Towards Context-aware Interactive Quality of Experience Evaluation for audiovisual Multiparty Conferencing

S. Marwin [1], R. Judith [12], C. Pablo [12]

[1] CWI: Centrum Wiskunde & Information

[2] Delft University of Technology

1

Previous Work

- Investigate the impact of **bitrate and packet-loss** on QoE during multiparty video conferencing.

- The authors argue that QoE is not only a result of system factors, but largely depends on **user and context factors (session)**.

- The initial analysis showed that differences between groups play a big role. Yet a closer observation indicates one group seemed to have a different experience.

Study Design

- Organize video conferencing sessions with 4 participants.
- 7 groups with total of 28 subjects (18 female).
- Choose a task require visual interaction, building Lego (ITU –T P.920)

Self-view
shown in highest quality
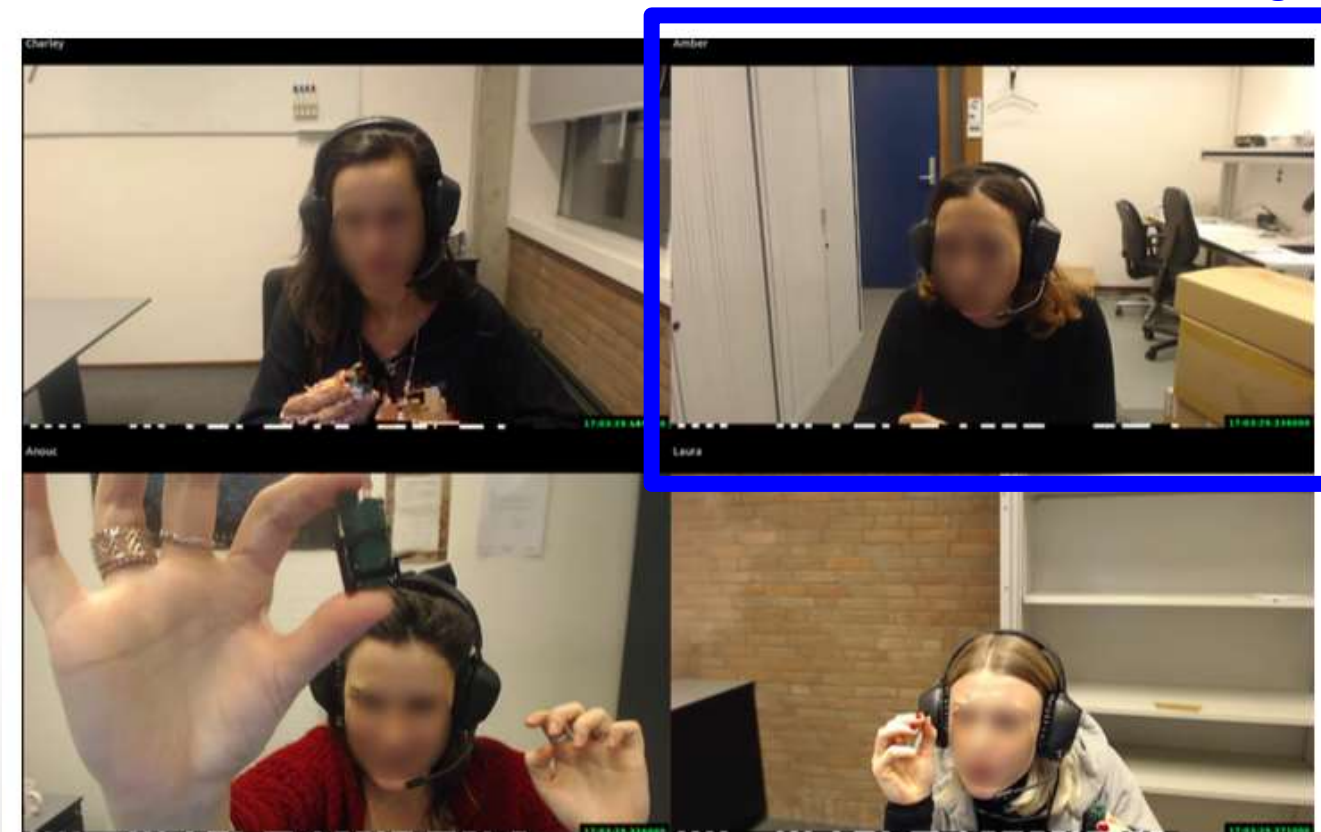
3

Study Design

- Choose bitrate and packet-loss rate as system factors. (bitrate: 256kbps, 1024kbps, 4096kbps; loss: 0%, 0.5%)
- Each group experienced 4 of the 6 possible conditions (counterbalanced).

4

Counterbalancing

- Between subjects vs. within subjects.
  - Within subjects – all participants try all conditions.
  - Between subjects – each participants tries some conditions.
- For between subjects, participants' performance may improve with practice as they progress from one conditions to another.
- To compensate, the order of presenting conditions is **counterbalanced**.
- Participants are divided into groups, and a different order of conditions is used for different group.

5

Counterbalancing

- The order can be govern by Latin Square when there is too many conditions.

- Examples:

3 X 3 Latin Square

A	B	C
B	C	A
C	A	B

4 x 4 Latin Square

A	B	C	D
B	C	D	A
C	D	A	B
D	A	B	C

A row is a test group →

4 x 4 Balanced Latin Square

| A	B	C	D
B	D	A	C
D	C	B	A
C	A	D	B

In a balanced Latin Square each condition both precedes and follows each other condition an equal number of times.

Study Design

- Choose bitrate and packet-loss rate as system factors. (bitrate: 256kbps, 1024kbps, 4096kbps; loss: 0%, 0.5%)

- Each group experienced 4 of the 6 possible conditions (counterbalanced).

- All participants filled in a questionnaire including audio and video quality evaluation questions, conversational dynamics, and enjoyment of task (based on Absolute Category Rating scale).

- Encode video with H.264 and encode audio with AMR narrowband codec. Audio was unimpaired.

Analysis
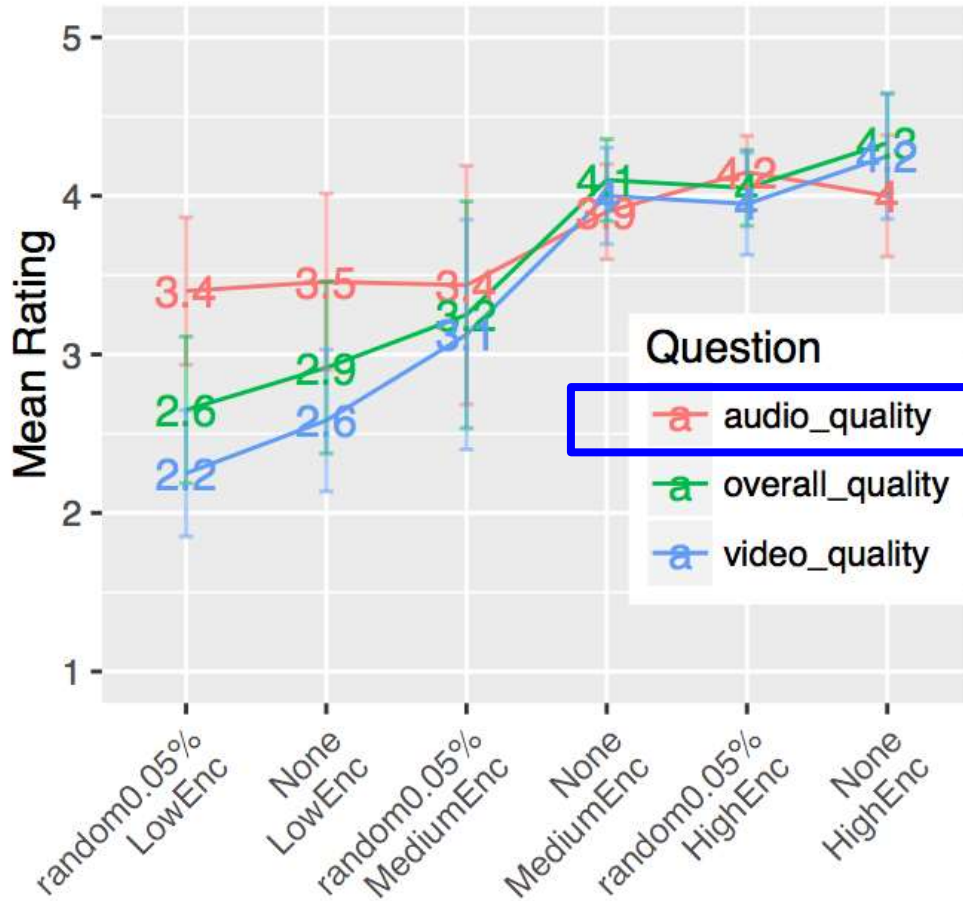
The previously reported analysis showed that the manipulation of video quality had a small effect on audio quality.

Analysis

- Linear mixed effect model is extension of linear regression model for data that are collected and summarized in groups.

- The authors count bitrate and loss as fix effects and test groups and individual participants as random effects.

$$y = X\beta + Z\gamma + \epsilon$$

random effect matrix

dependent variable

fix effect matrix

- (m1) audio quality ~ bitrate + loss + (bitrate | Group/User)

Analysis

- Evaluate the goodness of fit of the models by R^2
  - Marginal R^2 quantifies the explained variance due to the fixed factor (R^2 : 8.45%).
  - Conditional R^2 quantifies the explained variance considering the random effects (R^2 : 73.69%).

- The results point out that most of the ratings variance could be explained by the **characteristics of the individual user**.
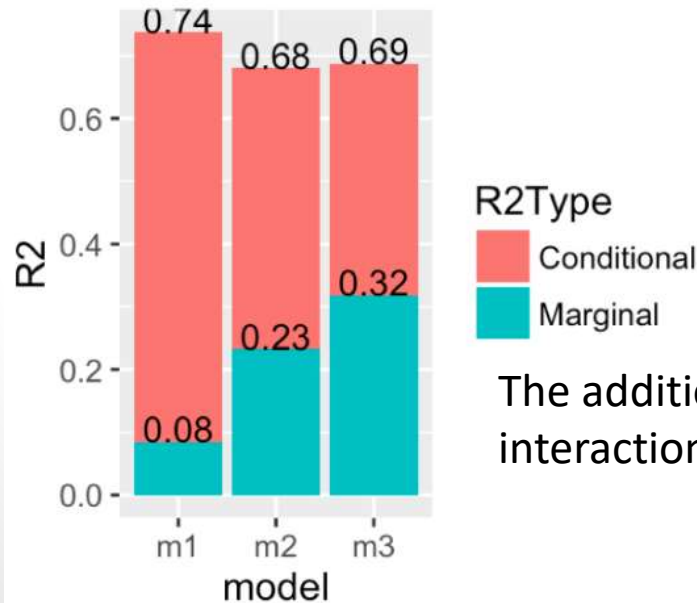
Individual QoE

- Clustered participants according to their average audio quality rating.
- An elbow-plot reveals that 2 clusters give the best ratio of explained variance to number of clusters.

11
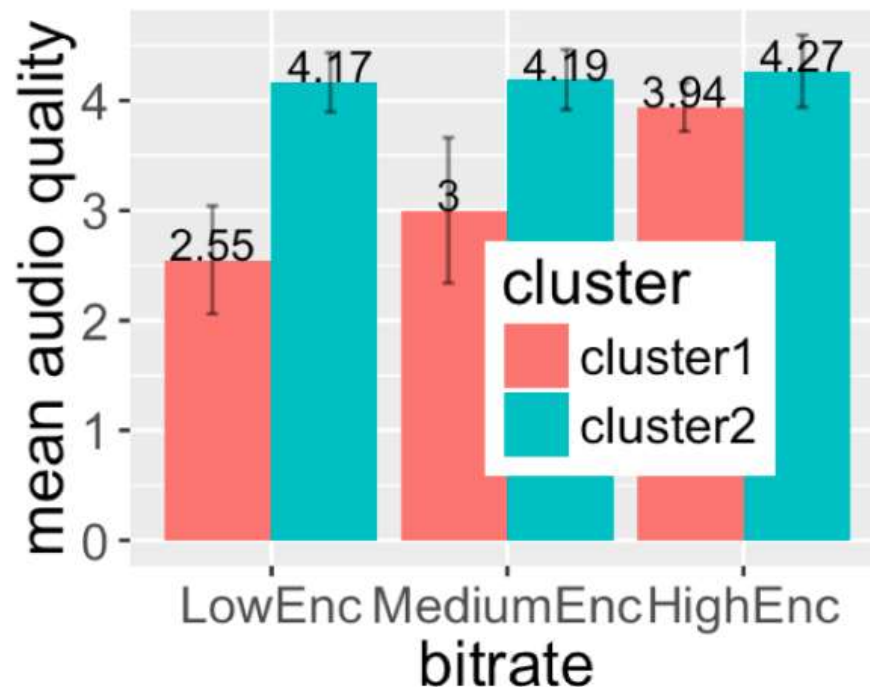
Individual QoE

- (m2) audio quality ~ (bitrate + loss) +cluster+(bitrate| Group/User)

- (m3) audio quality ~ (bitrate + loss) *cluster+(bitrate| Group/User)

- Use Likelihood Ratio Test (LRT) to compare for two models if the improvement of fit of the model.

The addition of cluster fixed factor in interaction with other two improves R^2.

Individual QoE

- The overall difference between the two clusters is significant (p < 0.001).
- The paired comparisons reveal that the difference gets stronger as the quality degradations get stronger.

| cluster	Encoding			Loss	
	Low-High	Low – Medium	Medium - High	None –0.5%	
cluster1	<0.01	0.18	0.67	0.99	
cluster2	0.92	0.99	0.92	0.93	
	Low	Medium	High	None	0.5%
Cluster1-Cluster2	<0.001	0.13	0.69	<0.001	<0.001

Individual QoE

- In the plot of perceived video quality, the author observes that the cluster 1 participants also rate the quality more critical than cluster 2 participants.
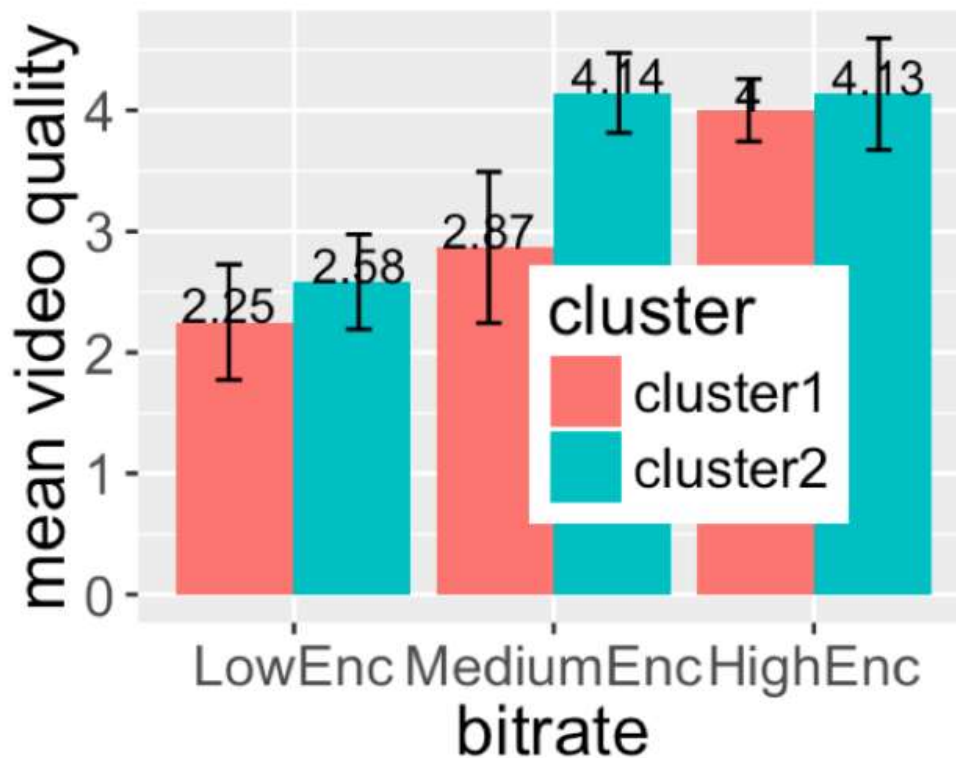
Individual QoE

- Segment the audio data in on-off speaking pattern.
- The author found the difference in the average time participants were involved in two or more people speaking at the same time (double talk).
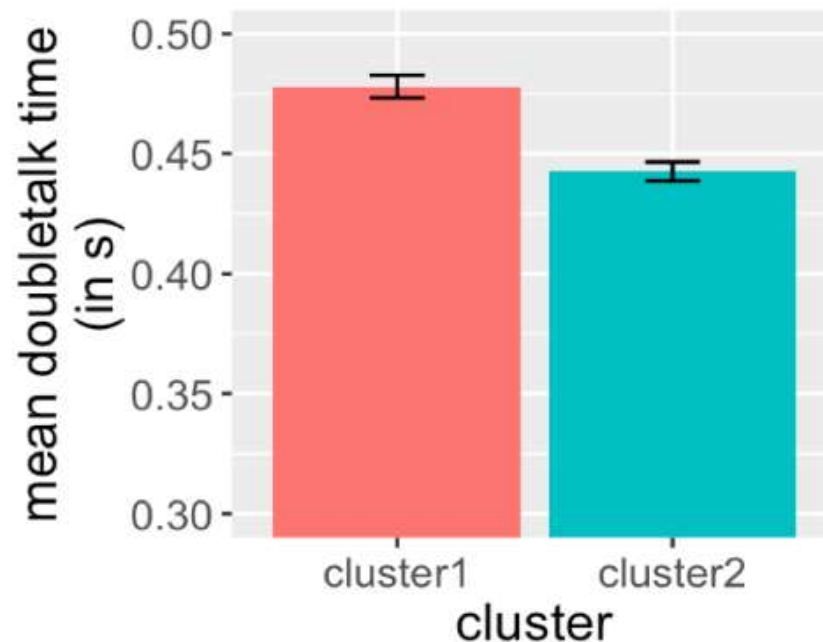
Individual QoE

- The test shows a significant difference in enjoyment of the study and in the rating of their own video quality (which is unimpaired during the whole experiment).

Table 3 P-values of Wilcoxon Rank Sum Test for the final questionnaire on 5 point likert-like scale (end labels in parenthesis)

Question	p-value
In enjoyed participating in this study (**enjoyment; Not at all –very much**)	<0.01
I liked the task of playing with Lego. (**likelego; Not at all –very much**)	0.63
How would you rate the quality of your own video? (**ownvideo; bad –excellent**)	<0.01
I noticed delay in the connection and it was: (**delay;very annoying – imperceptible**)	0.1
Did you have problems determining which participant was speaking? (**problemsspeaking; Never-very often**)	0.1
I am very experienced in using video-conferencing systems. (**priorexp; Very unexperienced-Very experienced**)	0.37
Age	0.61

16

Conclusion

- The analysis of the perceived audio quality showed users could be differential into groups.

- User factors are important, service that can gather long-term information about users would be able to create better services, personalizing delivery strategies.