Acceptability and Annoyance of Video Quality via Crowdsourcing

Ali Ak, Patrick Le Callet
Nantes University
Abhishek Gera, Denise Noyes, Hassene Tmar, Ioannis Katsavounidis
META

VQEG Meeting
Klagenfurt, July 2024

Acceptability & Annoyance

- Video quality alone is not enough to define Quality of Experience

- User expectations has a major impact on user satisfaction

- Acceptance/Annoyance is a measure of user satisfaction for video streaming services, online social media platforms, etc., and it takes user expectations and user profile into account.

Acceptability & Annoyance: Multi-Step vs Single-Step

Multi Step
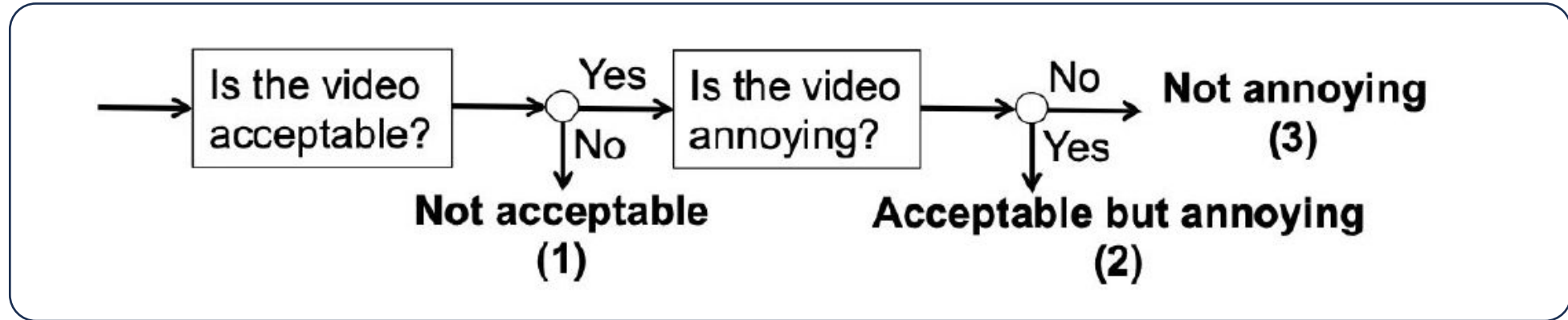
Single Step

Previously on Acceptability & Annoyance

1) Li et al.
   Basic vs Premium subscription
   Viewing on TV vs Tablet

2) Ak et al.
   Remaining Data
   Remaining Battery
   Signal Strength
   etc.,

1 - Jing Li, Lukas Krasula, Yoann Baveye, Zhi Li, and Patrick Le Callet, "Accann: A new subjective assessment methodology for measuring acceptability and annoyance of quality of experience," IEEE Transactions on Multi-media, vol. 21, no. 10, pp. 2589–2602, 2019
2 - Ali Ak, Anne Flore Perrin, Denise Noyes, Ioannis Katsavounidis, and Patrick Le Callet, "Video consumption in context: Influence of data plan consumption on qoe," in Proceedings of the 2023 ACM International Conference on Interactive Media Experiences, New York, NY, USA, 2023, IMX '23, p. 320–324, Association for Computing Machinery.

Nantes IPI-VUGC Dataset

- Two experiments with the same content: **AccAnn** and **ACR-HR**
- On Iphone 14 pro, in lab.

- 48 SRCs
  - 1080p resolution
  - 5 seconds
  - Varying fps (15-60)
  - Vertical orientation
  - Encoded with h264

- CRF:24 & Resolution: 512 × 288
- CRF:23 & Resolution: 640 × 360
- CRF:26 & Resolution: 960 × 540
- CRF:29 & Resolution: 960 × 540
- CRF:31 & Resolution: 1280 × 720
- CRF:34 & Resolution: 1920 × 1080

https://zenodo.org/doi/10.5281/zenodo.10475209

5

Can we replicate the same study via crowdsourcing?

- Crowdsourcing could provide,
  - Better reach
  - Understanding of user demographics on the user expectations

# Can we replicate the same study via crowdsourcing?

- Crowdsourcing could provide,
    - Better reach
    - Understanding of user demographics on the user expectations

- Few problems:
    - Is instruction based priming of user expectations suitable for crowdsourcing setting?

    - What is the influence of display and viewing environment on the video quality scores and AccAnn labels?

Crowdsourcing study

18 SRC from the IPI-VUGC dataset is selected for the crowdsourcing studies.
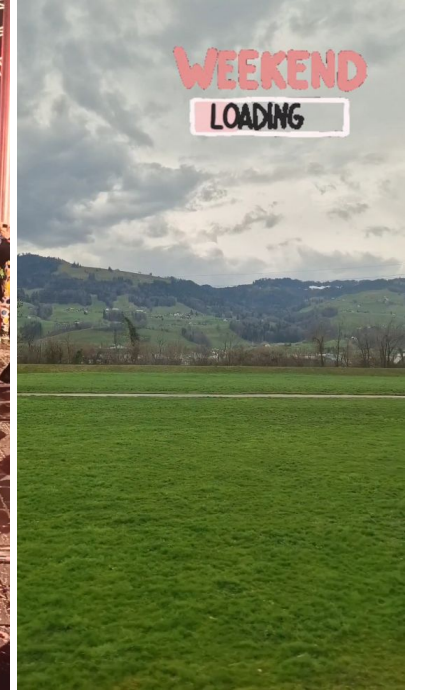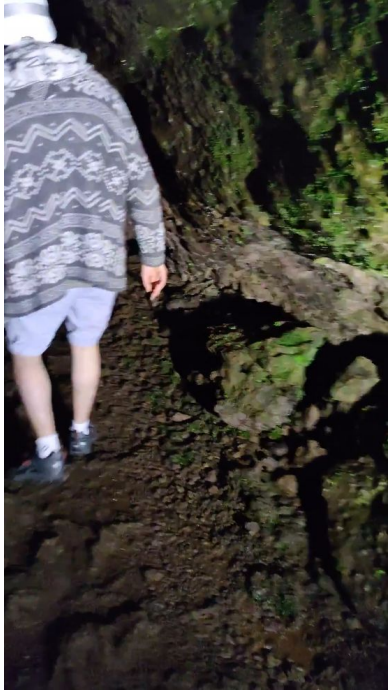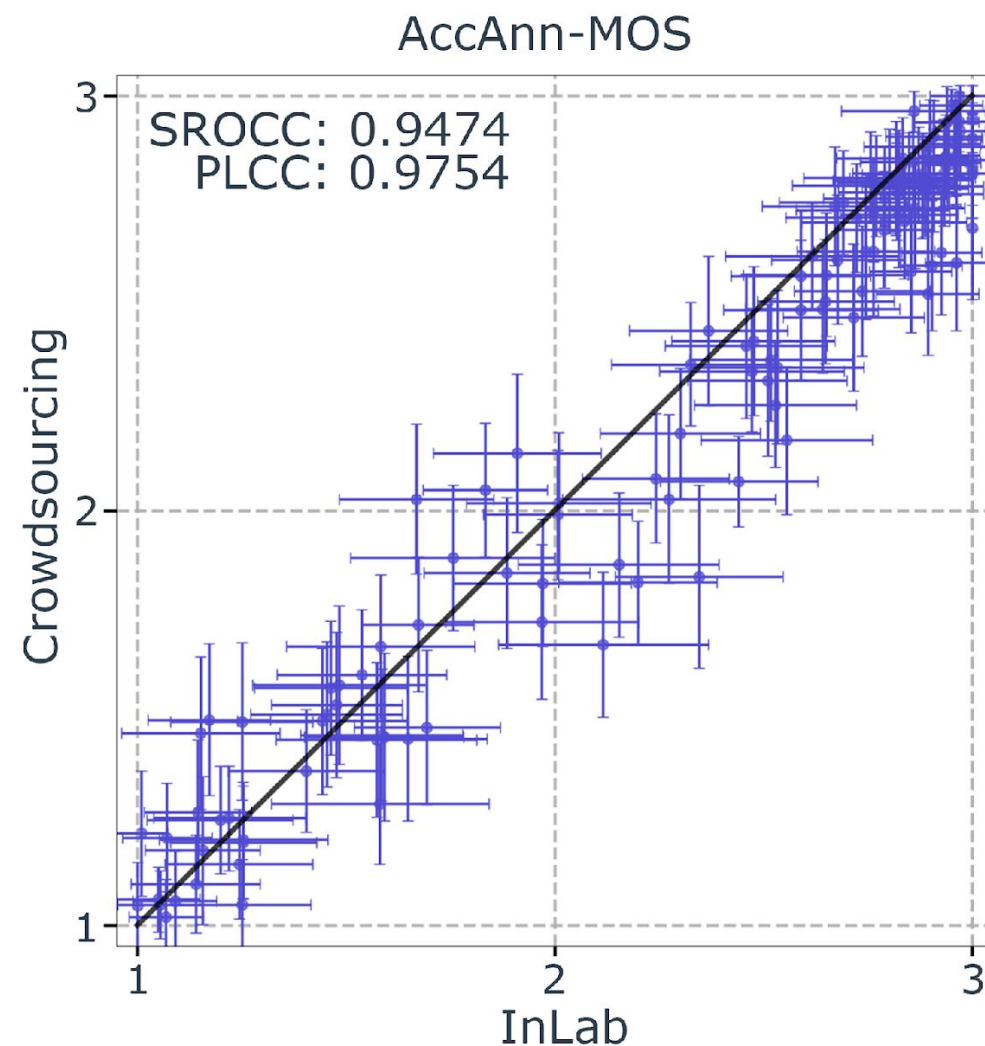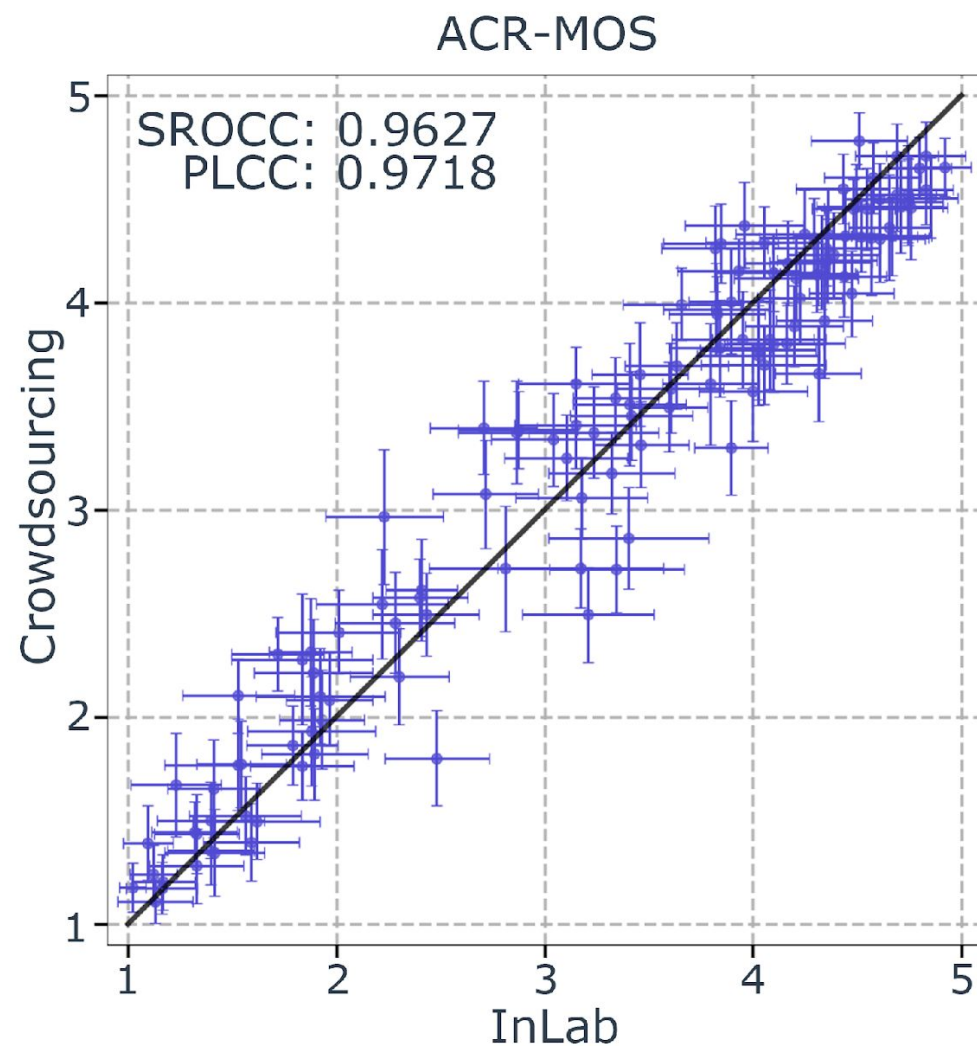
Crowdsourcing study

18 SRC from the IPI-VUGC dataset is selected for the crowdsourcing studies.

InLab

Crowdsourcing

Correlation of subjective opinions

ACR-MOS

SROCC: 0.9627
PLCC: 0.9718
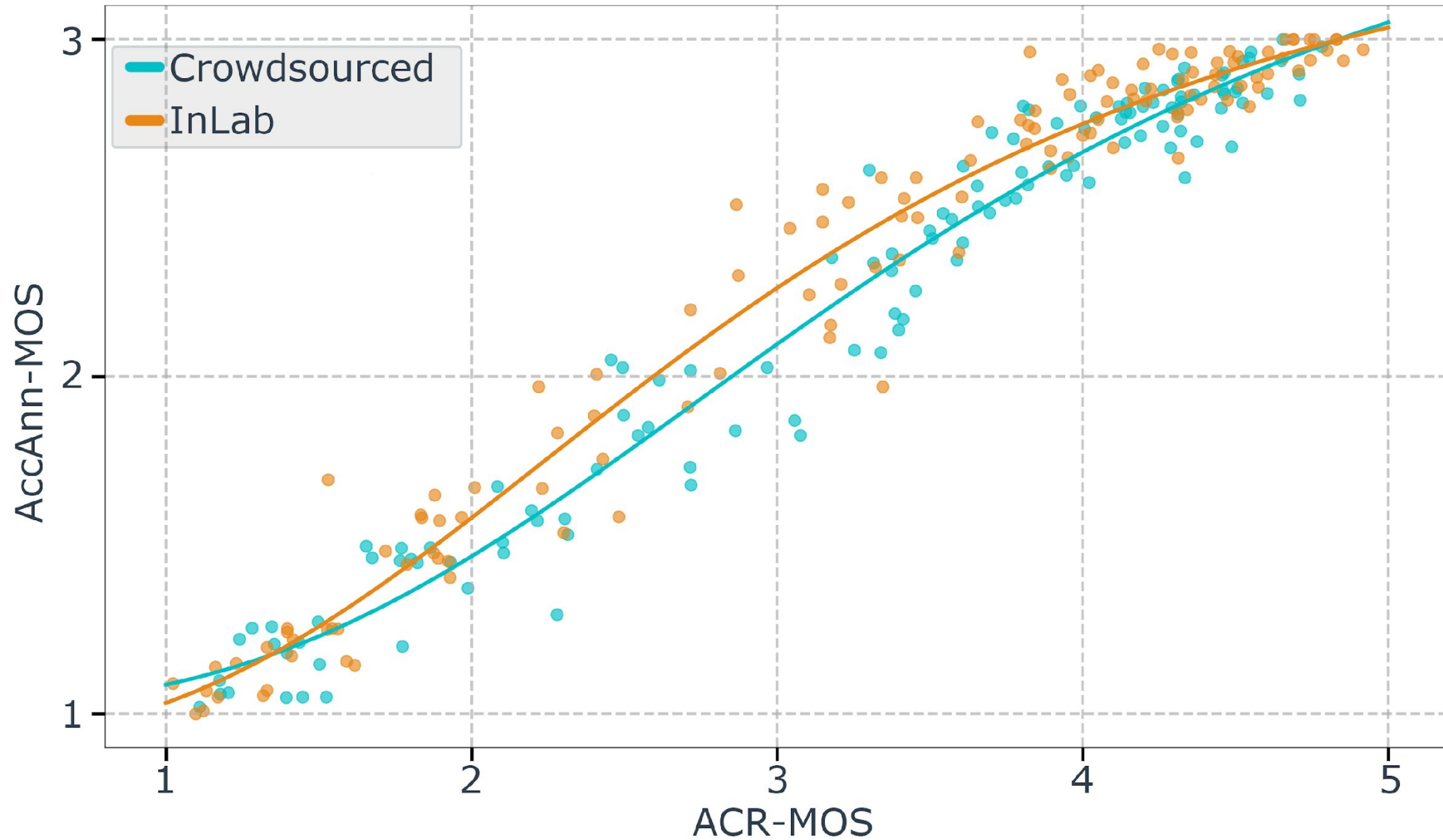
AccAnn-MOS

SROCC: 0.9474
PLCC: 0.9754

Lower correlation for AccAnn-MOS values between InLab and Crowdsourcing settings.
Slightly lower AccAnn-MOS values in crowdsourcing setting in mid/high quality range.

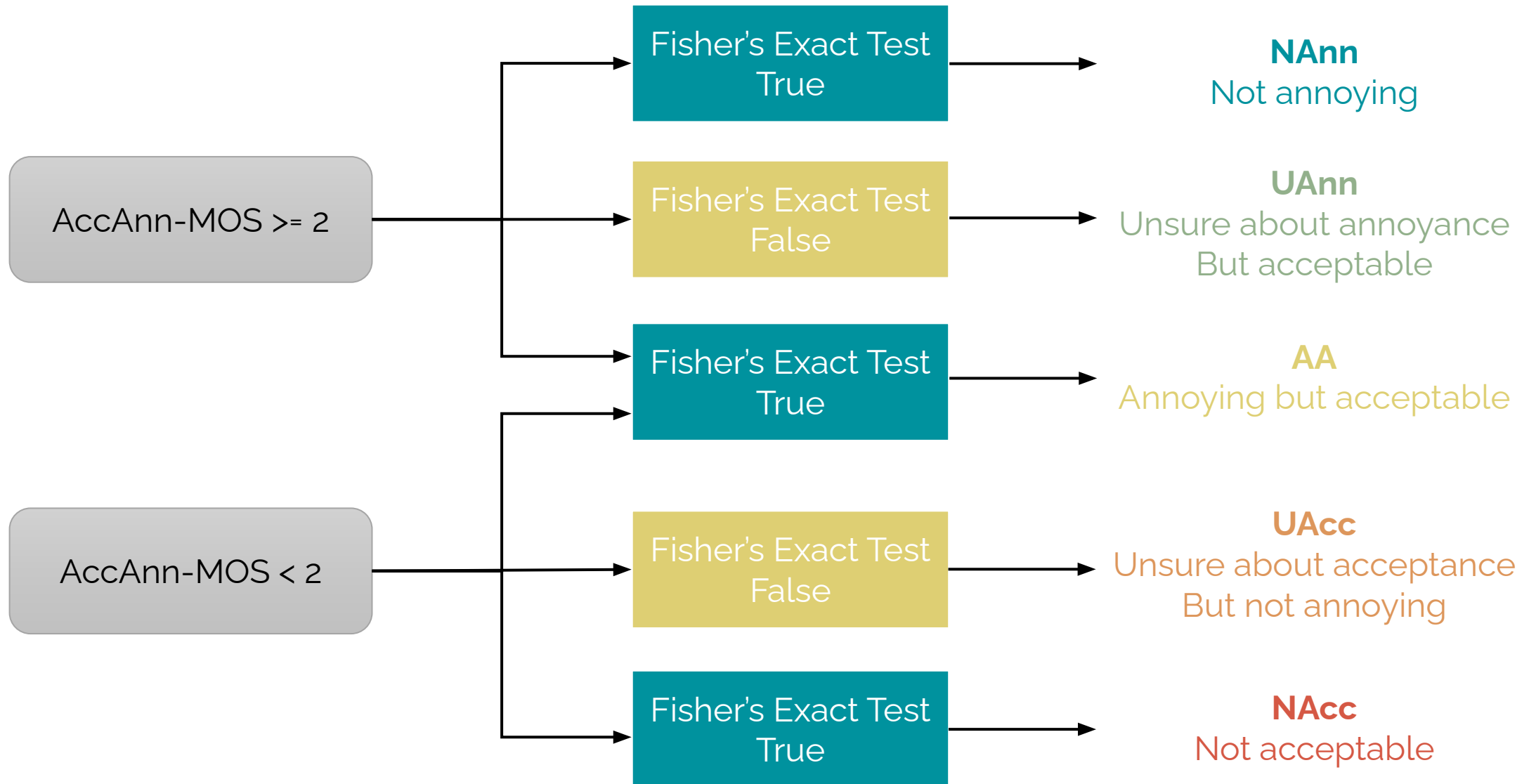# Discriminability and Mean 95% CI
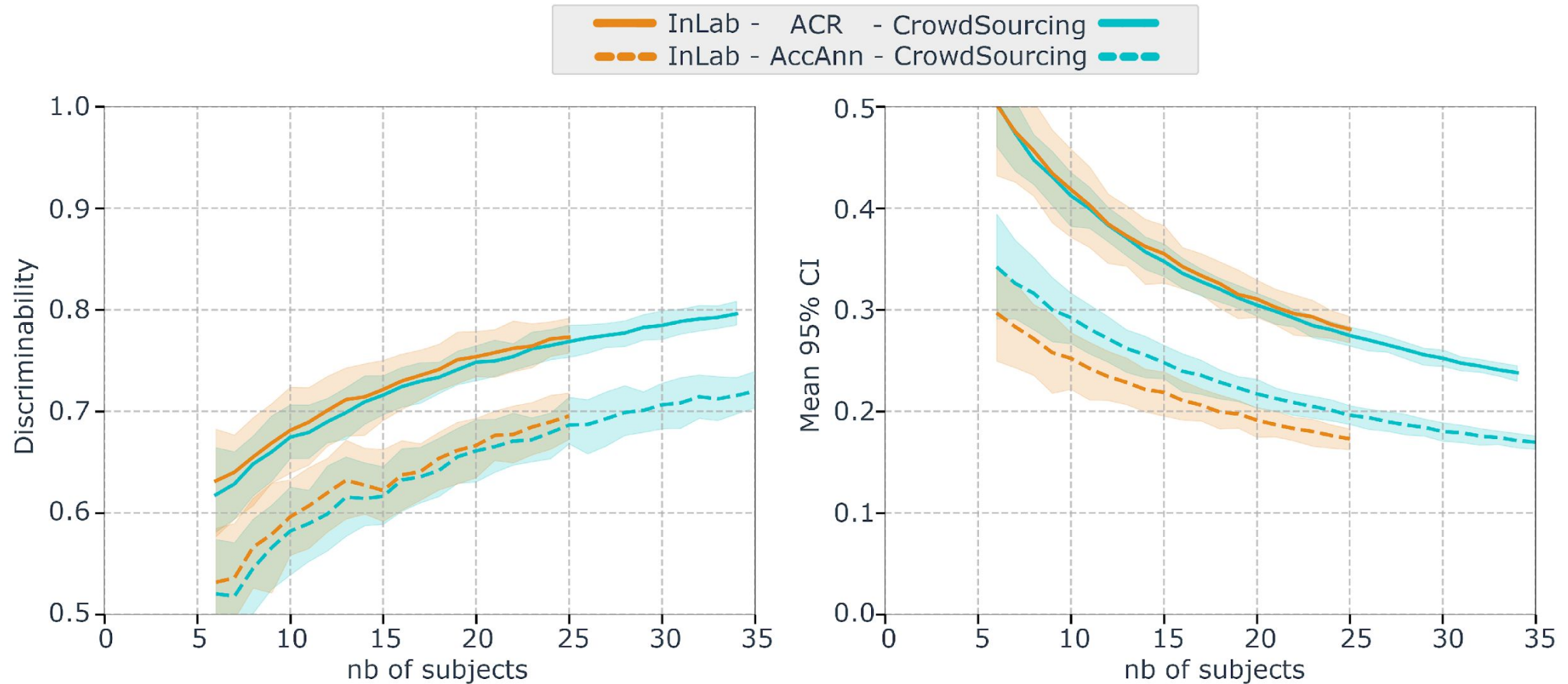
- ACR experiment has a higher discriminability than AccAnn experiment due to the variety in user expectations and the higher ambiguity of the acceptability & annoyance task.
- Lower CI indicates a higher agreement among subjects. AccAnn experiment InLab shows a higher agreement InLab than crowdsourcing.
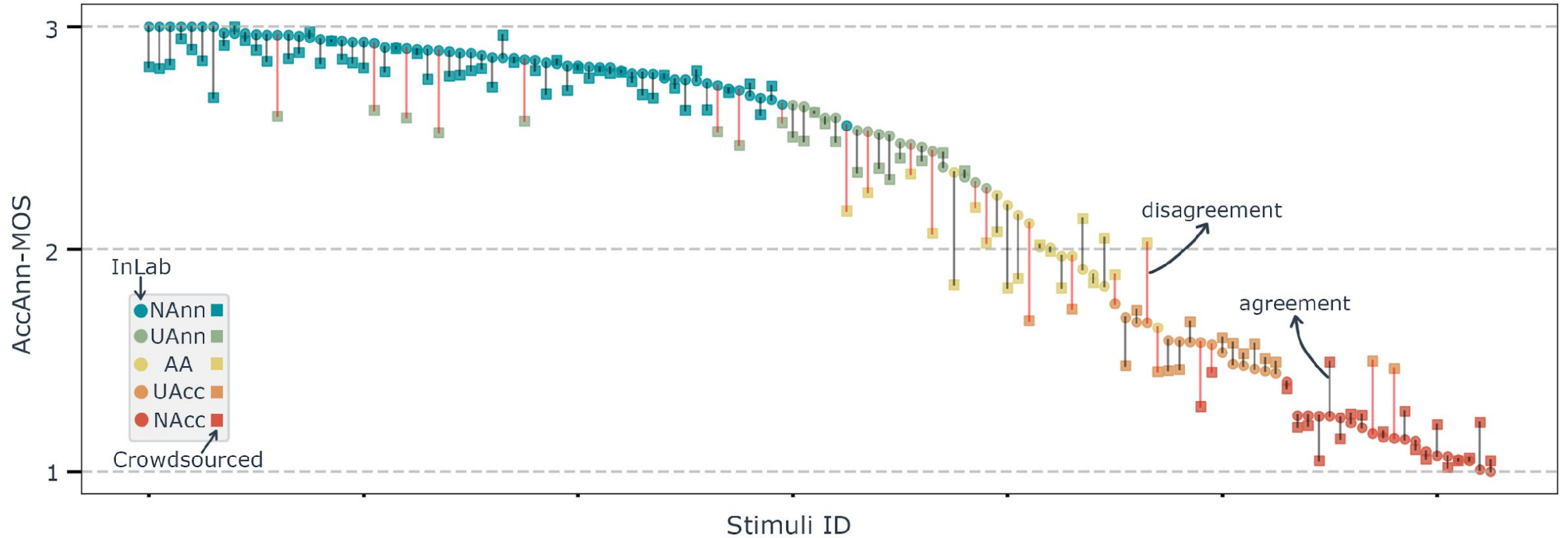
Relation between video quality and acceptability & annoyance

Determining acceptability & annoyance categories

AccAnn-MOS >= 2

Fisher's Exact Test True → **NAnn** Not annoying

Fisher's Exact Test False → **UAnn** Unsure about annoyance But acceptable

Fisher's Exact Test True → **AA** Annoying but acceptable

AccAnn-MOS < 2

Fisher's Exact Test False → **UAcc** Unsure about acceptance But not annoying

Fisher's Exact Test True → **NAcc** Not acceptable

Comparison of acceptability & annoyance categories

Acceptability & annoyance thresholds in terms of ACR-MOS and UVQ

Table 1. Acceptability and annoyance thresholds in terms of ACR-MOS and UVQ. ACR-MOS and UVQ values has the theoretical range of [1, 5].

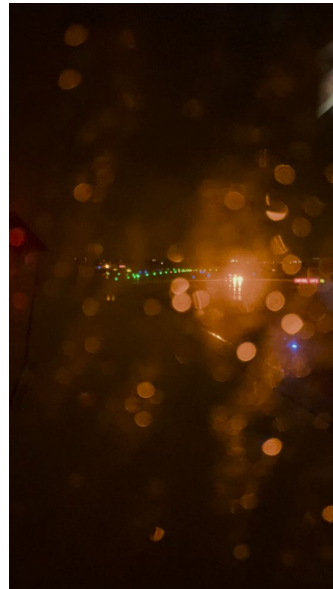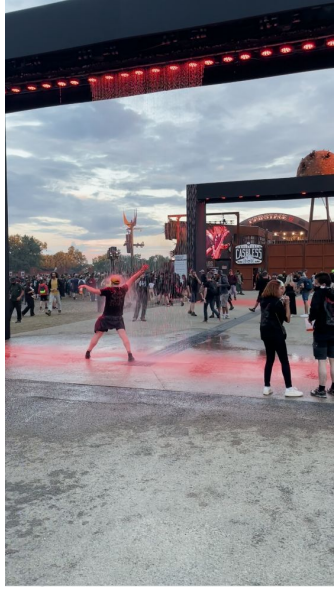|	Acceptability		Annoyance	
	InLab	crowdsourcing	InLab	crowdsourcing
ACR-MOS	1.9804	2.1157	3.4001	3.6543
UVQ	3.1891	3.1616	3.6508	3.6898

Conclusion

- Transferring acceptability & annoyance to crowdsourcing is not straightforward.

- Our findings indicate a more linear relationship between video quality and QoE in the crowdsourcing setting.

- In crowdsourcing settings, subjects exhibit a limited ability to adjust their expectations in response to instructions, unlike in controlled laboratory experiments.

- Simple transfer of the same methodology is not enough, and possibly a more sophisticated training session is required to enforce engagement of the crowdsourcing participants.

Nantes-MobileHDRVQA Dataset

https://zenodo.org/doi/10.5281/zenodo.11190759

- Recently published Nantes-MobileHDRVQA Dataset and its Acceptability & Annoyance pair
  - 60 SRC + 300 compressed videos (AV1)
  - User generated Iphone HDR videos
  - Vertically oriented
  - 1080p resolution
  - Varying fps (30-60)
  - 5 Seconds long

  - Experiment conducted InLab with Iphone 15 pro.

Nantes-MobileHDRVQA Dataset

Relation between video quality and acceptability & annoyance
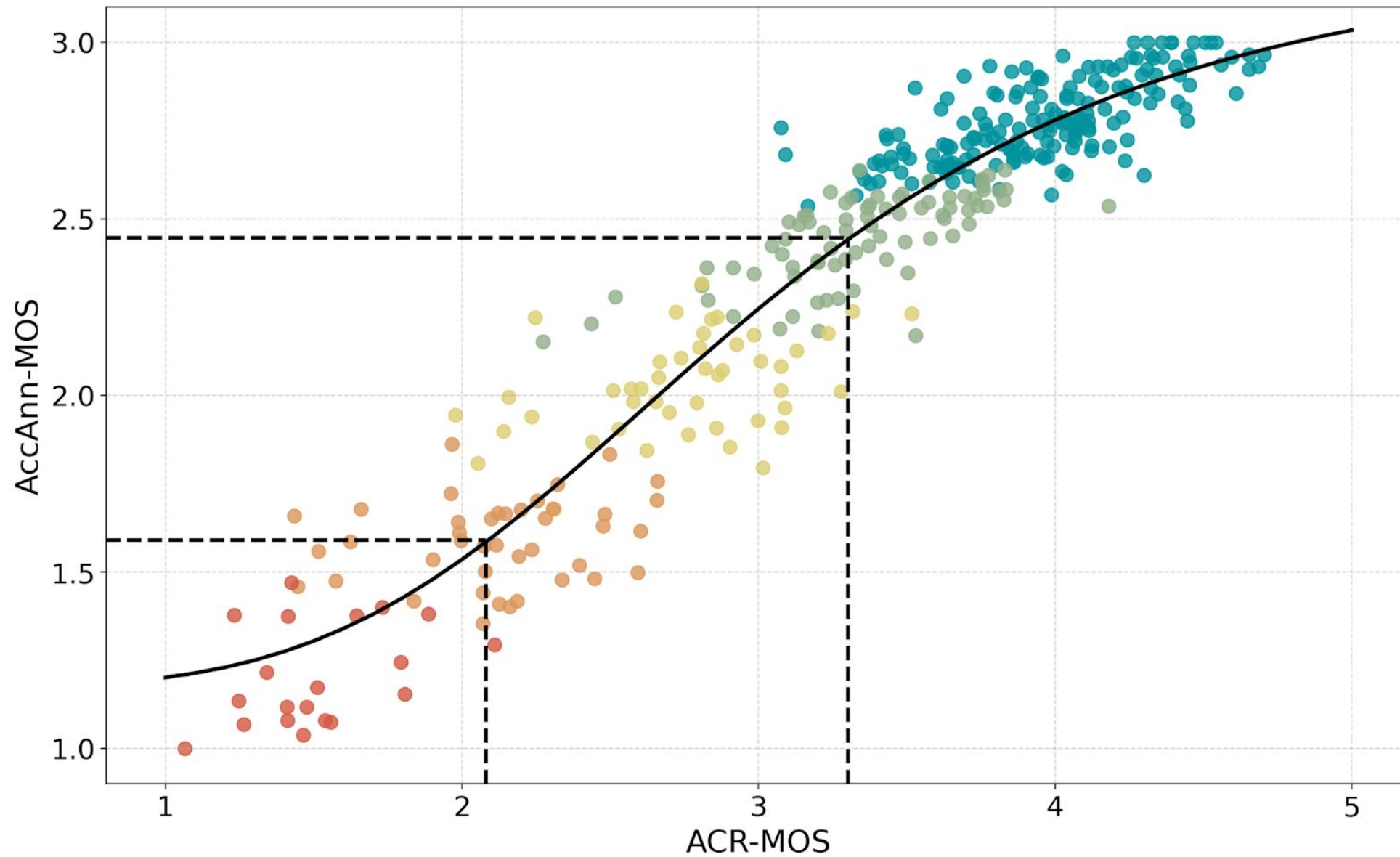
VMAF Performance

model	SRCC	PCC	RMSE
scores.baseline.vmaf.4k.v0.6.1	0.8109170	0.7871537	14.45074
scores.baseline.vmaf.float.4k.v0.6.1	0.8107823	0.7870906	14.44568
scores.baseline.vmaf.float.b.4k.v0.6.3	0.8070943	0.7793961	25.29217
scores.baseline.vmaf.v0.6.1	0.8054036	0.7776190	25.40265

phase	SRCC	PCC	RMSE
training	0.8734333	0.8580000	15.24153
testing (excluding references)	0.8319515	0.8236842	11.26952