Application of Sensory Evaluation Methods in Film

How Do Expert and Non-Expert Evaluations Differ?

Dominik Keller [1], **Tamara Seybold** [2], **Janto Skowronek** [1], **Rakesh Rao Ramachandra Rao** [1], **Alexander Raake** [1]

[1] Audiovisual Technology Group, Technische Universität Ilmenau, Germany

[2] Image Science, Arnold & Richter Cine Technik (ARRI), Germany

Contact: dominik.keller@tu-ilmenau.de

Motivation

sharp

red-tinged

How good?

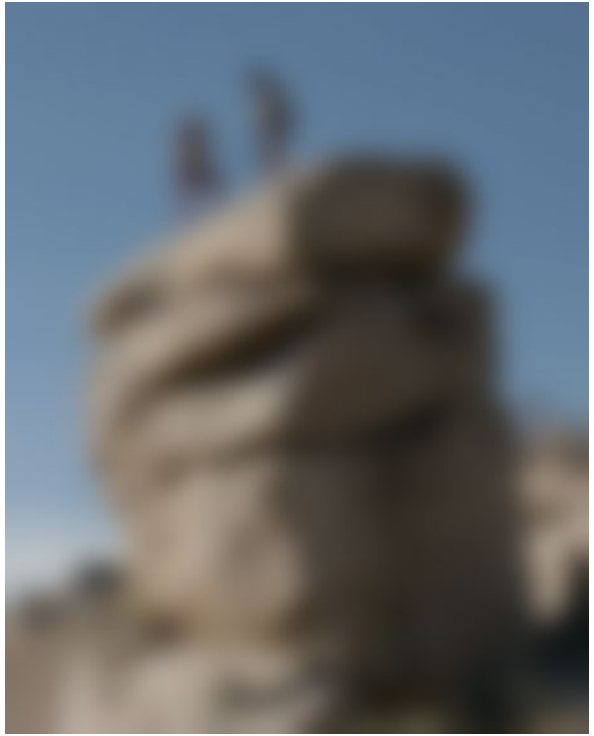
real

dark

sterile

grainy

Motivation



Which one is better?

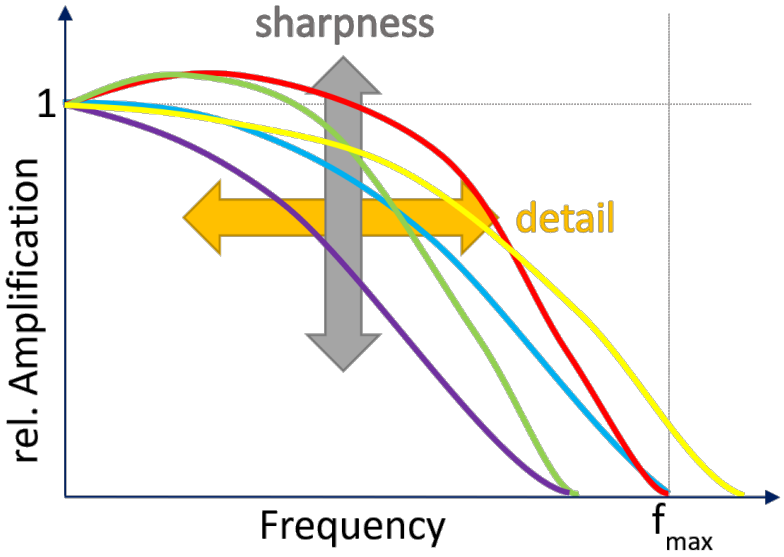
How different are they?

Motivation

- Established methods assessing error-related quality
  - Little explanation of what properties are perceptible and decisive
- Sensory evaluation methods
  - Origin: assessment of properties and quality of food [LawlessHeymann2013]
  - reveal perceptual dimensions to find optimal technical settings
  - are done with expert assessors
  - Discrimination, description and acceptance
  - Relevant insights in speech / audio [WierstorfEtAl2013; BechZacharov2006; WältermannEtAl2010]

Texture

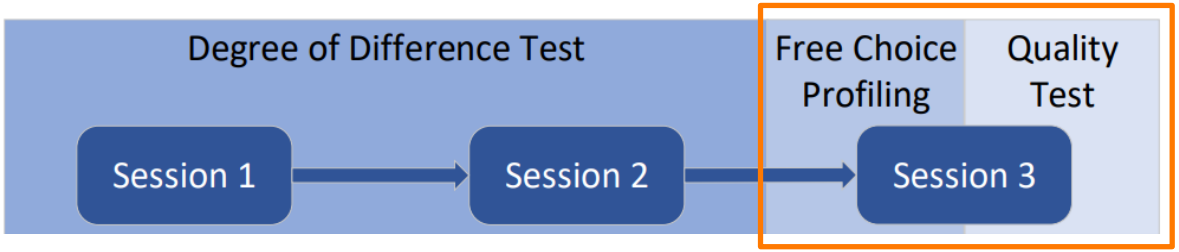
- *Look*, visual / technical parameters
- Sharpness and Detail: transmission behaviour of spatial frequencies
- Noise
  - artificial ARRI ALEXA noise and grain
  - varying size, gain, saturation
- Very subtle differences between stimuli



Sharpness and Detail

Testing Procedure

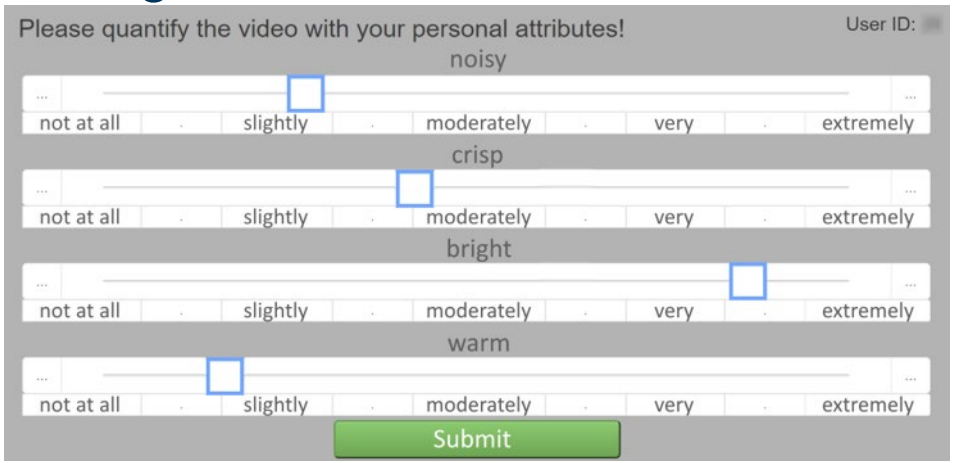
- Mixed-method approach [StrohmeierEtAl2010; BechEtAl1996; KaplanisEtAl2017]
- 25 expert subjects per test; in total 41 experts (incl. overlap)
- 18* UHD1 stimuli of ≈19s; static scenery
- Test setup acc. to ITU-T Rec. P.910



* Only 16 stimuli in Degree of Difference test caused by time reasons

Free Choice Profiling (FCP) and Generalized Procrustes Analysis (GPA)

- FCP related to [Lorho2005]
  - Individual elicitation of ≤7 attributes [Miller1956]
  - Rating on scales

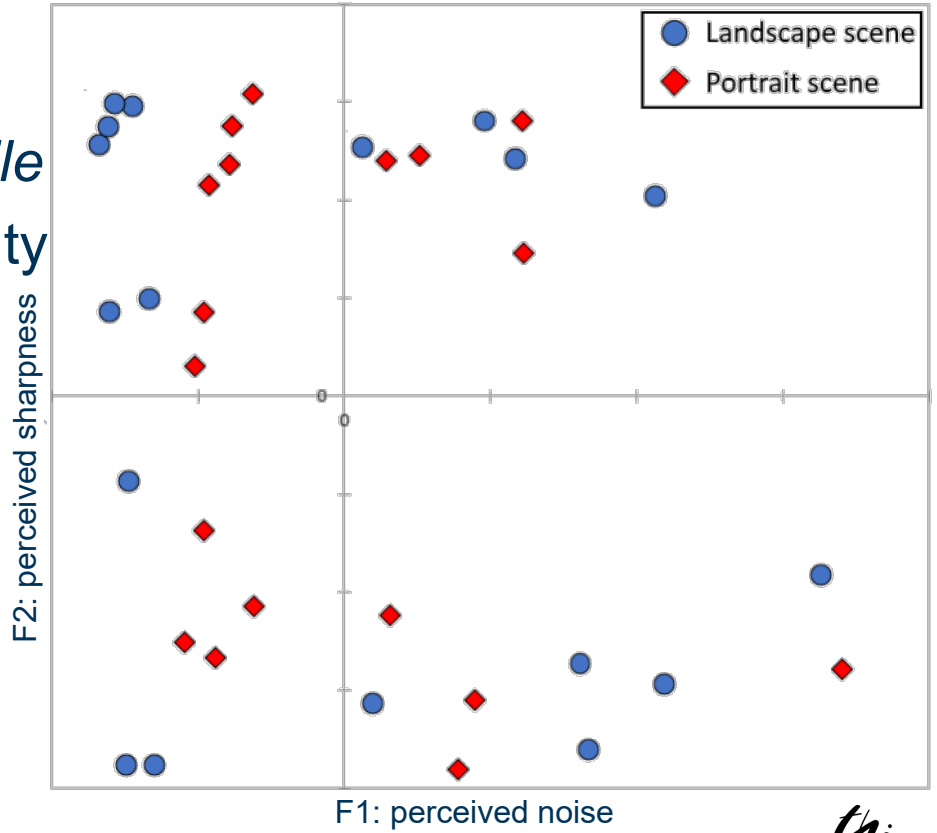


Free Choice Profiling (FCP) and Generalized Procrustes Analysis (GPA)

- FCP related to [Lorho2005]
  - Individual elicitation of ≤7 attributes [Miller1956]
  - Rating on scales
- GPA [Gower1975]
  - Fits individual data to consensus through transformations
  - Attribute values as Euclidian distances

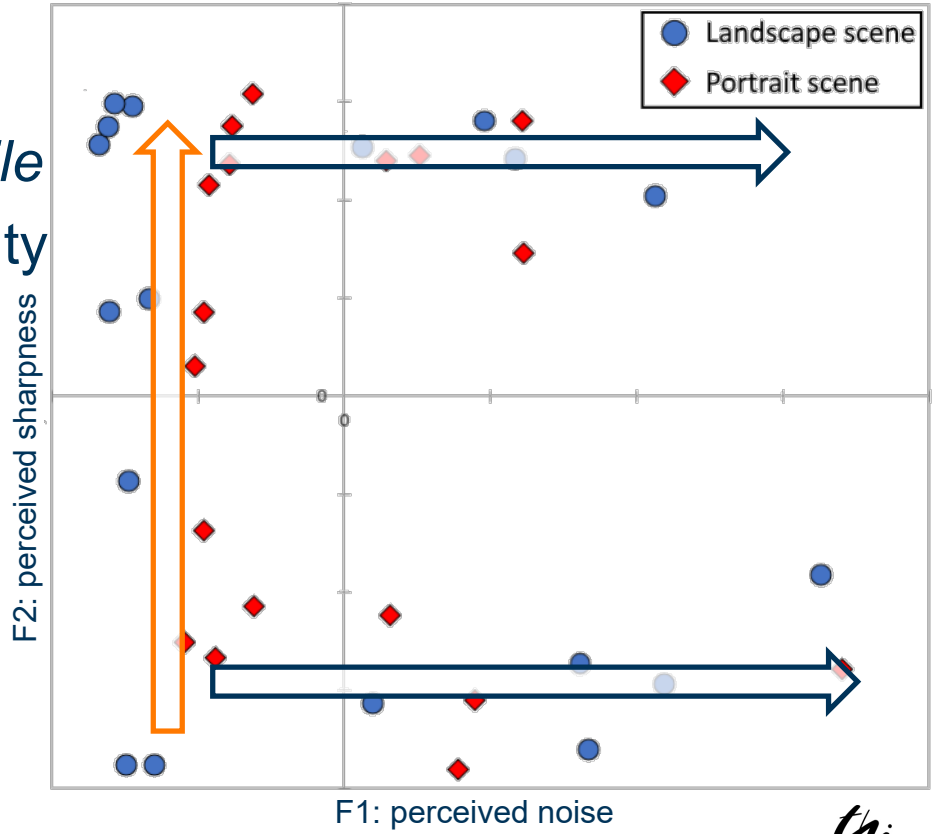
Free Choice Profiling and GPA

- Exemplary elicitated attributes:
  noisy, *active*, *sharp*, *detailed*, *sterile*
- Depictions cover >90% of variability
- Varimax-rotated



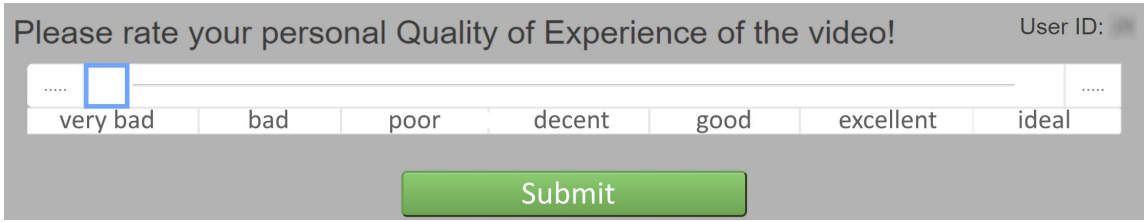
Free Choice Profiling and GPA

- Exemplary elicitated attributes:
  noisy, *active*, *sharp*, *detailed*, *sterile*
- Depictions cover >90% of variability
- Varimax-rotated
- F1: perceived noise
  F2: perceived sharpness
  ⇨ increasing noise level
  ⇨ increasing sharpness level



Experts: Quality Test, Preference Mapping

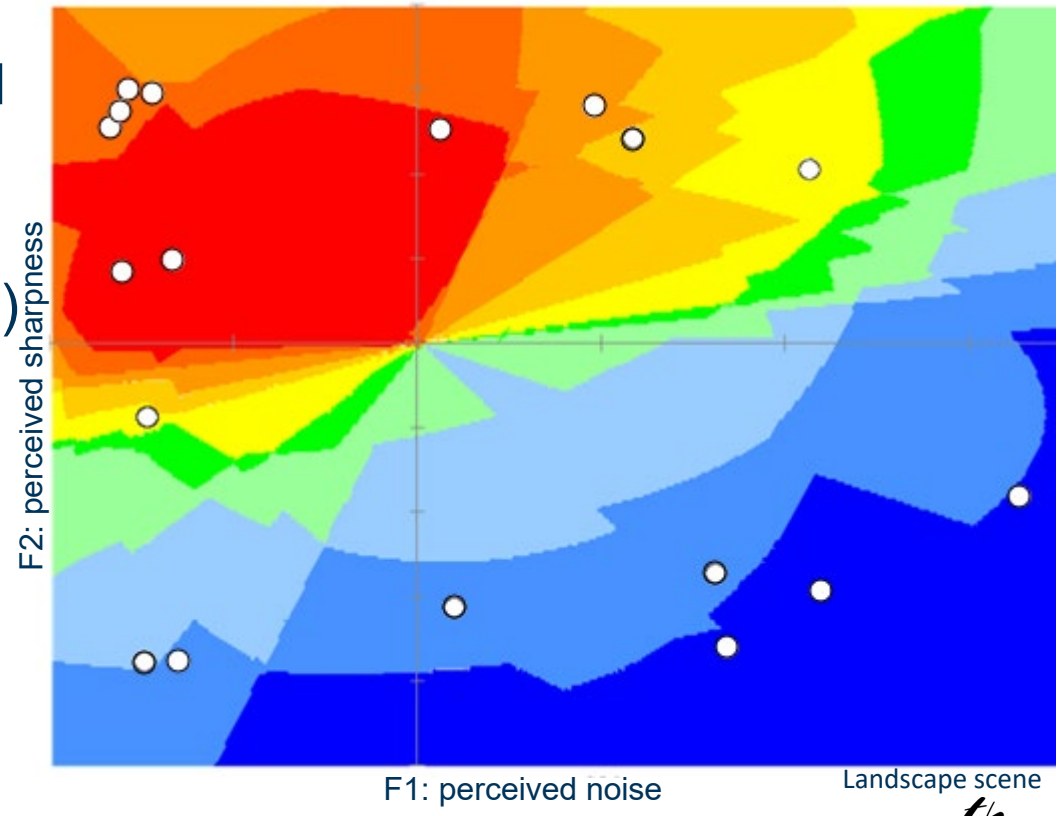
- Quality test
  - Rating on extended continuous scale [BoddenJekosch1996]



- External Preference Mapping (PM) [GreenhoffMacFie1994]
  - Visualizing quality in form of stimulus maps
  - External PM using colors / contours

Combination: External PM

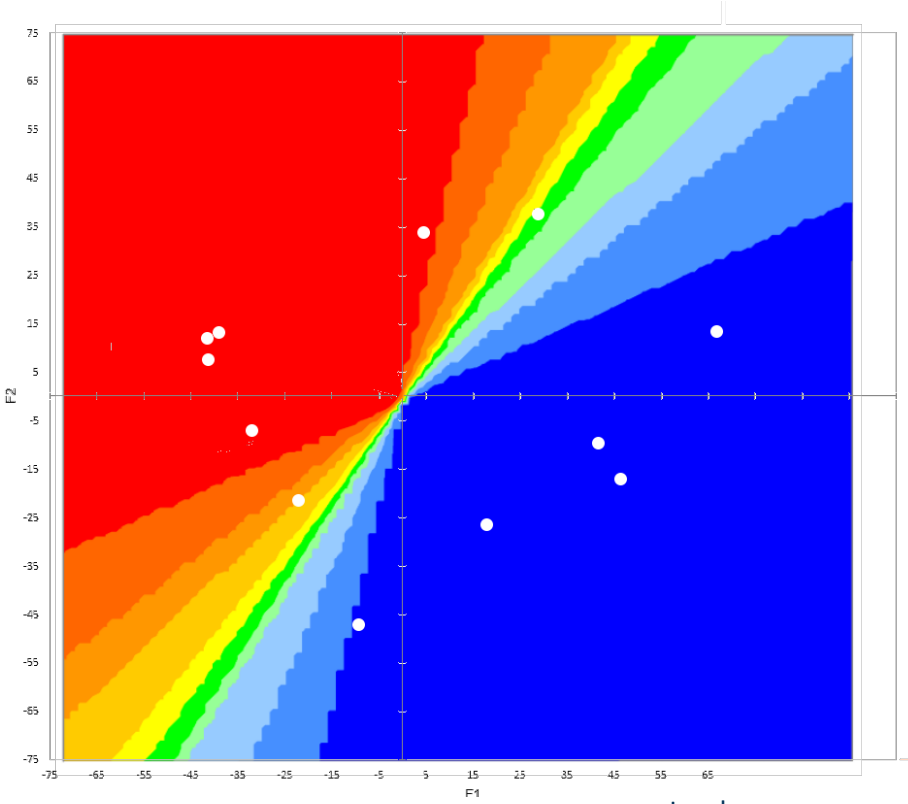
- FCP: Coordinate system and placement of stimuli
- Quality test: colored satisfaction level (low to high)
- Analytic dimensions mapped with hedonic judgements
- Optimum configuration of technical parameters



Non-Experts: Quality Test & PM

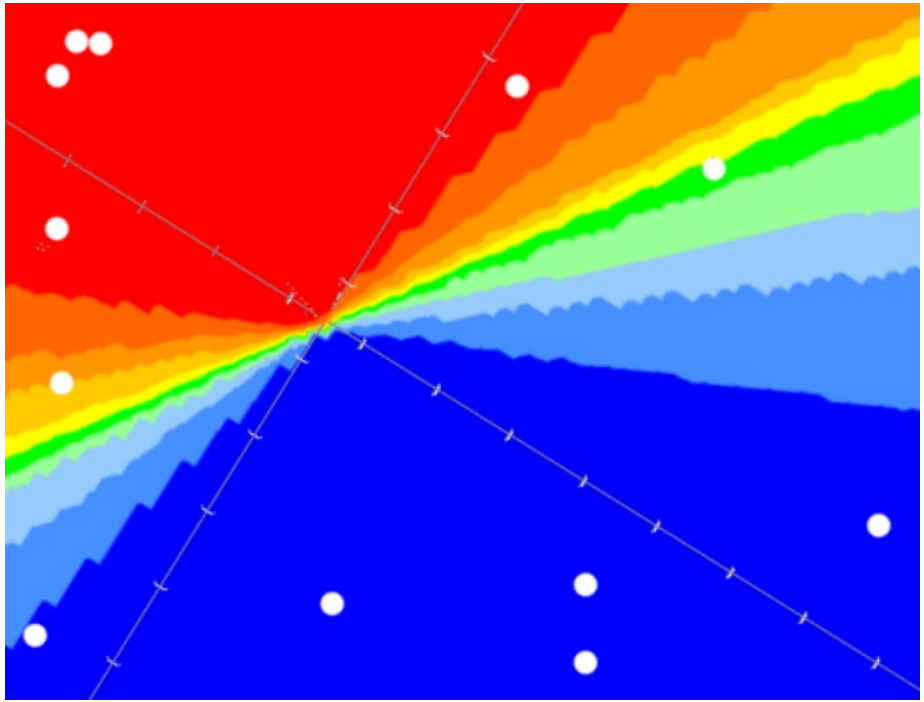
Naive test

- Quality test
  - Paired preference ratings: what part did you prefer?
  - Smaller subset of videos
  - Transferred to ACR ratings via Bradley-Terry model



Landscape scene

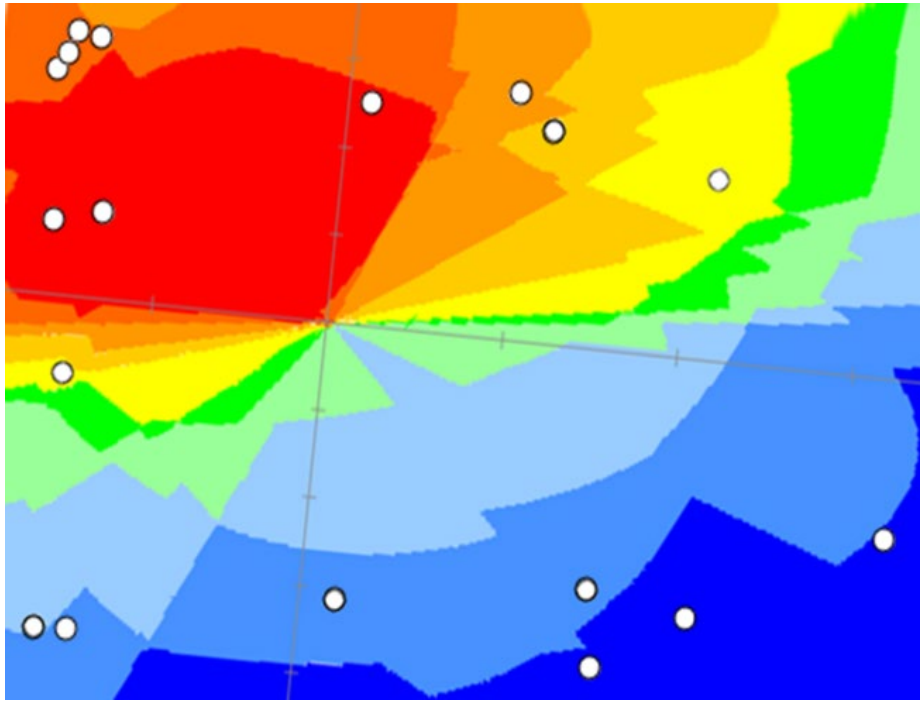
Comparing naïve and expert ratings



Landscape scene

Naive test

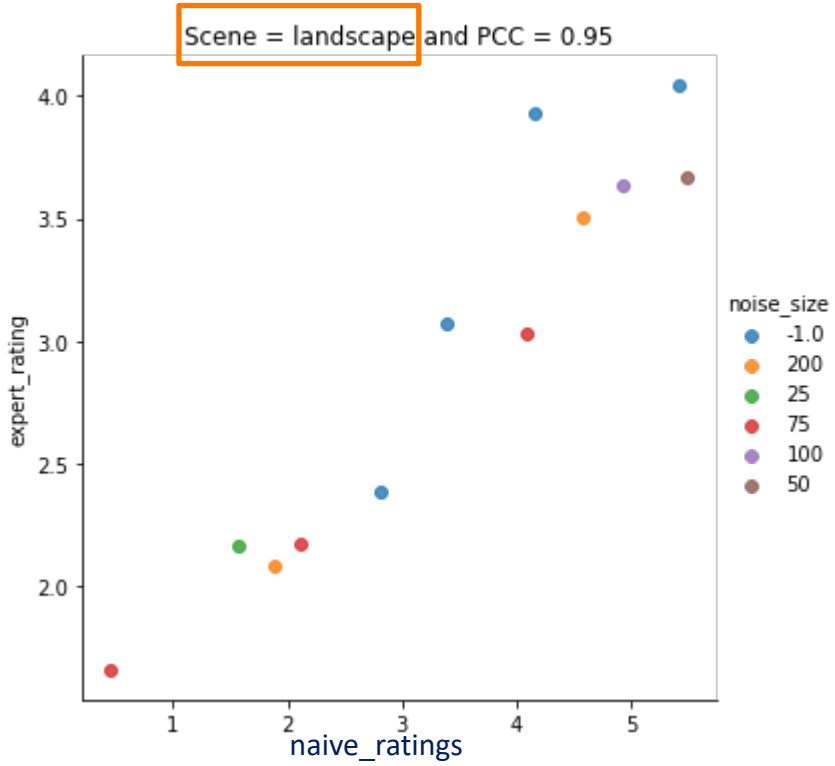
Less stimuli due to pairwise comparison

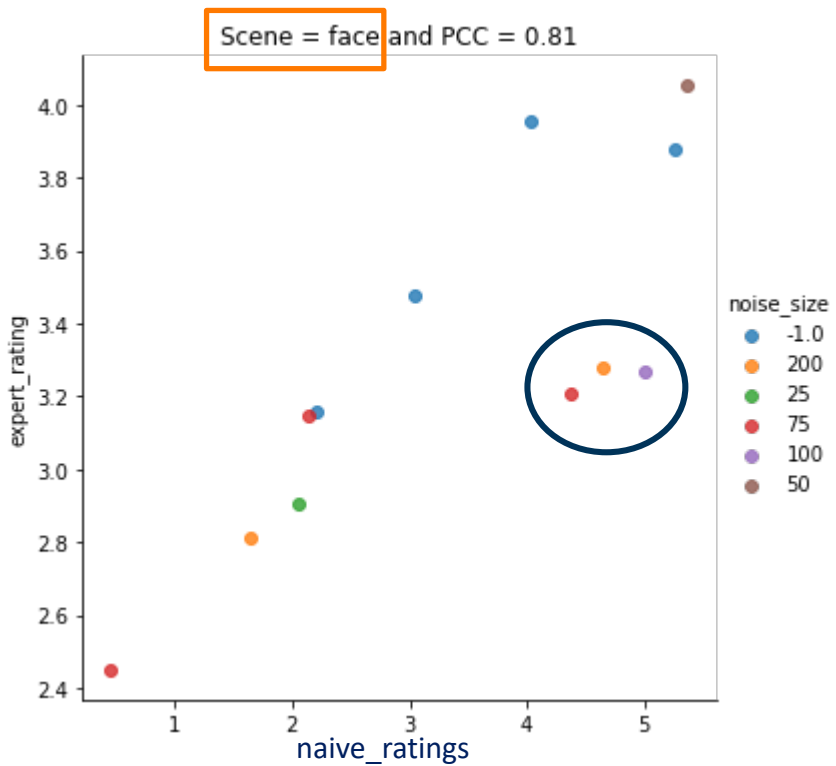


Landscape scene

Expert test

Detailed Comparison Between Scenes





Conclusion and Outlook

- Sensory evaluation methods work well with high quality film material
- Sharpness and noise mostly separable perceptive dimensions
- Expert ratings more detailed and more balanced between dimensions
- Extension
  - Contents
  - Technical effects: optics, color, products
  - Holistic attempt: coding, display

Thank you very much for your attention!

Contact details

dominik.keller@tu-ilmenau.de

ARRI

Audiovisual Technology Group

TECHNISCHE UNIVERSITÄT ILMENAU

Sources

[BechEtAl1996] S Bech et al.: Rapid perceptual image description (RaPID) method

[BechZacharov2006] S Bech, N Zacharov: Perceptual audio evaluation: theory, method and application

[BoddenJekosch1996] M Bodden, U Jekosch: Entwicklung und Durchführung von Tests mit Versuchspersonen zur Verifizierung von Modellen zur Berechnung der Sprachübertragungsqualität

[Carroll1972] J Carroll: Individual differences and multidimensional scaling

[Gower1975] J Gower: Generalized procrustes analysis

[GreenhoffMacFie1994] K Greenhoff, H MacFie: Preference mapping in practice

[LawlessHeymann2013] H Lawless, H Heymann: Sensory evaluation of food: principles and practices

[Lorho2005] G Lorho: Individual vocabulary profiling of spatial enhancement systems for stereo headphone reproduction

[Miller1956] G Miller: The magical number seven, plus or minus two: Some limits on our capacity for processing information

[WierstorfEtAl2013] H Wierstorf, A Raake, M Geier, S Spors: Perception of focused sources in wave field synthesis

[WältermannEtAl2010] M Wältermann, A Raake, S Möller: Quality dimensions of narrowband and wideband speech transmission

[StrohmeierEtAl2010] D Strohmeier, S Jumisko-Pyykkö, K Kunze: Open profiling of quality: A mixed method approach to understanding multimodal quality perception

[KaplanisEtAL2017] N Kaplanis et al.: Perception of reverberation in small rooms: a literature study

[ZangoEtAl2011] Y Zango et al.: Identification of perceptive dimensions of speech and audio codecs subjective quality