https://doi.org/10.2352/EI.2023.35.8.IQSP-261

Proposing More Ecologically-Valid Experiment Protocol Using YouTube Platform

Wielgus Gabriela, *AGH University of Science and Technology Cracow, Poland*
Janowski Lucjan, *AGH University of Science and Technology Cracow, Poland*
Koniuch Kamil, *AGH University of Science and Technology Cracow, Poland*
Leszczuk Mikołaj, *AGH University of Science and Technology Cracow, Poland*
Figlus Rafał, *AGH University of Science and Technology Cracow, Poland*

Abstract

Video streaming is becoming increasingly popular, and with platforms like YouTube, users do not watch the video passively but seek, pause, and read the comments. The popularity of video services is possible due to the development of compression and quality prediction algorithms. However, those algorithms are developed based on classic experiments, which are non-ecologically valid. Therefore, classic experiments do not mimic real user interaction. Further development of the quality and compression algorithms depends on the results coming from ecologically-valid experiments. Therefore, we aim to propose such experiments. Nevertheless, proposing a new experimental protocol is difficult, especially when there is no limitation on content selection and control of the video. The freedom makes data analysis more challenging. In this paper, we present an ecologically-valid experimental protocol in which the subject assessed the quality while freely using YouTube. To achieve this goal, we developed a Chrome extension that collects objective data and allows network manipulation. Our deep data analysis shows a correlation between MOS and objectively measured results such as resolution, which proves that the ecologically-valid test works. Moreover, we have shown significant differences between subjects, allowing for a more detailed understanding, of how the quality influences the interaction with the service.

Introduction

YouTube platform ranks second among the most visited websites globally [9]. High demand for streaming video, makes quality assurance a challenging task. Subjective assessments of video quality are an important part of the optimization process of modern video streaming services. This area of study is called ”Quality of Experience”(QoE). Typically, researchers use standardized laboratory protocol to study users' QoE [1]. On the other hand, researchers use the crowdsourcing method for the assessment of QoE in different contexts [8, 20, 21]. Those approaches differ in terms of internal and external validity.

To obtain results that correspond to the real world, it is important to consider both internal and external validity in the experimental design. Internal validity refers to the accuracy and appropriateness of the experimental design in measuring the factors being studied [16]. External validity, on the other hand, refers to the extent to which the results can be generalized to other populations and contexts [16]. The aspect of external validity describing the generalizability of experimental data to real-life contexts is called ecological validity.

Having the above in mind, we designed an ecologically-valid experimental protocol to receive quality scores from users interacting with a video service naturally.In order, to obtain ecological validity in our experiment, testers had the freedom to use the YouTube platform almost without limitations. No limitation on content selection and control of the video playback allows us to get a closer view of a tester's behavior, and thus obtain results showing the actual use of the service. To collect quality scores and additional data related to user activity, we developed a Chrome extension for recording YouTube statistics, network manipulation, and gathering participants' ratings. Since the proposed methodology is a new test design, we decided to record the user screen and face. The recordings were additional, and important, source of information.

In this paper, we show the results of a study investigating the relationship between Mean Opinion Score (MOS) and objective video quality (resolution). Moreover, we focus on the impact of stalling on subjective qulity ratings. The results demonstrate a high correlation between MOS and resolution, which proves the reliability of ecological validity experiments. Additionally, the study shows stalling was a less negative factor in determining QoE than we assumed. Moreover, during the experiment, participants presented various behaviors including seeking, pausing, manual resolution change, and viewing the comment sections or description of the video.

The rest of the paper is organized as follows. In Section Experiment Design, we describe how the experiment protocol Further, this section includes a description of the developed Chrome extension. In Section Data analysis we describe collected data and user behaviors We grouped participants based on their behaviors and the correlations between MOS and resolution. We also show the influence of stalling on subjective quality ratings. The main contribution of this paper is a new experimental protocol that allows users for natural behavior during the test procedure. A high correlation between resolution and MOS proves the reliability of the conducted experiment. Moreover, the lower influence of stalling suggests that ecologically valid studies might

Related Work

In classical video QoE standardized experiments, 10 seconds long, soundless video clips are used as stimuli[1]. Subjective quality ratings are gathered following the presentation of each stimulus with the Absolute Category Rating (ACR) scale. The same content is displayed multiple times on different degradation levels(e.g.[21]). In our experiment, we use the ACR scale to calculate Mean Opinion Score (MOS) as recommended in ITU-T P.800.1[2]. The main difference between the experiments carried out so far and the new experimental protocol we are proposing is the ecological validity of the experimental setup. Our experimental setup is carried out in the laboratory while allowing the use of the YouTube platform freely. On the other hand, researchers use crowdsourcing as a way to approach ecological validity.

Conducting experiments in a crowd setting can increase the realism of the study and make the results more representative of real-world situations the example of related work is [8] which focuses on the QoE study, including estimation the MOS according to ITU-T P.1203 [3]. Researchers discuss the use of crowdsourcing to assess the quality of experience (QoE) and engagement while watching streaming videos. The study aims to examine the effectiveness of the crowdsourcing method in evaluating the factors that affect QoE and engagement, such as video quality, playback experience, and user interactions. Additionally, researchers demonstrate that the minor rise in YouTube scores at MOS can be attributed to stalling incidents. If there is a stalling event present, the video will not achieve the maximum MOS score.

In the Data analysis section, we detailedly described the observation of various behaviors based on screen recordings of the subjects. Example of related research in the field of user behavior testing can be found in the studies: [6, 10, 11, 12, 13]. The study's conclusion is that the experimental biases may occur when conducting user behavior and QoE assessments in a laboratory setting. The researchers suggested that future studies should be conducted in real-world settings to minimize these biases and to provide more accurate results.

Moreover, studies conducted in the laboratory have extensively investigated the phenomenon of stalling, a sample study is [20]. In the mentioned paper authors present a chart to illustrate the correlation between the frequency of stalling on YouTube and MOS. The chart highlights that even a single occurrence of stalling causes a drastic drop in the subjective evaluation, plummeting to a score of 3.

In order to improve the ecological validity of the results, combining the traditional quality of experience (QoE) experiment with an analysis of behaviors and a comprehensive examination of the network is important. This approach can provide a more comprehensive understanding of the complex relationships between stalling, behavior, and network conditions, and help to generate results that are more representative of real-world situations. By integrating these different aspects, we can gain a deeper insight into the underlying causes of stalling.

Experiment Design

In order, to improve external validity, we designed an experiment based on ecological validity. In this section, we describe the laboratory environment based on ITU-R recommendation and the description of the developed software.

Laboratory Environment

The laboratory environment was prepared according to ITU-R BT.500-14[1]. We specified peak luminance (200cd/m^2), screen illuminance (200 lux), and viewing distance from the screen (1.6H).

The experiment scenario allows the participant to freely use YouTube, while software limits the network. In addition, we recorded screen, mouse movements, and "Statistics for Nerds" similarly as [22]. The duration of the experiment is approximately 90 minutes, during which participants watch videos according to their preferences without any restrictions. A transparent window with the question about interest and quality on a 5-point scale appears every 2.5 minutes. The interest rating levels relate to the question "To what extent are you interested in the content?", where 1 is "boring" and 5 is "interesting". The Absolute Category Rating (ACR) scores describe quality as: 1 – "bad", 2 – "poor", 3 – "fair", 4 – "good", and 5 – "excellent". At the end of the experiment, participants are asked to complete a questionnaire, which we will discuss in more detail in the section Data analysis.

Software

The developed extension consists of two parts: Chrome extension and REST API which runs locally. For each experiment scenario, the extension introduces random bandwidth at constant intervals. Bandwidth manipulation is to simulate conditions as close as possible to the home conditions of people participating in the experiment. The Chrome extension displays a semi-transparent panel with questions about interest, followed by a quality rating. Moreover, the plugin collects data from "Statistics for Nerds" by YouTube, limited bandwidth, as well as responses of interest and quality. The second part of the extension which is a REST API saves all data in a relational database from where are generated to the final results. Communication between the Chrome extension and REST API is based on HTTP requests.

Data analysis

We tested 15 participants, 8 female, and 7 male. The minimum age for the tester was 20 years and the maximum was 46 years. Each participant confirmed that they had used the platform recently, where 27% are intensive users - watching YouTube many times a day, 40% are frequent users - using YouTube several times a day, the rest are regular and occasional users - and used YouTube daily or occasionally.

During the experiment, we collected data using our Chrome extension and screen recordings. Using the Chrome extension, we obtained data directly related to the video playback including viewport, resolution, video timestamp, video ID, and bandwidth. Based on the analysis of the screen recordings, we gain a deeper insight into the tester's behavior. We measured the following behaviors: seeking forward/backward, pause, manual resolution change, changing the size of the player, and viewing the comment section.

First, we are describing the behavior from a high-level perspective, and then we look at the data in more detail to illustrate groups of data based on their behaviors.

Behavior

Initially, we analyzed the behavior of the participants using YouTube Statistics for Nerds, which was captured by our Chrome extension in 0.5-s intervals. The first analysis showed intense participants' interaction with the YouTube service. Thus, we used screen recordings to investigate the efficacy of Statistics for Nerds for capturing behavior. Observational analysis showed important differences between the data captured by our extension and the screen recordings. These differences were generated by behaviors that occurred more frequently than 0.5 s. Therefore, we use observational studies to detect a large number of the behaviors described below.

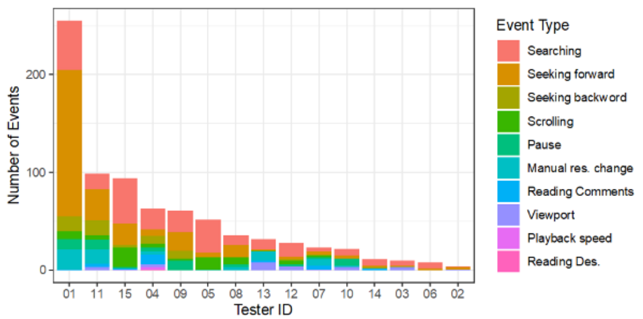

Figure 1. *Behaviors in detail.*

Seeking By seeking, we understand the ability to scroll through the video forward and backward. As seen in Figure 1, each of the testers exhibited the behavior described. Testers more often seek the video forward to review the entire material.

Pause Pausing in most cases resulted from two reasons: the first is the preference to view comments or read the description, and the second is the inclination to search for the video using the search button or from the list of proposed.

Viewport Changing By this type of behavior, we mean a change in the YouTube video player. Three sizes depended on the native resolution of the device used. In our laboratory environment, the viewport is defined as: *small* (1280×720), *medium* (1467×825), and *full-screen* (1920×1080). The dominant value of the viewport in the player was *small*.

Manual Change of Resolution YouTube provides the ability to manually change the resolution for users. Based on the recordings, we observed that most of the results were dictated by the desire to improve video resolution due to the deterioration of network bandwidth.

Viewing Comments and Description Through this behavior, we understand the actual browsing of the comments section, as well as reading the description provided by the creator under the video. We observed that reading comments and descriptions of the video is an inseparable element of using YouTube.

User vs Net

Further, the observed behaviors and video resolutions were divided into two categories. The "User" category directly corresponds to the behaviors that are generated by the user, and the "Net" category relates to the occurring network behaviors such as stalling and initial buffering.

In Figure 2, we illustrate the number of behaviors exhibited by each of the testers, broken down into net and user categories. Based on the analysis, we noticed that the number of network events increases with the number of user behaviors.

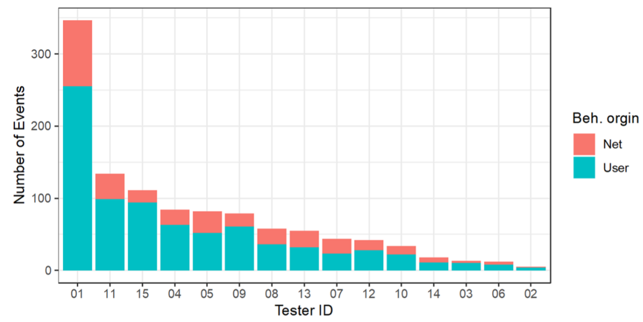

Figure 2. *Behaviors divided into their origin.*

Data Groups

Further, the analysis of the video recordings showed variations in terms of the number of behaviors exhibited by the participants. This means that some participants showed more behaviors than others during the testing process. This variability in the behavior of the participants is an important aspect to consider when interpreting the results of the study (see fig. 3).

Therefore, we decided to group the data obtained from the Chrome extension according to the behavior of the participants. The groups are as follows:
A: a group of data that relates only to constant video watching, without any additional interaction with the platform, such as scrolling, pausing, or manually changing the resolution.
B: a group of data that relates to the behaviors exhibited, such as scrolling, pausing, or manually changing the resolution without any events related to the network.

In addition, the data has been divided based on the events that occur in the network between ratings, such as stalling and initial buffering:
C: a group of data that relates to the occurring stalling.
D: a group of data that relates to initial buffering.

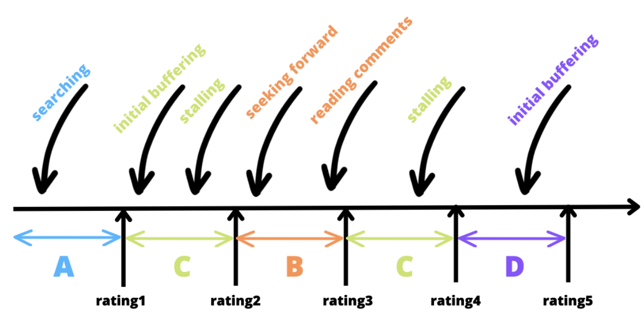

Figure 3. *Classification of data into groups based on observed behaviors.*

Correlation

Finally, we analyzed the correlation of resolution with MOS. The previous division of the data into groups allowed for a more detailed correlation analysis.

For ungrouped data, we obtained a correlation value of 0.81 for all MOS scores and a resolution of $\sqrt{res_x \times res_y}$, as shown in Figure 4. This high correlation value led us to the conclusion that an ecologically valid experiment can provide reliable results.

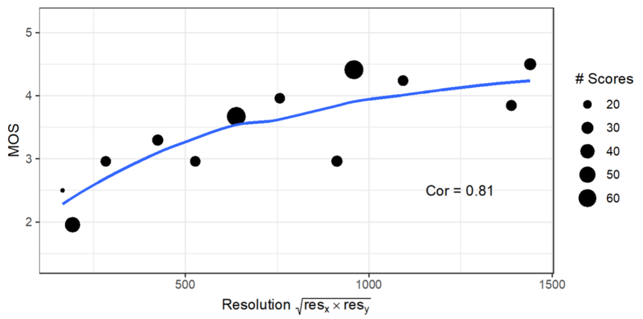

Figure 4. *Correlation of resolution with MOS for all data.*

Then, in Figure 5we present the correlation of resolution with MOS as a function of time. The blue line represents Group A (without occurring behaviors between ratings), and the red line describes Group C (data group with occurring stalling). The aim was to study the correlation between resolution and ratings depending on the time preceding the quality ratings. As can be seen from the chart 5 for group A the correlation is almost constant over time and is about 0.6. On the other hand, data including stalling is inconsistent. At around 30 seconds it stabilizes to around 0.2 and further increases to 0.4.

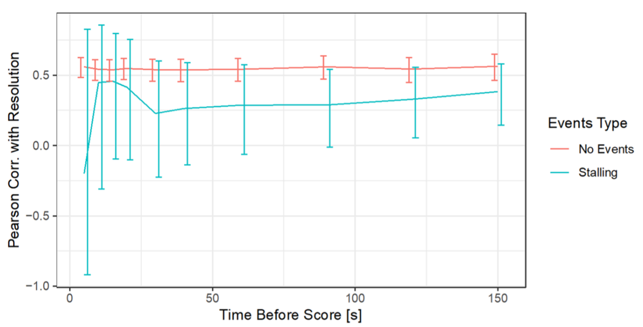

Figure 5. *Correlation of resolution with MOS in time, along with confidence intervals.*

Furthermore, by analyzing Figure 5, it is clear that the intervals of the correlation data for the stalling and the group without events overlap. Consequently, we have decided to delve deeper into the correlation coefficients' differences 6. This implies that there is no statistical significance in the difference between the groups with stalling and without any events.

In light of the above-discussed results, we decided to investigate the occurrence of stalling in intervals between ratings in relation to the number of periods. In Figure 7, we present the corresponding statistics. As seen, single-stalling incidents were the most frequent while the maximum number of stalling was 9.

As we have described in the Experiment Design section, in addition to MOS, we also collected data on the level of interest using a five-point scale. This additional data provides us with a more complete understanding of the subjects' engagement during the experiment and allows us to evaluate the effectiveness of the environmental design in terms of capturing the subjects' attention. Measuring interest was an important objective of our study and the results show that on average, the interest level was 4.63, indicating a high level of engagement among the participants.

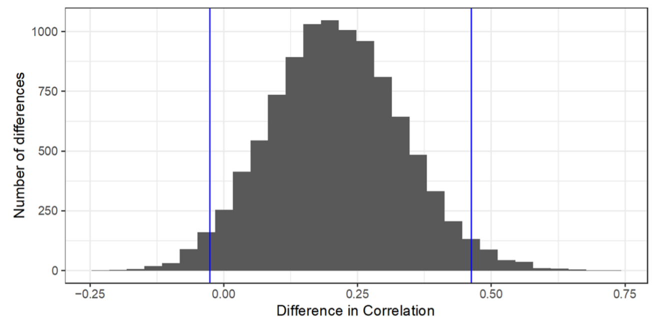

Figure 6. *Difference in Correlation with Resolution with/without Stalling and stalling statistics.*

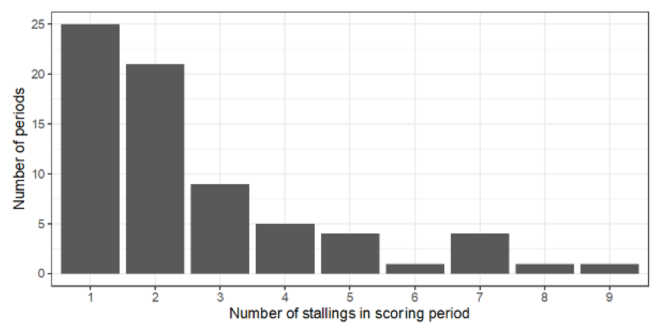

Figure 7. *Number of stallings in scoring period in relation to the number of periods.*

Questionnaire

Additionally, we conducted the questionnaire at the end of each test to measure the ecological validity of our experiment. Participants indicated that the experiment was different from their everyday experiences, 46% of them referred to a lack of multitasking and different use cases. The results of the questionnaire indicate that 54% of the testers paid attention to the room settings, such as lightning.

Discussion

In this paper, we present the results of an exploratory study of a new experimental protocol. The measured method aims to increase the ecological validity of the experiment with the maximum possible preservation of reliability. An exploratory study conducted on 15 participants indicates the usefulness of the proposed experimental protocol.

During the experiment, we collected screen recordings and YouTube's "Statistic for Nerds" was captured every 0.5 s by our Chrome extension . Initially, we planed to use data from the "Statistic for Nerds" for analysis of the behavior of the participants. However, in our initial analysis, we measure the continuity of watching by selecting time intervals without any behavior occurring. Reflecting on the continuity of the videos watched, we ask ourselves "Has the participant watched the currently selected video?". Due to the doubts raised by a large number of behaviors, we decided to investigate screen recordings with an observational study.

The screen recordings of a participant showed that the data from "Statistics for Nerds" *did not* clearly reflect the interaction with the video player. Our analysis of recordings indicated that the user's interactions with the service can change more often than 0.5 seconds and that some events can overlap in time. An example of this is seeking forward to a video and quick scrolling to the comment section. This observational analysis gave us a broader view of the actual use of the YouTube platform, which not only serves to watch the video but imposes additional interactions that influence the experience (e.g. reading the comment section).

Furthermore, examination of the video recordings revealed disparities in the number of behaviors between the participants. Some participants interacted with the player way more frequently than others. To better understand these disproportions in the frequency of participants' behavior, we divided the data into two groups. The "User" group corresponds to the behaviors generated by the participants, while the "Network" group represents the events generated by our manipulation of the network. Based on this, we observed that the amount of network events depends on the number of events generated by the participants. That means that the participants could have been causing the network events themselves. We observed that some participants showed a tendency to change the resolution at the time of stalling, which generated more stallings.

Subsequently, we made the decision to classify the data obtained from the Chrome extension into 4 groups A, B, C, and D as explained in the Data analysis section. The division of data into groups was designed to provide a more accurate examination of the correlation between subjective quality ratings and resolution. We used data from group A (without events or behaviors) and group C (with stalling events) for further analysis.

We started the the analysis by calculating the correlation between MOS and resolution for ungrouped data. In the dataset containing all of the data groups (A, B, C, and D) the correlation between MOS and resolution was high (0.81). The fact that the correlation result is high shows that the experimental protocol is effective and might deliver reliable results.

Next, we analyzed the correlation between resolution and MOS as a function of time. The purpose was to investigate the correlation between resolution and ratings based on the time preceding quality ratings. We used data groups A and C to analyze the influence of stalling on ratings. By breaking into data groups, we revealed that when stalling was absent, the correlation remained steady at 0.6, but when stalling was present, it fluctuated and ultimately settled at 0.2. However, the coefficient intervals overlapped, indicating that the differences between the data groups are statistically insignificant.

Because we did not observe significant differences between the groups with and without stalling, we decided to investigate the frequency of stallings. For that purpose, we analyzed the number of stallings between quality ratings. The results showed that single stalling incidents were the most prevalent, reaching a maximum of 9 occurrences. These findings suggest that in real-world scenarios, the influence of stalling on user satisfaction is lower.

As we described in the Experiment Design section, we not only collected MOS but also measured information about the level of interest on a five-point scale. This supplementary data offer a more complete understanding of the subjects' engagement during the experiment. Most of the participants were highly invested in the chosen video content, supporting our initial hypotheses. The high level of interest demonstrates the ecological validity of our experimental protocol. Despite the fact that participants were in the laboratory, they choose the content according to their interests.

On the other hand, the questionnaire suggests that 54% of the testers noticed the room settings. This result is supported by the subjects' statements that the luminance value was significantly different from the lighting conditions they typically experience while watching videos on YouTube. As described in the Experiment Design section, the preparation of the room for the experiment was in accordance with the BT-500 recommendation. Therefore, the recommendation does not refer to everyday users' environments and thus should not be used in experiments aiming for ecological validity. Additionally, the questionnaire showed that 46% of the respondents normally do not watch YouTube videos fully engaged. Participants claimed that they rather use YouTube during daily activities and treat the video as an addition to the tasks performed. Those answers suggest that we could raise the ecological validity of our protocol by using more natural lighting. Even more natural would be allowing participants to multitask during the experiment.

Conclusions

In conclusion, YouTube is not only used to watch videos, but it can serve multiple purposes. Users interact with each other by reading the comments and video descriptions. In addition, they interact intensely with the service by searching for interesting videos or rewinding to the interesting part. Moreover, network problems can evoke new behaviors like changing video resolution or reading comments during stalling events.

To measure all those interactions more ecologically valid experimental protocols are needed. In our study, unrestrained interactions with the video service enabled behavioral measurements. Moreover, we could use events and behaviors that occurred during the experiment for grouping the data.

Above described data analysis indicate that when considering data that is representative of real-world scenarios, there is no longer a notable negative effect on the quality when stalling occurs. These findings provide valuable insights into the relationship between objective metrics and perceived video quality and highlight the importance of continuous improvement of methodologies for video streaming studies.

Acknowledgment

The research leading to these results has received funding from the Norwegian Financial Mechanism 2014-2021 under project 2019/34/H/ST6/00599.

Figure 8. *IS&T logo.*

References

[1] ITU-R, 500-14 (10/2019): Methodologies for the Subjective Assessment of the Quality of Television Images, ITU: Geneva, Switzerland (2020).

[2] ITU-T, P.800.1 (07/16): Mean opinion score (MOS) terminology, ITU: Geneva, Switzerland (2016).

[3] ITU-T, P.1230 (10/17): Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport,ITU: Geneva, Switzerland (2017).

[4] Schmuckler, Mark A. "What is ecological validity? A dimensional analysis." Infancy 2.4 (2001): 419-436.

[5] Pérez, Pablo, et al. "Ecological validity through gamification: an experiment with a mixed reality escape room." 2021 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR). IEEE, 2021.

[6] Robitza, Werner, et al. "On the experimental biases in user behavior and QoE assessment in the lab." 2016 IEEE Globecom Workshops (GC Wkshps). IEEE, 2016.

[7] Brown, GDA and Sinclair, RD,On the nature of ecological validity, 1995, The Quarterly Journal of Experimental Psychology.

[8] Robitza, Werner, et al. "Are you still watching? streaming video quality and engagement assessment in the crowd." 2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX). IEEE, 2020.

[9] Analytics, Verto, Most popular websites worldwide as of November 2021, by total visits (in billions)[Graph], Statista. Retrieved from https://www.statista.com/statistics/1201880/most-visited-websites-worldwide/ (2022).

[10] Moldovan, Christian, Florian Wamser, and Tobias Hoßfeld. "User behavior and engagement of a mobile video streaming user from crowdsourced measurements." 2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX). IEEE, 2019.

[11] Krishnan, S. Shunmuga, and Ramesh K. Sitaraman. "Video stream quality impacts viewer behavior: inferring causality using quasi-experimental designs." Proceedings of the 2012 Internet Measurement Conference. 2012.

[12] Laiche, Fatima, Asma Ben Letaifa, and Taoufik Aguili. "QoE-aware traffic monitoring based on user behavior in video streaming services." Concurrency and Computation: Practice and Experience (2021): e6678.

[13] Robitza, Werner, and Alexander Raake. "(Re-) actions speak louder than words? A novel test method for tracking user behavior in Web video services." 2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX). IEEE, 2016.

[14] Kieffer, Suzanne, Ugo Braga Sangiorgi, and Jean Vanderdonckt. "Ecoval: A framework for increasing the ecological validity in usability testing." 2015 48th Hawaii International Conference on System Sciences. IEEE, 2015.

[15] Robson, Colin, and Kieran McCartan. Real world research. Wiley Global Education, 2016.

[16] Campbell, Donald T., and Julian C. Stanley. Experimental and quasi-experimental designs for research. Ravenio books, 2015.

[17] Salmon, Catherine A., and Jessica A. Hehman. "The importance of ecological validity, ultimate causation, and natural categories." Behavioral and Brain Sciences 45 (2022).

[18] Andrade C. Internal, External, and Ecological Validity in Research Design, Conduct, and Evaluation. Indian J Psychol Med. 2018 Sep-Oct;40(5):498-499. doi: 10.4103/IJPSYM.IJPSYM_334_18. PMID: 30275631; PMCID: PMC6149308.

[19] Hossfeld, Tobias, Christian Keimel, and Christian Timmerer. "Crowdsourcing quality-of-experience assessments." Computer 47.9 (2014): 98-102.

[20] Søgaard, Jacob, et al. "On subjective quality assessment of adaptive video streaming via crowdsourcing and laboratory based experiments." Multimedia tools and applications 76.15 (2017): 16727-16748.

[21] Wang, Kun, et al. "Subjective evaluation of HDTV stereoscopic videos in IPTV scenarios using absolute category rating." Stereoscopic Displays and Applications XXII. Vol. 7863. SPIE, 2011.

[22] SEUFERT, Michael, et al. A wrapper for automatic measurements with YouTube's native Android app. In: 2018 Network traffic measurement and analysis conference (TMA). IEEE, 2018. p. 1-8.

Author Biography

Gabriela Wielgus *is a student of Computer Science Engineering at the Cracow University of Technology. She is a member of the AGH Video Quality of Experience (QoE) team. Her research focuses on QoE.*

Lucjan Janowski *received the Ph.D. degree in telecommunications from the AGH University of Science and Technology, Krakow, Poland, in 2006. In 2007, he was a Postdoctoral Researcher with the Laboratory for Analysis and Architecture of Systems, Centre National de la Recherche Scientifique, Paris, France. From 2010 to 2011, he was a Postdoctoral Researcher with the University of Geneva, Geneva, Switzerland. From 2014 to 2015, he was a Postdoctoral Researcher with the Telecommunications Research Center Vienna, Vienna, Austria. He is currently an AGH Professor with the Institute of Telecommunications, AGH University of Science and Technology. His research interests include statistics and probabilistic modeling of subjects and subjective rates used in QoE evaluation.*

Kamil Koniuch *is a PhD student at the Institute of Communication Technologies, AGH University of Science and Technology since 2020. He is a member of AGH Video Quality of Experience (QoE) team. His research focuses on the ecological validity of QoE measurements. He received his master's degree in psychology in 2019. He is a graduate of the Institute of Psychology, Jagiellonian University. ORCID: 000-0002-8243-7155*

Mikołaj Leszczuk*, DSc, is an associate professor at the AGH Department of Telecommunications. He is scientifically interested in multimedia subjects, especially video applications. He is specialized in the accurate description and standardization of methods of QoE/UX analysis. He is a board member of the Video Quality Experts Group (VQEG), chairing the Quality Assessment for Computer Vision Applications (QACoViA) subgroup. The VQEG group was established in 1997 and brings together experts involved in both subjective and objective assessment of video quality. Furthermore, several times, M. Leszczuk was invited to work for the Video Quality in Public Safety Working Group (VQiPS). The VQiPS Working Group, supported by the US Department of Homeland Security, has been developing public safety video surveillance applications. M. Leszczuk is also a senior member of the Institute of Electrical and Electronics Engineers (IEEE).*

Rafał Figlus *is a highly motivated individual with a Bachelor of Science degree in Information and Communication Technology. Before completing his studies, he started working as a software developer at TUFIQOE where he gained practical experience in the field. He is deeply interested in software development and web technologies, and his passion and education make him a valuable asset in the industry.*