Precyzja testów subiektywnych - modele i nowe propozycje szacowania dokładności eksperymentów w sieciach telekomunikacyjnych

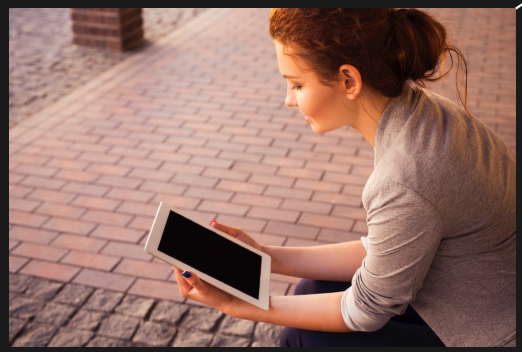
VI posiedzenie Sekcji Telekomunikacji PAN

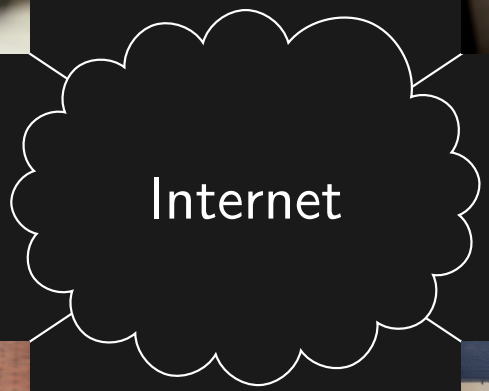
Lucjan Janowski, AGH

May 24, 2023

Outline

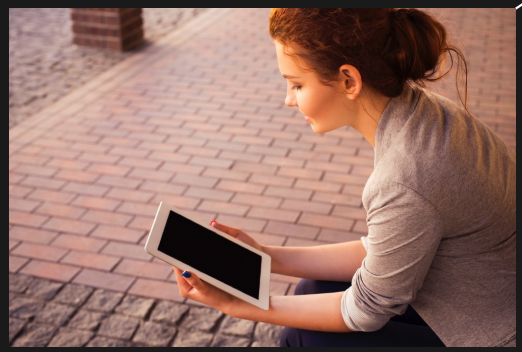
Internet



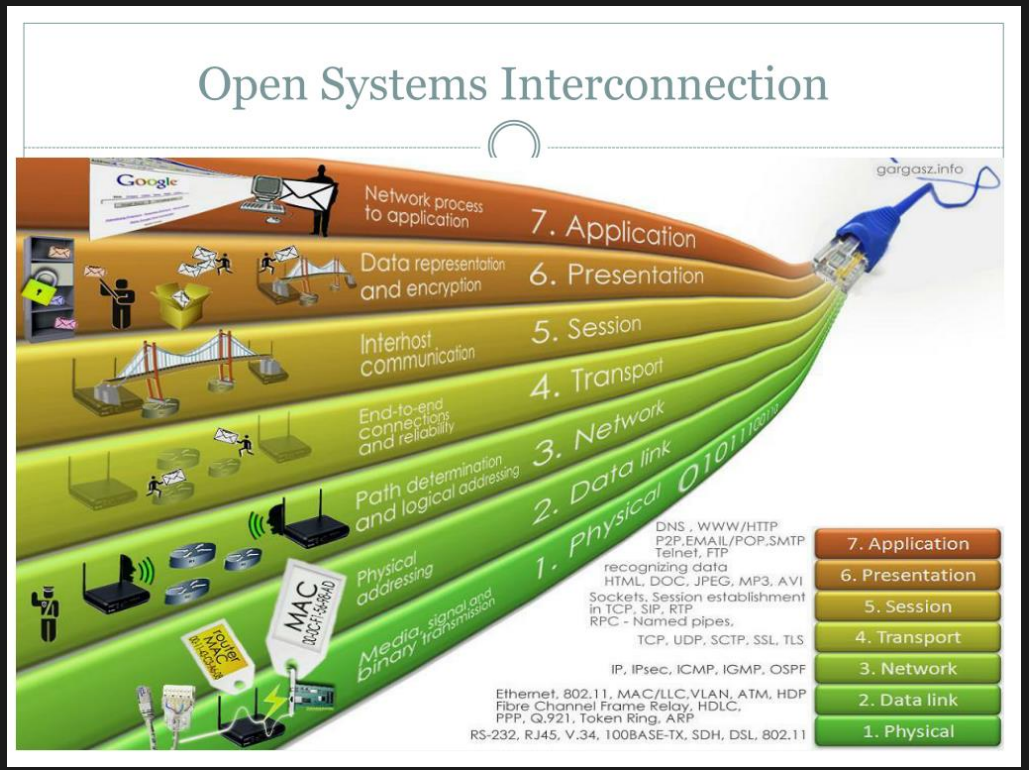


Quality of Service

Packets



ISO OSI

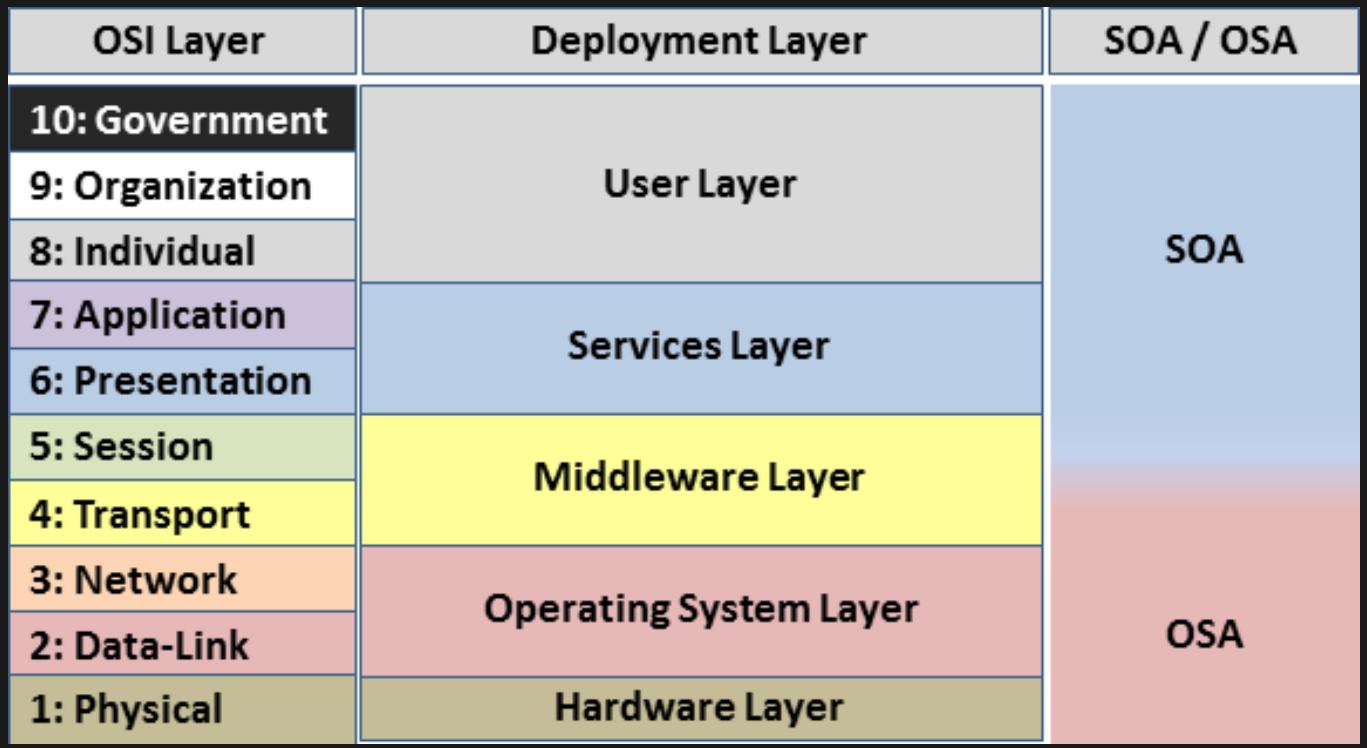


Source: `https://www.openworldlearning.org/understanding-the-importance-of-network-layers-in-telecommunications/`

Quality of Experience

Needs

Full Layers Model



Source: ByGvseostud-Ownwork, CCBY-SA3.0, https://commons.wikimedia.org/w/index.php?curid=29156115

Outline

What is Quality of Experience?

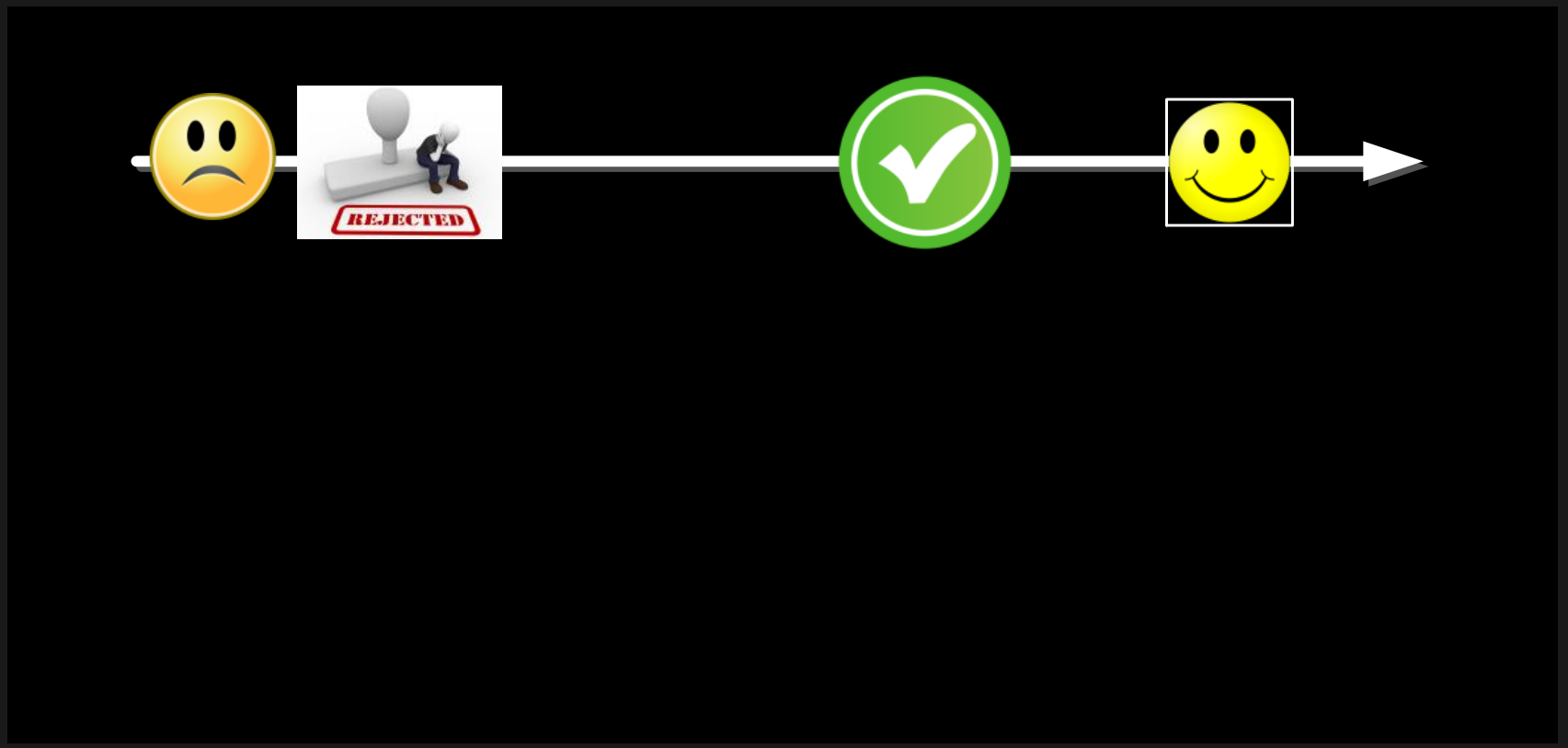
QoE Definition - ITU

ITU-T Study Group 12 (Geneva, 16-25 January 2007)
Ref. : TD 109rev2 (PLEN/12)

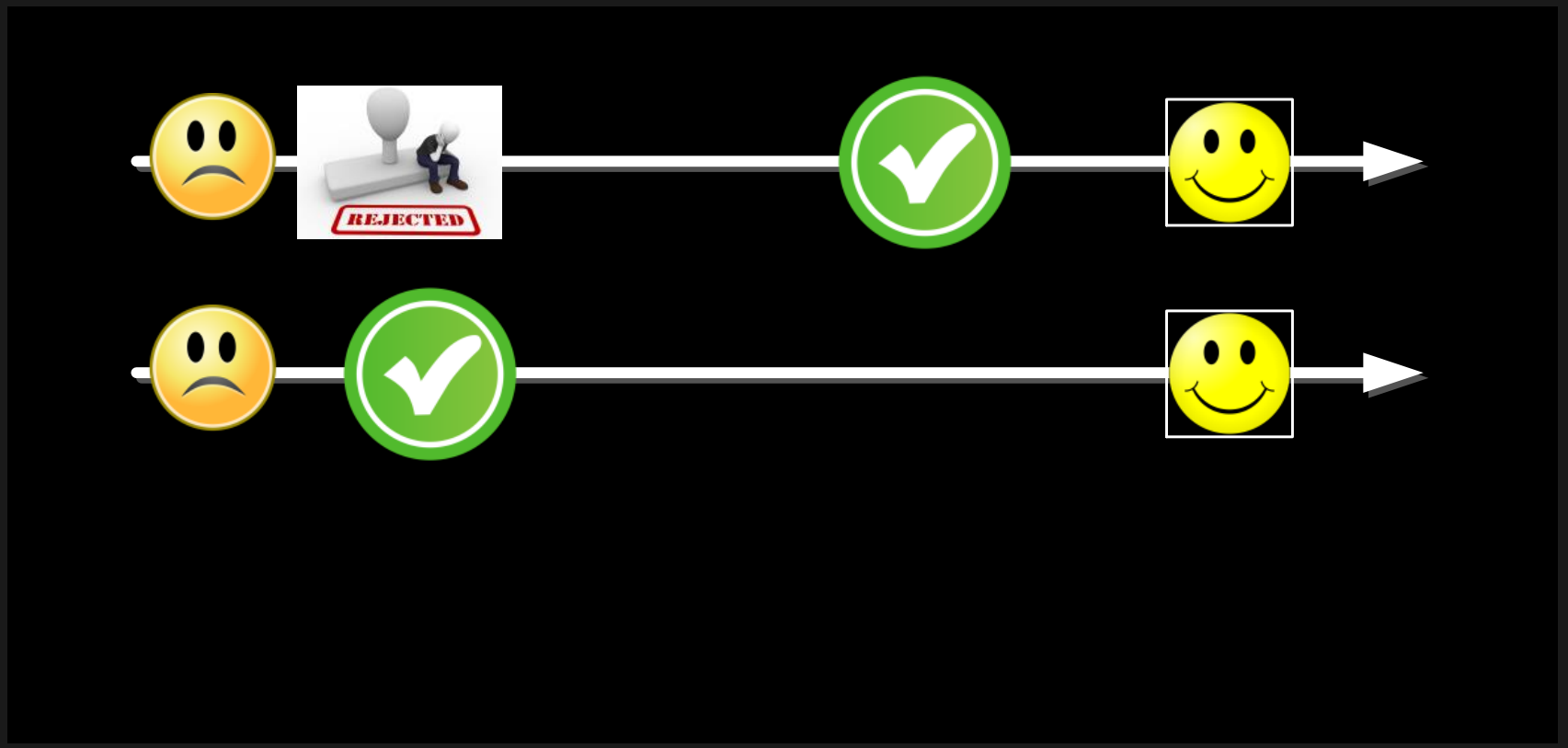
Quality of Experience (QoE)
The overall acceptability of an application or service, as perceived subjectively by the end-user.
NOTES

1. Quality of Experience includes the complete end-to-end system effects (client, terminal, network, services infrastructure, etc).
2. Overall acceptability may be influenced by user expectations and context.

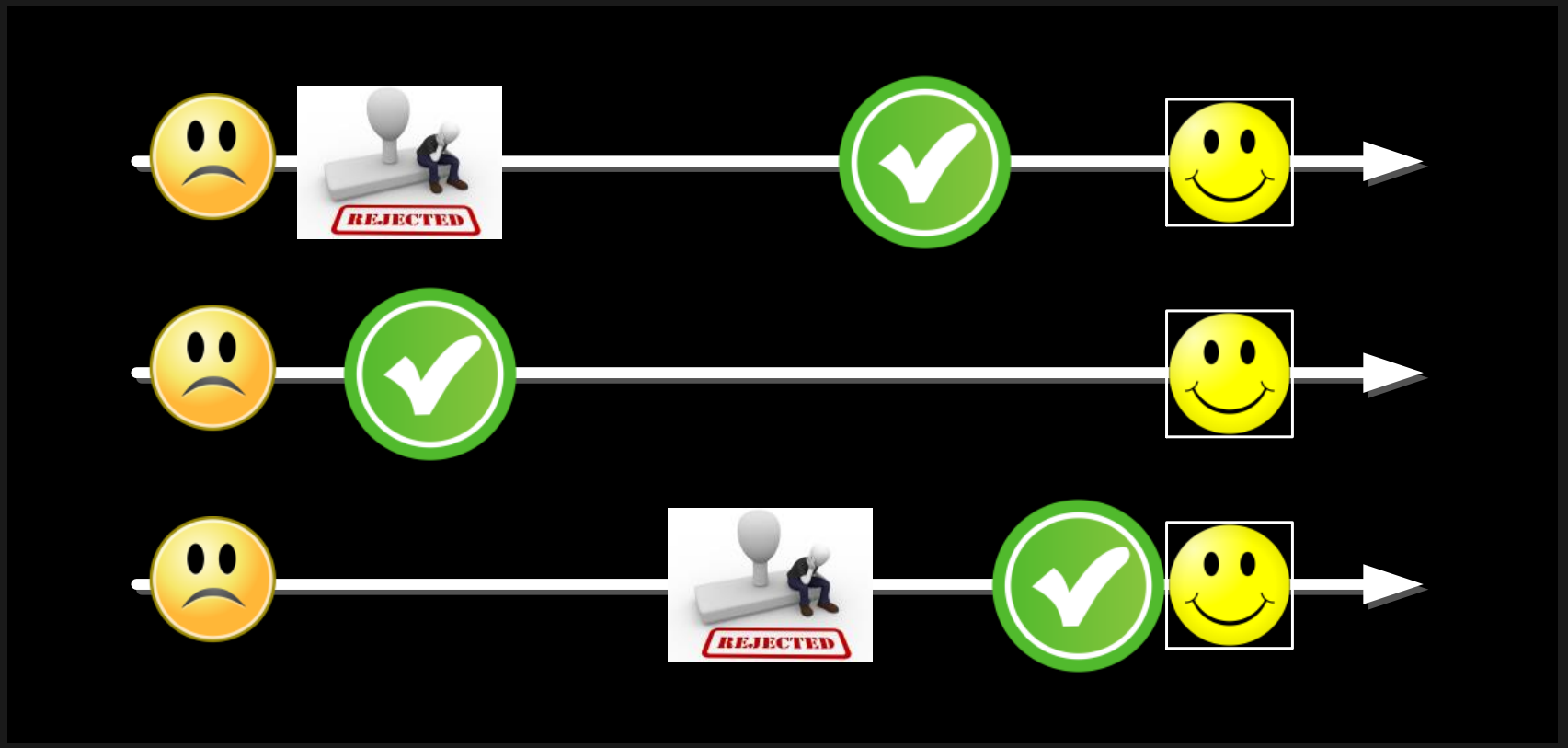
QoE $\neq$ Acceptability



QoE $\neq$ Acceptability



QoE $\neq$ Acceptability



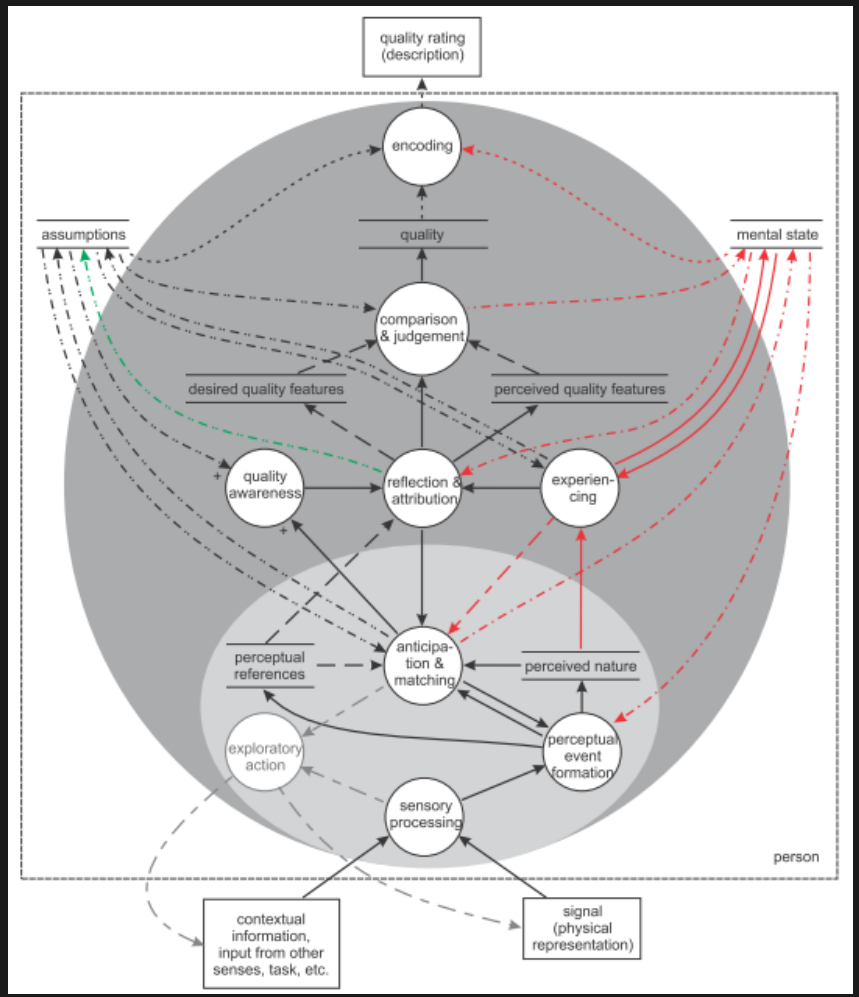
New QoE Definition

Quality of Experience (QoE) "is the degree of delight or annoyance of the user of an application or service. It results from the fulfillment of his or her expectations with respect to the utility and / or enjoyment of the application or service in the light of the user's personality and current state."

The first versions (paper): "Qualinet White Paper on Definitions of Quality of Experience" (2012)

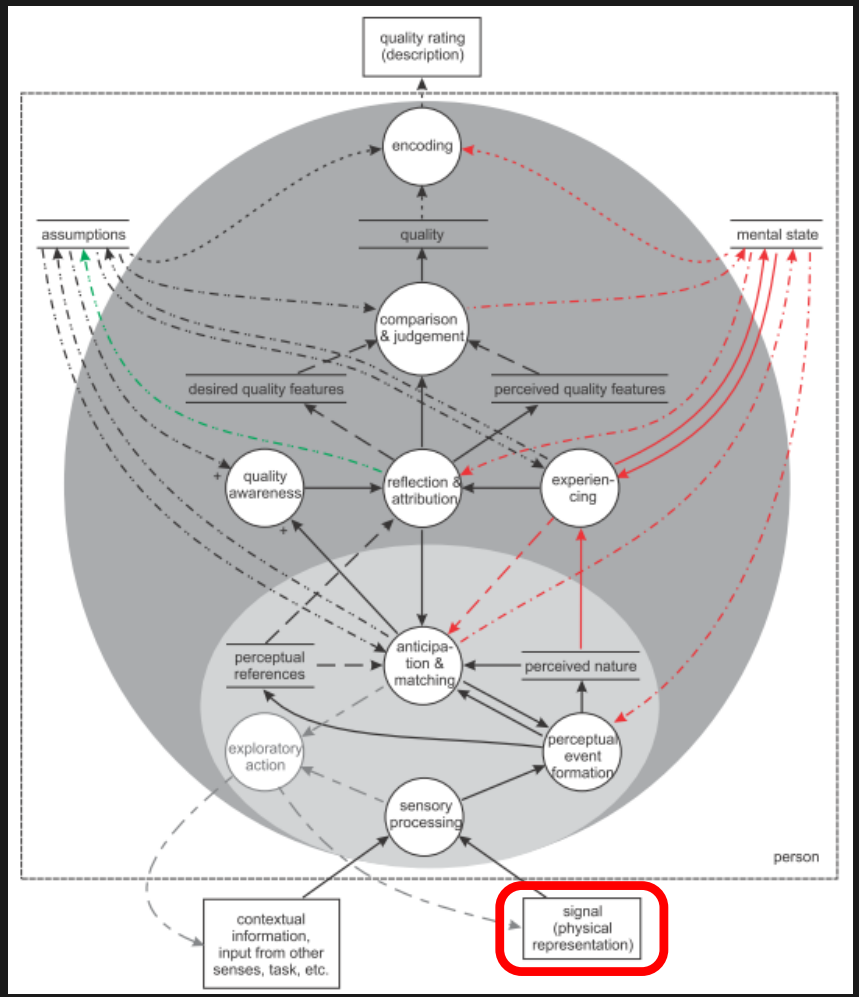
The second version (full book): "Quality of Experience: Advanced Concepts, Applications and Methods" (2014)

Answering Process



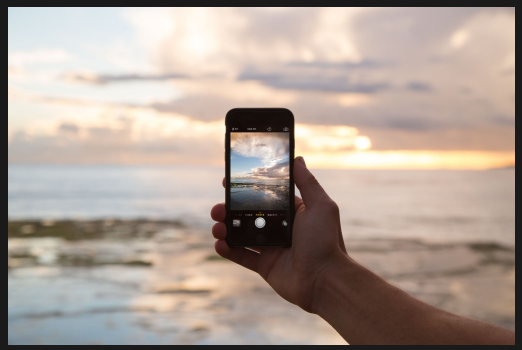
"Quality formation process during active experiencing." Copy from book "Quality of Experience: Advanced Concepts, Applications and Methods"

Answering Process

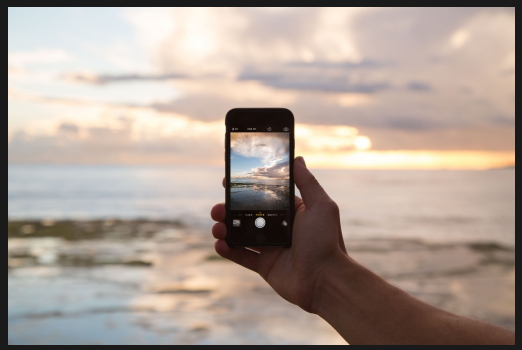


"Quality formation process during active experiencing." Copy from book "Quality of Experience: Advanced Concepts, Applications and Methods"

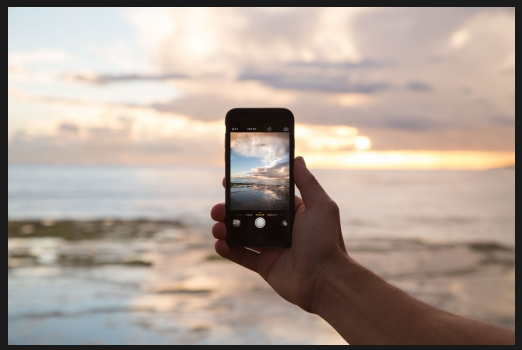
Who Is the System User?



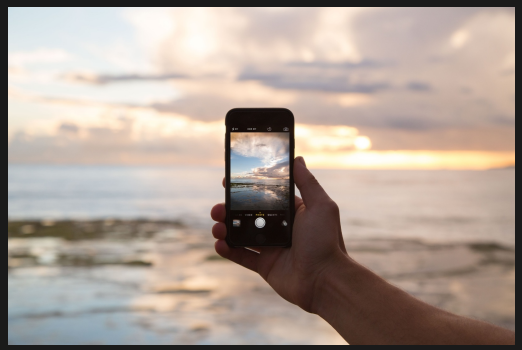
Who Is the System User?

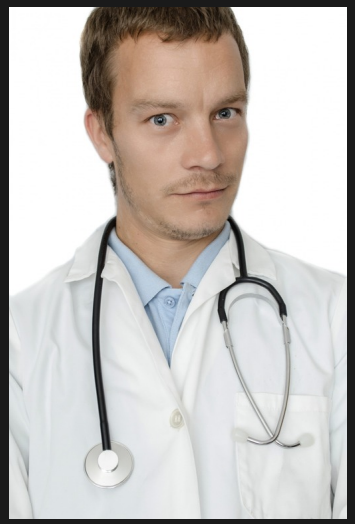


Who Is the System User?

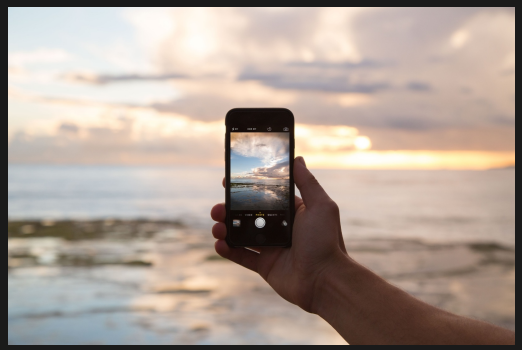


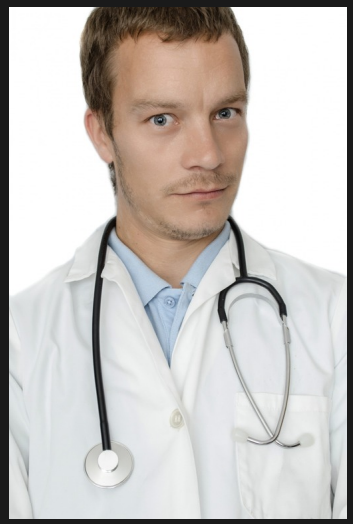
Who Is the System User?

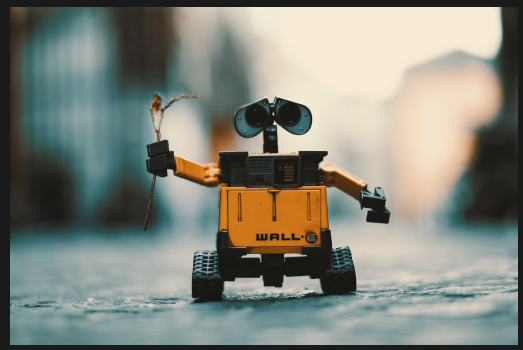




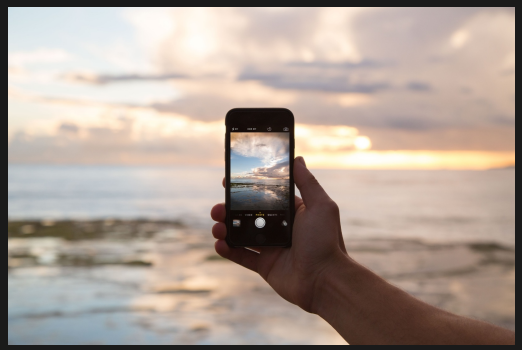
Who Is the System User?

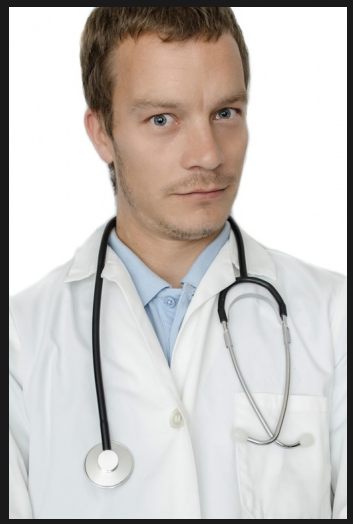


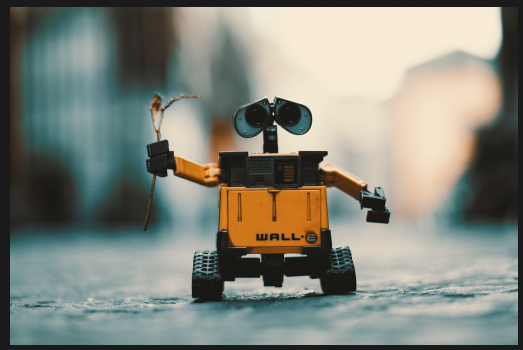




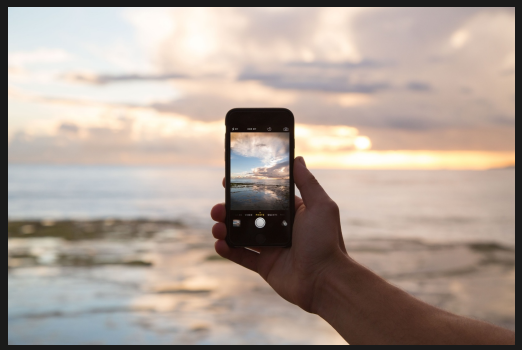
Subjects, Entertainment







Subjects, Entertainment



Outline

Subjects

Test 1

Answer Test 1

- 5 - Excellent
- 4 - Good
- 3 - Average
- 2 - Poor
- 1 - Bad

- 5 - 非常に良い
- 4 - 良い
- 3 - 普通
- 2 - 悪い
- 1 - 非常に悪い

Test 2

Answer Test 2

- 5 - Excellent
- 4 - Good
- 3 - Average
- 2 - Poor
- 1 - Bad

- 5 - 非常に良い
- 4 - 良い
- 3 - 普通
- 2 - 悪い
- 1 - 非常に悪い

Test 3

Answer Test 3

- 5 - Excellent
- 4 - Good
- 3 - Average
- 2 - Poor
- 1 - Bad

- 5 - 非常に良い
- 4 - 良い
- 3 - 普通
- 2 - 悪い
- 1 - 非常に悪い

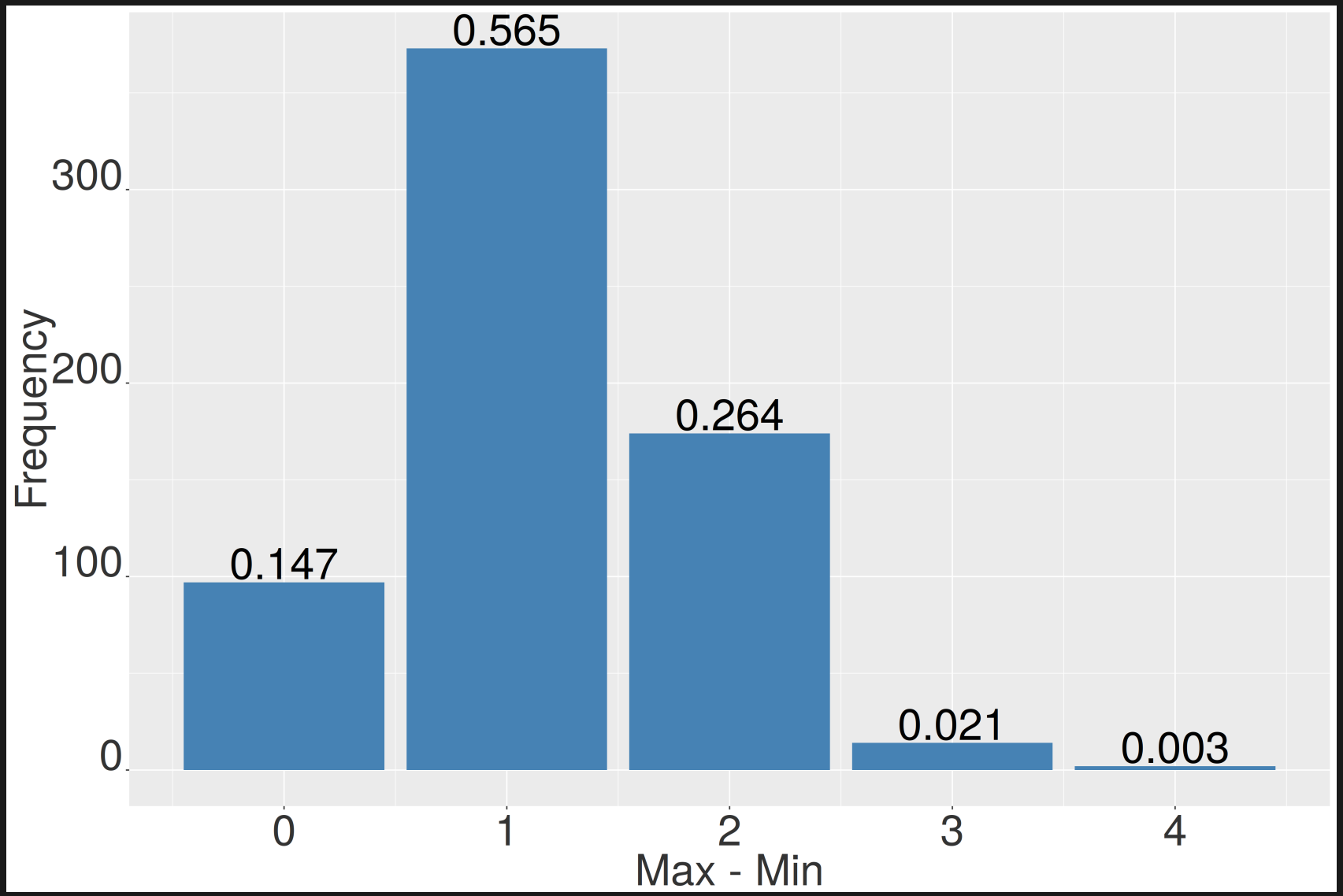
Summary

Scale Problem

Answer “4”: 9 out of 10

Answer “3”: 1 out of 10

Repetition


0.565
0.264
0.147
0.021
0.003
300
200
100
0
Frequency
0
1
2
3
4
Max - Min

Physical Measurements

Image Quality

Image Quality

Image Quality

Image Quality

Conclusions

- We cannot trust subjects' opinions
- Physical measurements are more precise
- FR metrics see small differences

Am I Right?

Am I Right?

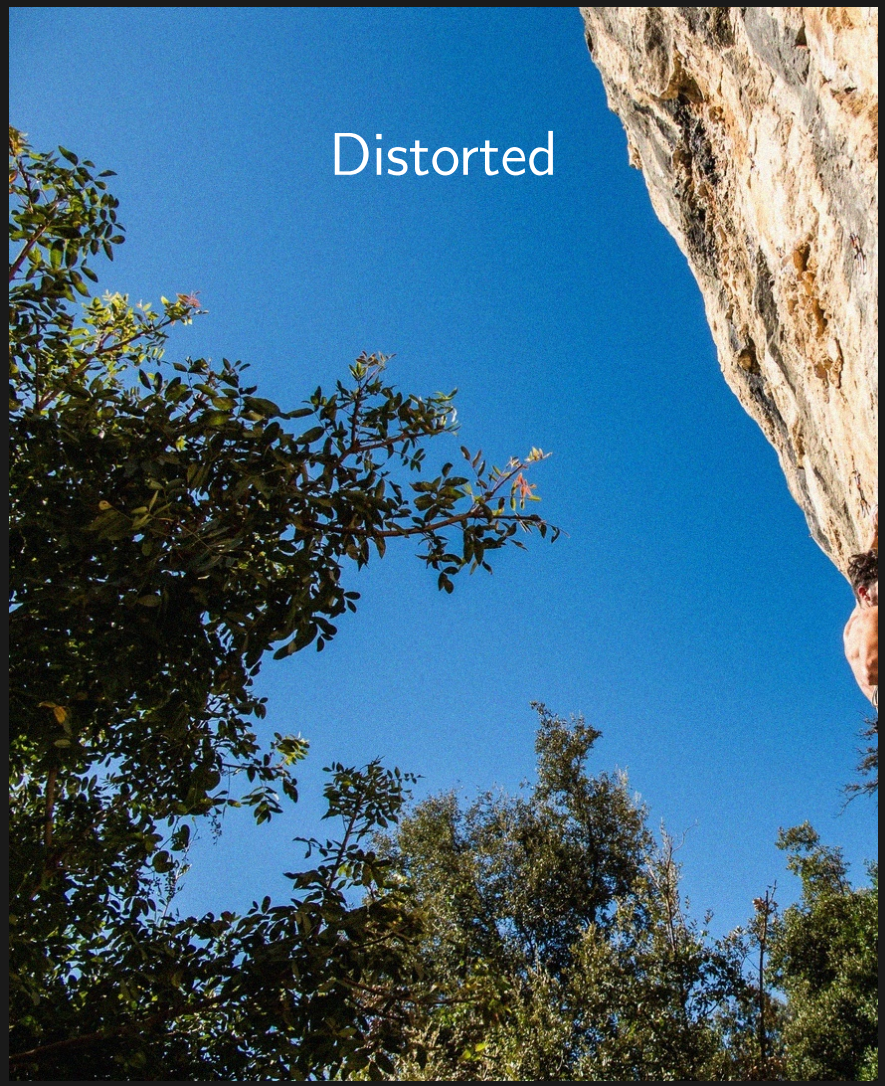
Can we remove human subject from the equation?

Should We be Precise?

Orginal

Distorted

Should We be Precise?

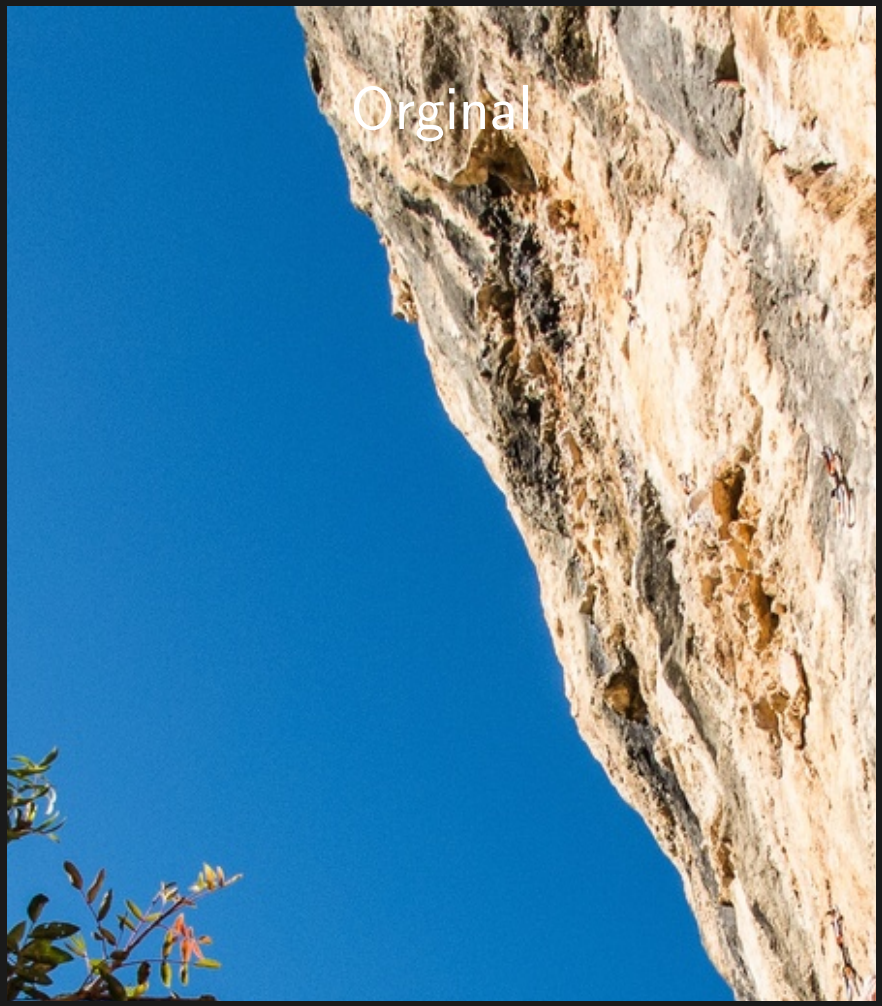


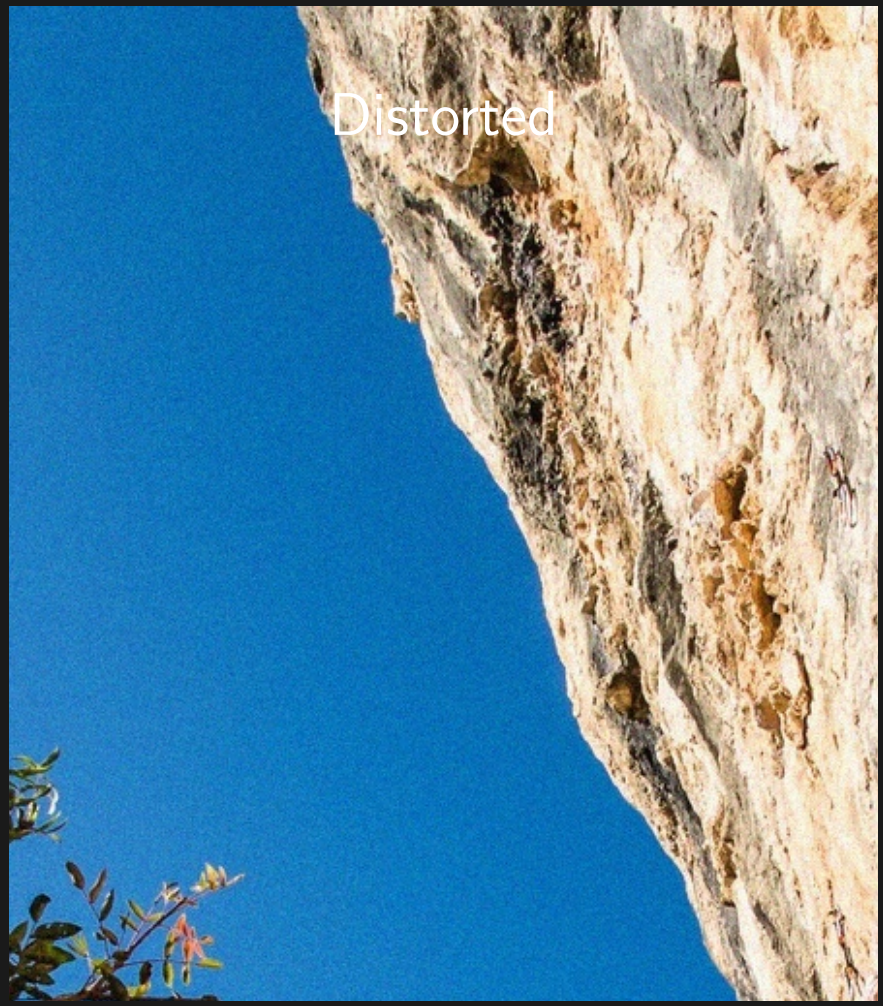
Should We be Precise?

Orginal

Distorted

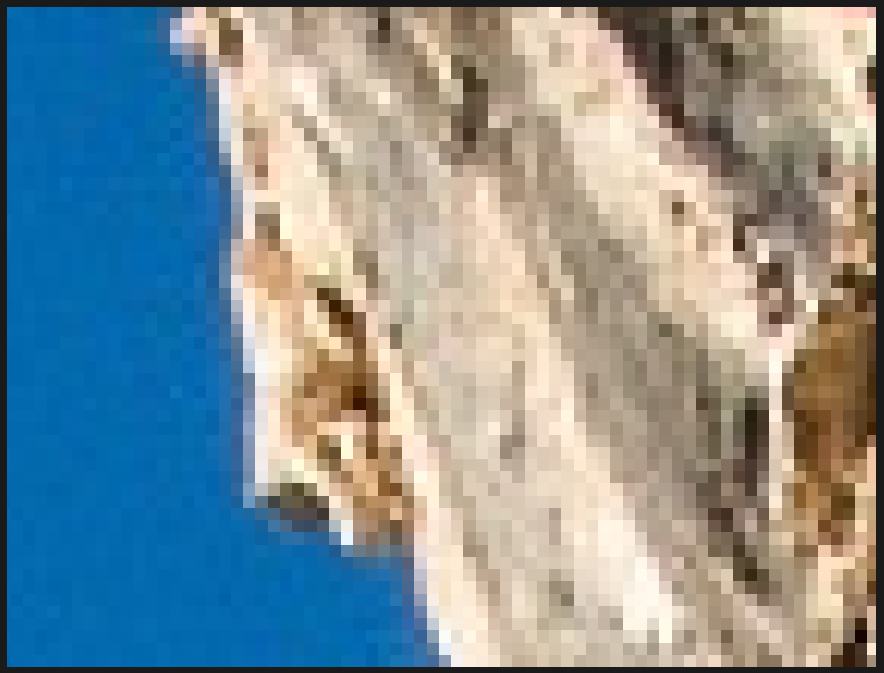
Should We be Precise?



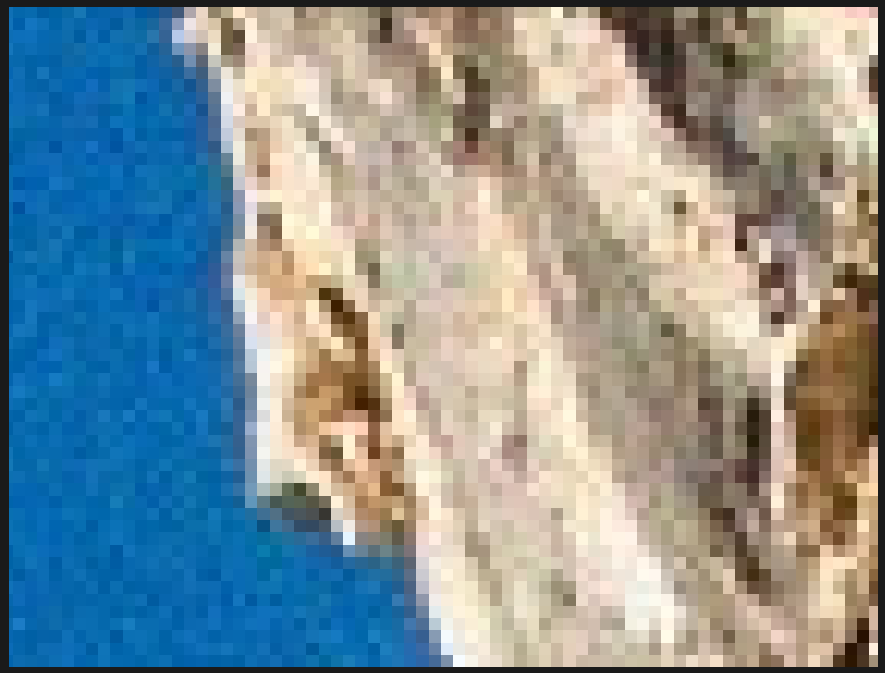


Should We be Precise?

Orginal



Distorted



Should We be Precise?

Orginal

Distorted

Specific Cases

Conclusion

We should teach our metrics to be less precise

Conclusion

We should teach our metrics to be less precise

To do so we need subjective experiments

www.vqeg.org

AGH

VQEG (Video Quality Expert Group) works on validating metrics.
You can join us!

Some projects:

- SAM (Statistical Analysis Methods) - I am the chair
- IMG (Immersive Media Group)
- No Reference Metrics (NORM)

www.vqeg.org

AGH

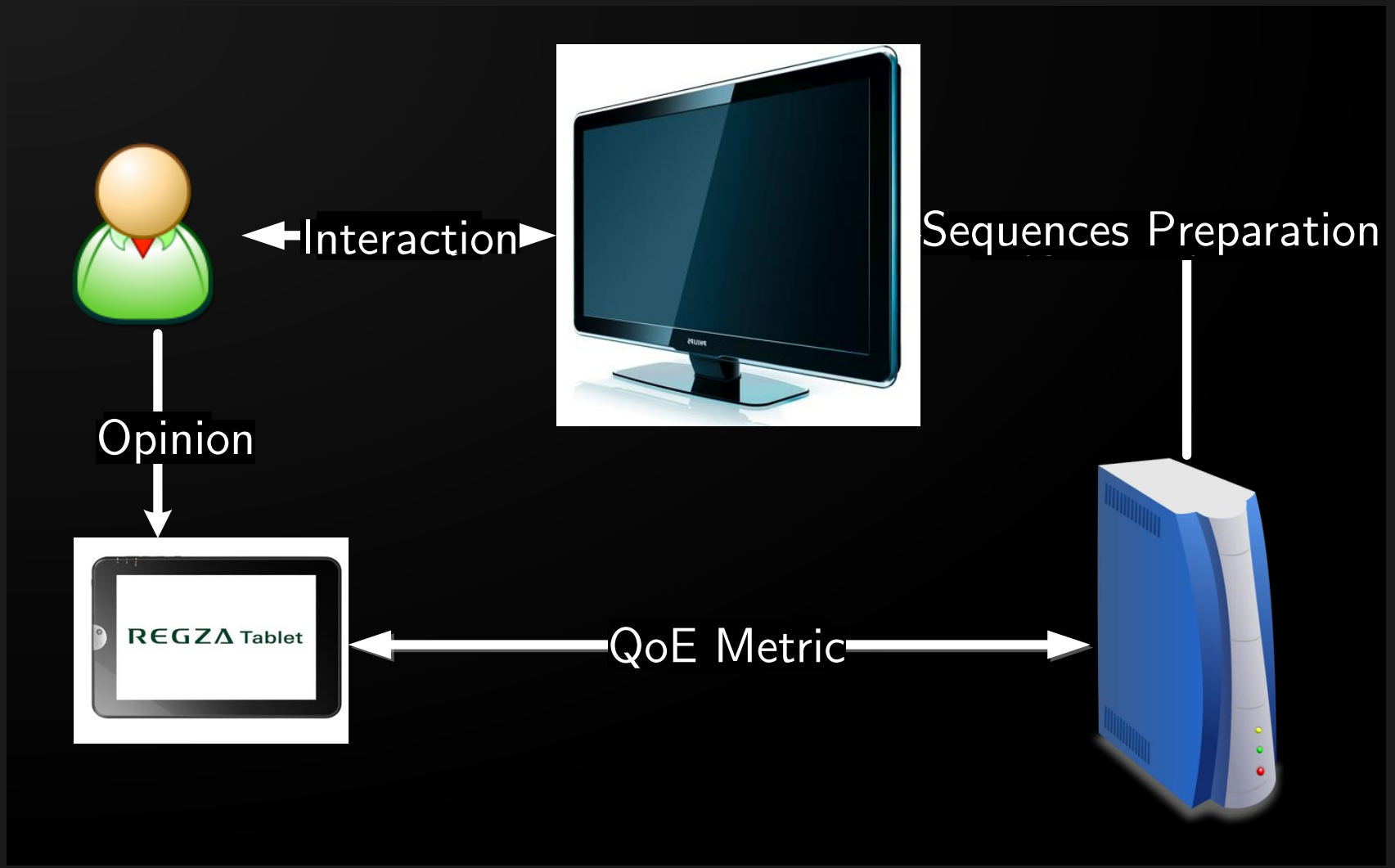
VQEG (Video Quality Expert Group) works on validating metrics.
You can join us!

Some projects:

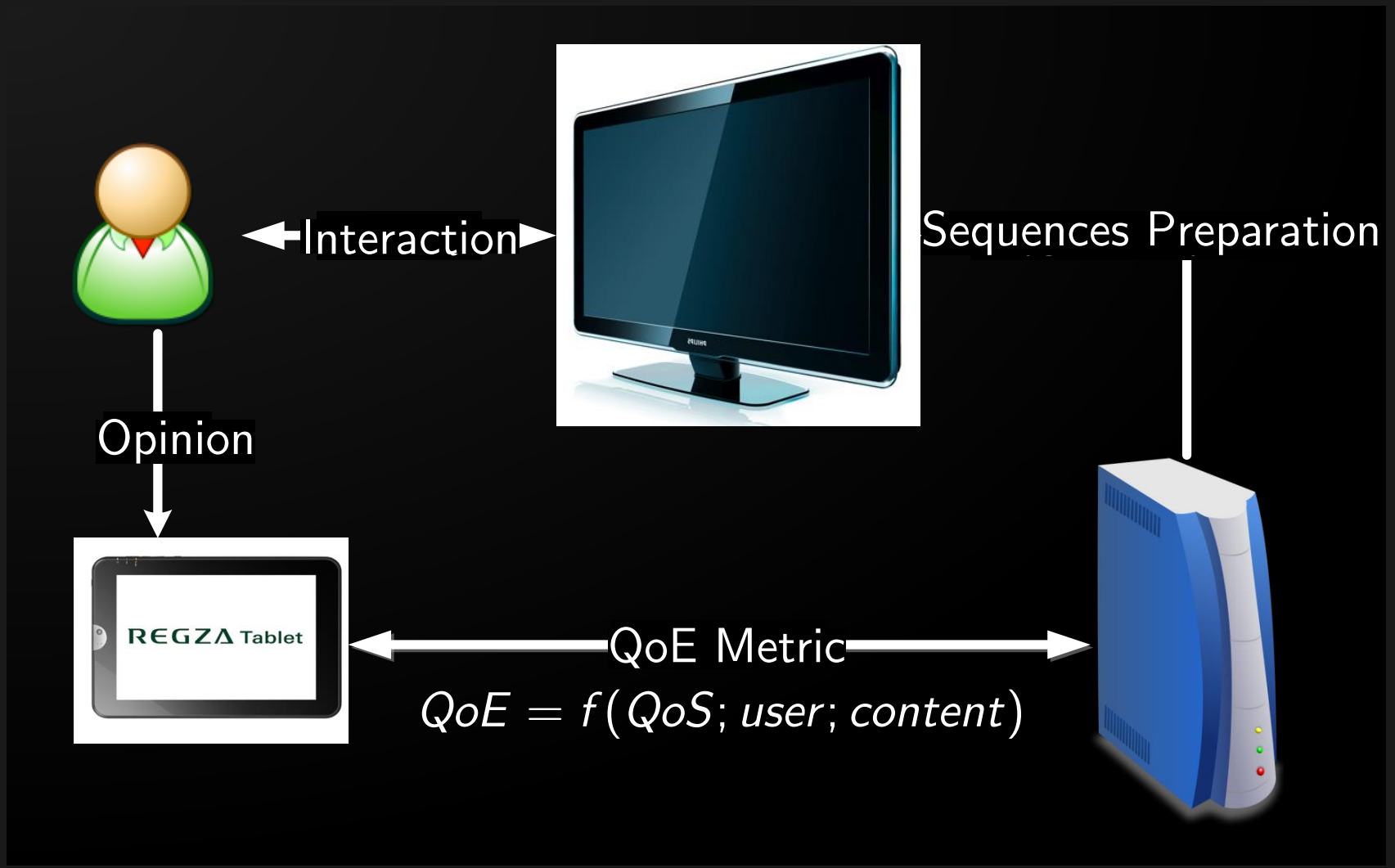
- SAM (Statistical Analysis Methods) - I am the chair
- IMG (Immersive Media Group)
- No Reference Metrics (NORM)

For pixel quality do not use PSNR, use VMAF from Netflix instead!

Subjective Experiment



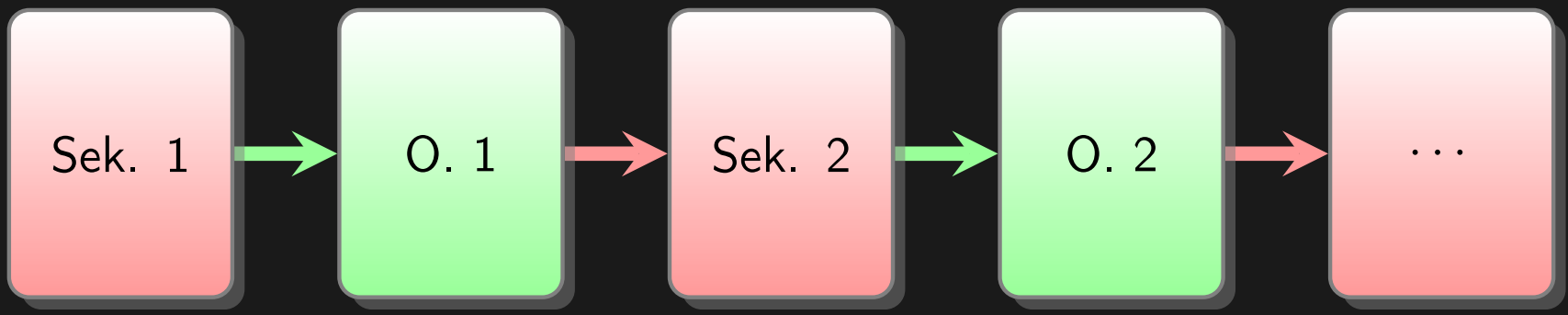
Subjective Experiment



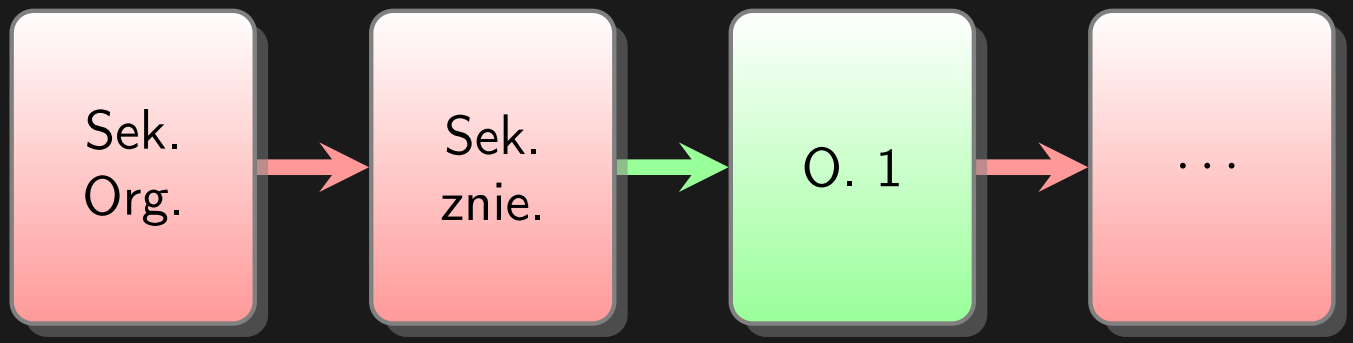
Experiment Preparation

	SRC1	SRC2	SRC3
HRC1			
HRC2			
HRC3			

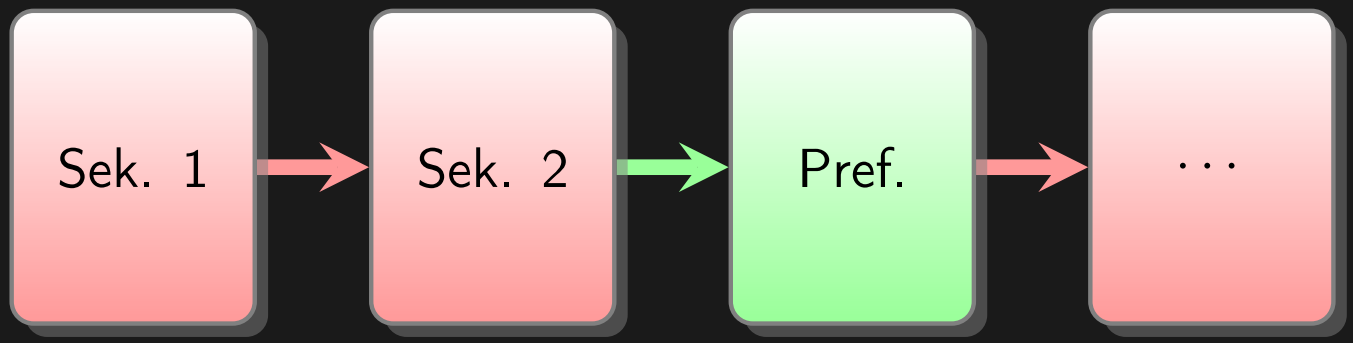
ACR (Absolute Category Rating)



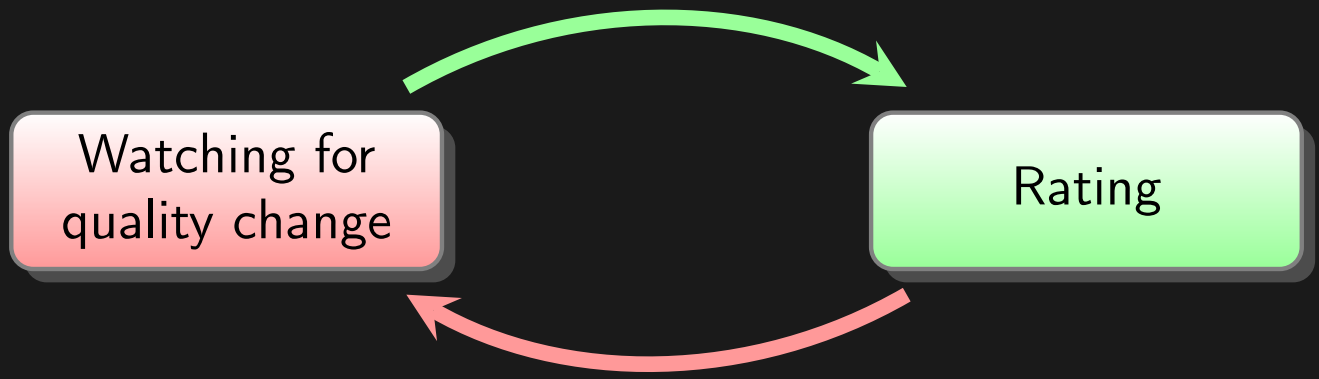
DCR (Degradation Category Rating)



PC (Pair Comparison)



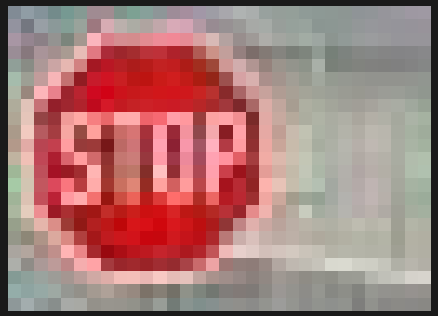
SSCQE (Single Stimulus Continuous Quality Rating)



Subjective Experiment

1,2,3,4,5

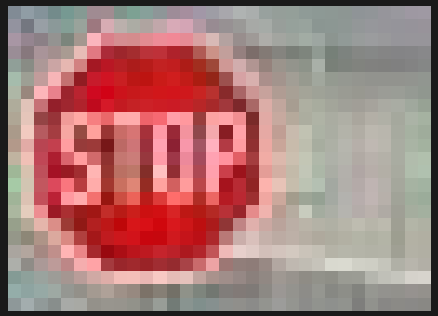
Subjective Experiment



1,2,3,4,5

1,2,3,4,5

Subjective Experiment



Quality 4.6

↑

1,2,3,4,5

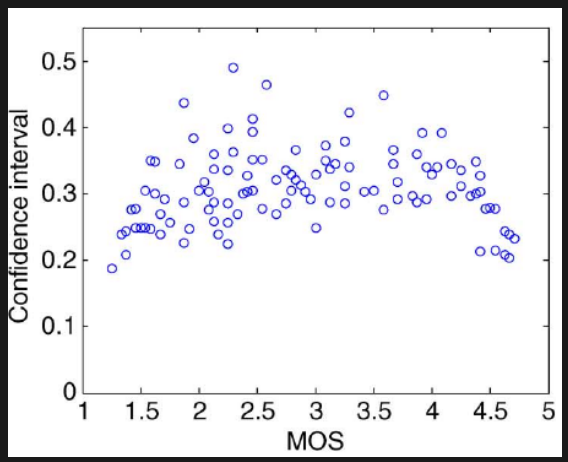
Quality 2.3

↑

1,2,3,4,5

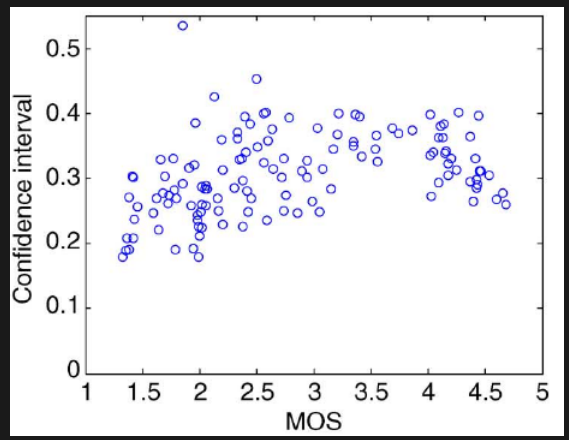
Different Scales

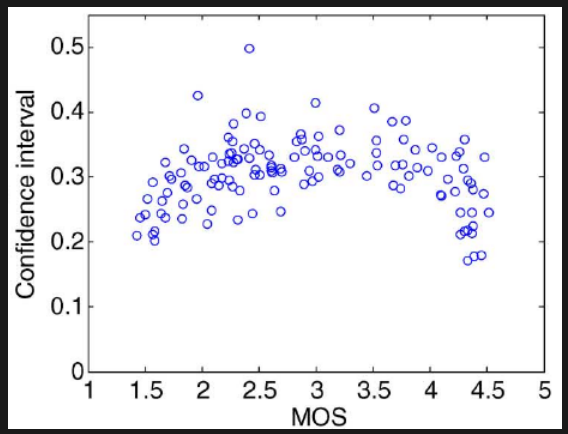
Q. Huynh-Thu, M. Garcia, F. Speranza, P. Corriveau and A. Raake, "Study of Rating Scales for Subjective Quality Assessment of High-Definition Video," in IEEE Transactions on Broadcasting, vol. 57, no. 1, pp. 1-14, March 2011.



$\leftarrow \{1, 2, 3, 4, 5\}$

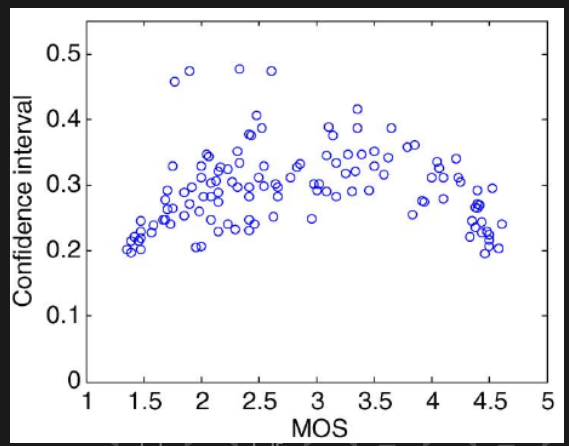
$(0, 10) \rightarrow$





$\leftarrow (1, 5)$

$\{1, \cdots, 9\} \rightarrow$



ACR versus PC

T. Tominaga, T. Hayashi, J. Okamoto and A. Takahashi, "Performance comparisons of subjective quality assessment methods for mobile video," 2010 Second International Workshop on Quality of Multimedia Experience (QoMEX), Trondheim, 2010, pp. 82-87.

DSCQS

ACR1

ACR2

ACR versus PC

T. Tominaga, T. Hayashi, J. Okamoto and A. Takahashi, "Performance comparisons of subjective quality assessment methods for mobile video," 2010 Second International Workshop on Quality of Multimedia Experience (QoMEX), Trondheim, 2010, pp. 82-87.

DSCQS

Similar

ACR1

ACR2

ACR versus PC

T. Tominaga, T. Hayashi, J. Okamoto and A. Takahashi, "Performance comparisons of subjective quality assessment methods for mobile video," 2010 Second International Workshop on Quality of Multimedia Experience (QoMEX), Trondheim, 2010, pp. 82-87.

DSCQS

Similar

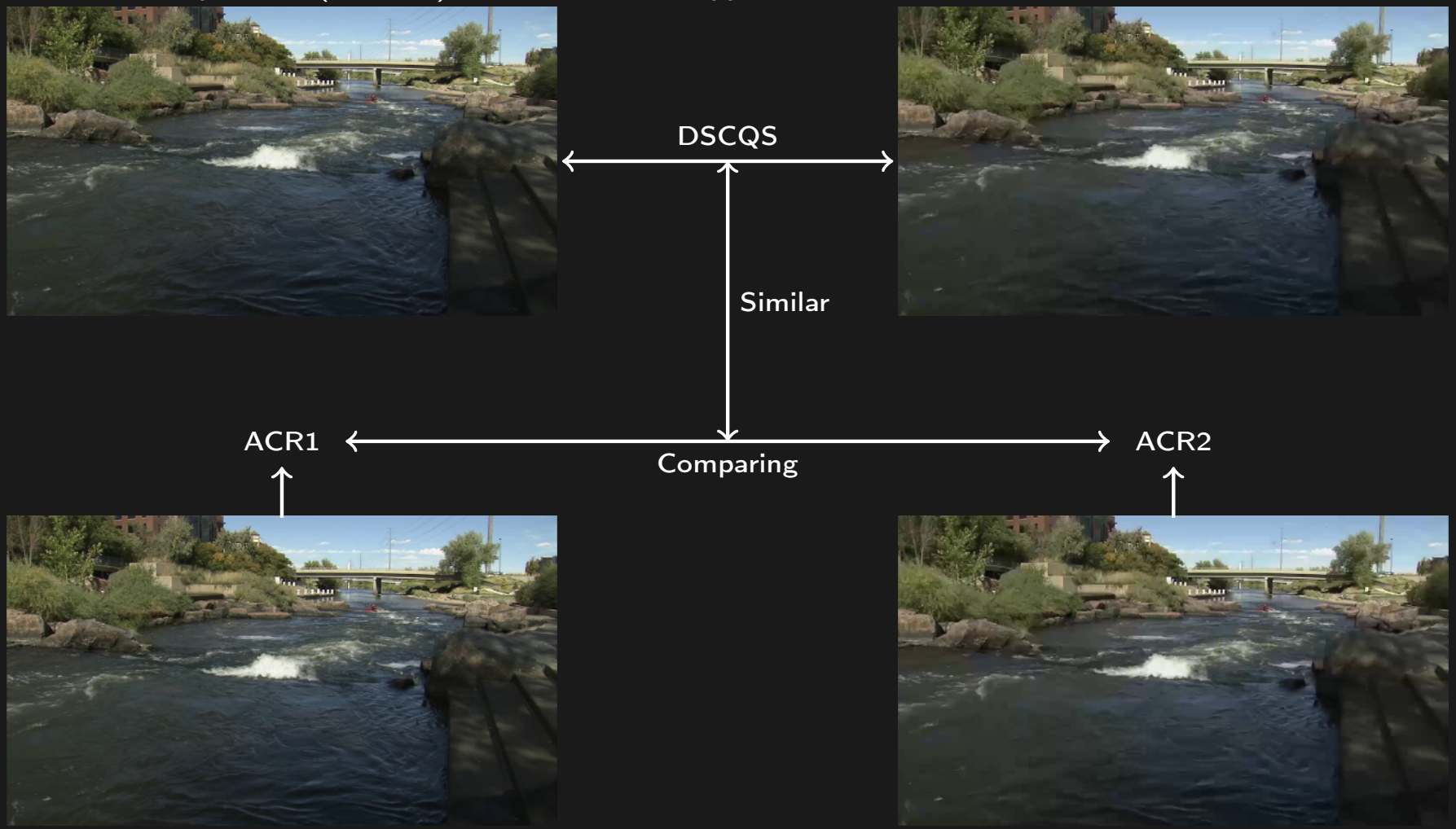
ACR1

ACR2

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review, 63(2), 81–97.

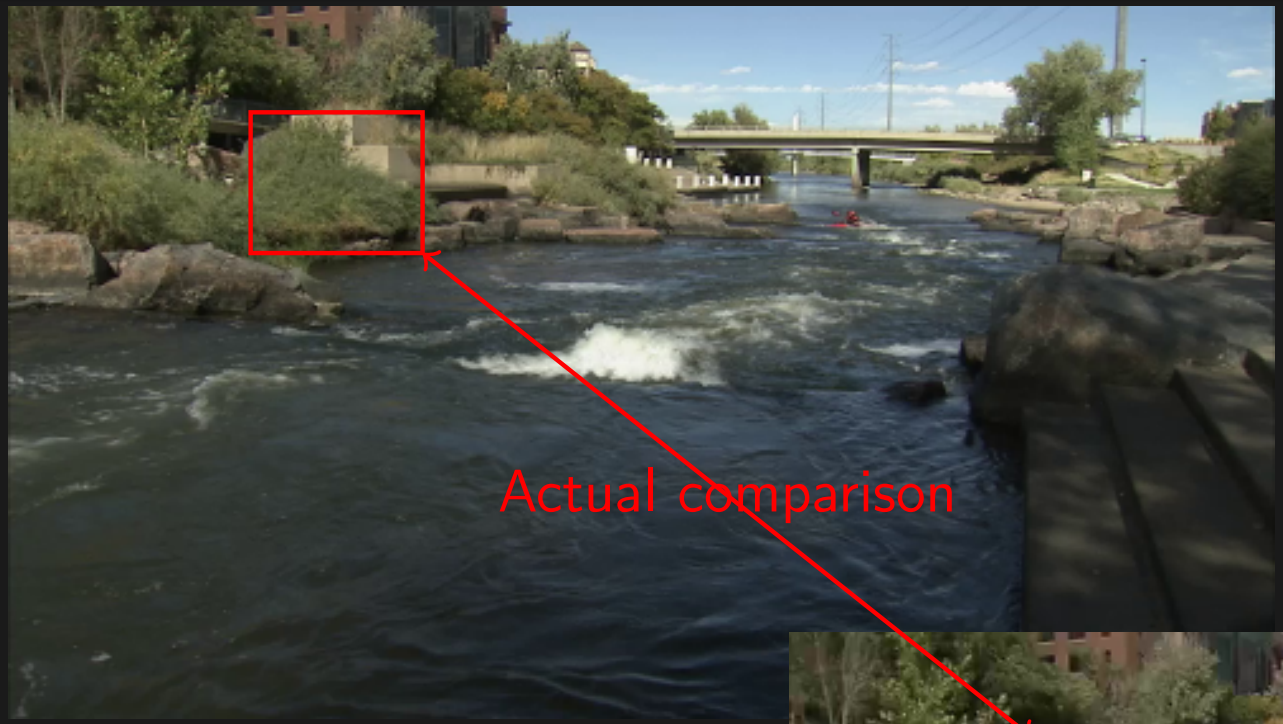
ACR versus PC

T. Tominaga, T. Hayashi, J. Okamoto and A. Takahashi, "Performance comparisons of subjective quality assessment methods for mobile video," 2010 Second International Workshop on Quality of Multimedia Experience (QoMEX), Trondheim, 2010, pp. 82-87.



Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review, 63(2), 81–97.

Focusing on Region

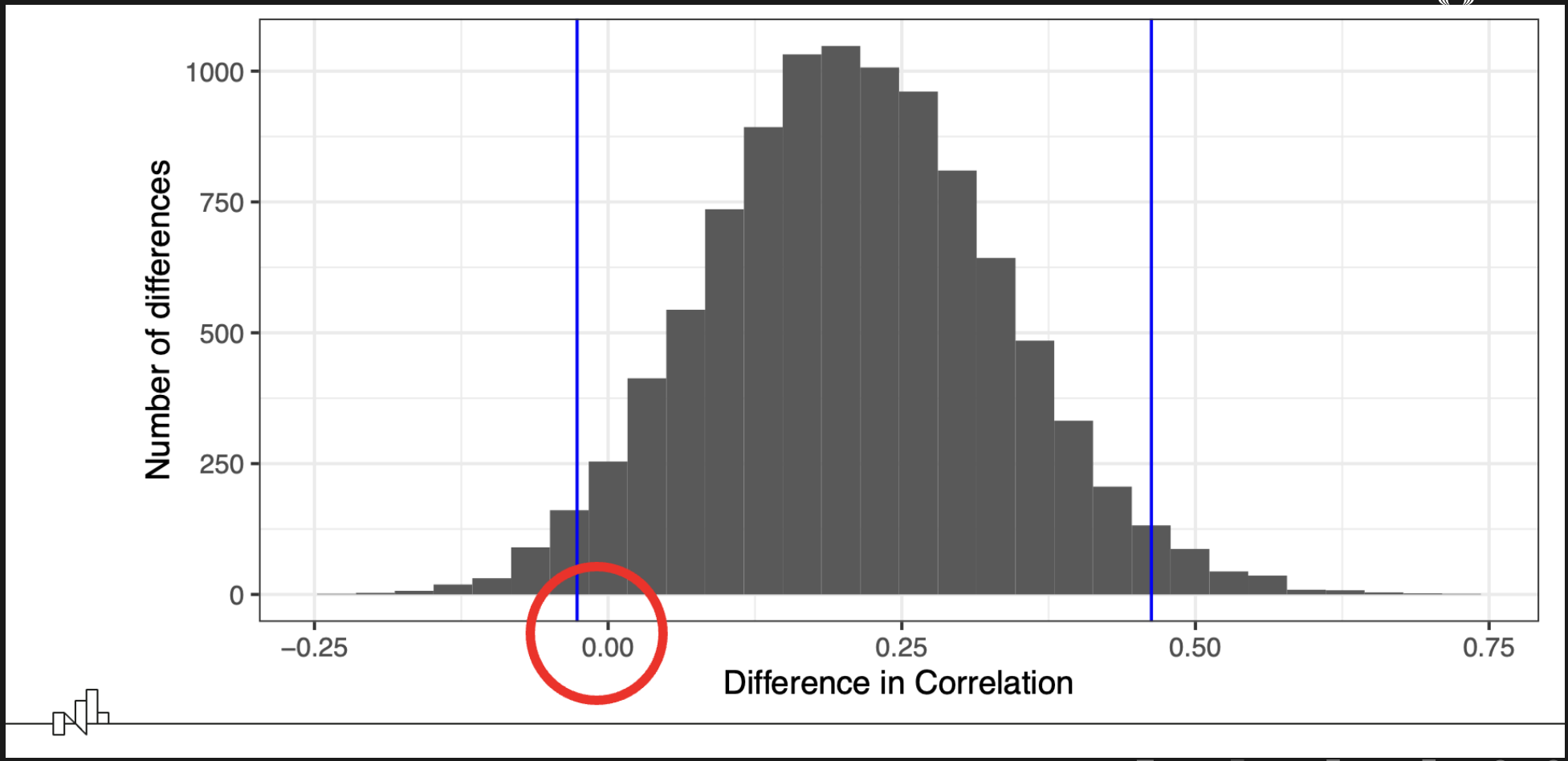


More Ecologically Valid Experiment Design

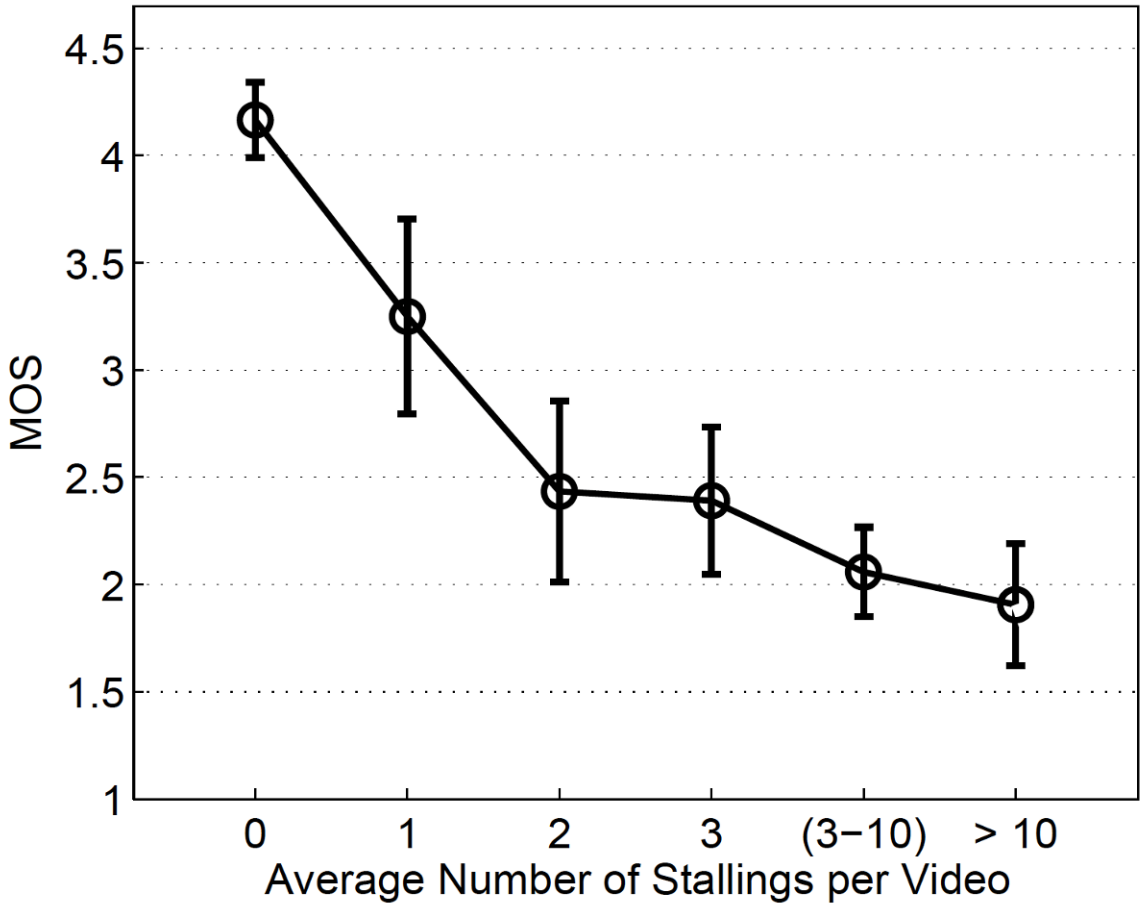
At AGH we run project: “Towards Better Understanding of Factors Influencing the QoE by More Ecologically-Valid Evaluation Standards”

Norway
grants

Results for Stalling

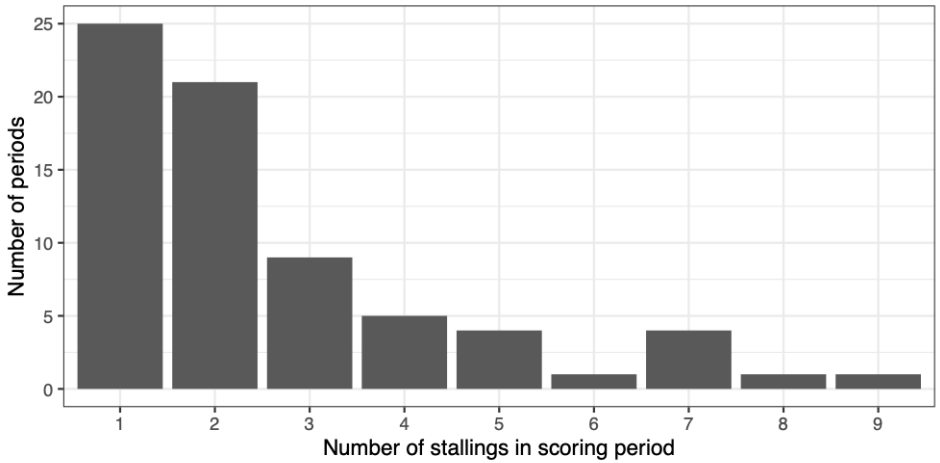


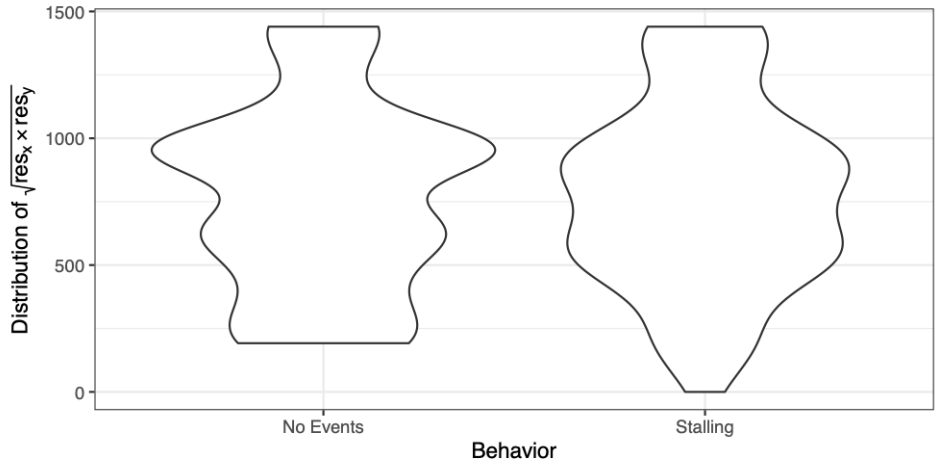
Results for Stalling - Literature



Casas, P., Sackl, A., Egger, S., & Schatz, R. (2012, December). YouTube & Facebook Quality of Experience in mobile broadband networks. In 2012 IEEE Globecom Workshops (pp. 1269-1274). IEEE.

Results for Stalling - Real Observations





Outline

Subject Model

Continuous description of a subjective score

Discrete description of a subjective score

[1] Lucjan Janowski, Margaret Pinson. Subject bias: Introducing a theoretical user model. Quality of Multimedia Experience (QoMEX), 2014 Sixth International Workshop on, pages 251–256. IEEE, 2014.
[2] Jakub Nawała, Lucjan Janowski, Bogadn Ćmiel, Krzysztof Rusek and Pablo Pérez, "Generalized Score Distribution: A Two-Parameter Discrete Distribution Accurately Describing Responses From Quality of Experience Subjective Experiments" in IEEE Transactions on Multimedia, 2022

Subject Model

Continuous description of a subjective score

$$o_{ij} = \psi_j$$

Discrete description of a subjective score

$$u_{ij} = \lfloor \psi_j \rceil$$

[1] Lucjan Janowski, Margaret Pinson. Subject bias: Introducing a theoretical user model. Quality of Multimedia Experience (QoMEX), 2014 Sixth International Workshop on, pages 251–256. IEEE, 2014.
[2] Jakub Nawała, Lucjan Janowski, Bogadn Ćmiel, Krzysztof Rusek and Pablo Pérez, "Generalized Score Distribution: A Two-Parameter Discrete Distribution Accurately Describing Responses From Quality of Experience Subjective Experiments" in IEEE Transactions on Multimedia, 2022

Subject Model

Continuous description of a subjective score

$$o_{ij} = \psi_j + \epsilon$$

Discrete description of a subjective score

$$u_{ij} = F(\psi_j, \rho)$$

[1] Lucjan Janowski, Margaret Pinson. Subject bias: Introducing a theoretical user model. Quality of Multimedia Experience (QoMEX), 2014 Sixth International Workshop on, pages 251–256. IEEE, 2014.
[2] Jakub Nawała, Lucjan Janowski, Bogadn Ćmiel, Krzysztof Rusek and Pablo Pérez, "Generalized Score Distribution: A Two-Parameter Discrete Distribution Accurately Describing Responses From Quality of Experience Subjective Experiments" in IEEE Transactions on Multimedia, 2022

Subject Model

Continuous description of a subjective score

$$o_{ij} = \psi_j + \Delta_i + \epsilon$$

Discrete description of a subjective score

$$u_{ij} = G(\psi_j, \Delta_i, \rho)$$

[1] Lucjan Janowski, Margaret Pinson. Subject bias: Introducing a theoretical user model. Quality of Multimedia Experience (QoMEX), 2014 Sixth International Workshop on, pages 251–256. IEEE, 2014.
[2] Jakub Nawała, Lucjan Janowski, Bogadn Ćmiel, Krzysztof Rusek and Pablo Pérez, "Generalized Score Distribution: A Two-Parameter Discrete Distribution Accurately Describing Responses From Quality of Experience Subjective Experiments" in IEEE Transactions on Multimedia, 2022

Subject Model

Continuous description of a subjective score

$$o_{ij} = \psi_j + \Delta_i + \mathcal{N}(0, \sigma)$$

Discrete description of a subjective score

$$u_{ij} = G(\psi_j, \Delta_i, \rho)$$

[1] Lucjan Janowski, Margaret Pinson. Subject bias: Introducing a theoretical user model. Quality of Multimedia Experience (QoMEX), 2014 Sixth International Workshop on, pages 251–256. IEEE, 2014.
[2] Jakub Nawała, Lucjan Janowski, Bogadn Ćmiel, Krzysztof Rusek and Pablo Pérez, "Generalized Score Distribution: A Two-Parameter Discrete Distribution Accurately Describing Responses From Quality of Experience Subjective Experiments" in IEEE Transactions on Multimedia, 2022

Subject Model

Continuous description of a subjective score

$$o_{ij} = \psi_j + \Delta_i + \mathcal{N}(0, \sigma)$$

$$u_{ij} = \lfloor o_{ij} \rceil$$

Discrete description of a subjective score

$$u_{ij} = G(\psi_j, \Delta_i, \rho)$$

$$P(U = s)$$

[1] Lucjan Janowski, Margaret Pinson. Subject bias: Introducing a theoretical user model. Quality of Multimedia Experience (QoMEX), 2014 Sixth International Workshop on, pages 251–256. IEEE, 2014.
[2] Jakub Nawała, Lucjan Janowski, Bogadn Ćmiel, Krzysztof Rusek and Pablo Pérez, "Generalized Score Distribution: A Two-Parameter Discrete Distribution Accurately Describing Responses From Quality of Experience Subjective Experiments" in IEEE Transactions on Multimedia, 2022

GSD

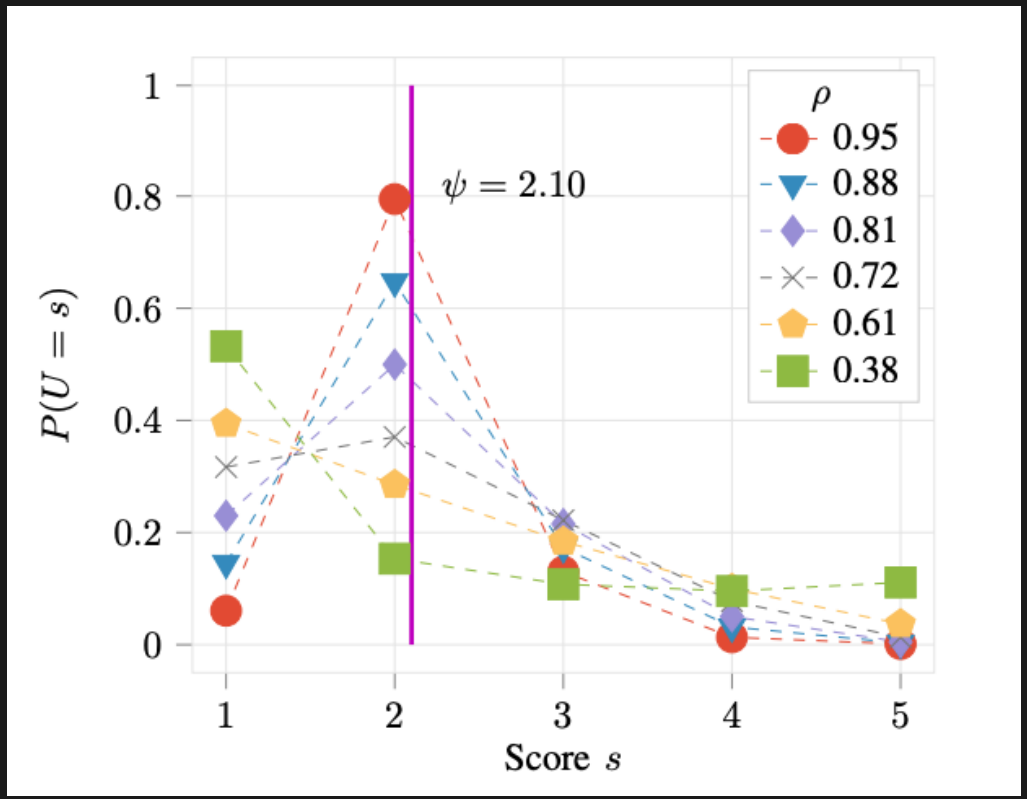
$$P_{F_\rho}(\epsilon = k - \psi) = \frac{\rho - C(\psi)}{1 - C(\psi)}[1 - |k - \psi|]_+ +$$
$$\frac{1-\rho}{1-C(\psi)}\binom{M-1}{k-1}\left(\frac{\psi-1}{M-1}\right)^{k-1}\left(\frac{M-\psi}{M-1}\right)^{M-k},$$

where $\rho \in [C(\psi), 1]$

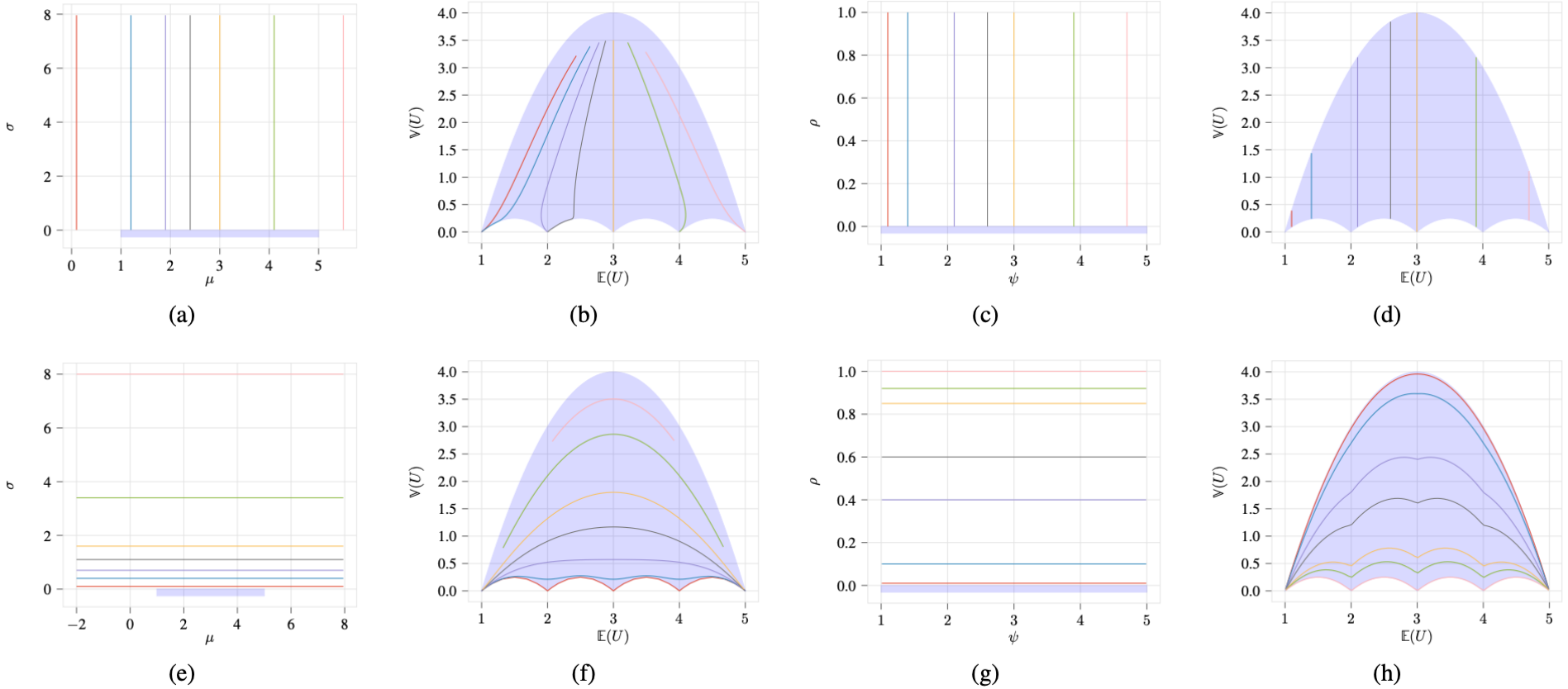
$$P_{G_\rho}(\epsilon = k - \psi) =$$
$$\binom{M-1}{k-1}\frac{\prod_{i=0}^{k-2}\left(\frac{(\psi-1)\rho}{(M-1)} + i(C(\psi)-\rho)\right)\prod_{j=0}^{M-k-1}\left(\frac{(M-\psi)\rho}{(M-1)} + j(C(\psi)-\rho)\right)}{\prod_{i=0}^{M-2}(\rho + i(C(\psi)-\rho))},$$

where $\rho \in [0, C(\psi)]$

GSD


$\psi = 2.10$
ρ
0.95
0.88
0.81
0.72
0.61
0.38
$P(U = s)$
Score s
0
0.2
0.4
0.6
0.8
1
1
2
3
4
5

Why Discretization is a Problem



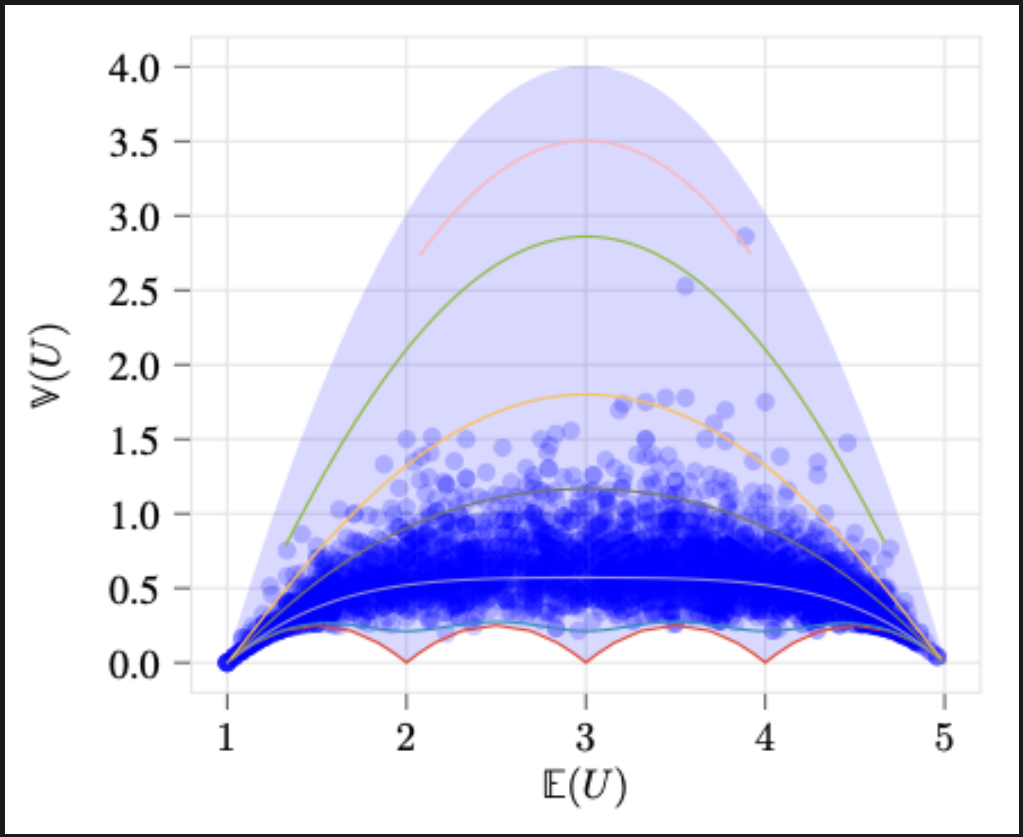
Outline

Definition

Definition: The term experiment precision provides a measure that quantifies the dispersion of the user ratings across different stimuli in a subjective experiment.

[3] Lucjan Janowski, Jakub Nawała, Tobias Hoßfeld, Michael Seufert, "Experiment Precision Measures and Methods for Experiment Comparisons" Quality of Multimedia Experience (QoMEX), 15th International Workshop on, 2023

Why Simple Variance Does Not Work?



Our Proposal

We proposed three metrics:

- $\ell = \frac{1}{N} \sum_{i=1}^{N} \hat{\sigma}_i$
- $g = \frac{1}{K} \sum_{j=1}^{K} \hat{\rho}_j$
- SOS parameter a

$$a = \frac{\sum_{j=1}^{K} (5 - \psi_j) \cdot (\psi_j - 1) \cdot \sigma_j}{\sum_{j=1}^{K} (5 - \psi_j)^2 \cdot (\psi_j - 1)^2}$$

Results for Real Data

Table: Experiment Precision Measures For QoE Subjective Experiments of 3 Types: VR(VR), Speech (S), and Video (V-n).

Exp.	$\ell \downarrow$	SE(ℓ)	$g \uparrow$	SE(g)	$a \downarrow$	SE(a)
V-6	0.574	0.014	0.908	0.0050	0.137	0.0020
V-1	0.583	0.011	0.891	0.0068	0.149	0.0022
V-4	0.610	0.020	0.826	0.0056	0.224	0.0021
V-3	0.613	0.016	0.863	0.0066	0.188	0.0021
V-5	0.627	0.019	0.871	0.0059	0.190	0.0021
V-2	0.627	0.022	0.867	0.0070	0.191	0.0021
S	0.953	0.028	0.744	0.0083	0.281	0.0015
VR	1.059	0.037	0.692	0.0093	0.335	0.0040

Detecting Fault Experiments

Table: Raw Experiment Precision Measures Results For Two Image QoE Experiments—VIME1 (I-V) and CCRIQ2 (I-C).

Exp.	$\ell \downarrow$	SE(ℓ)	$g \uparrow$	SE(g)	$a \downarrow$	SE(a)
I-V	1.053	0.0330	0.717	0.0085	0.314	0.0025
I-C	1.100	0.0316	0.683	0.0103	0.347	0.0030

Typical image and video experiments correspond to a between 0.0377 and 0.2116.

Outline

Conclusions

- Subjective experiments are crucial to include users perspective
- We need very careful experiment design since many aspects influence the final answer
- New analysis methods can help with better understanding the obtained results

Any Questions?