used as framework to perform the online surveys and crowdsourcing. The crowdsourcing *process* consisted of content used for assessment that was hosted in Google Forms. Participants were randomly assigned to each experience through a script developed in a web hosting service provided by GitHub [58], which supplied a link to shuffle the item order and the experiment type. YouTube [59] was used to deliver the 360º video renderings of the AV simulation. Besides not requiring user accounts to watch the hosted videos, the platform is well-known for internet users and was chosen to facilitate the delivery of the crowdsourced task: take part in the AV ride. Both *training phase* and *testing phase* consisted of 360º videos in the same format and quality for each of the groups and were hosted in the platform. However, the *training phase* was designed to guarantee that all participants would know how to navigate in the scenario and would receive content with the same performance in the test video. For that reason, basic navigation instructions were prompted in a mock scenario followed by the display of letters calibrated to be seen only within the range of resolution quality accepted for the experiment (Fig. 3). Participants were allowed to run the training video more than once to calibrate their devices to the optimal aspects and were requested to inform if they had navigation issues caused by computer or network.

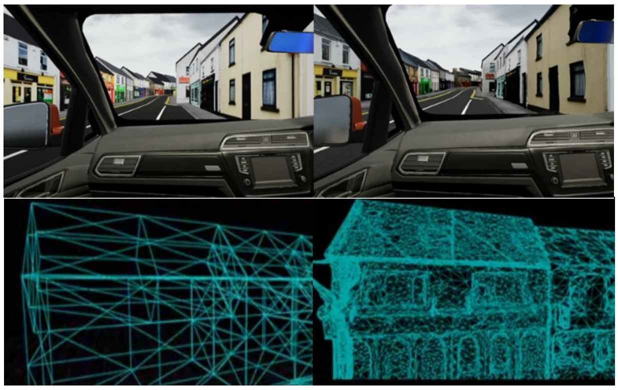

Fig. 2- Low-poly vs high-poly addressing the component "realism" of the VR simulation.

B. *Reliability Mechanisms*

Reliability Mechanisms are part of the crowdsourcing methodology to reliably collect data from remotely delivered applications. It refers to all the development focused on guaranteeing the replicability of the study and it is intended to filter out unreliable data used in QoE crowdsourcing evaluation [25]. These mechanisms were arranged in "safety checks" and considered on each step of the test design (Table 1). To rate the data quality outsourced by the crowd after undertaking the experiment and considering the safety checks, the Reliability Score (1) is proposed:

$$RS = \sum_{n=1}^{gp} \frac{v_c c_t (R_{ES} + \rho)}{c_p} \qquad (1)$$

where:

- v_c= vision check, v_c={0,1};
- c_t= contradictions, c_t={0,1};
- R_{ES} = resolution, R_{ES}={6, 4, 0};
- ρ = training score, ρ={4, 2, 1, 0, -5};
- c_p = computer performance, c_p={2, 1};

TABLE I. SAFETY CHECKS

Condition	Reliability Mechanisms
	Safety check description
Competency	Participants were asked in the consent form if they did not have any pre-existing condition that could interfere the completion of the tasks.
Visual acuity	Participants were asked in the screening phase to confirm if they have passed the visual acuity check and colour vision check.
Navigation and screen resolution	Participants had to respond the correct letters they have seen during the training section. Participants had to inform the chosen resolution and if they had performance issues.
Contradiction	Participants had to confirm they have understood the instructions before proceeding to the next page. Questions were asked more than once with inverted meanings.
Context	Participants were asked to proceed with the experiment only if they were in a quiet environment. They were asked to close any other application.
System performance	Participants were asked to evaluate their network/computer performance based on the experience during the training section and respond about any issues.

The detection of an outlier null sets the vision check and/or contradiction variables, zeroing the Reliability Score (RS) as it means the participant's vision acuity or contradictory inputs might interfere in overall perceived QoE of the simulation. For instance, it was asked if participants know the real place where the simulation took place and if they know the town centre, in two different questions. It would be impossible to personally know the place, but do not know the town (as the place is inside the town) and that null sets the contradiction variable, for example.

The other metrics designed to be used in the RS could only reduce the overall mark of the participant. The application was designed to be delivered in only two computer resolution options: "WQHD" (as an option for low-performance devices) and 2160s60 "4K" (as standard choice). The resolution variable "R_{ES}" was included to ensure that the inputs from the participants about computer performance matched with the expected output of the application. In addition, two other variables were addressed: "ρ" as training score to detect if participants were quick enough to click-and-drag on the screen, follow the prompts and inform which letters they saw (representing elements of the environment they could get involved during the *testing phase*) and "c_p" as computer performance score to divide the RS by two if participants deliberately informed that they had performance issues during the training section (representing choppy video playback due to internet or computer issues). The training score could range from positive to negative values if participants cheated about the letters they have seen (ρ = -5).

The summation index of the RS represents the participant, while the upper limit represents the maximum number of subjects on each group. Participants with a low score (0 < RS ≤ 4) were arranged into "Group 3" (N= 20) while participants with a medium score (5 ≤ RS ≤ 7) were arranged into "Group 2" (N=33) and participants with a high score (8 ≤ RS ≤ 10) were arranged into "Group 1" (N=62). Participants with a fail score (RS ≤ 0) were eliminated from the statistical analysis (N= 20).

Fig. 3- Training section and zoomed letters representing the minimum resolution requirements (1) and outside limits (2).

IV. RESULTS

The data collected from explicit metrics addressing QoE were statistically analyzed across groups. A one-way ANOVA test was conducted to compare the 3 groups based on the RS results. The question (a) *"From 0 to 10 how much would you like to try AV technology in real life?"* addressed the technology acceptance factor on the perceived experience; (b) *"How do you feel? (dominance)"* addressed emotions based on the Self-Assessment Manikin (SAM) questionnaire [56] and (c) represents the number of times each participant watched the training video *versus* the score they have obtained (Fig. 4). To make sure that the homogeneity of variances is met, these groups were first compared to have equal population variances. The Levene Statistic results held for each of the conditions tested confirms that the variables do not violate the homogeneity of variances assumption: (a) $F(2,113) = 0.244$, $p = 0.784$; (b) $F(2,113) = 0.532$, $p = 0.589$; (c) $F(2,108) = 1.836$, $p = 0.164$.

There was a statistically significant difference between groups on (a), determined by one-way ANOVA ($F(2,113) = 3.995$, $p = 0.021$). At 0.95 confidence level, a post-hoc test (Bonferroni) was conducted and highlighted the difference between Group 1 (MD = 1.543, S.E. = 0.575, p = 0.025) and Group 2 (MD = 1.583, S.E. = 0.635, p = 0.043) when compared with Group 3. Results indicate that participants with a higher reliability score had better overall QoE on expressing their wishes in terms of trying the technology in real life one day, despite the visual quality they have tried in simulation. That might indicate that the awareness about the tasks that comprise the crowdsourcing process is as important as the quality of the application that they receive.

The Self-Assessment Manikin scale showed significant statistical data when evaluated with the non-parametric Mann-Whitney U Test for independent variables (U = 1338.5, Z = -2.023, p = 0.043) with higher dominance ratings for the *high-poly* group (M = 64.06) *vs* the *low-poly* one (M = 52.34). When compared with the RS groups, there were no statistically significant differences between group means as determined by one-way ANOVA ($F(2,113) = 1.488$, $p = 0.230$). However, a significant number of outliers in this category was detected in the *low-poly* group and can be observed in (b). It is possible to spot that participants with a higher reliability score had consistent dominance values.

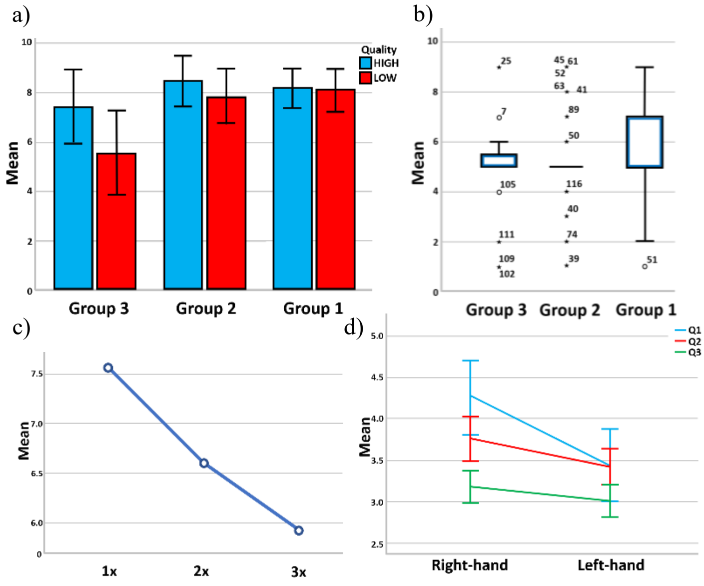

Fig. 4- Means across reliability groups: a) willingness to adopt AV technology; b) dominance after the simulation; c) RS vs training video tryouts; d)right-hand drive vs left-hand drive in Q1,Q2 and Q3.

It is important to highlight a statistical significance on (c) (F(2, 113) = 3.411 with p = 0.036) in the association of the RS values with the number of times participants have tried the training video. Even though there is no association with the number of times participants tried out the *training phase* with the equation (1), there was a strong correlation with the statistical findings, which stands as positive evidence for the validity of the RS. It can be inferred from the results obtained that participants are likely to have a lower RS if they spend more time watching videos in the training section. That might represent the number of adjustments in resolution/video cache time because of device performance/network issues.

The link between the quality-formation process and the perceptual references that are present in different levels of memory [22] can be noticed in (d) (Fig. 4). Questions like Q1: *"Somehow I felt that the virtual world surrounded me"* had higher mean ranks for people used to right-hand driving style (M = 66.09 *vs* M = 52.75) when evaluated with the Mann-Whitney U Test for independent variables (U = 1209.5, Z = -2.493, p = 0.012). The same happened when participants were asked Q2: *"What is your opinion about this statement: this simulation seems to be a reliable method for me to have an idea about how AV will go on real streets one day"*, (U = 1270.5, Z = -2.298, p = 0.021, "Right-Hand" M = 66.09, "Left-Hand" M 52.75) and Q3: *"In my opinion, the Autonomous Vehicle (AV) during the simulation worked, in terms of performance"* (U = 1276.5, Z = -2.124, p = 0.034). That indicates a tendence of higher acceptance over AV technologies when they are presented in familiar memory-based concepts, highlighting the impact of human IFs.

The same evidence can be observed when participants were asked to opine about if they *"had the feeling the car was driving in the wrong-way on the two-way street"* when compared with the group of participants that know the town, and those that do not. The clustered representation of the reliability groups show more consistent data and smaller error (at 0.95 confidence level) for Group 1 (Fig. 5). It is assumed from [23] that the memory-related HIFs alters the perception of participants that are used to the right-hand driving style

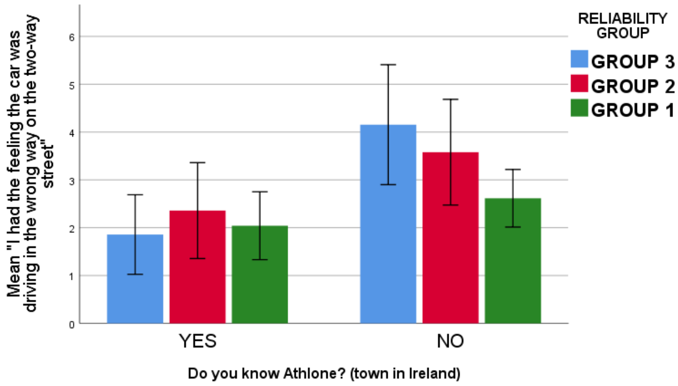

Fig. 5- Mean across the group of participants that know the town vs those that do not know (impact of memory-based HIFs).

(town streets represented in the simulation) instead of the left-hand one (used in other countries, causing the feeling that the car in the simulation was driving in the wrong way).

V. Conclusion

Crowdsourcing research is challenging because of the number of uncontrolled variables that exist. However, it has proven to facilitate the access to an infinite number of people. When the evaluation involves quantities that depend on state, performance and context of multiple objects, the reliability of the research might get affected specially from a remote assessment perspective. Results indicate that participant's performance is highly linked with the proposed individual RS scores, determined by System Influence Factors or SIFs (including user's device resolution and computer performance) and Human Influence Factors or HIFs (including screening results, contradictions when answering the questionnaires, low interaction input in training section and long-term memory effects, mainly the perceptually referenced ones). It indicates that the development of metrics based on safety checks during the design of remotely-delivered experiments can be used to filter out unreliable data from outliers, besides highlighting an influence of QoE on AV technology acceptance. Future work will investigate the results obtained in this paper with the laboratory approach on the same AV simulations, in order to validate the findings with objective metrics and exploit the effects of gender, age-related differences and affinity with VR and real AV technology in both traditional and crowdsourcing means. In addition, immersion and cybersickness will be further analysed by delivering the experiment using HMDs and comparing results with the current data.

References

[1] M. Nicola and e. al, "The socio-economic implications of the coronavirus pandemic (COVID-19): A review," *International Journal of Surgery,* vol. 78, pp. 185-193, 2020.

[2] A. F. Schiopu, R. I. Hornoiu, M. A. Padurean and A.-M. Nica, "Virus tinged? Exploring the facets of Virtual Reality use in tourism as a result of the COVID-19 pandemic," *Telematics and Informatics 60,* 2021.

[3] A. O. J. Kwok and S. G. M. Koh, "COVID-19 and Extended Reality (XR)," *Current Issues in Tourism,* 2020.

[4] D. Sarkady, L. Neuburger and R. Egger, "Virtual Reality as a travel substitution tool during COVID-19," in *ICT in Tourism 2021*, 2021.

[5] F. Pallavicini, A. Pepe and M. E. Minissi, "Gaming in Virtual Reality: what changes in terms of usability, emotional response and sense of presence compared to non-immersive video games?," *Simulation & Gaming,* vol. 50, no. 2, pp. 136-159, 2019.

[6] V. S. Pantelidis, "Reasons to use Virtual Reality in education and training courses and a model to determine when to use Virtual Reality," *Themes in Science and Technology Education,* vol. 2, no. 1, 2010.

[7] J. Psotka, "Immersive training systems: Virtual Reality and education and training," *Instructional Science,* vol. 23, pp. 405-431, 1995.

[8] R. C. Chick, G. T. Clifton, K. M. Peace, B. W. Propper, D. F. Hale, A. A. Alseidi and T. J. Vreeland, "Using technology to maintain the education of residents during the COVID-19 pandemic," *Journal of Surgical Education,* vol. 77, no. 4, 2020.

[9] T. Chen, L. Peng, X. Yin, J. Rong, J. Yang and G. Cong, "Analysis of user satisfaction with online education platforms in China during the COVID-19 pandemic," *MDPI Healthcare,* vol. 8, no. 3, 2020.

[10] S. Government, "Coronavirus (COVID-19): guidance for laboratories and research facilities," [Online]. Available: https://www.gov.scot/publications/coronavirus-covid-19-guidance-for-laboratories-and-research-facilities/. [Accessed January 2021].

[11] U. K. Governament, "Working safely during coronavirus," [Online]. Available: https://www.gov.uk/guidance/working-safely-during-coronavirus-covid-19/labs-and-research-facilities#national-lockdown. [Accessed January 2021].

[12] H. University, "Planning for reopening research facilities," [Online]. Available: https://provost.harvard.edu/research-laboratory-reentry. [Accessed January 2021].

[13] K. Votis, Z. Kozlakidis and P. Katsaounis, "CoronaBio: using crowdsourcing for biomedical research on COVID-19 to manage a pandemic," *Innovations in Digital Health, Diagnostics and Biomarkers,* vol. 1, no. 1, 2021.

[14] C. Elphick, A. Stuart, R. Philpot, Z. Walkington, L. Frumkin, M. Zhang, M. Levine, B. Price, G. Pike, B. Nuseibeh and A. Bandara, *Altruism and anxiety: engagement with online community support initiatives during COVID-19 lockdown in the UK and Ireland,* 2020.

[15] H. P. Dijkstra, E. Ergen, L. Holtzhausen, I. Beasley, J. M. Alonso, L. Geertsema, C. Geertsema, S. Nelis, A. S. H. Ngai, I. Stankovic, S. Targett and T. E. Andersen, "Remote assessment in sport and exercise medicine (SEM): a narrative review and teleSEM solutions for and beyond the COVID-19 pandemic," *British Journal of Sports Medicine,* vol. 54, no. 19.

[16] J. Howe, "The rise of crowdsourcing," *Wired magazine,* 2006.

[17] "Amazon Mechanical Turk," Amazon, [Online]. Available: https://www.mturk.com/. [Accessed January 2021].

[18] P. Bazilinskyy, M. Kyriakidis and J. d. Winter, "An international crowdsourcing study into people's statements on fully automated driving," *Procedia Manuacturing,* vol. 3, pp. 2534-2542, 2015.

[19] P. Bazilinskyy, S. Petermeijer, V. Petrovych, D. Dodou and J. d. Winter, "Take-over requests in hight automated sriving: a crowdsourcing survey on auditory, vibrotactile and visual displays," *Transportation Research Part F: Traffic Psychology and Behaviour,* vol. 56, pp. 82-98, 2018.

[20] K.-T. Chen, C.-J. Chang, C.-C. Wu, Y.-C. Chang and C.-L. Lei', "Quadrant of Euphoria: a crowsourcing platform for QoE assessment," *IEEE Network,* 2010.

[21] O. Alonso, D. E. Rose and B. Stewart, "Crowdsourcing for Relevance Evaluation," *ACM Special Interest Group on Information Retrieval,* vol. 42, no. 2, 2008.

[22] A. Raake and S. Egger, "Chapter 2: Quality and Quality of Experience," in *Quality of Experience - Advanced Concepts Applications and Methods*, Springer, 2014, pp. 22-26.

[23] U. Reiter, K. Brunnström, K. D. Moor, M.-C. Larabi, M. Pereira, A. Pinheiro, J. You and A. Zgank, "Chapter 4: Factors influencing Quality of Experience," in *Quality of Experience - Advanced Concepts Applications and Methods*, Springer, 2014, pp. 64-67.

[24] M. Roser, H. Ritchie and E. Ortiz-Ospina, "Internet," 2015. [Online]. Available: https://ourworldindata.org/internet. [Accessed 2020].

[25] T. Hoßfeld and C. Keimel, "Chapter 21: Crowdsourcing in QoE Evaluation," in *Quality of Experience - Advanced Concepts Applications and Methods*, Springer, 2014.

[26] E. Estellés-Arolas and F. González-Ladrón-de-Guevara, "Towards an integrated crowdsourcing definition," *Journal of Information Science,* 2012.

[27] International Telecommunication Union, "Recommendation E.812," ITU-T, Geneva, 2020.

[28] G. L. Vecchia and A. Cisternino, "Collaborative workforce, business process crowdsourcing as an alternative of BPO," in *ICWE*, 2010.

[29] J. Beyer and S. Möller, "Chapter 25 - Gaming," in *Quality of Experience - Advanced Concepts Applications and Methods*, Springer, 2013.

[30] S. Möller, M. Wältermann and M.-N. Garcia, "Chapter 5: Features of Quality of Experience," in *Quality of Experience - Advanced Concepts, Applications and Methods*, 2014.

[31] E. Coelingh and J. Nilsson, "Creating driving tests for self-driving cars," Spectrum IEEE, 2018. [Online]. Available: https://spectrum.ieee.org/transportation/self-driving/creating-driving-tests-for-selfdriving-cars. [Accessed September 2020].

[32] T. Sulkowski, P. Bugiel and J. Izydorczyk, "In search of the ultimate autonomous driving simulator," in *International COnference on Signals and Electronic Systems*, 2018.

[33] D. A. Bowman and R. P. McMahan, "Virtual Reality: how much immersion is enough?," *IEEE Computer,* vol. 40, no. 7, pp. 36-43, 2007.

[34] U.S. Army Research Laboratory, "Army researchers explore benefits of immersive technology for soldiers," 2019. [Online]. Available: https://www.sciencedaily.com/releases/2019/01/190116115523.htm. [Accessed October 2020].

[35] A. J. Snoswell and C. L. Snoswell, "Immersive Virtual Reality in health care: systematic review of technology and disease states," *JMIR Biomedical Engineering,* vol. 4, no. 1, 2019.

[36] J. Koo, J. Kwac, W. Ju and M. Steinert, "Why did my car just do that? Explaining semi-autonomous driving actions to improve driver understanding, trust, and performance," *International Journal for Interactive Design and Manufacturing,* vol. 9, no. 4, 2014.

[37] S. Maurer, R. Erbach, I. Kraiem and S. Kuhnert, "Designing a guardian angel: giving an Automated Vehicle the possibility to override its driver," in *10th International Conference of Automotive User Interfaces (AutoUI2018)*, 2018.

[38] M. Walch, T. Sieber, P. Hock and M. R. K. Baumann, "Towards cooperative driving: involving the driver in an Autonomous Vehicle's decision making," in *AutomotiveUI'16*, 2016.

[39] B. D. Seppelt and J. D. Lee, "Keeping the driver in the loop: dynamic feedback to support appropriate use of imperfect vehicle control automation," *International Journal of Human-Computer Studies,* vol. 125, pp. 66-80, 2019.

[40] J. J. L. Jr., "A discussion of cybersickness in virtual environments," *ACM SIGCHI Bulletin,* vol. 32, no. 1, pp. 47-56, 2000.

[41] A. Stevenson and C. Soanes, Oxford Dictionary of English, 2003.

[42] S. Shah, D. Dey, C. Lovett and A. Kapoor, "AirSim: High-Fidelity visual and physical simulation for Autonomous Vehicles," *Microsoft Research,* pp. 621-635, 2018.

[43] E. Dogan, M.-C. Rahal, R. Deborne, P. Delhomme, A. Kemeny and J. Perrin, "Transition of control in a partially automated vehicle: effects of anticipation and non-driving-related task involvement," *Transportation Research Part F,* vol. 46, pp. 205-215, 2017.

[44] A. Koohestani, P. Kebria, A. Khosravi and S. Nahavandi, "Drivers' performance evaluation using physiological measurement in a driving simulator," in *Digital Image Computing: Techniques and Applications (DICTA)*, 2018.

[45] D. Sportillo, A. Paljic, M. Boukhris, P. Fuchs, L. Ojeda and V. Roussarie, "An immersive Virtual Reality system for semi-autonomous driving simulation: a comparison between realistic and 6-DoF controller-based interaction," in *9th International Conference on Computer and Automation Engineering*, 2017.

[46] W. B. Thompson, P. Willemsen, A. A. Gooch and S. H. Creem-Regehr, "Does the quality of the computer graphics matter when judging distances in visually immersive environments," *Presence Teleoperators & Virtual Environments,* vol. 13, no. 5, pp. 560-571, 2004.

[47] C. Klimmit and T. Hartmann, "Dimensions and determinants of the enjoyment of playing digital games: a three level model," in *Level Up - Digital Games Research Conference*, 2006.

[48] O. Dabeer, W. Ding, R. Gowaiker, S. K. Grzechnik, M. J. Lakshman and S. L. e. al., "An end-to-end system for crowdsourced 3D maps for Autonomous Vehicles: the mapping component," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017.

[49] S. Melax, "A simple, fast and effective polygon reduction algorithm," *Game Developer Magazine,* November 1998.

[50] N. Statham, "Use of photogrammetry in video games: a historical overview," *Games and Culture,* vol. 19, no. 1, 2018.

[51] SurveyCircle, 2016. [Online]. Available: https://www.surveycircle.com. [Accessed 2020].

[52] Zeiss, "ZEISS Online Visions Screening," [Online]. Available: https://www.zeiss.com/vision-care/int/better-vision/zeiss-online-vision-screening-check.html. [Accessed August 2020].

[53] T. Schubert, F. Friedmann and H. Regenbrecht, "The experience of presence: factor analytic insights," *Presence Teleoperators & Virtual Environments,* vol. 10, pp. 266-281, 2001.

[54] B. Laugwitz, T. Held and M. Schrepp, "Construction and evaluation of a User Experience Questionnaire," in *HCI and Usability for Education and Work*, Graz, 2008.

[55] R. S. Kennedy, N. E. Lane, K. S. Berbaum and M. G. Lilienthal, "Simulator Sickness Questionnaire: an enhanced method for quantifying simulator sickness," *The International Journal of Aviation Psychology,* vol. 3, no. 3, pp. 203-220, 2009.

[56] J. D. Morris, "Observations SAM: The Self-Assessment Manikin an efficient cross-cultural measurement of emotional response," *Journal of Advertising Research,* vol. 35, no. 6, pp. 63-68, 1995.

[57] Google Inc., "How to use Google Forms," [Online]. Available: https://support.google.com/docs/answer/6281888?hl=en&ref_topic=6063584. [Accessed August 2020].

[58] Microsoft, "GitHub," [Online]. Available: http://www.github.com. [Accessed 2020].

[59] Google Inc., "G Suite," [Online]. Available: http://gsuite.google.com. [Accessed August 2020].