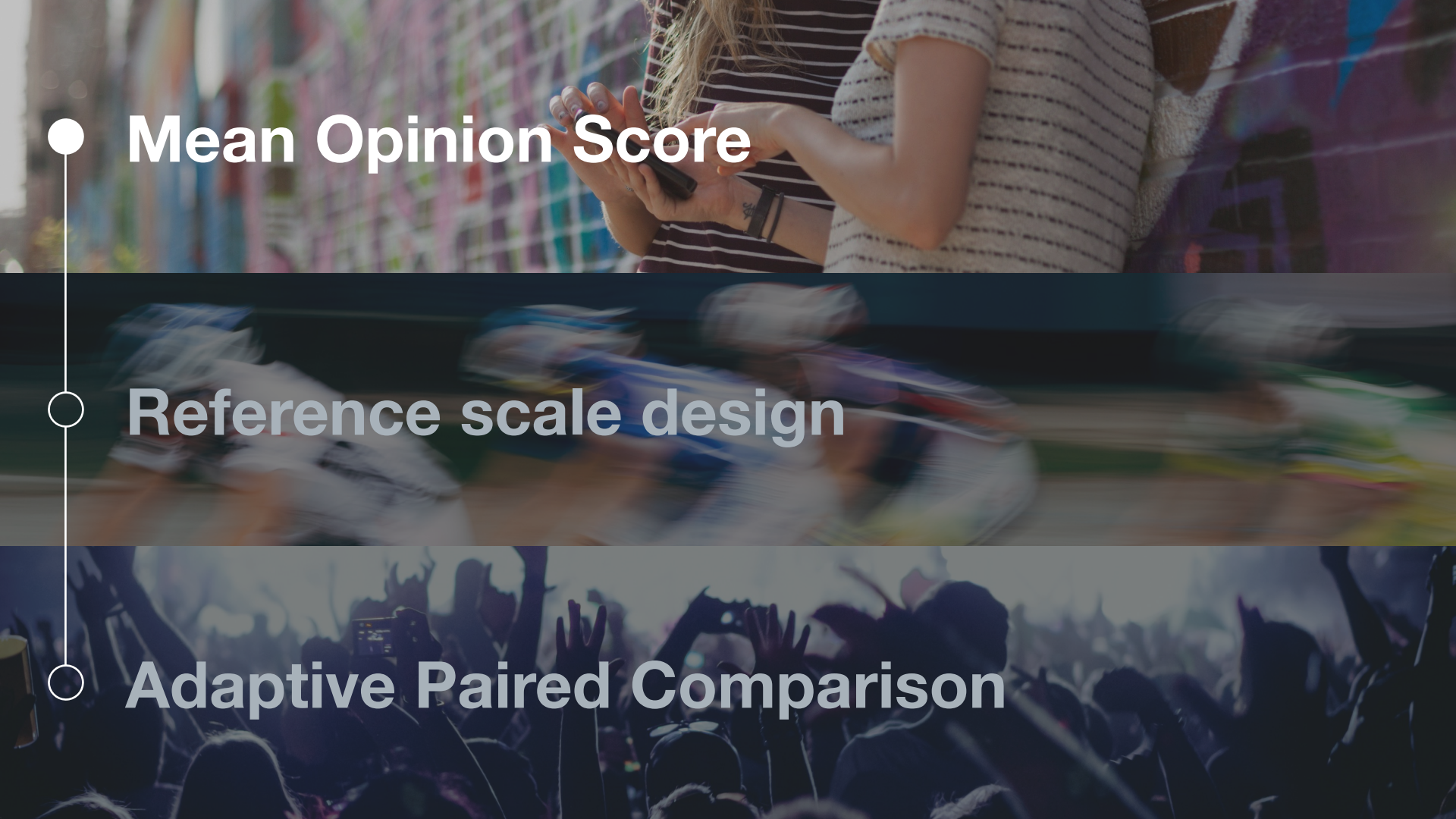
Subjective video quality assessment for mobile device

#MHV2019

@svleuven

How to measure the perceived video quality mobile devices?



- Mean Opinion Score
- Reference scale design
- Adaptive Paired Comparison

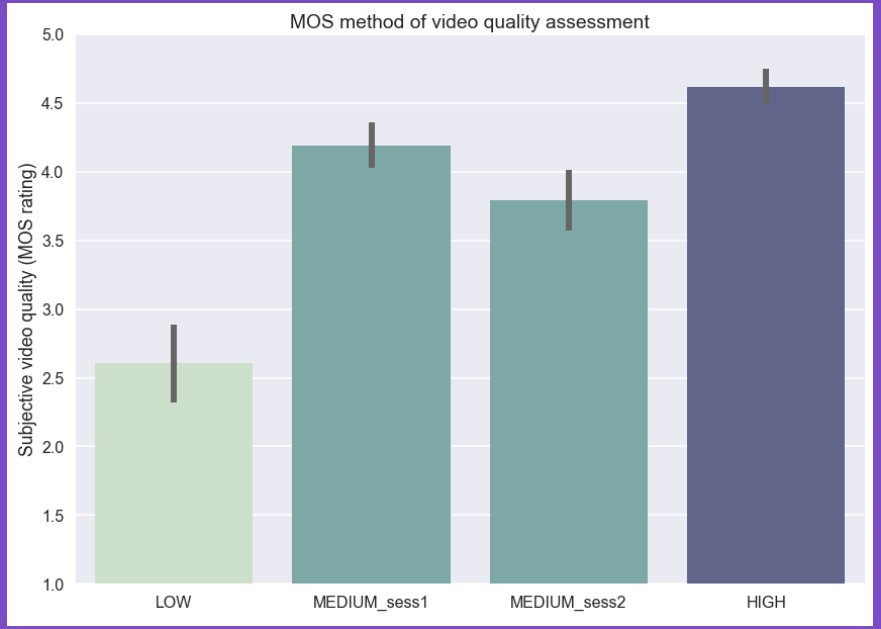
#MeanOpinionScore

Absolute Category Rating

1. Show a random video
2. Rate video on 1-5 scale
3. Repeat

#MeanOpinionScore

- \+ Easiest Method
- \- Inaccurate Results
- \- Inconsistency between tests
- \- Requires a lot of raters
- \- Non-linear results

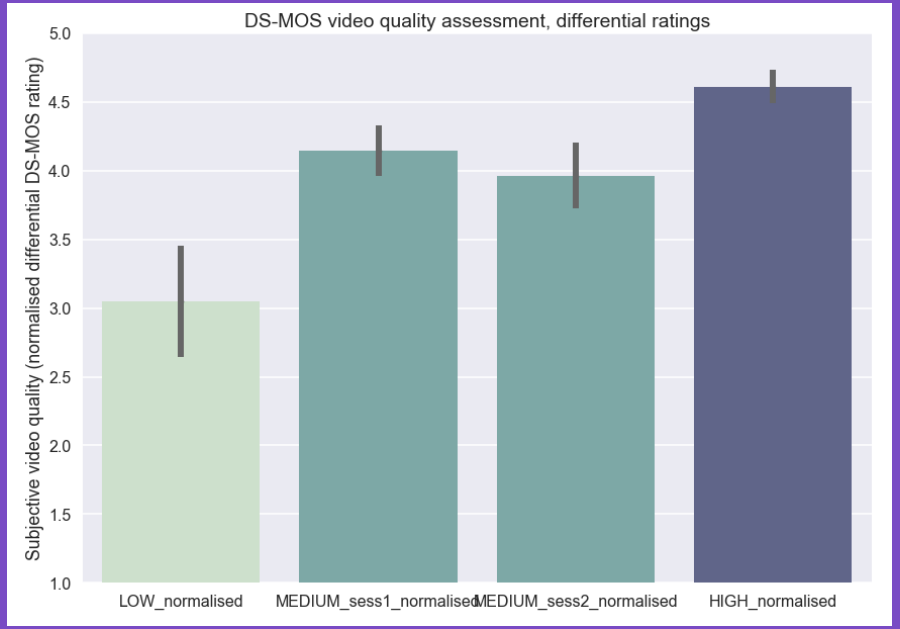


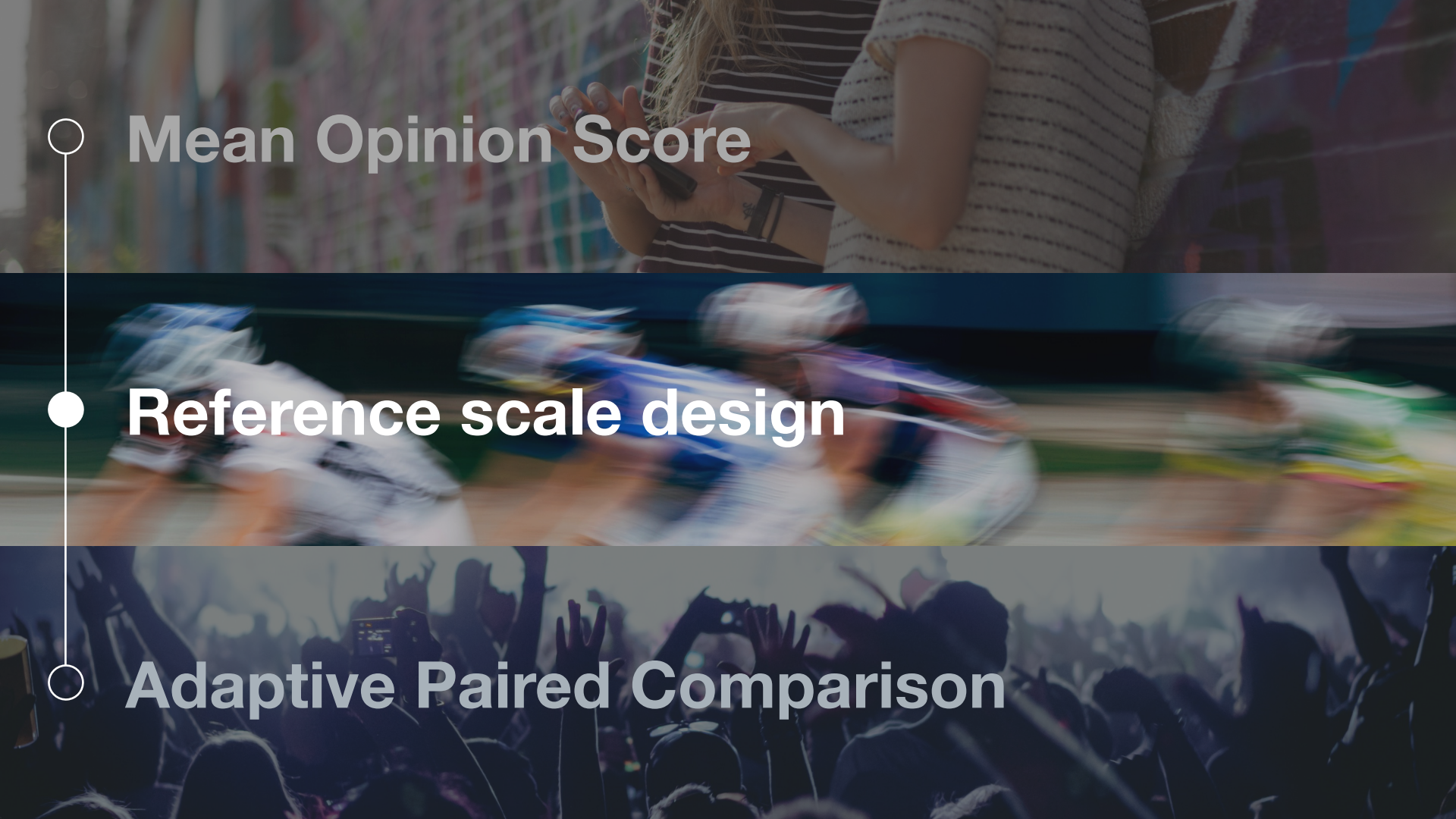
#DoubleStimulusMOS

1. Select a random video
   - Show original version
   - Show impaired version
2. Rate video on 1-5 scale
3. Repeat

#DoubleStimulusMOS

- \+ More consistent than MOS
- \- Inaccurate Results
- \- Not consistent enough
- \- Requires a lot of raters
- \- Non-linear results




Mean Opinion Score
Reference scale design
Adaptive Paired Comparison

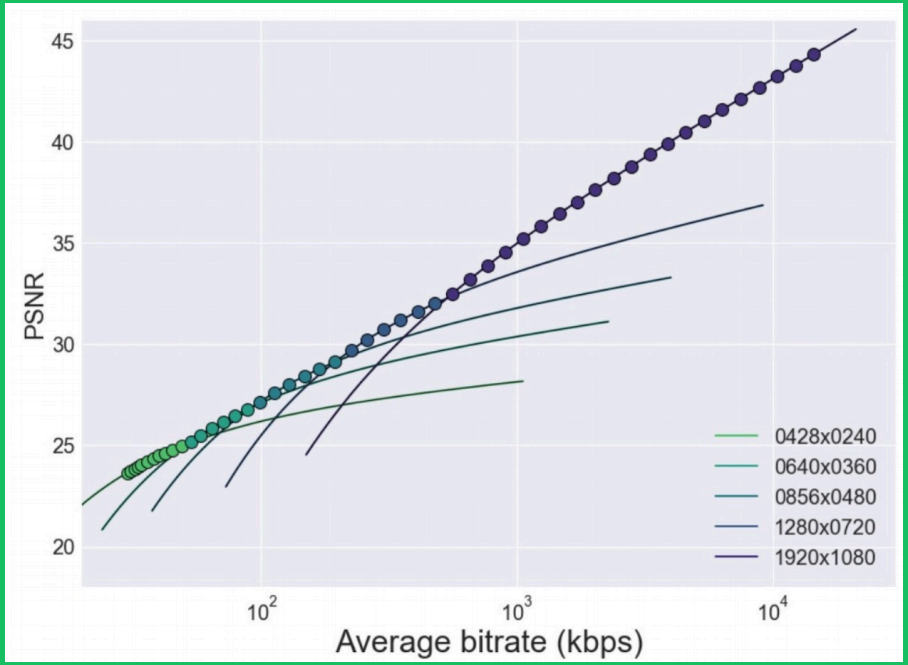
Reference scale design

GOAL: Create a linear monotonically increasing perceptual quality scale

1. Create 50 objectively linearly spaced samples
2. Subjects rate pairs of linear samples
3. Resample based on linear subjective quality

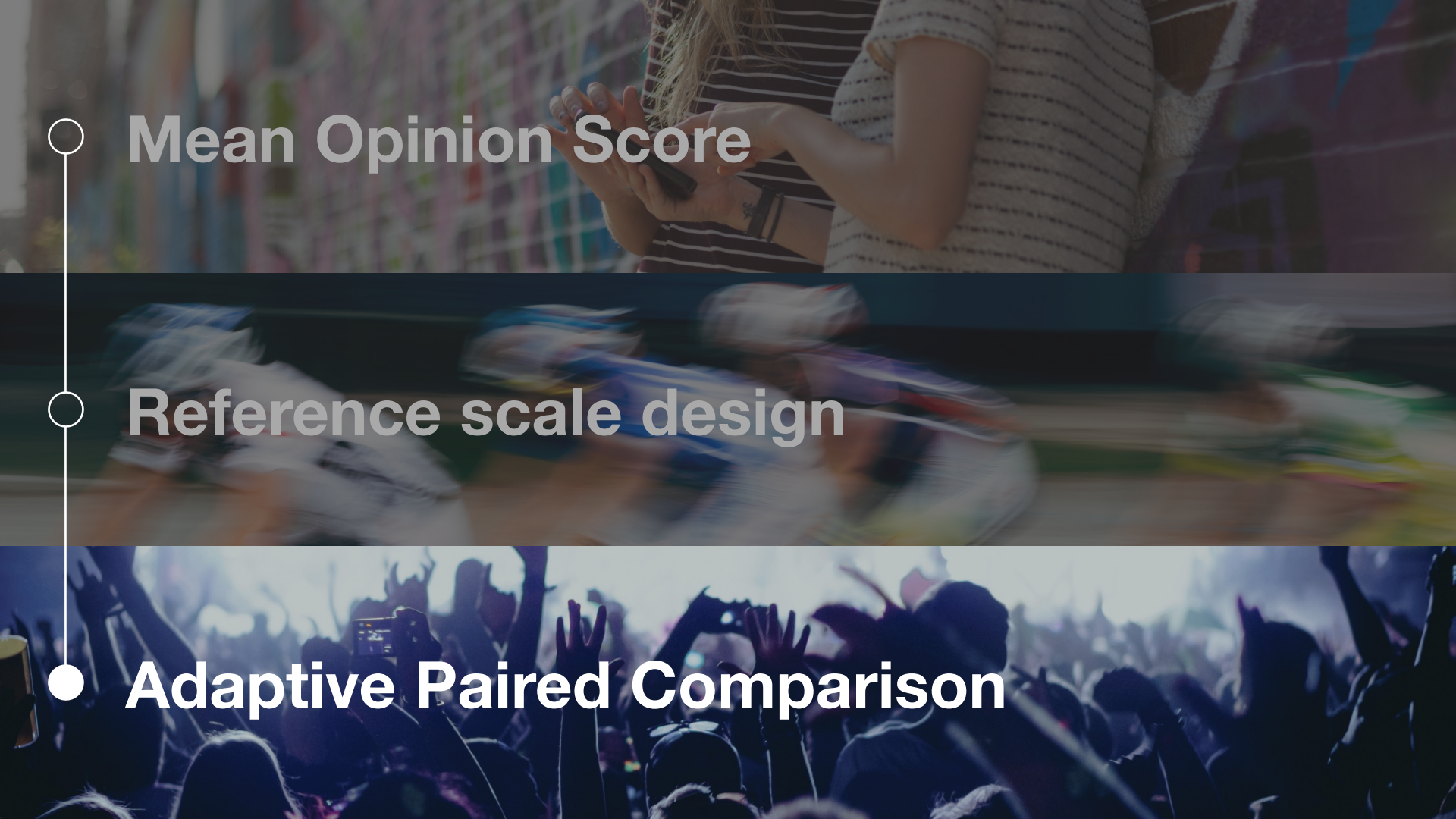
Create initial reference set

1. Encode corpus with different resolution over range of bitrates
2. Create convex hull based on PSNR values
3. Select 50 logarithmically spaced rate points
4. Encode new samples at intermediate locations



Refine reference set

1. Subjects evaluate pairs from the initial reference set
2. Use gradient decent to infer relative perceptual quality
3. Select 50 linearly spaced samples from perceptual quality axis
4. Encode new samples at intermediate locations


Mean Opinion Score
Reference scale design
Adaptive Paired Comparison

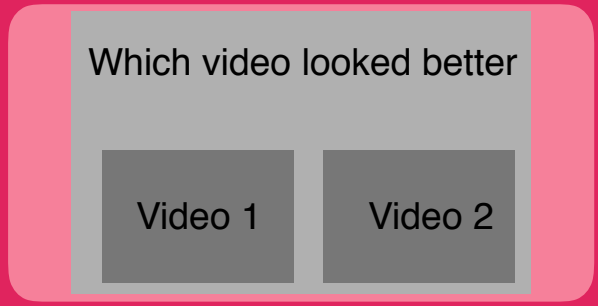
#AdaptivePairedComparison

- Large video corpus
- Show each sample once
- Pair sample and reference
- Adaptively select new reference level

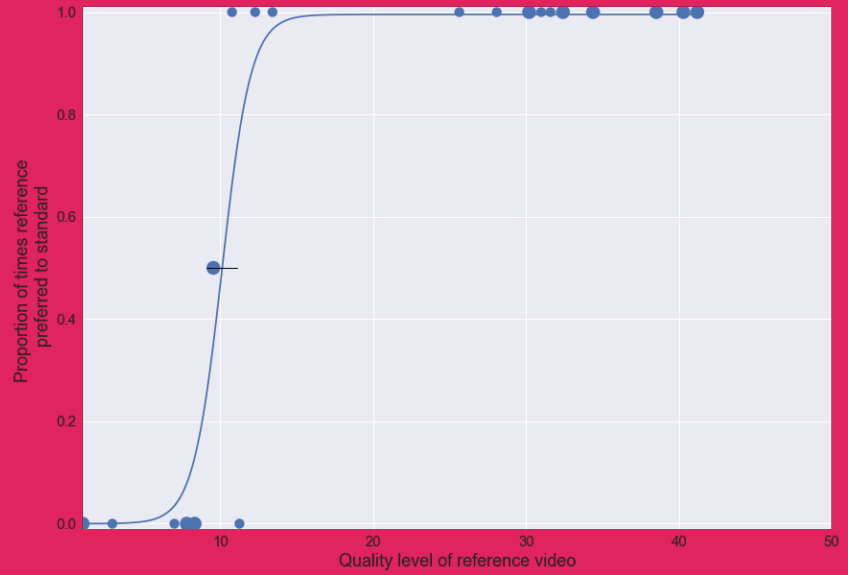
- 32 samples per encoding experiment
- 50 references on a quality scale
    - very bad to extremely good quality

#AdaptivePairedComparison

1. Select a random video
2. Particle filtering algorithm selects a reference quality level
3. Show impaired and reference sample
4. Ask which looks better
5. Update probability distribution over subjective qualities



#AdaptivePairedComparison



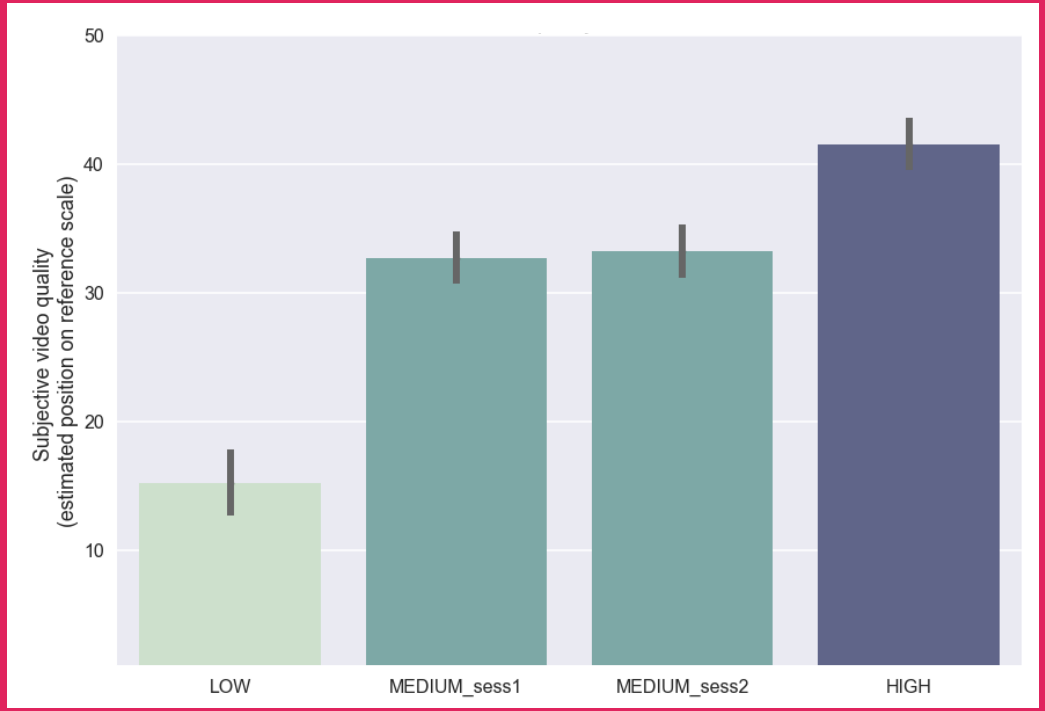
After rating all samples:

Estimate if a reference level looks better

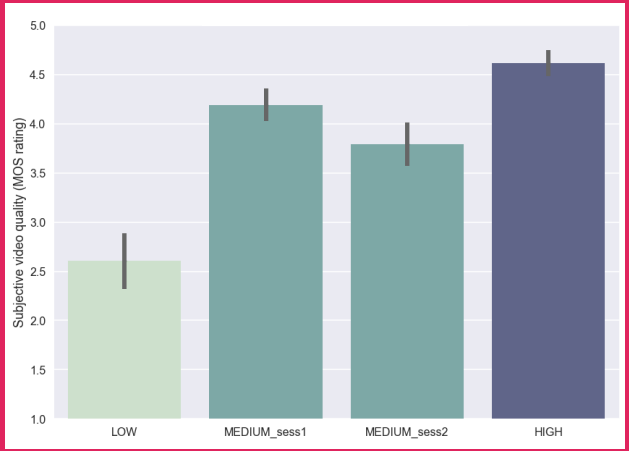
Fit sigmoid to data points

Intersection with 50% probability is the observed quality level

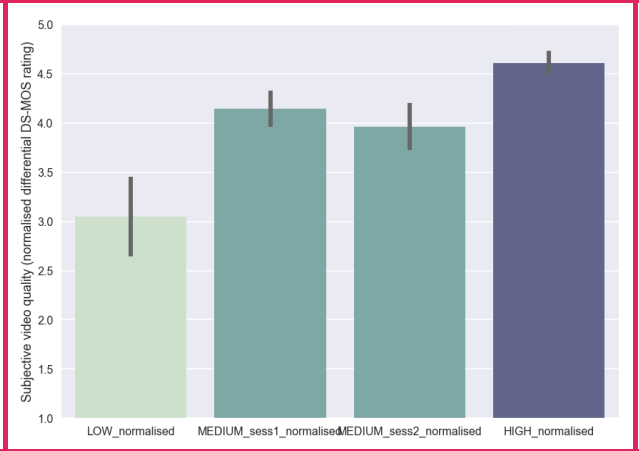
#AdaptivePairedComparison



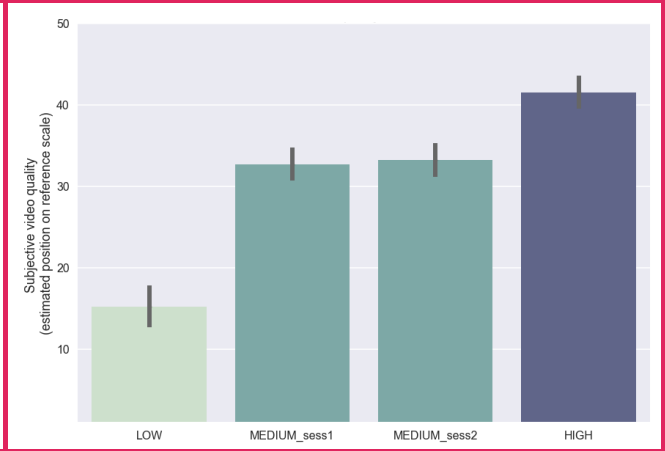
Mean Opinion Score



Double Stimulus - Mean Opinion Score



Adaptive Paired Comparison



Blogpost:

#MHV2019

@svleuven