Leonardo Peroni, Sergey Gorinsky, Farzad Tashtarian, and Christian Timmerer. 2023. Empowerment of Atypical Viewers via Low-Effort Personalized Modeling of Video Streaming Quality. The definitive version is published in *Proc. ACM Netw.*, 1, CoNEXT3, Article 17 (December 2023), https://doi.org/10.1145/3629139.

I ARTIFACTS: IQOE DATASET AND CODE

I.1 Abstract

This appendix describes our artifacts and their usage for reproducing the reported experimental results. We support the reproducibility on two levels characterized by different resource requirements and time commitments. Level A (LA) recreates all figures and tables, mostly by running our Python code. The reproduction performs limited calculations based on data in intermediate representations, e.g., scores by a rater for a personalized series of experiences. Level B (LB) involves advanced computations, such as model training, and carries out in-depth replication from raw data, allowing also for new data from independent subjective studies and simulations. While LA is primarily for validating the results of this paper, LB targets the long-term impact of the artifacts through their reuse in future research.

I.2 Artifact checklist

- **Algorithm:** Algorithm 1.
- **Programs:** (LA) Python 3.7 or above with supporting libraries; (LB) FFmpeg, ffmpeg-debug-qp, Park, MP4Box, BiQPS, HTML, JavaScript, and CSS.
- **Models**: SVR, RF, GP, and XGB.
- **Datasets:** (LA) iQoE and Waterloo-IV; (LB) network traces, video chunks, and experience set.
- **Hardware:** no particular requirements.
- **Output:** (LA) all figures and tables of the paper; (LB) Fig. 2-13 and 15-18 and Tables 1 and 2.
- **Approximate disk space requirement:** (LA) 4 GB; (LB) 24 GB.
- **Approximate experiment completion time:** (LA) under one hour; (LB) about one week.
- **Availability:** public on GitHub.

I.3 Description

I.3.1 Artifact access. Our code and datasets are publicly available on GitHub[1]. We structure the GitHub repository in folders dedicated to a figure or table. Each folder contains LA and/or LB subfolders corresponding to the two levels of result reproducibility. The subfolders hold all the data and code required for generating the respective results.

I.3.2 Hardware dependencies. There are no specific hardware requirements except for support of Python. The machine used in our experiments is an Intel i7 with six cores, 2.6-GHz CPUs, 16-GB RAM, and Windows 10.

I.3.3 Software dependencies. The requirements.txt file in the repository lists the software dependencies, with readme.md files in individual subfolders supplying any further details and clarifications.

I.3.4 Models. SVR, RF, GP, and XGB are from Python's scikit-learn and xgboost libraries.

I.3.5 Datasets. In addition to the experiment-specific data in each subfolder, the repository includes a standalone Datasets folder with our iQoE dataset in its three versions used by different experiments. These versions appear in the .xlsx format in subfolders dataset_34, dataset_120, and dataset_128 where 34, 120, and 128 refer to the number of raters in the dataset version. The results

[1]https://github.com/Leo-rojo/iQoE_Dataset_and_Code

Author's address; leonardo.peroni@imdea.org.

from the motivation and simulation sections leverage the Waterloo-IV dataset[2]. The in-depth LB reproducibility utilizes network traces[3], video chunks[4], and experience set[5].

I.3.6 Extra code for the LB reproducibility. To support independent subjective tests, the GitHub repository includes a standalone Subjective_assessments folder. This folder provides the code of our iQoE website and also enables real-world subjective assessments with the code that creates training and testing experiences. To facilitate new simulations, the repository contains a Synthetic_raters folder with the code that generates synthetic raters, synthetic experiences, and scores of the experiences by the synthetic raters.

I.4 Installation

The LA reproducibility involves installation of Python 3.7 or above with libraries as described in the requirements.txt file. For the in-depth LB replication, the additional installations include Park[6] and FFmpeg[7], with Codex[8] as the recommended FFmpeg version, to generate experiences and calculate PSNR, SSIM, VMAF, and other metrics. The computation of the quantization parameter leverages ffmpeg-debug-qp[9]. The creation of video chunks uses MP4Box[10]. The experiments with baseline model L require installation of BiQPS[11]. Reproduction of our subjective studies from scratch involves deployment of the iQoE website developed in HTML, JavaScript, and CSS.

I.5 Evaluation and expected results

As mentioned in Appendix I.3.1, the repository allocates a separate self-contained folder for reproducing the results of each figure or table on the LA and/or LB levels, with readme.md files providing specific instructions.

The LA reproducibility of the results in Fig. 2-13 and 15-18 as well as Tables 1 and 2 largely consists in running the provided Python code on the associated dataset. Fig. 2-4 utilize Waterloo-IV data. Fig. 5, 6, 8, and 15 and Table 1 rely on the dataset_120 version of our iQoE dataset. Fig. 7 leverages the extended dataset_128 version. The validation of the synthetic raters in Fig. 9 entails the dataset_34 version of the iQoE dataset. The replication of Fig. 9-13 and 16-18 and Table 2 on the LA level leverages the data already recorded in our simulations, e.g., scores by the synthetic raters.

For completeness of the LA reproducibility, the GitHub repository includes the .pptx source of our diagrams in Fig. 1 and supports replication of the screenshots in Fig. 14 by providing access to our original iQoE website[12] with anonym and iQoE_92 as the username and password, respectively.

Comprehensive replication of our results on the LB level involves additional existing and newly developed software. The code in the Subjective_assessments folder enables a researcher to recreate the subjective studies by deploying an own version of our iQoE website and preparing own training and testing experiences so as to reproduce Fig. 5-9 and 15 and Table 1. The Synthetic_raters folder empowers the researcher to engender new synthetic raters and conduct independent simulations for the LB reproducibility of Fig. 9-13 and 16-18 and Table 2.

[2] https://dx.doi.org/10.21227/j15a-8r35
[3] https://doi.org/10.6084/m9.figshare.24460084.v1
[4] https://doi.org/10.6084/m9.figshare.24460078.v2
[5] https://doi.org/10.6084/m9.figshare.24460081.v1
[6] https://github.com/park-project/park
[7] https://ffmpeg.org/download.html
[8] https://www.gyan.dev/ffmpeg/builds
[9] https://github.com/slhck/ffmpeg-debug-qp
[10] https://github.com/gpac/gpac/wiki/MP4Box
[11] https://github.com/TranHuyen1191/BiQPS
[12] https://iqoe.itec.aau.at