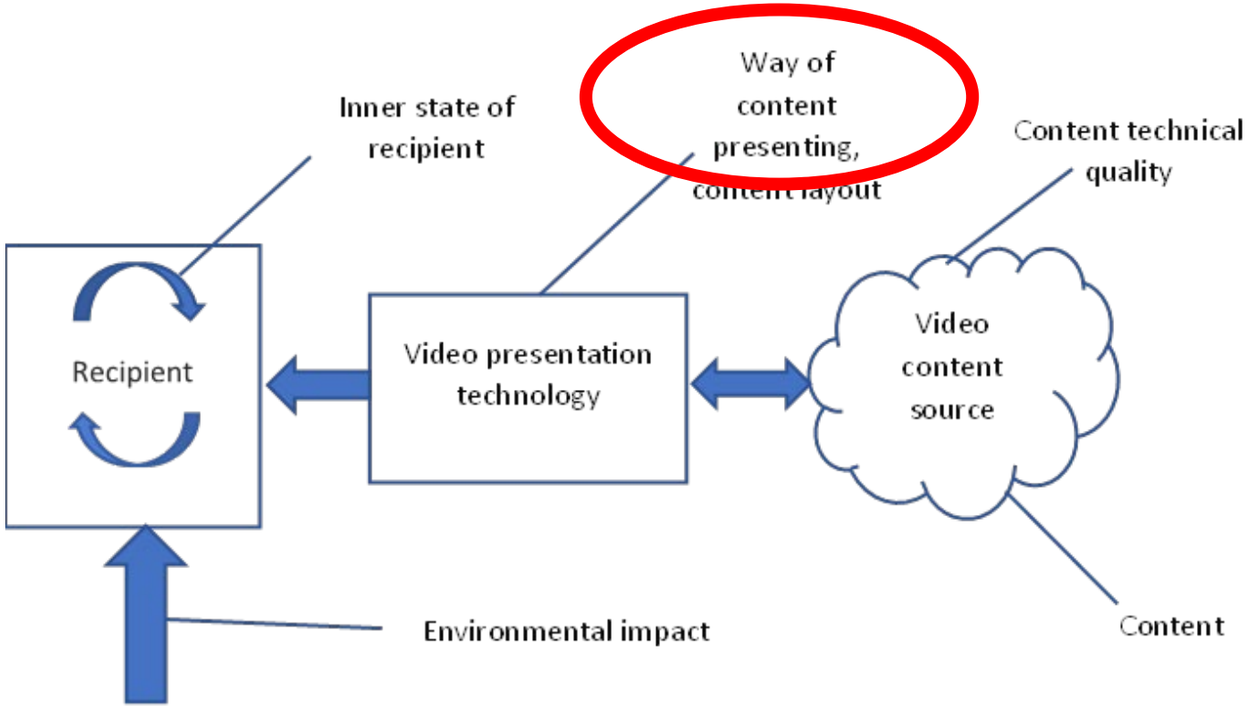
Stimulus presentation structure impact on subjective assessments in QoE

Tomasz Konaszyński
AGH University of Krakow

Video Quality Experts Group (VQEG) meeting, July 3rd, 2024

Factors impacting Recipient's perception



Stimulus presentation structure impact on QoE
Experiment assumptions

- Basic aim of an experiment - a study of 2D video quality perceived by user with verification if and how the order and structure of the video sequence (research stimulus presentation structure) affects the subjective assessment

- Impact of content variability/repeatability and quality of the preceding sequence was analyzed

Research hypotheses

- **Hypothesis 1: There is a relationship between the number of views of a video sequence and its subjective rating**
  - the average subjective rating for the first viewing of a given video sequence is statistically significantly different from the average subjective rating for the third viewing of a given sequence

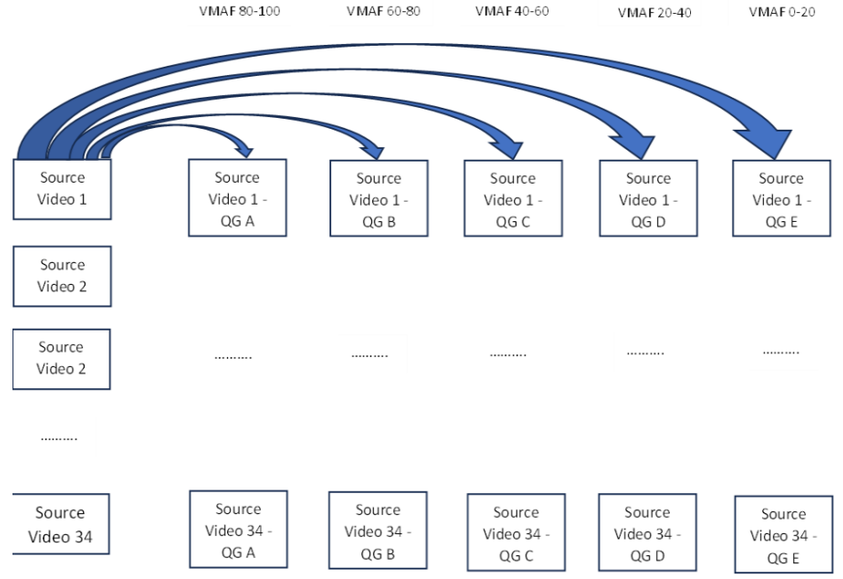
- **Hypothesis 2: There is a relationship between the objective quality of the video immediately preceding a given video sequence and its subjective rating**
  - the average subjective rating for a video whose direct predecessor was a video of higher objective quality differs statistically significantly from the average subjective rating for a video whose direct predecessor was a video of lower objective quality

Experiment design

- The aim - to obtain subjective ratings for a set of PVS sequences displayed in the appropriate order and number of views
- Experiment in accordance with ITU-T P.913 - Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment
- Video presentation sessions divided into two groups: (1) REGULAR group, in which each tester viewed each PVS only once; (2) REPEAT group where each tester viewed each PVS sequence three times.
- Assessment in 5-point Absolute Category Rating (ACR) scale
- In the REGULAR group - each tester was shown a total of 170 PVS sequences, consisting of a randomly selected combination of all 34 unique SRC sequences selected for the experiment, at 5 selected levels of quality degradation - each SRC sequence with a specific level of quality degradation was displayed in sequence random only once
- In the REPEAT group, each tester was shown a total of 180 videos consisting of a randomized combination of 60 unique PVS sequences (12 SRC sequences selected out of 34 SRC videos in 5 quality degradation levels), so that each video sequence was shown to the tester 3 times in random order

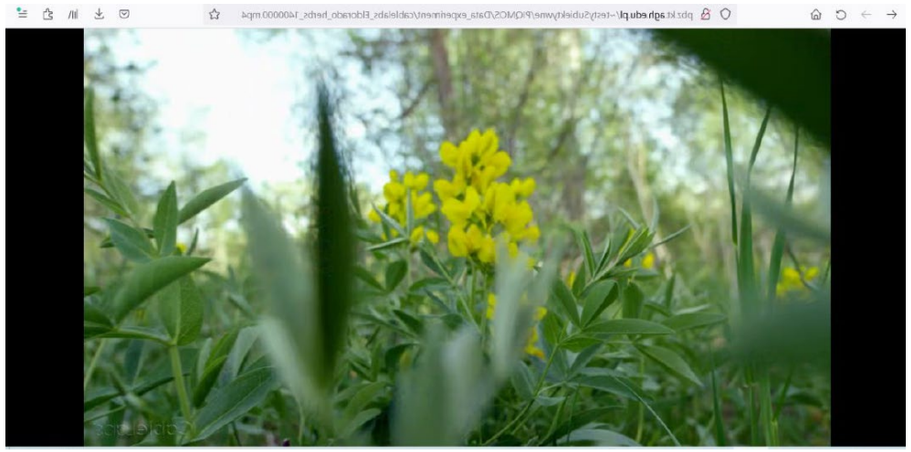
Dataset

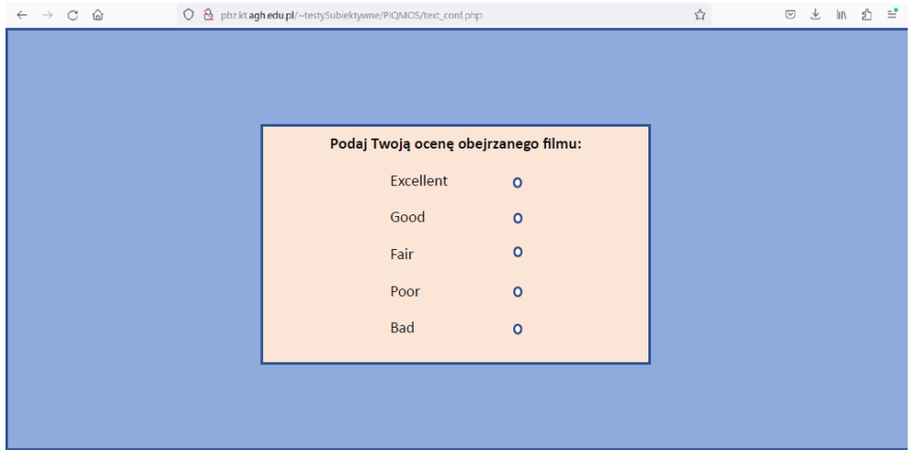
- Set of video sequences selected from sequences made publicly available in the Netflix, CableLabs, SJTU Media Lab and Xiph.org Video Test Media databases
- Full HD video sequences in MPEG-4 standard, with a resolution of 1080p: 1920 x 1080 pixels, with a playback speed of 60 frames/s and with a variable bit rate
- PVS sequences selected for the experiment were created by processing 34 unique SRC video sequences, undifferentiated in terms of characteristics, generating corresponding videos for each sequence at five levels of quality degradation and assigning the videos to appropriate quality groups
- For each SRC sequence, files created at bitrate levels varying between 100,000 bps and 21,000,000 bps were generated
- For each of the 34 selected SRC sequences, one PVS sequence was selected to the appropriate quality group (from A to E), whose objective VMAF assessment was closest to the value of 90 (Group A), 70 (Group B), 50 (Group C), respectively 30 (Group D) and 10 (Group E). Thus, there were 34 unique video sequences in each of the 5 quality groups
- Each PVS sequence lasted 10 s



Data collection interface

- Experiment carried out in the computer laboratory of the AGH Institute of Telecommunications
- Data collection with use of dedicated software - a test platform, created for the purpose of the experiment, available on the university server

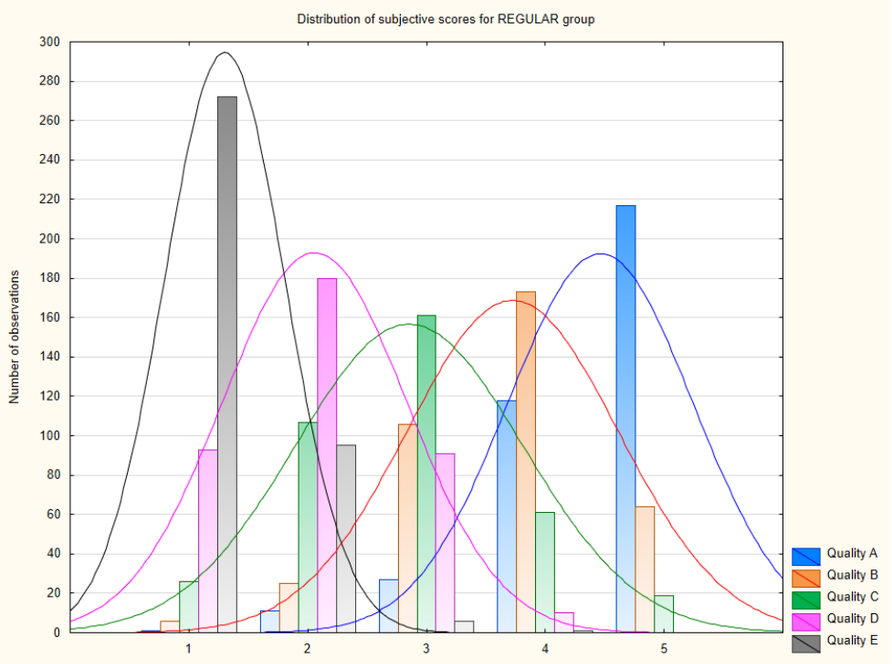


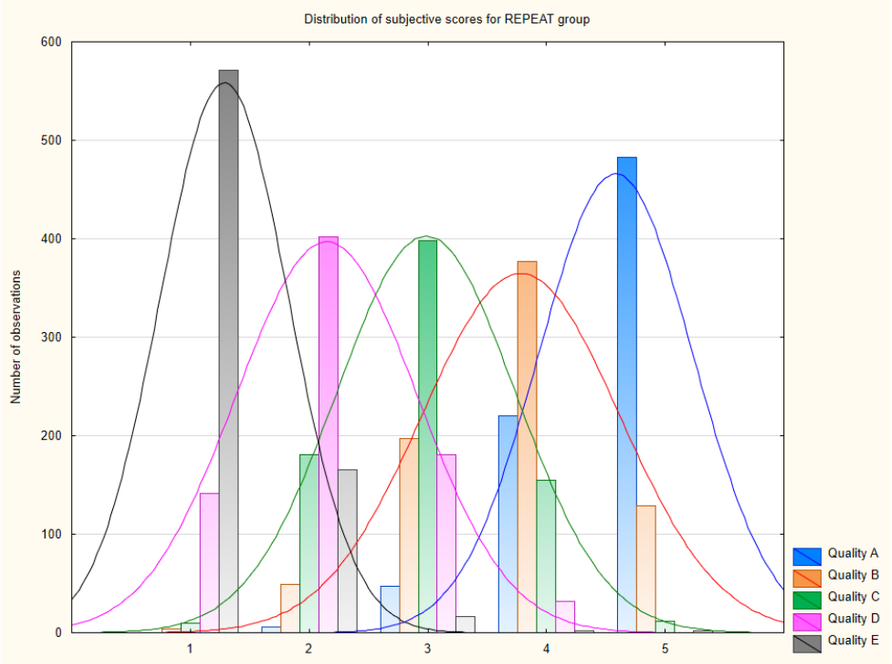


Testers

- Testers' pool was selected out of the population of technical university students interested in multimedia
- The group of testers was homogenous because the selection of testers did not include gender allocation or any other characteristics differentiating the testers
- No analyzes were planned in the experiment regarding the differentiation of any characteristics of the testers
- Experiment involved 35 testers, including 7 women and 28 men, divided into two groups - REGULAR group consisted of 12 people, and the REPEAT group - 23 testers
- Assignment of testers to particular groups was based on the evenness or oddness of the last digit of the ID used by the tester to log into the test platform

Obtained results - histograms





Outliers verification

- Verification method in accordance to ITU-T P.913 recommendation

- Method examines the correlation of tester average scores with the corresponding average scores of all testers using Pearson's linear correlation coefficient

$$\text{LPCC}(x, y) = \frac{\sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}}{\sqrt{\left(\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}\right)\left(\left(\sum_{i=1}^{n} y_i^2 - \frac{(\sum_{i=1}^{n} y_i)^2}{n}\right)\right.}}$$

where xi - average MOS score from all PVS tests, yi: average MOS score for a given tester for PVS, n - total number of PVS

- To calculate LPCC on individual stimuli (i.e., per PVS), compute:

$r_1(x,y)$=LPCC(x,y)

where: x_i: MOS of all subjects per PVS, y_i: individual score of one subject for the corresponding PVS, n: total number of PVSs, I: PVS sequence number

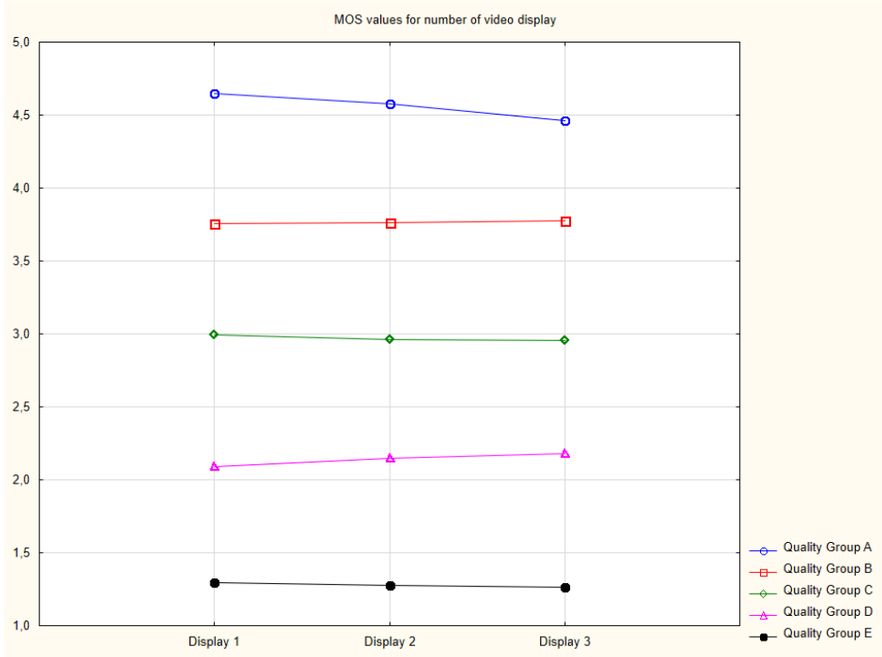
- For setting up a rejection criteria screening analysis is performed per PVS only. Subjects are rejected if r_1 falls below a set threshold. A discard threshold of ($r_1 < 0.75$) is recommended for ACR tests

- Finally, as a result of verification, the scores received from 3 testers were rejected

Hypothesis 1

Tab.4. Average MOS values depending on the multiplicity of views of a given video (PVS) for the REPEAT group

	View 1	View 2	View 3
QG A	4,65	4,57	4,46
QG B	3,75	3,76	3,78
QG C	2,99	2,96	2,96
QG D	2,09	2,15	2,18
QG E	1,30	1,28	1,26



- With statistical significance α=0.05, for the highest level of objective quality, the dependence of the subjective assessment MOS on the number of views of given video was demonstrated

Interpretation

- In the group with the highest objective quality (A), the average MOS score for the 1st view was statistically significantly higher than for the 3rd view

- High-quality PVS ratings deteriorate with subsequent views, which may indicate an increase in testers' criticism of objectively high-quality content

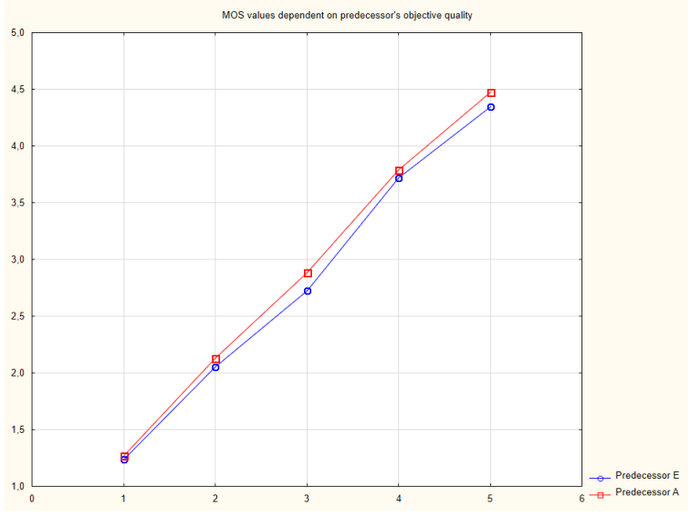
Hypothesis 2

Tab.5. Mean MOS values depending on the objective quality of the antecedent for the REGULAR group

	QG of preceding video				
QG of current video	A	B	C	D	E
A	4,48	4,41	4,39	4,58	4,35
B	3,79	3,78	3,79	3,45	3,72
C	2,88	2,99	2,74	2,87	2,72
D	2,13	2,13	1,95	1,96	2,05
E	1,27	1,30	1,29	1,37	1,24

Tab.6. Mean MOS values depending on the objective quality of the antecedent for the REPEAT group

	QG of preceding video				
QG of current video	A	B	C	D	E
A	4,53	4,47	4,54	4,42	4,31
B	3,74	3,70	3,61	3,67	3,69
C	3,03	2,98	2,93	2,95	2,92
D	2,34	2,19	2,14	2,04	2,07
E	1,44	1,31	1,36	1,28	1,34



- With statistical significance α=0.05, for the highest level of video quality, the dependence of the subjective assessment on the quality of the previously displayed sequence was demonstrated - the average rating of the video in the case where the preceding video belonged to the highest quality group is higher than in the case where the preceding video was of very low quality

Results interpretation

- In each quality group of both research samples (REGULAR/REPEAT), the average MOS score for the preceding videos from the highest quality group (A) was higher than for the preceding videos from the lowest quality group (E)

- Although statistical significance was obtained only in the REPEAT group for quality groups A and D, repeatable higher assessments obtained in the group with a very high-quality predecessor indicate the desirability of further research in this area

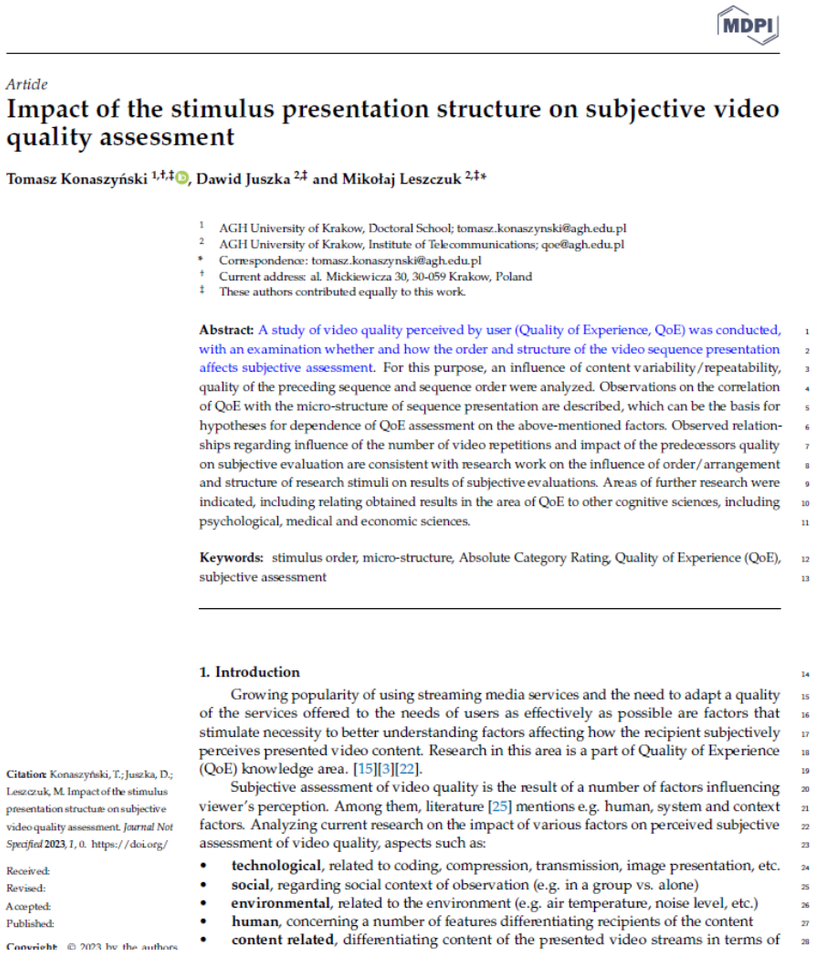
Future directions

- analysis of the results obtained in the context of the state of knowledge and existing research results, both in the area of multimedia Quality of Experience and correlation to other cognitive sciences, which are affected by the aspect of the impact of the stimuli presentation structure on outcome results (medicine, psychology, economics, etc.)
- randomization of the layout and structure of multimedia content presentation in a way that eliminates or minimizes the impact of the phenomena listed above on the results of subjective research
- Analyzing impact of actions contrary to randomization, i.e. intentional strengthening of an "order bias" effects to maximize perceived subjective quality through appropriate selection of the tested stimuli presentation structure
- deepening future analyzes towards taking into account the correlation of the obtained phenomena with other factors, in particular of a human and systemic nature

Publications referred to an experiment

- Konaszyński, T., Juszka, D., & Leszczuk, M. (2023). Impact of the Stimulus Presentation Structure on Subjective Video Quality Assessment. Electronics, 12(22), 4593

- Konaszyński, T., Juszka, D., & Leszczuk, M. (2023). Analiza wpływu mikro-struktury prezentacji treści na subiektywną ocenę jakości wideo i jej stabilność. Przegląd Telekomunikacyjny + Wiadomości Telekomunikacyjne



MDPI

Article

Impact of the stimulus presentation structure on subjective video quality assessment

Tomasz Konaszyński [1,†,‡], Dawid Juszka [2,‡] and Mikołaj Leszczuk [2,‡,*]

[1] AGH University of Krakow, Doctoral School; tomasz.konaszynski@agh.edu.pl
[2] AGH University of Krakow, Institute of Telecommunications; qoe@agh.edu.pl
* Correspondence: tomasz.konaszynski@agh.edu.pl
† Current address: al. Mickiewicza 30, 30-059 Krakow, Poland
‡ These authors contributed equally to this work.

Abstract: A study of video quality perceived by user (Quality of Experience, QoE) was conducted, with an examination whether and how the order and structure of the video sequence presentation affects subjective assessment. For this purpose, an influence of content variability/repeatability, quality of the preceding sequence and sequence order were analyzed. Observations on the correlation of QoE with the micro-structure of sequence presentation are described, which can be the basis for hypotheses for dependence of QoE assessment on the above-mentioned factors. Observed relationships regarding influence of the number of video repetitions and impact of the predecessors quality on subjective evaluation are consistent with research work on the influence of order/arrangement and structure of research stimuli on results of subjective evaluations. Areas of further research were indicated, including relating obtained results in the area of QoE to other cognitive sciences, including psychological, medical and economic sciences.

Keywords: stimulus order, micro-structure, Absolute Category Rating, Quality of Experience (QoE), subjective assessment

1. Introduction

Growing popularity of using streaming media services and the need to adapt a quality of the services offered to the needs of users as effectively as possible are factors that stimulate necessity to better understanding factors affecting how the recipient subjectively perceives presented video content. Research in this area is a part of Quality of Experience (QoE) knowledge area. [15][3][22].

Subjective assessment of video quality is the result of a number of factors influencing viewer's perception. Among them, literature [25] mentions e.g. human, system and context factors. Analyzing current research on the impact of various factors on perceived subjective assessment of video quality, aspects such as:

- **technological**, related to coding, compression, transmission, image presentation, etc.
- **social**, regarding social context of observation (e.g. in a group vs. alone)
- **environmental**, related to the environment (e.g. air temperature, noise level, etc.)
- **human**, concerning a number of features differentiating recipients of the content
- **content related**, differentiating content of the presented video streams in terms of

Citation: Konaszyński, T.; Juszka, D.; Leszczuk, M. Impact of the stimulus presentation structure on subjective video quality assessment. *Journal Not Specified* **2023**, *1*, 0. https://doi.org/

Received:
Revised:
Accepted:
Published:

Thank you for attention

Tomasz Konaszynski

AGH University of Krakow