



HAL
open science

Sustainable Video Streaming using Acceptability and Annoyance Paradigm

Ali Ak, Abhishek Gera, Hassene Tmar, Denise Noyes, Ioannis Katsavounidis,
Patrick Le Callet

► **To cite this version:**

Ali Ak, Abhishek Gera, Hassene Tmar, Denise Noyes, Ioannis Katsavounidis, et al.. Sustainable Video Streaming using Acceptability and Annoyance Paradigm. 2024 32nd European Signal Processing Conference (EUSIPCO 2024), EURASIP, Aug 2024, LYON, France. hal-04615336

HAL Id: hal-04615336

<https://hal.science/hal-04615336v1>

Submitted on 18 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Sustainable Video Streaming using Acceptability and Annoyance Paradigm

Ali Ak

Nantes Université, École Centrale Nantes, CNRS, LS2N
UMR 6004, F-44000 Nantes, France
ali.ak@univ-nantes.fr

Patrick Le Callet

Nantes Université, École Centrale Nantes, CNRS, LS2N
UMR 6004, F-44000 Nantes, France
Institut Universitaire de France (IUF)
patrick.le-callet@univ-nantes.fr

Abhishek Gera
Meta, USA

Hassene Tmar
Meta, USA

Denise Noyes
Meta, USA

Ioannis Katsavounidis
Meta, USA

Abstract—The energy consumption of information and communications technologies is increasing rapidly, contributing to the climate crisis. Video streaming technologies are particularly impactful, accounting for over half of this energy use. Recent advancements in video streaming technologies have improved user experience by delivering higher quality videos but at the expense of significantly increased data volumes. In pursuit of a more sustainable approach to video streaming, we investigate greener methods ways of integrating Quality of Experience (QoE) measures into video streaming pipelines. QoE defines to users’ satisfaction while consuming video content. User satisfaction with video content is largely influenced by their expectations. Depending on the context, users may have varying expectations regarding video quality. To address this, the acceptability & annoyance paradigm is proposed to measure user satisfaction as a function of both user expectations and video quality. By understanding user expectations and satisfaction together, video streaming services can reduce energy costs associated with processing, storing, and transmitting video content while maintaining the same level of QoE. In this study, we focus on the acceptability and annoyance of short-form vertical user-generated video content commonly found on online social media platforms. We analyze the relationship between video quality and acceptability & annoyance, demonstrating how this paradigm can be leveraged to achieve significant bitrate savings while preserving user satisfaction.

Index Terms—acceptability and annoyance, sustainability, video streaming, quality of experience

I. INTRODUCTION

Video streaming includes a variety of platforms and services, such as online social media platforms (*e.g.*, Facebook, TikTok, etc.), video streaming services (*e.g.*, Netflix, Amazon Prime, etc.), user generated streaming platforms (*e.g.*, Youtube, Vimeo, etc.), video conferencing and chat applications (*e.g.*, Zoom, Skype, etc.), and gaming-related applications (*e.g.*, Twitch, GeForce Now, etc.).

Over the past decade, with higher bandwidth and technological advancements in consumer devices, we’ve witnessed a significant surge in the popularity of video streaming services and platforms. Video streaming has become the predominant source of internet traffic, receiving considerable interest in assessing its environmental impact [1]. As a result, several

efforts have been made to estimate the carbon footprint of video streaming [2], [3]. The models proposed in the literature covers a broad spectrum of carbon emission sources within the video streaming chain such as device manufacturing, content production, processing, transmission, storage, and display.

In an ideal setting with unlimited resources, online social media platforms and video streaming services would aim to consistently provide the highest possible visual quality. However, the practicalities of video streaming at scale necessitates a trade-off between the delivered video quality and the resources required (*e.g.*, compute, bandwidth, storage, *etc.*) to achieve it. It is inefficient to allocate substantial resources to deliver marginally improved content without a corresponding increase in user satisfaction [4]. Therefore, determining the boundaries of user satisfaction is crucial for online video streaming services.

While numerous datasets and objective metrics exist for evaluating video quality, determining video quality alone may not be the primary goal for video streaming service providers. QoE defines the observer’s level of satisfaction and how well their expectations are met while viewing the video content [5]. QoE may vary even for the same video stimuli depending on the context in which it is consumed [6]. To this end, acceptability & annoyance paradigm has been proposed and studied [7]–[11]. IPI-VUGC dataset ¹ is the largest publicly available dataset with acceptability & annoyance labels and video quality ratings for 336 videos, each with 5 seconds duration and in vertical orientation.

Previous research indicates that the video quality does not directly correlate with the acceptability and annoyance of the video, as the thresholds at which a video becomes annoying or unacceptable are heavily influenced by user expectations [6]. Therefore, classical subjective image/video quality (*i.e.*, ACR, DSIS) methodologies [12] may not adequately capture the acceptability & annoyance. To bridge this gap, a mapping from video quality to the acceptability & annoyance space is necessary, which should be tailored to the use-case and

¹IPI-VUGC Dataset URL: <https://zenodo.org/doi/10.5281/zenodo.10475209>

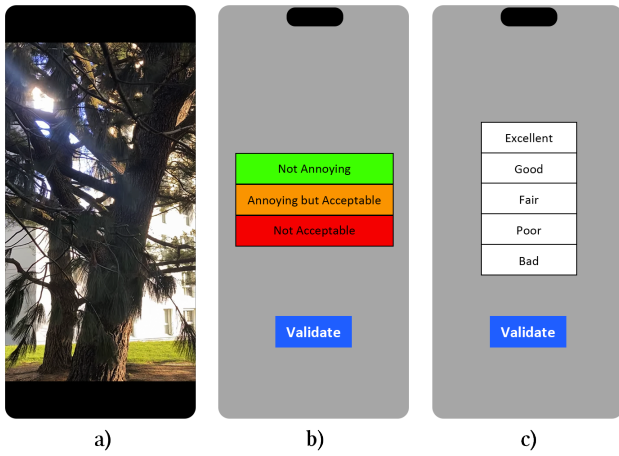


Fig. 1. Screen captures illustrating user interface, in a) during stimuli presentation, in b) during voting in AccAnn experiment, and in c) during voting in ACR-HR experiment.

the context [7], [10], [13]. Recent studies have explored specific contexts that impact user expectations, such as the type of service profile subscribed to (*e.g.*, premium or basic), which significantly influences the video quality required for user satisfaction [7]. Additionally, it has been observed that low remaining data in a monthly data plan can alter users' expectations regarding video quality [8]. All of which are valuable in optimizing the energy consumption and carbon footprint of the video streaming services and platforms.

This study investigates the acceptability & annoyance levels associated with short-form user-generated video content (UGC) in the context of online social media platforms. We conduct a detailed analysis of acceptability and annoyance thresholds in relation to video quality, utilizing the publicly available IPI-VUGC dataset. We propose a 4-parameter logistic function to accurately model the relationship between video quality and acceptability & annoyance of UGC videos on online social media platforms. Our findings demonstrate that leveraging the acceptability & annoyance paradigm can lead to substantial bitrate savings without compromising QoE.

II. IPI-VUGC DATASET

As stated, this work relies on the publicly available IPI-VUGC dataset [14]. In this section, we give the necessary introduction to the dataset. Interested readers are invited to refer to the dataset repository for more information.

To gather acceptability & annoyance, and video quality labels, a set of subjective experiments were conducted in controlled laboratory conditions. Identical video content is utilized in both experiments to examine the relationship between to examine the relationship between video quality and acceptability & annoyance. The experiments are conducted using an iPhone 14 Pro² device with the native player (AV Player³). The device is held freely by the subjects at a comfortable distance

during the experiment. The screen brightness is fixed at the maximum brightness to minimize any inter-subject differences. To account for the display's maximum brightness, the ambient room brightness is set to a comfortable 50 lux (on average from multiple measurements), mimicking a standard living-room environment. The remaining aspects of the experimental setup follows the ITU recommendations [12].

A. Content

A total of 48 UGC videos serve as source content (SRC) in the experiments. This collected content is captured using various smartphones by different individuals over the past 4 years. Each video is cropped to a duration of 5 seconds and subsequently re-encoded with h.264 [15]. Audio was removed from each SRC. Figure 1-a illustrates the rendered video screen during the experiment. The content is rendered at the native screen size and resolution, relying on the phone's native resampling algorithm. The gaps occurring at the top and bottom sides of the screen, resulting from display dimensions, are filled with black color.

6 Processed Video Sequences (PVS) are created by compressing each SRC using h.264 codec with varying resolution (ranging from 288p to 1080p) and constant rate factors (CRF), as detailed in the dataset. This results in a dataset comprising 336 videos where each video is evaluated by 25 unique naive subjects.

B. AccAnn and ACR-HR Experiment Details

We employed the acceptability & annoyance paradigm with a single-step methodology [7], [8]. The acceptability & annoyance scale, accompanied by provided instructions, enables the measurement of QoE as a combination of video quality and users' expectation of video quality. Three color-coded scales, namely "Not Annoying", "Annoying but Acceptable", and "Not Acceptable" were utilized, as illustrated in Figure 1-b. The collected labels are denoted as AccAnn-MOS and numerically represented in [1,3] range where 3 indicates a higher QoE (*i.e.*, "Not Annoying").

The instructions given to participants play a crucial role in aligning subjects with the intended profile and managing their expectations. In IPI-VUGC dataset, the instructions are carefully crafted, drawing upon recommendations from prior studies [7], [8] and are presented below:

"You are going to participate in an experiment determining the acceptability and annoyance of videos. You will need to imagine yourself scrolling through your preferred social media platform (*e.g.*, Facebook, Instagram, TikTok, etc.) and encountering these videos. Based on your expectations of the video quality in these encounters, you will need to rate the quality of the video in terms of acceptability and Annoyance.

- The video is not annoying when its quality satisfies or exceeds your expectations.
- The video is annoying but acceptable when its quality is acceptable but not completely satisfies your expectations.

²<https://www.apple.com/go/2022/iphone-14-pro/>

³<https://developer.apple.com/documentation/avfoundation/avplayer/>

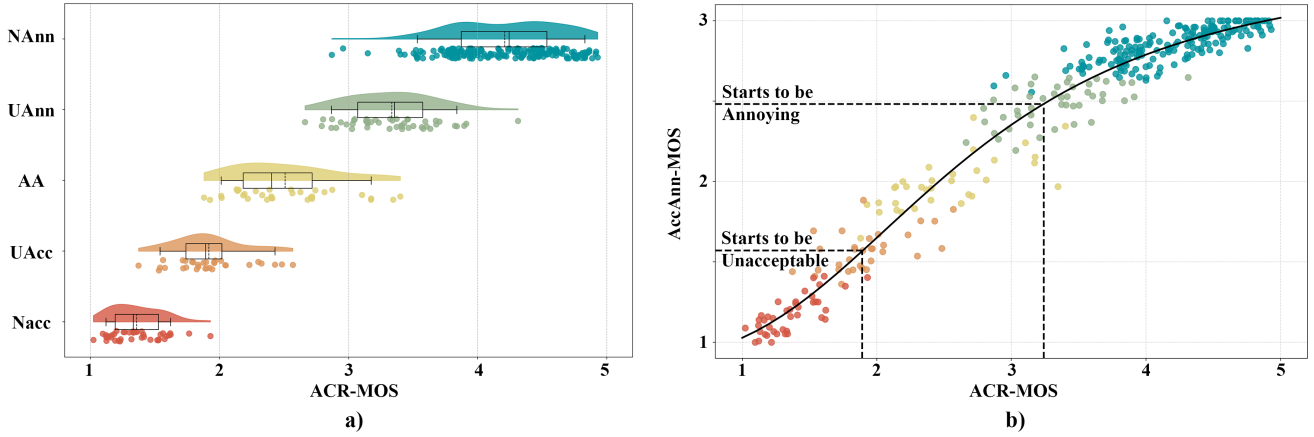


Fig. 2. Visualisation of the relation between ACR-MOS and AccAnn-MOS, as well as between ACR-MOS and acceptability & annoyance categories. In a), raincloud plot illustrates the distribution of videos’ ACR-MOS within each acceptability & annoyance category. In b), relation between the ACR-MOS and AccAnn-MOS is depicted through a scatter plot. Categories are color-coded and the mapping function given in Equation 1 is overlaid on the scatter plot.

- The video is not acceptable when its quality does not meet your expectations. Such video quality makes you think about skipping to the next video.”

An Absolute Category Rating experiment with a Hidden Reference (ACR-HR) is also conducted using the classical scale (“Bad”, “Poor”, “Fair”, “Good”, “Excellent”), as illustrated in Figure 1-c. Similar to AccAnn-MOS, the collected labels in ACR-HR experiment are referred as ACR-MOS and numerically represented in the range $[1, 5]$, where 5 corresponds to highest video quality (e.g., “Excellent”). Additionally, ACR-DMOS is utilized in our analyses. It is defined as the mean of differences between the subjects’ opinion score of a stimulus and its SRC.

In contrast to the AccAnn experiment, no specific instructions regarding social media experience are provided to subjects in the ACR experiment. Instead, participants are instructed to evaluate the visual quality without any contextual information.

III. RELATION BETWEEN VIDEO QUALITY AND ACCEPTABILITY & ANNOYANCE

In this section, we determine the AccAnn categories from individual opinions and present the mapping function between the ACR-MOS and AccAnn-MOS values. Note that all subjective opinion scores were processed with ZREC [16], a MOS recovery algorithm, to mitigate the bias and inconsistencies associated with raw subjective opinions. Additionally, we conducted analyses using P910 [12] method, revealing similar MOS and 95% confidence intervals.

A. Determining AccAnn Category

We employ Algorithm-2 from [7] to determine the AccAnn categories of the videos. The abbreviations used are “NAnn” for “Not Annoying”, “UAnn” for “Unsure about Annoyance but sure about the Acceptability”, “AA” for “Annoying but Acceptable”, “UAcc” for “Unsure about Acceptability but sure

about the Annoyance”, and “NAcc” for “Not Acceptable”. In summary, each video is classified into one of the three main categories (NAnn, AA, and NAcc) or one of the two threshold categories (UAnn and UAcc). If individual opinion scores of a given video show statistically significant agreement, it belongs to one of the main categories based on the majority opinion. Otherwise, due to a lack of agreement between subjects, video is categorized into one of the threshold categories. UAnn category serves as the threshold for videos where the video quality starts to become annoying, while UAcc is the threshold for videos that start to become unacceptable.

Figure 2-a depicts raincloud plots associated with each category, with ACR-MOS values on the horizontal axis. Each raincloud plot includes a density plot, box plot, and the actual distribution. acceptability & annoyance categories are color-coded and indicated on the vertical axis. We can observe that the ACR-MOS values between the 3 main categories (NAnn, AA, and NAcc) are well separated, suggesting that the visual quality is a reliable proxy for understanding the acceptability & annoyance categories.

B. Mapping Visual Quality to Acceptability & Annoyance Range

In addition to the categorical analysis, we can explore the relation between visual quality and acceptability & annoyance in a continuous scale. Figure 2-b presents a scatter plot of video quality (as ACR-MOS in the horizontal axis) against the acceptability & annoyance values (as AccAnn-MOS on the vertical axis). The categories are color-coded as in Figure 2-a.

The “starts to be unacceptable” and “starts to be annoying” thresholds over the plot are determined by identifying the closest points on the mapping function to the mean AccAnn-MOS values of the stimuli in the UAcc and UAnn categories. 3.24 ACR-MOS is the threshold for content to begin becoming annoying, while 1.89 ACR-MOS corresponds

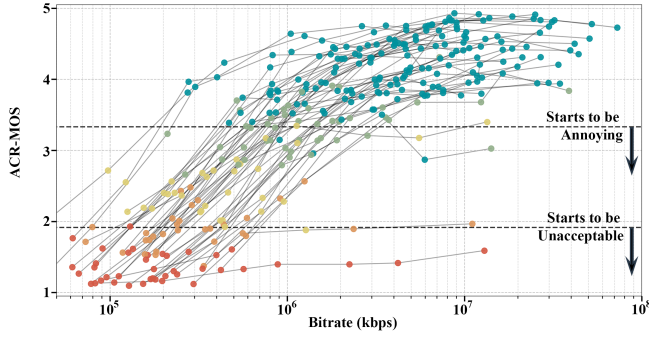


Fig. 3. Each point represents a video from the IPI-VUGC dataset. Horizontal axis shows the bitrates in kbps, vertical axis shows the ACR-MOS. Thresholds for acceptance & annoyance are plotted as dashed horizontal lines. PVS that is generated from the same SRC are plotted over a line. Acceptance & annoyance categories are color-coded.

to the threshold for content to start becoming unacceptable. Contents with an ACR-MOS higher than 3.24 ensures meeting or exceeding user expectations of visual quality on an online social platform. More crucially, to deliver at least an acceptable user experience and prevent user disengagement, the threshold of 1.89 ACR-MOS should not be crossed.

Moreover, we employed a 4-parameter logistic function to find a mapping between the ACR-MOS and AccAnn-MOS values. The function is plotted over the scatter plot in Figure 2-b and is defined as:

$$y = ((d - a)/(1 + (x/c)^b)) + a \quad (1)$$

where, $a=3.3357$, $b=2.9503$, $c=2.6233$, and $d=0.8942$. x represents the ACR-MOS and y represents the AccAnn-MOS.

IV. SAVING BITRATE WITHOUT SACRIFICING QOE

In this section, we demonstrate how the acceptability & annoyance paradigm can be employed to achieve bitrate savings while maintaining user satisfaction. As illustrated in Figure 2, a slight decrease in video quality may not necessarily lead to a decline in user satisfaction. For instance, even if the ACR-MOS rating drops from 5 to 4, the resulting video may still be deemed satisfactory (*i.e.*, Not Annoying) by users. Building on this observation, Figure 3 illustrates the bitrates plotted against the ACR-MOS values for each PVS in the IPI-VUGC dataset. Note that the bitrates are presented in a logarithmic scale, and consistent color coding is maintained for the AccAnn categories.

At high quality range (*i.e.*, $ACR-MOS \geq 3.5$), we observe that even a tenfold reduction in bitrate doesn't affect the AccAnn categories. However, this observation applies only to certain stimuli from the ACR-MOS perspective. Consequently, instead of pursuing marginal gains in ACR-MOS scale that come with a tenfold increase in bitrate costs, it's more prudent to rely on the acceptability & annoyance paradigm. Depending on the use-case, in relevant scenarios, stimuli above the annoyance threshold can be allocated a much lower bitrate, resulting in substantial resource savings.

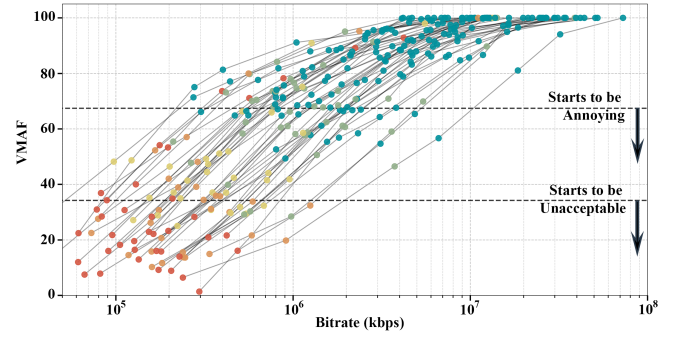


Fig. 4. Each point represents a video from the IPI-VUGC dataset. Horizontal axis shows the bitrates in kbps, vertical axis shows the VMAF scores. VMAF scores for SRCs are set to 100. Thresholds for acceptance & annoyance are plotted as dashed horizontal lines. PVS that is generated from the same SRC are plotted over a line. Acceptance & annoyance categories are color-coded.

However, implementing this approach requires the acceptability & annoyance categories of the videos, which is impractical without a reliable model for predicting these categories. As shown in Figure 2, there's a clear relationship between video quality and acceptability & annoyance, indicating that state-of-the-art video quality metrics can be utilized to predict acceptability & annoyance categories. Several studies have addressed this issue by establishing acceptability & annoyance thresholds within the prediction ranges of objective quality metrics for various contexts [7], [8], [17]. These thresholds are typically calculated as the mean of threshold categories (UAcc and UAnn, respectively). Notably, VMAF [18] is commonly employed due to its reliability and widespread adoption in the industry.

Figure 4 presents the VMAF scores for each video in the dataset, similar to Figure 3. For user-generated content, we observe inaccuracies in the classification of acceptability & annoyance categories with VMAF. Despite relying on VMAF for this analysis to maintain comparability with other studies, it's crucial to use a video quality metric suitable for the specific use-case. Without loss of generality, similar bitrate gains can be achieved using acceptability & annoyance thresholds obtained with objective quality metric predictions.

Previous studies have shown that acceptability & annoyance thresholds can vary significantly depending on the context provided to the observers. In the IPI-VUGC dataset, the context is defined as watching UGC content on online social media platforms. In a previous study [7], the authors experimented with two contexts: a premium and basic subscription to an online video streaming platform. Table I compares the obtained thresholds in the VMAF prediction range. Online social media thresholds are calculated on the IPI-VUGC dataset by us, meanwhile the Basic and Premium subscription thresholds are taken from the aforementioned study [7]. Acceptability threshold is calculated as the mean VMAF score of all stimuli in UAcc category, while annoyance threshold is calculated as the mean VMAF score of all stimuli in UAnn category. Same algorithm is used to determine the stimuli categories in

TABLE I
ACCEPTABILITY AND ANNOYANCE THRESHOLDS IN VMAF PREDICTION
RANGE IN DIFFERENT CONTEXTS

	Acceptability	Annoyance
Online Social Media [14]	34	67
Basic Subscription [7]	58	80
Premium Subscription [7]	71	87

both studies.

User expectations regarding video quality significantly changes as we compare the online social media platform use-case to a paid subscription to a video streaming service. Hence, we see a higher acceptability and annoyance thresholds in basic subscription plan in comparison to online social media platforms. Furthermore, thresholds further increases as users are instructed for a premium subscription to a video streaming service, instead of a basic subscription. In other words, users expect higher quality as they pay more for the service. Based on the annoyance thresholds estimated in these contexts, video streaming platforms can adjust their streaming pipelines to avoid wasting resources on marginal gains in video quality which does not impact the user satisfaction in a significant way. Considering the user satisfaction remains at similar levels above annoyance thresholds, service providers may utilize these thresholds to allocate their resource more efficiently.

V. CONCLUSION

We conducted comprehensive analyses on publicly available IPI-VUGC dataset comprising acceptability & annoyance and video quality labels for 336 UGC videos. Our aim was to comprehend the relationship between user satisfaction and video quality within the context of online social media platforms and provide recommendations towards a more sustainable video streaming pipeline. In this regard, we established a mapping between video quality and acceptability & annoyance spaces. Furthermore, we determined thresholds for the acceptability & annoyance of video content in terms of mean opinion scores and metric predictions, and compared these thresholds with those obtained in other contexts.

We showcased that video streaming services can adjust their video streaming pipelines to avoid pursuing minimal gains in QoE at the expense of significant compression, storage, and transmission costs. Acceptability & annoyance paradigm can be leveraged to set thresholds based on user expectations. Designing adaptive bitrate ladders with a QoE driven approach (more specifically acceptability & annoyance driven approach) may reduce the storage and bitrate costs associated with bitrate delivery pipelines.

While our study explored acceptability & annoyance as a context-dependent paradigm, further insights can be gained by adopting a content and context-dependent approach. User expectations are heavily influenced by content; for instance, users don't hold the same quality expectations for cinematic drone footage as they do for cute cat videos. By understanding user expectations and adapting video streaming pipelines

accordingly, the video streaming ecosystem can mitigate its impact on the climate crisis and reduce global greenhouse emissions.

REFERENCES

- [1] Cisco, "Cisco annual internet report (2018–2023) white paper," 2020.
- [2] S. Makonin, L. U. Marks, R. Przedpelski, A. Rodriguez-Silva, and R. ElMallah, "Calculating the carbon footprint of streaming media: Beyond the myth of efficiency," *Eighth Workshop on Computing within Limits 2022*, 2022.
- [3] J. Moulierac, G. Urvoy-Keller, M. Dinuzzi, and Z. Ma, "What is the carbon footprint of one hour of video streaming?," Apr. 2023.
- [4] S. Ling, Y. Baveye, and P. L. Callet, "Rate-distortion video coding and uncertainties: to be blindly chasing marginal improvement or to be greener," in *Applications of Digital Image Processing XLIII* (A. G. Tescher and T. Ebrahimi, eds.), vol. 11510, p. 115100F, International Society for Optics and Photonics, SPIE, 2020.
- [5] P. Le Callet, S. Möller, A. Perkis, *et al.*, "Qualinet white paper on definitions of quality of experience," *European network on quality of experience in multimedia systems and services (COST Action IC 1003)*, vol. 3, no. 2012, 2012.
- [6] S. Jumisko-Pyykkö and M. M. Hannuksela, "Does context matter in quality evaluation of mobile television?," in *Proceedings of the 10th International Conference on Human Computer Interaction with Mobile Devices and Services, MobileHCI '08*, (New York, NY, USA), p. 63–72, Association for Computing Machinery, 2008.
- [7] J. Li, L. Krasula, Y. Baveye, Z. Li, and P. Le Callet, "Accann: A new subjective assessment methodology for measuring acceptability and annoyance of quality of experience," *IEEE Transactions on Multimedia*, vol. 21, no. 10, pp. 2589–2602, 2019.
- [8] A. Ak, A. F. Perrin, D. Noyes, I. Katsavounidis, and P. Le Callet, "Video consumption in context: Influence of data plan consumption on qoe," in *Proceedings of the 2023 ACM International Conference on Interactive Media Experiences, IMX '23*, (New York, NY, USA), p. 320–324, Association for Computing Machinery, 2023.
- [9] M. Angela and H. Knoche, "Quality in context-an ecological approach to assessing qos for mobile tv," 01 2006.
- [10] A. Oeldorf-Hirsch, J. Donner, and E. Cutrell, "How bad is good enough? exploring mobile video quality trade-offs for bandwidth-constrained consumers," 10 2012.
- [11] W. Song and D. W. Tjondronegoro, "Acceptability-based qoe models for mobile video," *IEEE Transactions on Multimedia*, vol. 16, no. 3, pp. 738–750, 2014.
- [12] ITU-R, "Subjective video quality assessment methods for multimedia applications." ITU-R Recommendation P.910, 2023.
- [13] T. De Pessemier, K. De Moor, W. Joseph, L. Marez, and L. Martens, "Quantifying subjective quality evaluations for mobile video watching in a semi-living lab context," *IEEE Transactions on Broadcasting*, vol. 58, pp. 580–589, 12 2012.
- [14] A. Ak and P. Le Callet, "IPI-VUGC: Acceptance/Annoyance and Video Quality of Vertically Oriented User Generated Videos," Jan. 2024.
- [15] ITU-T, "Reference software for itu-t h.264 advanced video coding." ITU-T H.264-2, 2016.
- [16] J. Zhu, A. Ak, P. Le Callet, S. Sethuraman, and K. Rahul, "Zrec: Robust recovery of mean and percentile opinion scores," in *2023 IEEE International Conference on Image Processing (ICIP)*, pp. 2630–2634, 2023.
- [17] A. Ak, A. Gera, D. Noyes, F. Blouin, H. Tmar, I. Katsavounidis, and P. Le Callet, "Estimating Metric Thresholds for Acceptability and Annoyance of User Generated Video Content," in *Human Vision and Electronic Imaging 2024*, (San Francisco, United States), Jan. 2024.
- [18] Z. Li, A. Aaron, I. Katsavounidis, A. Moorthy, and M. Manohara, "Toward a practical perceptual video quality metric," *The Netflix Tech Blog*, 2016.