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

S. Marwin ¹, R. Judith¹², C. Pablo¹²

¹ CWI: Centrum Wiskunde & Information
² Delft University of Technology

PQS 2016 5th ISCA/DEGA Workshop on Perceptual Quality of System
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
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).
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.
Counterbalancing

• The order can be govern by Latin Square when there is too many conditions.
• Examples:

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.
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.

\[ y = X\beta + Z\gamma + \epsilon \]

• \((m1)\) audio quality \(\sim\) bitrate + loss + (bitrate \mid \text{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.
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 $R^2$. 

![Graph showing $R^2$ values for m1, m2, and m3 models](image)
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.

<table>
<thead>
<tr>
<th>cluster</th>
<th>Encoding</th>
<th>Loss</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low-High</td>
<td>Low - Medium</td>
</tr>
<tr>
<td>cluster1</td>
<td>&lt;0.01</td>
<td>0.18</td>
</tr>
<tr>
<td>cluster2</td>
<td>0.92</td>
<td>0.99</td>
</tr>
<tr>
<td>Cluster1-Cluster2</td>
<td>&lt;0.001</td>
<td>0.13</td>
</tr>
</tbody>
</table>


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)

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