AVrate Voyager: an open source online testing platform

Steve Göring, Rakesh Rao Ramachandra Rao, Alexander Raake

Audiovisual Technology Group,
Technische Universität Ilmenau, Germany;
Email: [steve.goering, rakesh-rao.ramachandra-rao, alexander.raake]@tu-ilmenau.de

Code: `https://bit.ly/3uEyPNI`
(paper submitted to MMSP)

#105

June 7, 2021

TECHNISCHE UNIVERSITÄT
ILMENAU

Motivation

▶ subjective testing required for quality assessment and research [16]

▶ usual tool: **lab-based tests**

    ○ controlled environment, e.g. ITU-R BT.500-13 [10], ITU-T Rec. P.913/910 [9, 23]

    ○ ensuring highly reliable results

    ○ not always possible

       ▷ non-accessibility of lab rooms or rooms not usuable, e.g. disallowed (e.g. COVID-19)

▶ crowdsourcing or online tests as alternative:

    ○ audio [14, 13], video [20, 2, 6, 18, 25, 28] or image quality assessment [8, 7, 3, 24].

    $\rightarrow$ How to implement such an online/crowd test?

AVRate Voyager: Our Framework
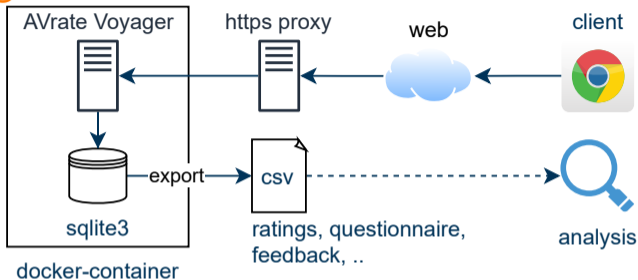
- ▶ based on AVRateNG
  - ○ usable for video, audio and othe lab tests [21, 17, 5, 19, 4, 22, 11, 15, 1, 26, 27, 12]
  - ○ similar tests possible for AVrate Voyager

- ▶ online/crowd nature requires adaption of test design
- ▶ scalable web technology (HTML5, CSS, bootstrap, docker)
- ▶ usable for: audio, image, and video tests or surveys

AVRate Voyager: Procedure
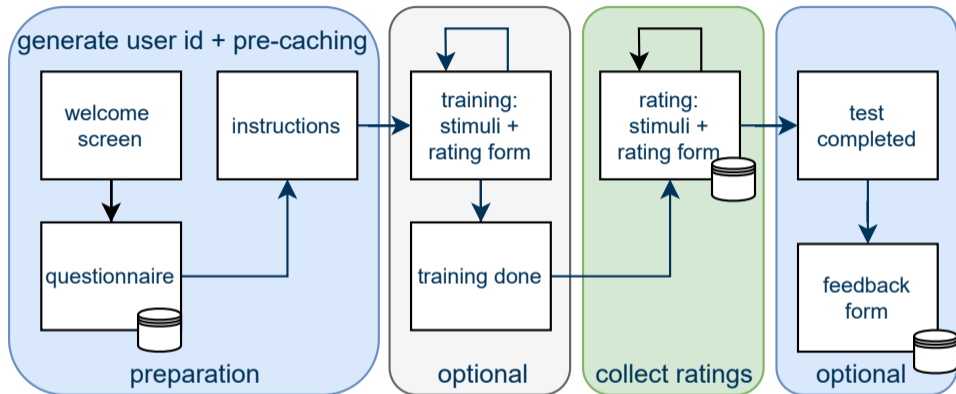
- ▶ configuration similar to AVRateNG (json file)
- ▶ adaption of templates required

Test instances – images

image quality test

▶ proof of concept; comparison to lab tests

▶ 1484 full-hd sized square image patches (195 MB)

▶ each participant rated 150 images (20 MB)

▶ ACR rating scheme; 15 minutes

▶ 238 participants: 35700 ratings

$\rightarrow$ 720-1080p max resolution for participants

Test instances – video

video quality, see *Rao, Göring, and Raake* [20] (QoMEX '21)

▶ $540p$ center crop, comparison to lab test

▶ 30 videos per user (70 MB), 180 in total (433 MB)

▶ ACR rating scheme; 15 minutes

▶ 175 participants: 4390 ratings

$\rightarrow$ requirement for pre-caching

Test instances – audio mixed

spaciousness of music, see *Stirnat and Göring* [29]

▶ replacement for a lab tests, small scale

▶ 41 stimuli (97 MB): 6 audio-only, 5 video-only, and 30 mixed

▶ multiple sliders for rating; 60 minutes

$\rightarrow$ window size per stimuli, play stimuli several times

Conclusion, Summary and Future Work

- ▶ overview of AVRate Voyager
  - ○ typical instantiation
  - ○ included procedure

- ▶ example tests briefly shown

- ▶ open and next steps:
  - ○ include other multimedia formats
  - ○ extend the framework

- ▶ **Demonstration**

References I

[1] Darkhan Ashimov, Maria G Martini, and Nabajeet Barman. "Quality assessment of gaming videos compressed via AV1". In: *QoMEX*. IEEE. 2020, pp. 1–4.

[2] Óscar Figuerola Salas et al. "Assessing internet video quality using crowdsourcing". In: *Proc. of the 2nd ACM int. workshop on Crowdsourcing for Multimedia*. 2013, pp. 23–28.

[3] Deepti Ghadiyaram and Alan C Bovik. "Massive online crowdsourced study of subjective and objective picture quality". In: *IEEE Trans. on Image Processing* 25.1 (2015), pp. 372–387.

[4] Steve Göring et al. "Analyze and predict the perceptibility of UHD video contents". In: *Electronic Imaging* 2019.12 (2019), pp. 215–1.

References II

[5]   Steve Göring et al. "Modular Framework and Instances of Pixel-based Video Quality Models for UHD-1/4K". In: *IEEE Access* 9 (2021), pp. 31842–31864.

[6]   T. Hoßfeld et al. "Quantification of YouTube QoE via Crowdsourcing". In: *IEEE ISM*. 2011, pp. 494–499.

[7]   V. Hosu et al. "KonIQ-10k: An Ecologically Valid Database for Deep Learning of Blind Image Quality Assessment". In: *IEEE Trans. on Image Processing* 29 (2020).

[8]   Vlad Hosu et al. "The Konstanz natural video database (KoNViD-1k)". In: *QoMEX*. IEEE. 2017.

[9]   ITU-T. *ITU-T Rec. P.913 (16/03)*. Tech. rep. Int. Telecommunication Union, 2016.

References III

[10] ITU-T. *Rec. ITU-R BT.500-13 – Methodology for the subjective assessment of the quality of television pictures*. Tech. rep. Int. Telecommunication Union, 2014.

[11] Dominik Keller et al. "Assessing texture dimensions and video quality in motion pictures using sensory evaluation techniques". In: *QoMEX*. IEEE. 2019.

[12] Dominik Keller et al. "Let the Music Play: An Automated Test Setup for Blind Subjective Evaluation of Music Playback on Mobile Devices". In: *QoMEX*. IEEE. 2020, pp. 1–4.

[13] Babak Naderi et al. "Effect of trapping questions on the reliability of speech quality judgments in a crowdsourcing paradigm". In: *16th annual conf. of the int. speech comm. association*. 2015.

10 / 16

References IV

[14] Babak Naderi et al. "Towards speech quality assessment using a crowdsourcing approach: evaluation of standardized methods". In: *Quality and User Experience* 6.1 (2020), pp. 1–21.

[15] Margaret H Pinson. *ITS4S3: A Video Quality Dataset With Unrepeated Videos, Camera Impairments, and Public Safety Scenarios.* Tech. rep. NTIA Technical Memorandum 19-538, 2019.

[16] Alexander Raake and Sebastian Egger. "Quality and Quality of Experience". In: *Quality of Experience: Advanced Concepts, Applications and Methods.* Ed. by Sebastian Möller and Alexander Raake. Cham: Springer International Publishing, 2014, pp. 11–33.

[17] Alexander Raake et al. "Multi-model standard for bitstream-, pixel-based and hybrid video quality assessment of UHD/4K: ITU-T P.1204". In: *IEEE Access* 8 (2020), pp. 193020–193049.

[18] Benjamin Rainer and Christian Timmerer. "Quality of Experience of Web-Based Adaptive HTTP Streaming Clients in Real-World Environments Using Crowdsourcing". In: *Proc. of the Workshop on Design, Quality and Deployment of Adaptive Video Streaming*. VideoNext '14. Australia: ACM, 2014.

[19] Rakesh Rao Ramachandra Rao et al. "Adaptive video streaming with current codecs and formats: Extensions to parametric video quality model ITU-T P. 1203". In: *Electronic Imaging* (2019).

References VI

[20] Rakesh Rao Ramachandra Rao, Steve Göring, and Alexander Raake. "Towards High Resolution Video Quality Assessment in the Crowd". In: QoMEX.

[21] Rakesh Rao Ramachandra Rao et al. "AVT-VQDB-UHD-1: A large scale video quality database for UHD-1". In: IEE ISM. IEEE. 2019, pp. 17–177.

[22] Rakesh Rao Ramachandra Rao et al. "A Large-scale Evaluation of the bitstream-based video-quality model ITU-T P.1204.3 on Gaming Content". In: 22st International Workshop on Multimedia Signal Processing (MMSP). IEEE. 2020, pp. 1–6.

[23] ITUT Recommendation. "P. 910,Subjective video quality assessment methods for multimedia applications," in: *Int. Telecommunication Union, Tech. Rep* (2008).

[24] Judith Alice Redi et al. "Crowdsourcing-based multimedia subjective evaluations: a case study on image recognizability and aesthetic appeal". In: *Proc. of the 2nd ACM int. workshop on Crowdsourcing for multimedia*. 2013, pp. 29–34.

[25] M. Shahid et al. "Crowdsourcing based subjective quality assessment of adaptive video streaming". In: *QoMEX*. 2014.

[26] Ashutosh Singla et al. "Assessing media qoe, simulator sickness and presence for omnidirectional videos with different test protocols". In: *Conf. on VR and 3D User Interfaces (VR)*. IEEE. 2019, pp. 1163–1164.

[27] Ashutosh Singla et al. "Subjective quality evaluation of tile-based streaming for omnidirectional videos". In: *Proc. of the 10th ACM Multimedia Systems Conference*. 2019, pp. 232–242.

[28] Z. Sinno and A. C. Bovik. "Large-Scale Study of Perceptual Video Quality". In: *IEEE Trans. on Image Processing* 28.2 (2019).

[29] Claudia Stirnat and Steve Göring. "Audiovisual Perception of Spaciousness of Music in an Online Study". In: *KLINGT GUT! KLG 2020/21* (). to appear.

Thank you for your attention

. . . . . . are there any questions?