TECHNISCHE UNIVERSITÄT
ILMENAU

Perceptual evaluation of Immersive Media –
From video quality towards a holistic QoE perspective

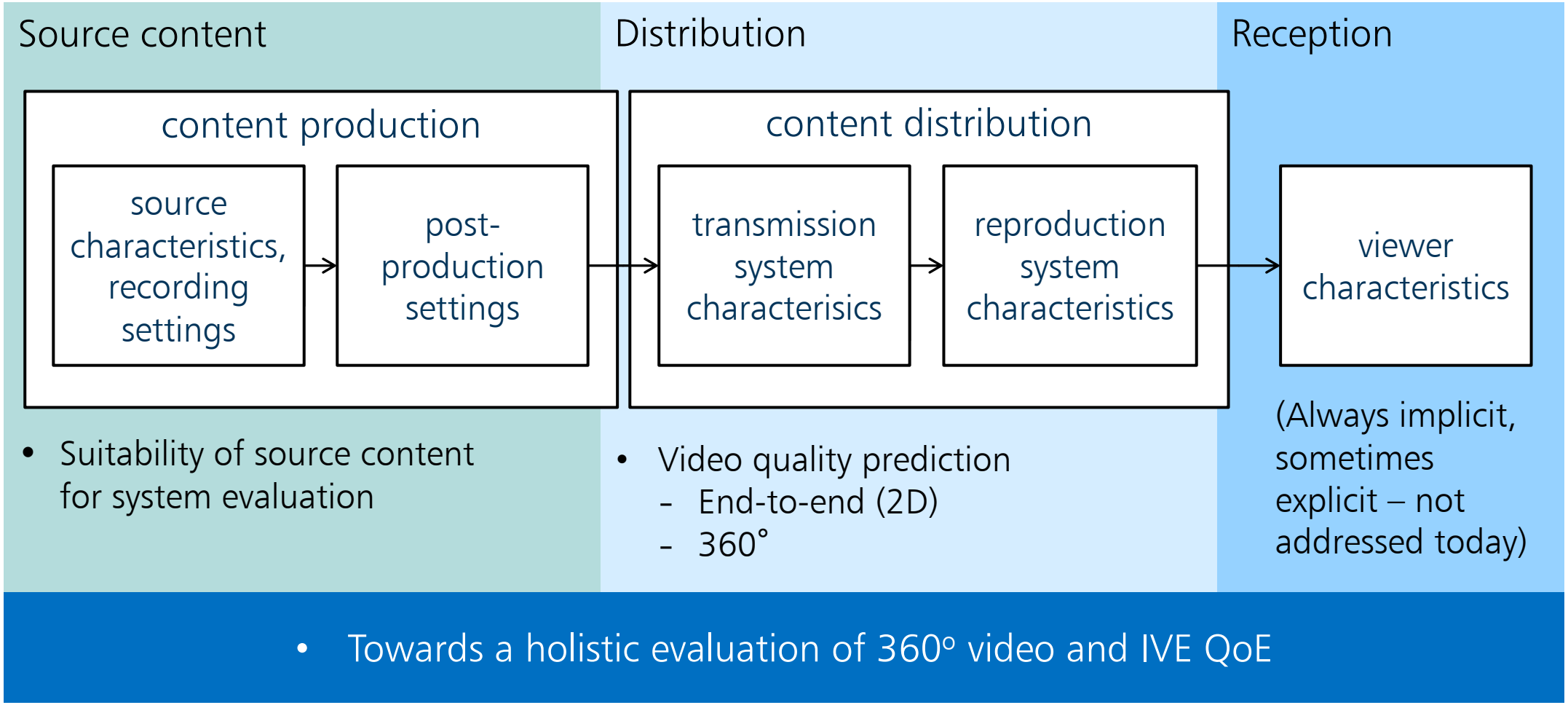
Alexander Raake & team
Audiovisual Technology (AVT) Group
Institute for Media Technology
TU Ilmenau (TUIL)
Germany

Audiovisual Technology Group

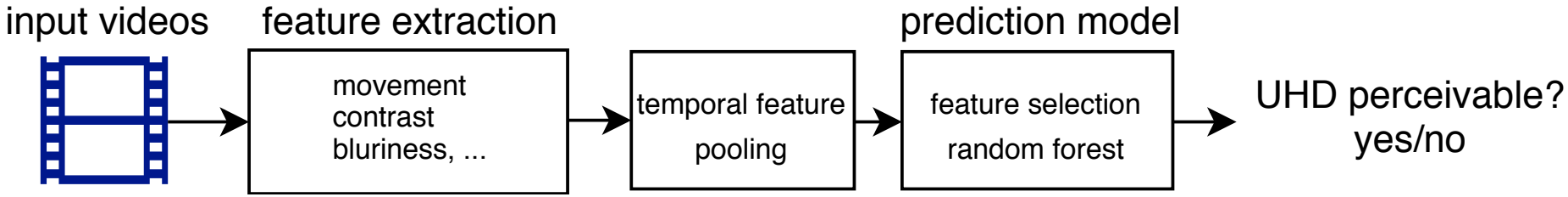
The SPIRIT of science

TECHNISCHE UNIVERSITÄT
ILMENAU

Outline talk – exemplary models



Determining suitability of content for 4K
Model (using synthetic dataset)



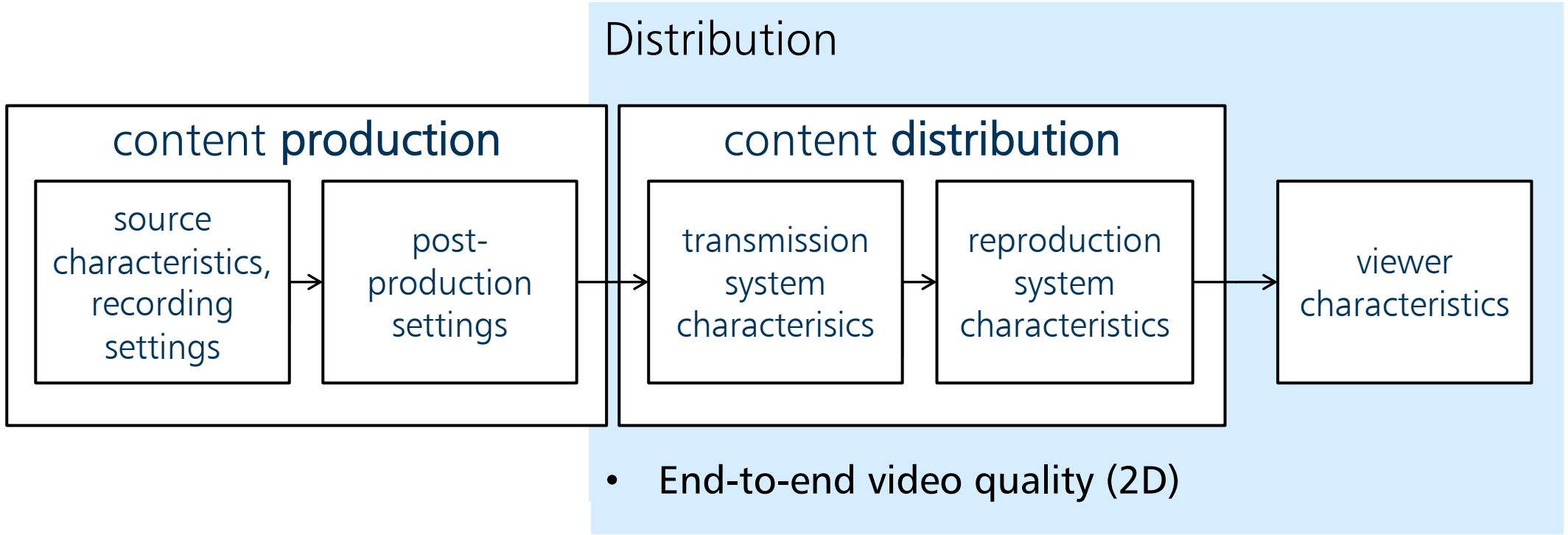
- Classification problem; analysis of different ML-algorithms
- Features: I=image, M=motion based
  - contrast^{I}, blur^{I}, uhdhdsim^{I} , temporal^{M}, blockmotion^{M}, movement^{M}, staticness^{M}
  - fft^{I} (Katsavounidis, Aaron & Ronca, SMPTE Conf. 2015)
  - niqe^{I} (Mittal, Soundararajan, and Bovik, IEEE SPL 2013)
  - SI^{I} , TI^{M} (ITU-T Rec. P.910, 2008)
  - colorfulness^{I} (Hasler and Suesstrunk, HVEI 2003)
  - tone^{I}, saturation^{I} (Aydın, Smolic, and Gross, IEEE TVCG 2015)

class	precision	recall	f1-score	support
avg / total	0.83	0.82	0.82	72

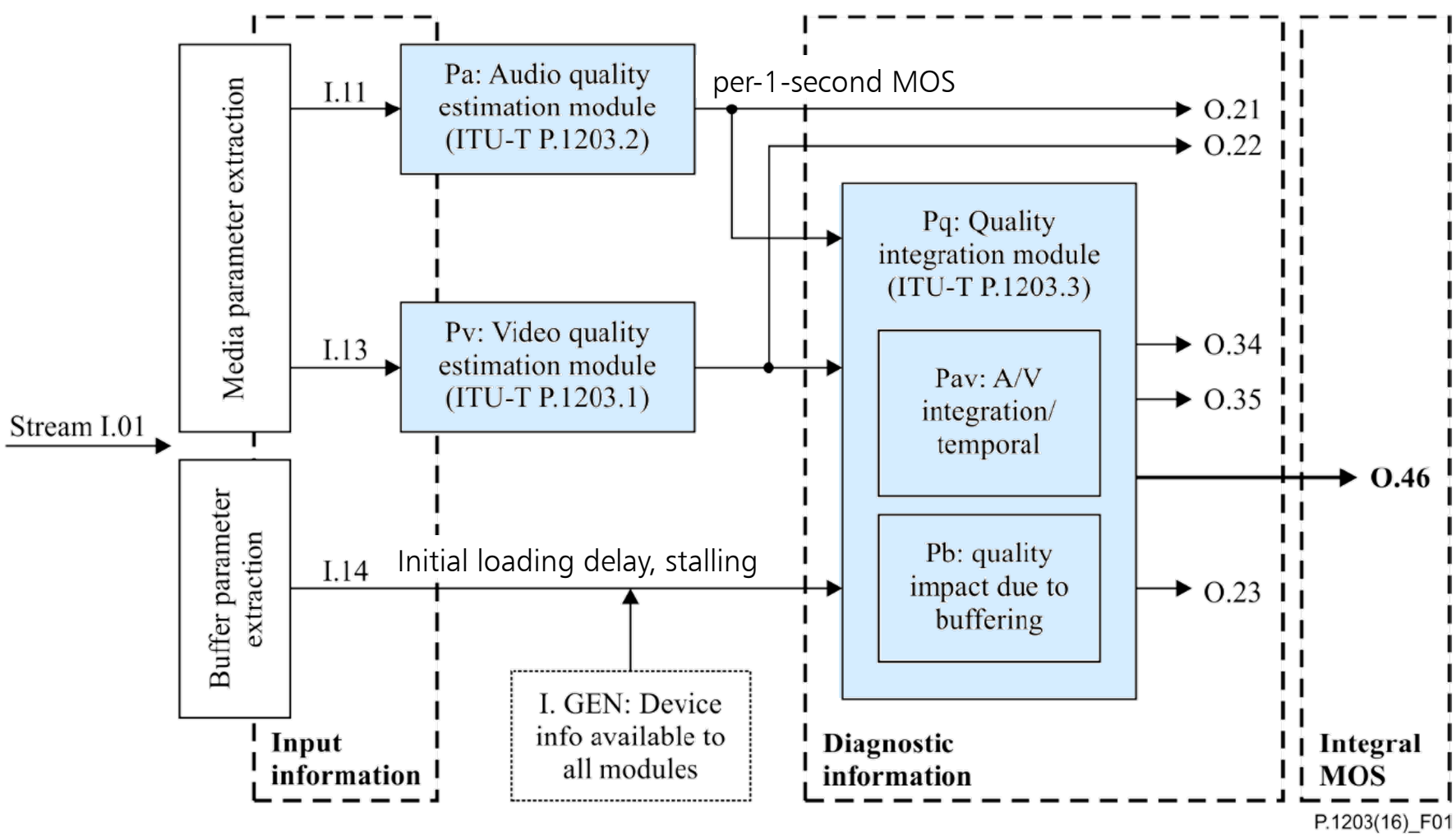
Open Source features:
https://github.com/Telecommunication-Telemedia-Assessment/quat

(Göring et al., HVEI 2019*)

Quality of content distribution



Context: ITU-T P.1203 model architecture (P.NATS Phase 1)

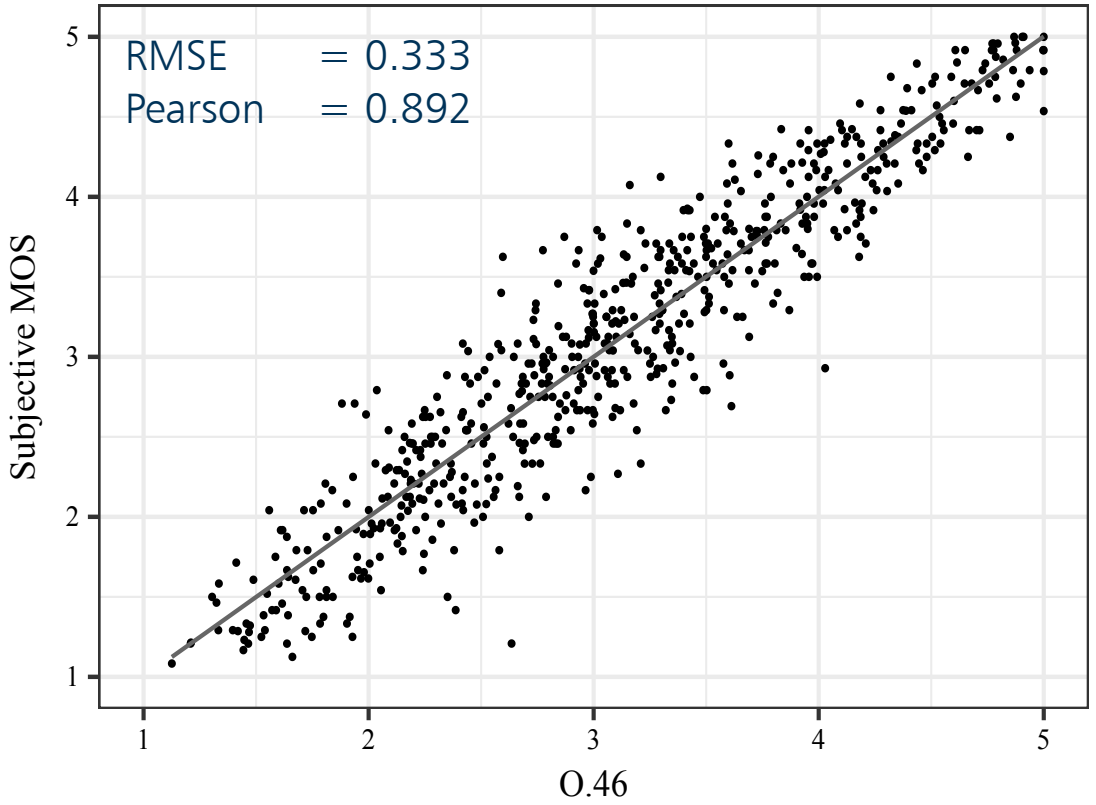


ITU-T Rec. P.1203

- P.NATS Phase 1: [1,5] min sequence QoE, audiovisual
- Trained and validated on 1064 audiovisual sequences, i.e. 30 databases with subjective ratings
- Only published model for HTTP Adaptive Streaming validated so extensively

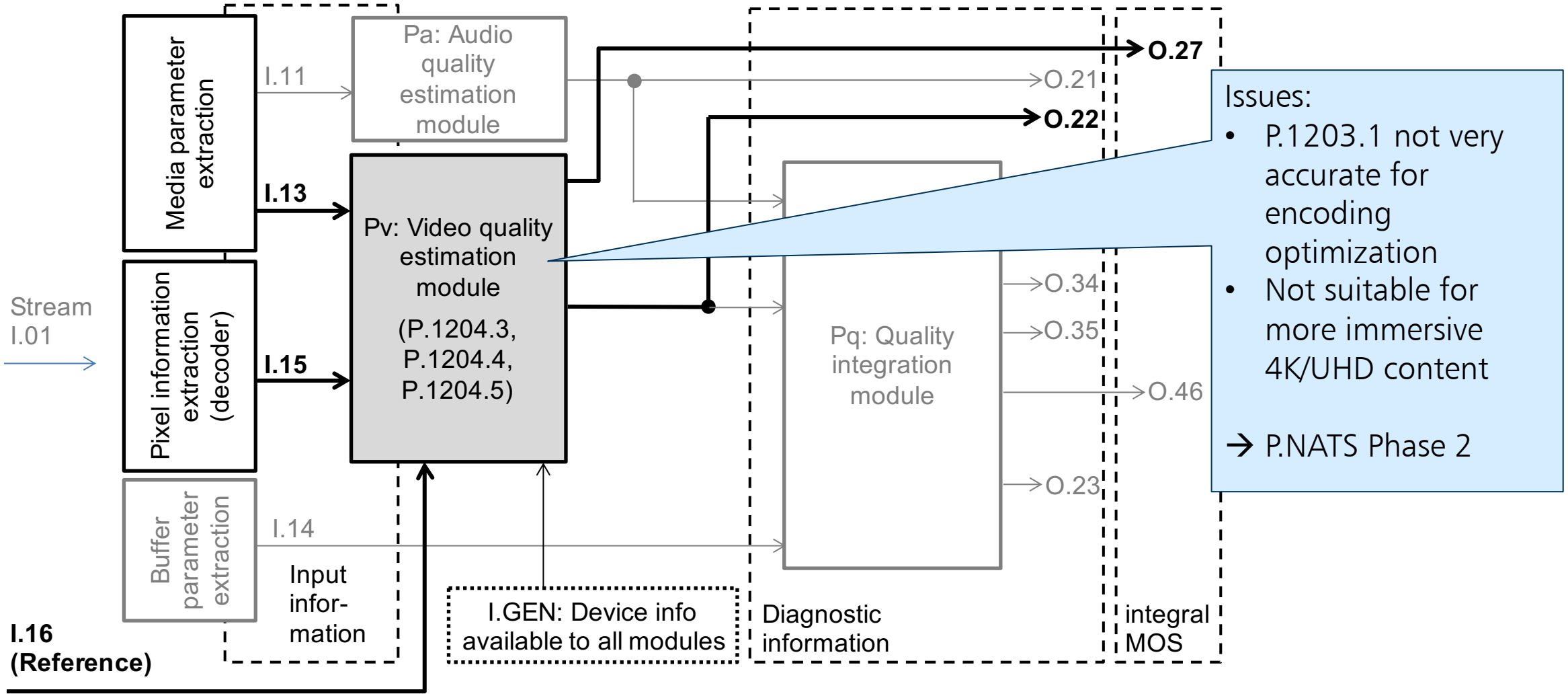
- Open Source:
  - → https://github.com/itu-p1203/itu-p1203
  - → Extensions and other info: https://telecommunication-telemedia-assessment.github.io/bitstream_based_models/

Example: P.1203 Mode 3



(Robitza et al., ACM MMSys 2018*;
Raake et al., QoMEX 2017*)

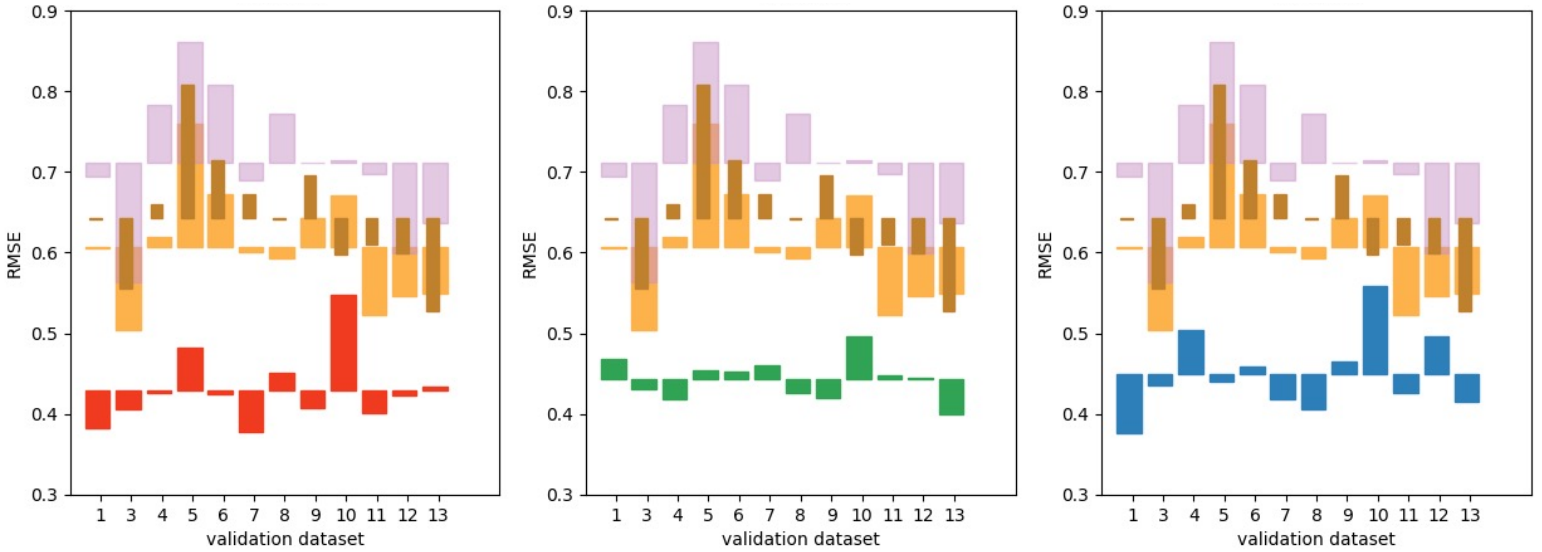
From P.1203 to P.1204



ITU-T Rec. P.1204
P.NATS Phase 2 / VQEG AVHD

- Short-term video quality prediction modules (5–10 second clips)
- New features beyond P.1203
  - Codecs: H.264, HEVC , VP9
    - Ongoing: AV1 extension
  - Resolutions: up to UHD-1 (2160p)
  - Framerate: up to 60 fps

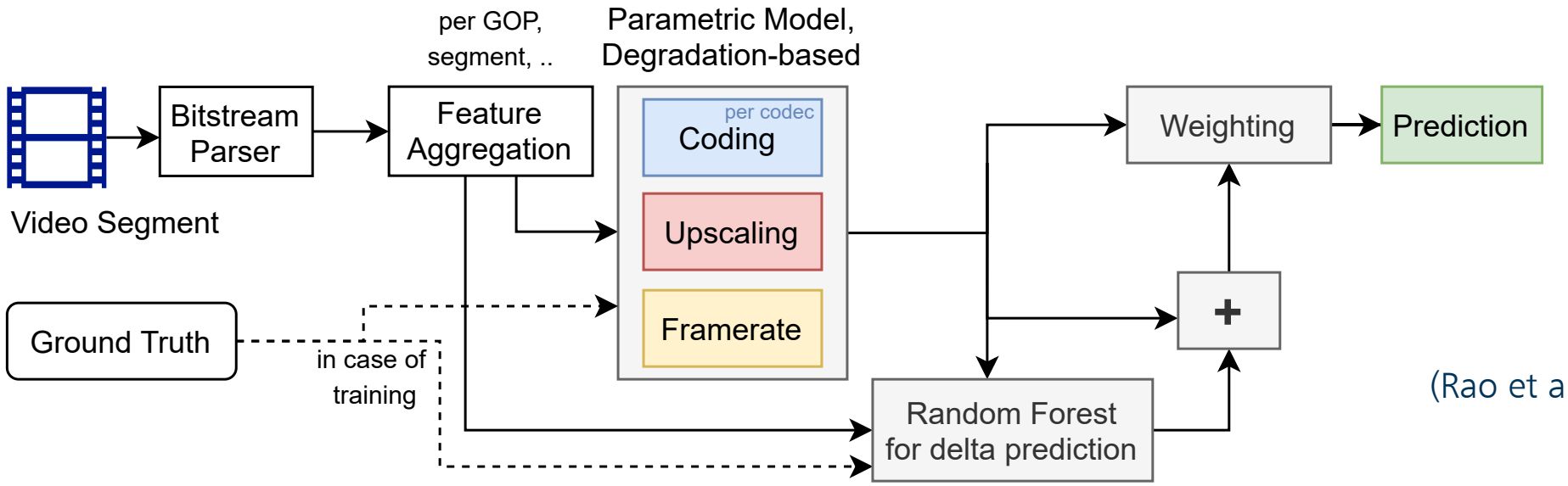
Bitstream-based (P.1204.3) Pixel RR/FR (P.1204.4) Hybrid metadata (P.1204.5)



Prediction error (RMSE) for submitted models (red, green, blue) compared to PSNR (purple), VMAF (orange), and SSIM (brown)

(Raake et al. IEEE Access 2020*)

Example: ITU-T Rec. P.1204.3 Model Algorithm



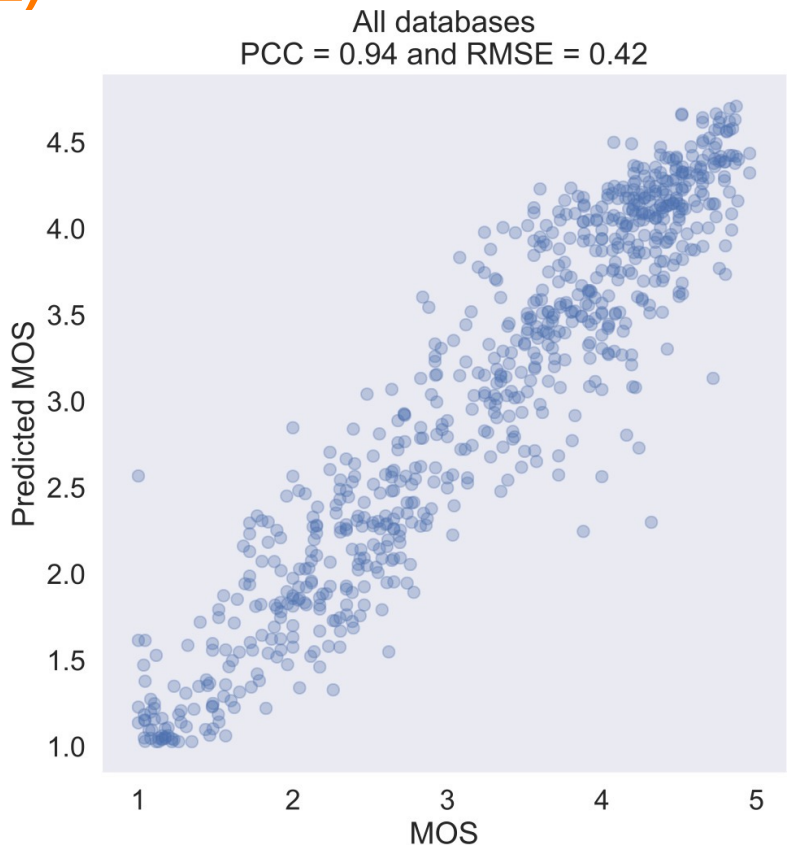
(Rao et al. IEEE QoMEX 2020*)

- Ensemble model: Curve fitting + machine learning (Random Forest)
- Feature used: QP, average motion statistics, frame size statistics, metadata
- Long-term integration for sequence durations [1,5] min based on P.1203.3 cf. P.1204.3, App. II
- Reference implementation: https://github.com/Telecommunication-Telemedia-Assessment/bitstream_mode3_p1204_3

ITU-T Rec. P.1204.3 Performance (TUIL)

- Open dataset for 4K/UHD subjective quality
  (cf. Rao et al. IEEE ISM 2019*)
  https://github.com/Telecommunication-Telemedia-Assessment/AVT-VQDB-UHD-1
- Comparison for cases w/o framerate variation
  (cf. Rao et al. IEEE QoMEX 2020*)

Metric	RMSE	Pearson	Spearman	Kendall	R^2 Score
P.1204.3	**0.4**	**0.962**	**0.948**	**0.804**	**0.874**
VMAF	0.531	0.880	0.889	0.721	0.774
BRISQUE	0.653	0.815	0.838	0.653	0.660
NIQE	1.009	0.432	0.445	0.301	0.187
PSNR	1.109	0.131	0.682	0.531	0.017
SSIM	0.956	0.520	0.761	0.569	0.270
MS-SSIM	0.896	0.599	0.752	0.563	0.358
ADM2	0.580	0.855	0.874	0.698	0.731
VIFP	0.757	0.736	0.756	0.562	0.542



- Ongoing: Validation of subjective test methods and model extensions based on "outside-the-lab" and crowd-testing, cf. e.g., (Rao et al., QoMEX 2021*)

P.1204.3 for gaming

- Development of full-HD extension of P.1204.3 (Rao et al., MMSP 2020*)
- Used 4 datasets (3 open + 1 own): GamingVideoSet[1], KUGVD[2], CGVDS[3], Own TU Ilmenau
- Note: Ongoing ITU-T SG12 Q14 Work Item P.BBQCG on Cloud Gaming Quality Monitoring, involving also AVT/TUIL's remote-testing software AVRateVoyager (Göring et al., MMSP 2021*)

GamingVideoSet

Model	RMSE	PCC	SROCC	Kendall	R^2
P.1204.3	0.45	0.88	0.87	0.69	0.77
FHD-mapped P.1204.3	**0.43**	**0.89**	**0.88**	**0.70**	**0.79**
PSNR	0.63	0.74	0.74	0.57	0.55
SSIM	0.57	0.80	0.80	0.61	0.62
VMAF	0.47	0.87	0.86	0.69	0.75
STRREDOpt	0.65	0.71	0.74	0.55	0.51
SPEEDQA	0.65	0.71	0.74	0.56	0.51
BRISQUE	0.84	0.49	0.46	0.31	0.20
BIQI	0.84	0.43	0.45	0.30	0.18
NIQE	0.64	0.77	0.71	0.53	0.52
MEON	0.87	0.35	0.30	0.20	0.13

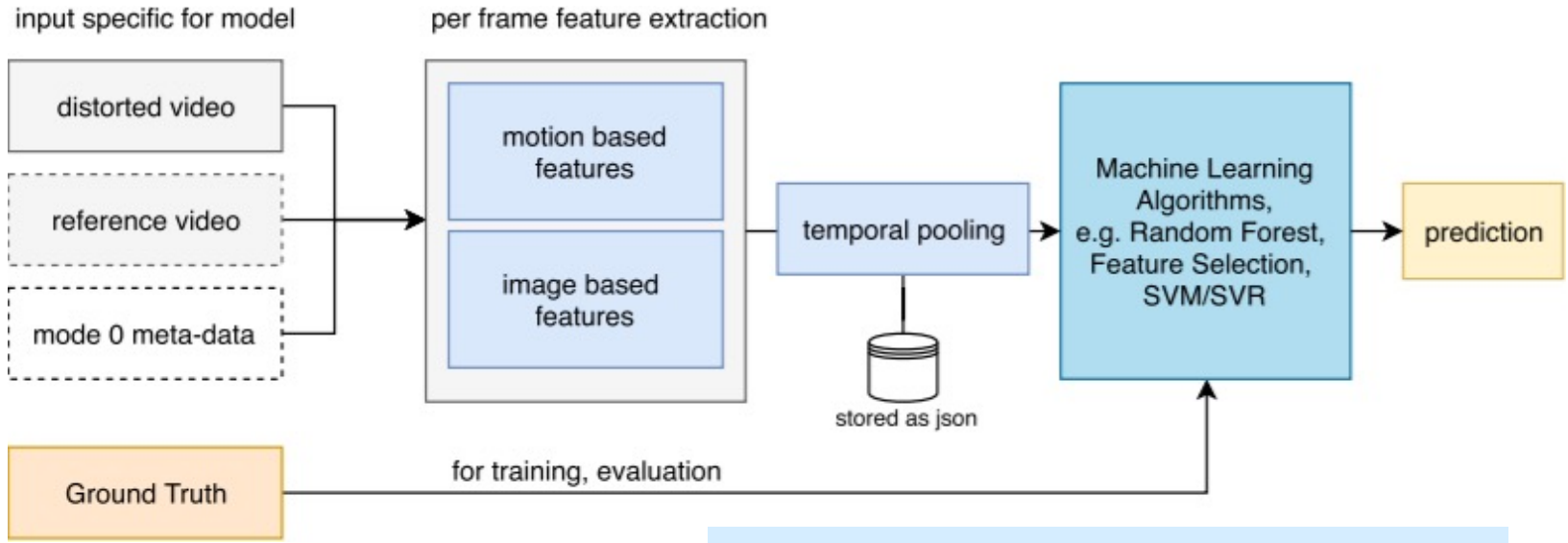
KUGVD

Model	RMSE	PCC	SROCC	Kendall	R^2
P.1204.3	0.39	0.93	0.92	0.77	0.86
FHD-mapped P.1204.3	**0.35**	**0.94**	**0.93**	**0.78**	**0.88**
PSNR	0.62	0.80	0.84	0.67	0.64
SSIM	0.48	0.89	0.91	0.74	0.79
VMAF	0.41	0.92	0.92	0.77	0.85
STRREDOpt	0.71	0.73	0.84	0.66	0.53
SPEEDQA	0.74	0.70	0.85	0.67	0.49
BRISQUE	0.81	0.63	0.60	0.42	0.39
BIQI	0.83	0.60	0.59	0.40	0.36
NIQE	0.55	0.85	0.84	0.66	0.72
MEON	0.94	0.44	0.39	0.26	0.19

[1](Barman et al., NetGames 2018), [2](Barman et al., IEEE Access 2019), [3](Zadtootaghaj et al., ACM MM 2020)

Note: Pixel-based and metadata models at TU Ilmenau

- Pixel-based models: Machine-learning (Göring et al., IEEE Access 2021*)
  - FUME: Full-reference
  - NOFU: No-reference
  - HYFR: Hybrid full-reference
  - HYFU: Hybrid no-reference



Reference implementation https://github.com/Telecommunication-Telemedia-Assessment/pixelmodels

- Metadata-based models: Based on P.1203.1 and P.1204.3 ("Mode 0", "Mode 1"; will be made available Open Source shortly), plus tools for practical deployment

360° video

Distribution

content **production**

source characteristics, recording settings

post-production settings

content **distribution**

transmission system characterisics

reproduction system characteristics

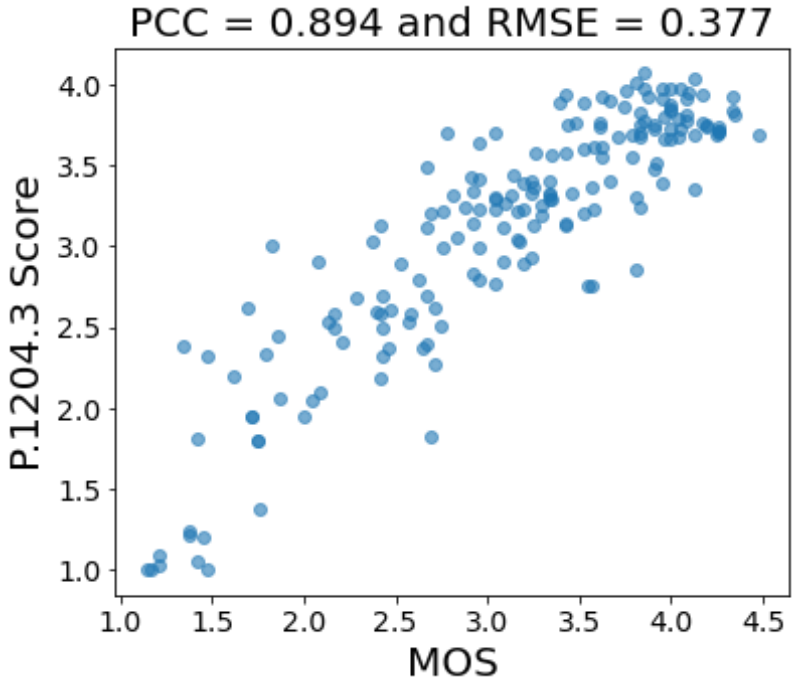
viewer characteristics

- P.1204.3 for 360° Video Quality?

360° video quality – P.1204.3 / bitstream model (Mode 3)

- 3 own tests with 7 / 8 source sequences each, mix of entertaining ARTE and YouTube contents

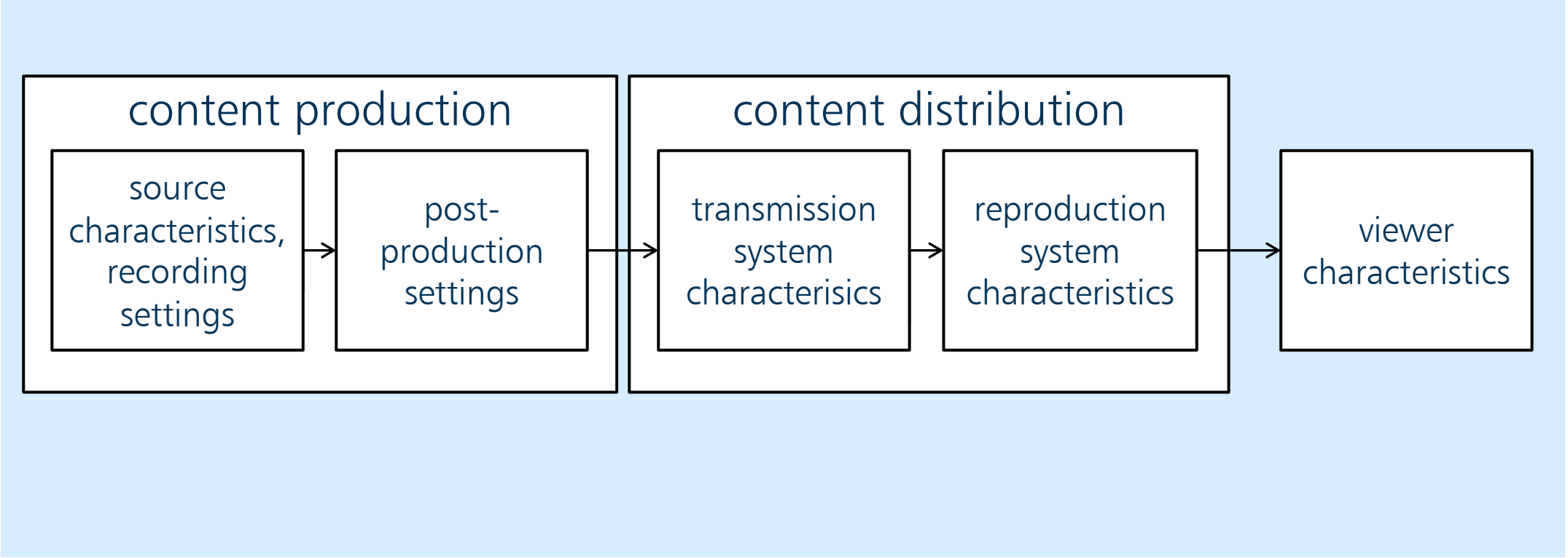
Model	RMSE ↓	PLCC ↑	SROCC ↑	Kendall ↑
Hybrid [2]	0.425	0.891	0.89	0.714
Mode 0 [2]	0.503	0.832	0.865	0.68
VMAF_cc [3]	0.384	0.898	0.872	0.700
VMAF [11]	0.431	0.870	0.834	0.664
ADM2 [5]	0.494	0.825	0.819	0.640
WS_SSIM	0.500	0.820	0.864	0.671
VIFP [15]	0.554	0.773	0.656	0.502
WS_PSNR	0.598	0.729	0.767	0.582
SSIM [18]	0.622	0.702	0.730	0.563
PSNR	0.762	0.489	0.627	0.469
P.1204.3 [14]	0.377	0.894	0.870	0.679



(Rao et al., paper in preparation, to be submitted 2021*)

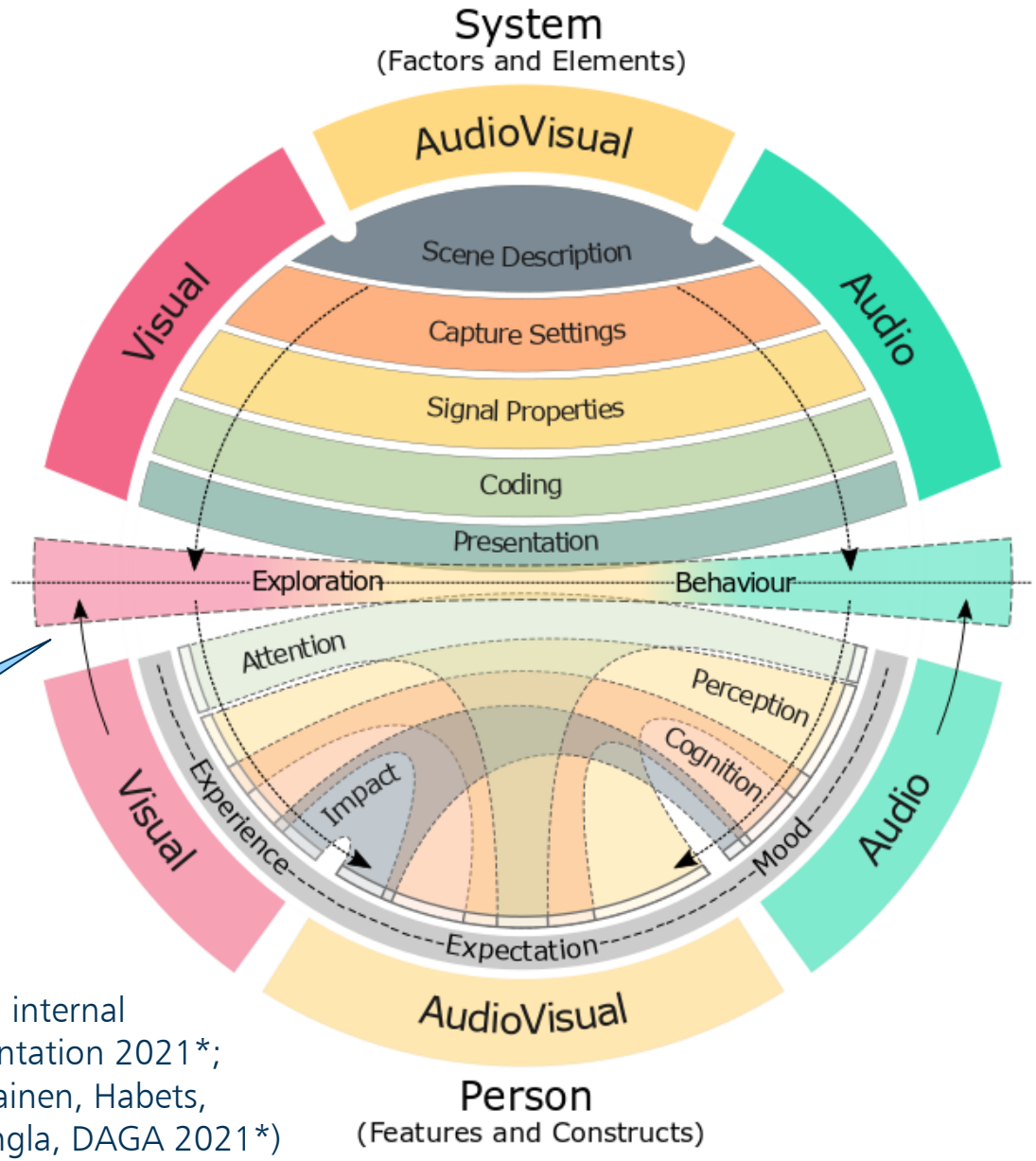
- *Note: Number of further studies by AVT/TUIL's on subjective and objective assessment of 360° video quality in Singla et al. (2017-today) and Fremerey et al. (2017-today), contributions VQEG, ITU-T SG12*
  - *Encoding settings, framerate & frame interpolation, HMD comparison, …*

Towards a holistic evaluation of 360o QoE & IVEs



QoE evaluation framework for audiovisual 360° and IVEs

- Project QoEvaVE – QoE Evaluation of Interactive Virtual Environments (IVEs) with Audiovisual Scenes
- Collaboration International Audio Labs Erlangen & TU Ilmenau
  - 3 DoF to 6 DoF, 360° video to CGI IVEs, visual + spatial audio (HoA)
- Taxonomy, QoE assessment framework, testing tool, labelled databases
- Open Science approach

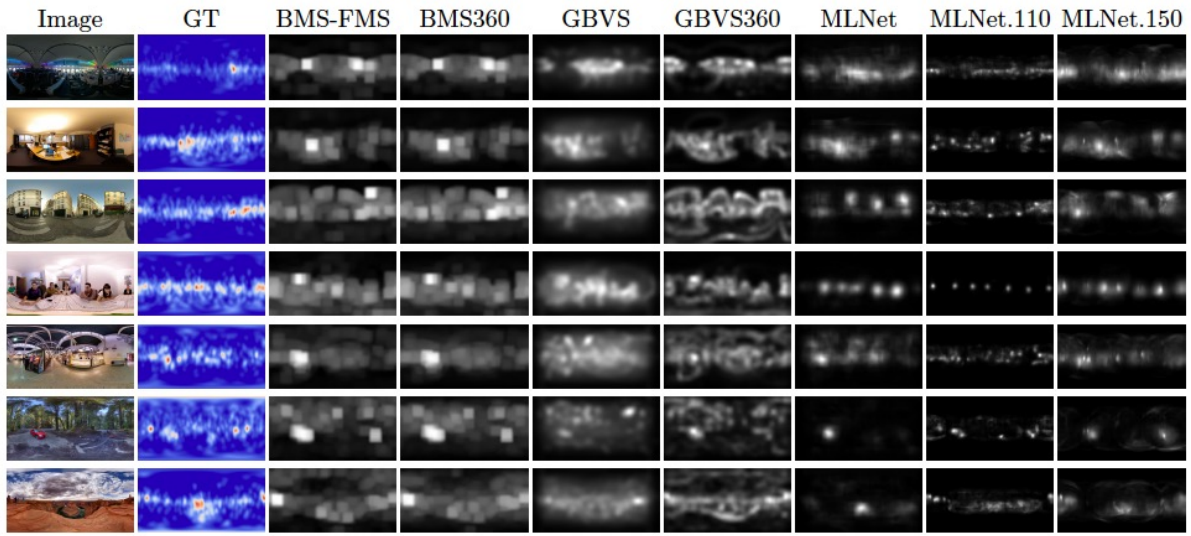


Prior work at TU Ilmenau
- 360° video viewing behavior
- Simulator Sickness assessment

(Robotham et al., internal project documentation 2021*; Raake, Rummukainen, Habets, Robotham & Singla, DAGA 2021*)

360° video viewing behavior assessment

- Head-saliency measurement
  - TUIL dataset in collaboration with ARTE (48 participants, 20 videos with 30 s duration, plus software for data capture AVTrack360; Fremerey et al., ACM MMSys 2018*)
- Saliency prediction
  - 360 images (Lebreton & Raake, ICME 2017*, J. Signal Proc.: Image Comm. 2018*)
  - 360 video (Lebreton, Fremerey & Raake, ICME 2018*)
  - Current: 360 audiovisual saliency with & w/o audio (project QoEvaVE)

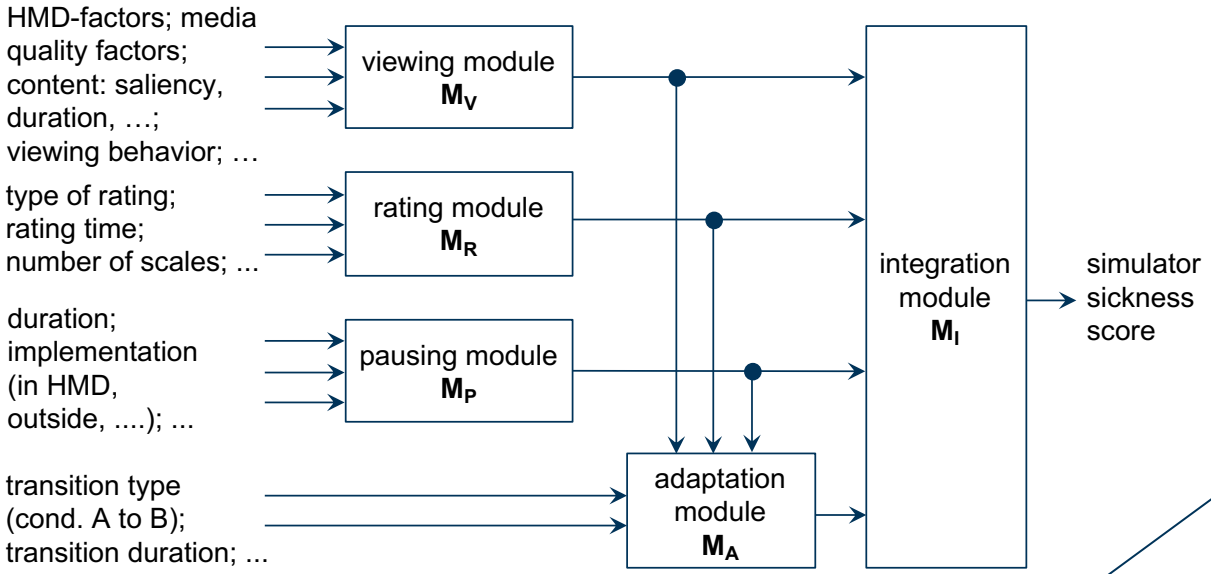


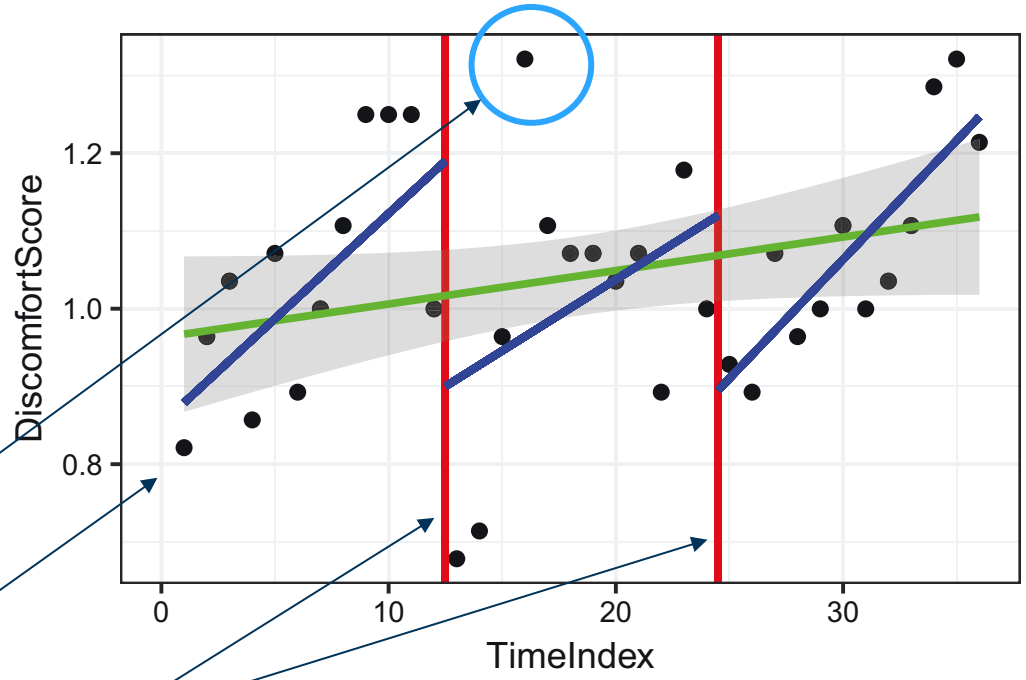
- *Note: Number of further studies including exploration behavior, simulator sickness, presence by Singla et al. (2017-today) and Fremerey et al. (2017-today), contributions VQEG, ITU-T SG12, cf. talk by Jesús Gutiérrez*

Towards a Simulator Sickness Model (SiSiMo)

- Test: Judge quality (ACR) and comfort on single-valued scale each
- Modelling temporal evolution of simulator sickness

Simulator Sickness Model (SiSiMo) Overview





Outliers: Remaining content effects

60s long videos

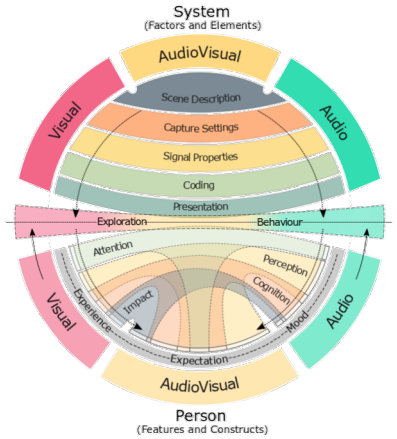
Breaks, several min.

(Raake et al. IEEE VR 2020*)

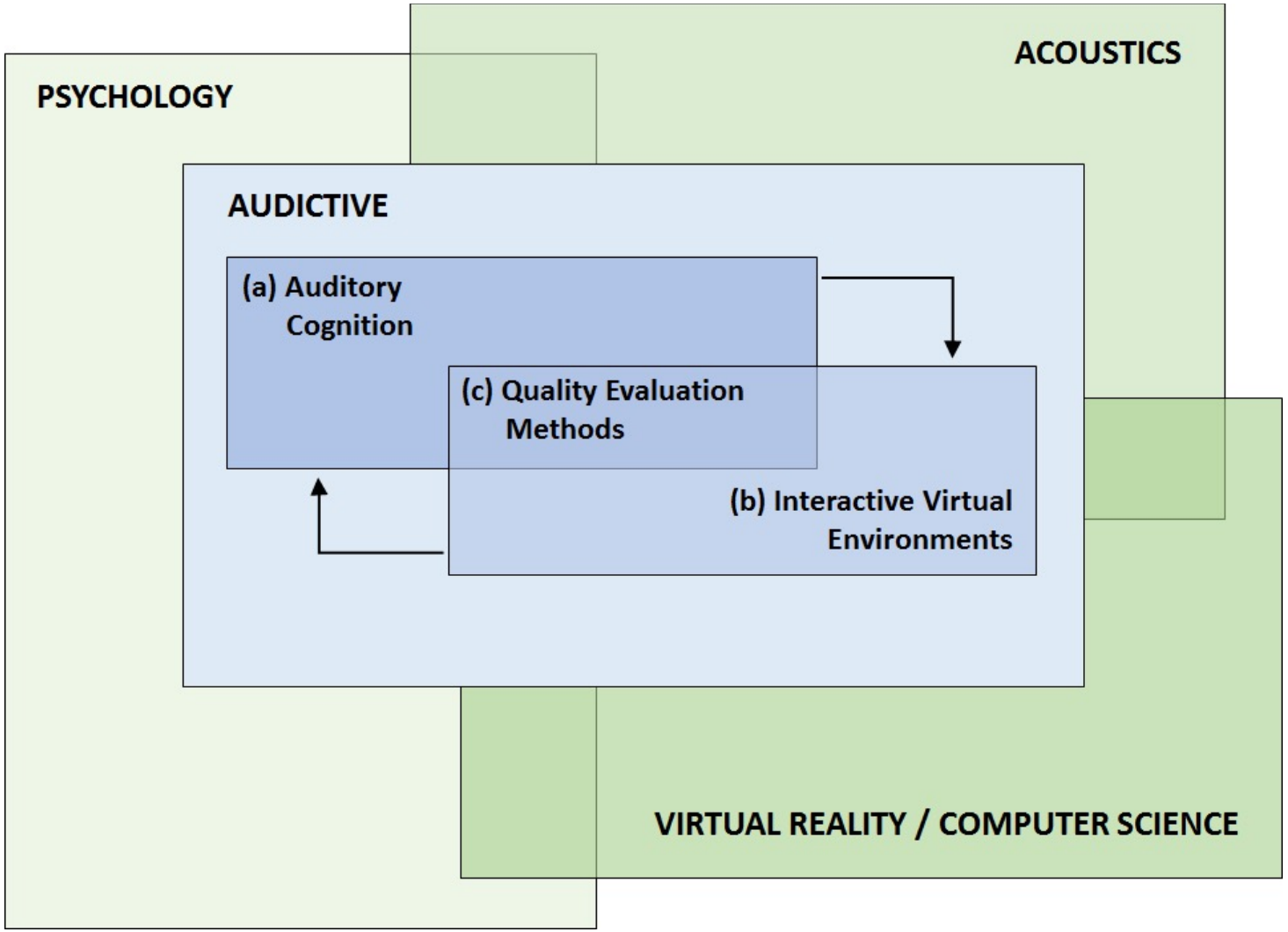
AUDICTIVE – Auditory Cognition in Interactive Virtual Environments

- Priority program SPP 2236
  German Research Foundation (DFG)
  (http://www.spp2236-audictive.de/)
- Phase 1: 2021 - 2024
- 11 interdisciplinary projects
  for 3 years (2-3 partners each)
  - Further TUIL projects:
    Cognitive load in classrooms, social XR communication and co-presence

→ Using cognitive performances as IVE QoE indicators



Example: QoEvaVE



Thank you for your attention!
I look forward to your questions.

AVT Group: https://www.tu-ilmenau.de/en/mt-avt/

This work was funded partially by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) and by the Free State of Thuringia, Germany, project CYTEMEX (FKZ: 2018-FGI-0019).

Funded by

DFG Deutsche Forschungsgemeinschaft German Research Foundation

EFRE EUROPE FOR THURINGIA EUROPEAN REGIONAL DEVELOPMENT FUND

EUROPEAN UNION