

Spring 2024 VQEG meeting July 1 to 5, 2024 in Klagenfurt, Austria

AVT-ECoClass-VR: An open-source audiovisual 360° video and immersive CGI multi-talker dataset to evaluate cognitive performance

Paper presented at 16th International Conference on Quality of Multimedia Experience (QoMEX) 2024. [3]

<u>Stephan Fremerey</u>*, Carolin Breuer**, Larissa Leist***, Maria Klatte***, Janina Fels** and Alexander Raake*

* Audiovisual Technology Group, Technische Universität Ilmenau, Germany
** Institute for Hearing Technology and Acoustics, RWTH Aachen University, Germany
*** Center for Cognitive Science, RPTU Kaiserslautern-Landau, Germany

03.07.2024 AVT-ECoClass-VR









Motivation

- Focus ECoClass-VR project: Cognitive performance evaluation in classroom-type settings for adults and children
- Increase realism of experimental procedures in terms of
 - Cognitive tasks used
 - Audiovisual representation

Existing test paradigms Transfer of paradigms using IVEs Adapted audiovisual paradigms

03.07.2024 AVT-ECoClass-VR











Background

- Cognitive performance studies often use simple auditory paradigms and acoustic reproductions
- Visual aspects are often overlooked
- Research objectives:
 - Transfer and test three auditory paradigms in advanced audiovisual IVEs (e.g., Ahrens et al.'s Audiovisual Scene Analysis [2])
 - Evaluate paradigms
 - Develop cognition-based QoE measures, impact of technical aspects of IVEs on cognitive performance





Test paradigm from Ahrens et al. [2]

03.07.2024 AVT-ECoClass-VR









General aspects

- Objective: Validate suitability of AV-SA paradigm for cognitive performance evaluation
- Method: subjective testing
- Setup of paradigm in ECoClass-VR
 - 360° video with real persons & CGI-based room model and 3D-scanned persons
 - Circle of 20 chairs, 18° steps
- Task: Assign stories to speakers quickly
- Procedure: Use an interaction wheel and ray to assign symbols to input fields above speakers
- Each story plays for 120s, then break if needed, then next scene















AVT-ECoClass-VR Dataset – Production Steps 360° Video



03.07.2024 AVT-ECoClass-VR

DEUTSCHE Forschungsgemeinschaft German Research Foundation TECHNISCHE UNIVERSITÄT





AVT-ECoClass-VR Dataset – Production Steps CGI













AVT-ECoClass-VR Dataset

- Contents of publicly available dataset
 - Dataset recorded with 20 different speakers in German language
 - 220 different video recordings & 360° image of the classroom
 - 200 different single-channel audio recordings
 - Rigged 3D models and 3D scans of 20 different speakers
 - 3D model of classroom
 - 360° implementation of the IVE
 - CGI implementation of the IVE
 - Example output data from pending subjective tests, 5 subjects for both IVEs (head rotation data and speaker-to-story mappings)



Left: 360° representation of the different speakers Right: CGI representation of the different speakers

R

Rheinland-Pfälzisch

03.07.2024 AVT-ECoClass-VR



TECHNISCHE UNIVERSITÄT ILMENAU



360° Video IVE



03.07.2024 AVT-ECoClass-VR







R TChische Universität R Endau Koiserslautern Londau

CGI IVE



03.07.2024 AVT-ECoClass-VR







R TUJ Rheinland-Pfälzische Technische Universität Koiserslautern Landau

AVT-ECoClass-VR Dataset – Evaluation (1)

- Subjective Experiments: Validation of AV-SA test paradigm
- Task: Assign stories to speakers using interaction wheel and ray, answer quickly and accurately
- Presentation modes
 - Video
 - » 360° vs. CGI (Experiments 1+2 vs. Experiment 3)
 - Audio
 - » Mono (Experiment 1)
 - » Binaural free field (Experiments 2+3)
 - Data Management
 - » Story-to-speaker mappings saved in JSON
 - » Viewing data recorded in CSV







03.07.2024 AVT-ECoClass-VR









AVT-ECoClass-VR Dataset – Evaluation (2)



03.07.2024 AVT-ECoClass-VR







R P Rheinland-Pfälzische Technische Universität Kaiserslautern Londau

AVT-ECoClass-VR Dataset in Numbers

- 360° Video IVE
 - 60h: recording
 - 440h: stitching & rendering
 - 1100h: video denoising
 - 440h: chroma keying, cutting & rendering
 - 1400h: processing time to render final videos (concatenating 21 8K video inputs)
 - → Total time spent: ~3440h
- CGI IVE
 - 600h: modelling CGI classroom & integration with Unity
 - 10h: 3D scanning
 - 60h: post-processing & rigging of 3D models
 - \rightarrow Total time spent: ~670h
- Overall dataset size about 3.7 TB









AVT-ECoClass-VR – Conclusion

- Development of 2 IVEs
 - 360° video
 - CGI
- Variety of contents
 - 360° video and mono audio of 20 speakers
 - Rigged 3D models
 - 3D scans
- Usage: evaluate cognitive performance in classroom-like settings
- Future work: Further subjective tests with both IVEs
 - Impact of factors like lip-sync, video resolution, etc. on cognitive performance



Link to AVT-ECoClass-VR Database on GitHub [1]









Literature

[1] <u>https://github.com/Telecommunication-Telemedia-Assessment/avt-ECoClass-VR</u>

[2] Ahrens, A., Lund, K. D., & Dau, T. (2019). Audio-visual scene analysis in reverberant multi-talker environments. Universitätsbibliothek der RWTH Aachen.

[3] Fremerey, S. et al. (2024). AVT-ECoClass-VR: An open-source audiovisual 360° video and immersive CGI multi-talker dataset to evaluate cognitive performance. 2024 16th International Conference on Quality of Multimedia Experience (QoMEX).

