Recent results on adaptive streaming

Jacob Søgaard, Jeevan Pokhrel, Muhammad Shahid, Kun Wang, Samira Tavakoli and Kjell Brunnström
OBJECTIVE

• Performing a **subjective** experiment to evaluate the Quality of Experience of adaptive video streaming

• Evaluate crowdsourcing as a method for performing subjective experiments
STUDY FACTORS

• Adaptive streaming by decreasing the video bitrate and rebuffering events

• Constant bitrate
  o H.264/AVC
  o Video Bitrate: 5Mbps, 3Mbps, 1Mbps, 600kbps

• Adaptation event
  o Rapid decreasing quality
  o Gradual decreasing quality

• Rebuffering events
  o 1 or 2 rebuffering  total time 2 sec
# VIDEO CONTENT TYPES

Source videos in 1280x720, 24 fps

<table>
<thead>
<tr>
<th>Video</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Movie 1</td>
<td>Action, some scene in smooth motion, some with groups of walking people, some camera panning</td>
</tr>
<tr>
<td>Movie 2</td>
<td>Drama, mostly with smooth motion in static background, some scene with groups of dancing people in bright ambient light</td>
</tr>
<tr>
<td>Movie 3</td>
<td>Action, Sci Fi, with cloudy skies</td>
</tr>
<tr>
<td>Documentary</td>
<td>Sport Doc., mostly with handheld camera</td>
</tr>
<tr>
<td>Sport</td>
<td>Soccer, average motion, with uniform camera panning</td>
</tr>
<tr>
<td>News</td>
<td>Spanish news, mostly static camera</td>
</tr>
<tr>
<td>Music</td>
<td>Music concert, extensive movement of the singer</td>
</tr>
</tbody>
</table>
# ADAPTATION SCENARIOS

<table>
<thead>
<tr>
<th>Code</th>
<th>Characteristic of scenarios</th>
</tr>
</thead>
<tbody>
<tr>
<td>DGR2</td>
<td>Gradually decreasing the quality (5-3-1-600) with 2 seconds chunk</td>
</tr>
<tr>
<td>DRP2</td>
<td>Rapid decreasing the quality (5-600) with 2 seconds chunk</td>
</tr>
<tr>
<td>1F3M</td>
<td>1 Freezing event for 2 seconds in the constant 3 Mbps video</td>
</tr>
<tr>
<td>2F3M</td>
<td>2 Freezing events for 1 second each in the constant 3 Mbps video</td>
</tr>
<tr>
<td>1F1M</td>
<td>1 Freezing event for 2 seconds in the constant 1Mbps video</td>
</tr>
<tr>
<td>N5</td>
<td>No degradation- The whole segment at 5Mbps</td>
</tr>
<tr>
<td>N3</td>
<td>No degradation- The whole segment at 3Mbps</td>
</tr>
<tr>
<td>N1</td>
<td>No degradation- The whole segment at 1Mbps</td>
</tr>
<tr>
<td>N600</td>
<td>No degradation- The whole segment at 600kbps</td>
</tr>
</tbody>
</table>
TEST METHODOLOGY

- Crowdsourcing
  - Put out work assignments on microworker.com
  - 215 microworker in total
  - Task about 10 min
  - Payment 1 $

- Pair comparison
  - Compare video clips pairwise (ITU-T Rec. P.910)
  - Simple task for subject: “Which one do you prefer?”
  - Control question was inserted
  - Model exist to convert into quality scale (Bradley-Terry)
  - Number of combination grows quickly (Optimized square design)
  - 126 pairs in total, split into 14 different set of 9 video pairs with 3 contents and 3 degradations
  - Screentests was performed initially
RESULTS

- First restricted to workers from US, UK, Canada, Australia, New Zealand
  - Very slow progress
- Restarted test for the whole world
  - Very fast progress
RESULTS – COMPARISON WITH LAB TEST

Linear Correlation: 0.75

- Fit
- 45-degree
- ClosetoEdge
- Darkhour
- Football
- News
- Pirates
- Rollingstone
- Streetdance

Crowdsourcing scores vs. Lab MOS

ACREO
RESULTS – QUALITY VS BITRATE
CONCLUSIONS

• Crowdsourcing test highly correlated to lab test
• Rebuffering lower quality in general
• Number of rebuffering events important
• 1 rebuffering at 3 Mbps same quality as the adaptive scenarios
CONTENT CHARACTERIZATION

• The Lab test subjective design was not full-matrix
• How could then the results be understood?
• One way would be to characterize the PVS
ADDITIONAL ANALYSIS OF LAB TEST

• Content characterization
  • Analysis of spatial- and temporal activity in the video (ITU-T Rec. P.910)

\[ SI = \max_{time} [\text{std}_{\text{space}}[\text{sobel}(f_n)]] \]
\[ TI = \max_{time} [\text{std}_{\text{space}}[M_n(i, j)]] \]

• Divide into 4 cases:
  • Low Spatial – Low Temporal (LS-LT)
  • High Spatial – High Temporal (HS-HT)
  • High Spatial – Low Temporal (HS-LT)
  • Low Spatial – High Temporal (LS-HT)
RESULTS – CONTENT CHARACTERISTICS

The graph illustrates the MOS (Mean Opinion Score) for different scenarios, categorized as increasing and decreasing. The x-axis represents the scenarios (LS-LT, HS-HT, HS-LT, LS-HT), and the y-axis represents the MOS. The data points are indicated with error bars to show variability. The increasing scenarios show a trend with MOS values ranging from 3 to 4, while the decreasing scenarios show a lower trend with MOS values ranging from 2 to 3.
RESULTS – CONTENT CHARACTERISTICS

![Graph showing MOS values for HS-HT and HS-LT conditions with different content characteristics.](image)
CONCLUSIONS

• High Spatial – Low Temporal (HS-LT) has significantly lower quality than other contents
DEMO – "STREAMINGKOLLEN"

- Service for the user to check their connection for media consumption
- QoE estimation:
  - Pause intensity (PI) model
    - Measures the durations of pause periods due to an empty buffer in relation to play periods and maps this to user QoE according to a specific formula.
  - Linear bitrate model
    - Measures changes in bitrates and maps this to user QoE according to a specific formula. The formula is based on the average and standard deviation of the bitrate during the video streaming.
ACCEPTED PAPERS TO QOMEX 2014 IN SINGAPORE

• CROWDSOURCING BASED SUBJECTIVE QUALITY ASSESSMENT OF ADAPTIVE VIDEO STREAMING
  • Muhammed Shahid, Jacob Søgaard, Jeevan Pokhrel, Kun Wang, Kjell Brunnström, Samira Tavakoli

• EFFECT OF CONTENT CHARACTERISTICS ON QUALITY OF EXPERIENCE OF ADAPTIVE STREAMING
  • Samira Tavakoli, Muhammed Shahid, Kjell Brunnström,