Human Speech Intelligibility Measurements over VoIP Channels

Laura Fernández Gallardo

Quality and Usability Lab, Technische Universität Berlin, Germany
Intelligibility of transmitted speech
Intelligibility of transmitted speech
Intelligibility of transmitted speech

• Previous studies
  • Benefits of extended bandwidths shown
  • Intelligibility tests:
    • [Spiegel et al., 1990][Teng and Kubichek, 2006]: NB codecs only
    • [Jokinen et al., 2015]: 16 kHz speech only
    • [Fernández and Möller, 2015]: G.722 (WB) higher intell than AMR-NB

• Contributions of this work
  • Effects of 23 channel degradations: NB, WB, SWB
  • Comparing to objective measures
    • POLQA-intelligibility predictions
    • POLQA MOS
Outline

- Motivation
- Speech material
- Intelligibility test
- Subjective vs. objective intelligibility
- Conclusions
Speech material

- 4 speakers (2m, 2f)
- 8 vowel-consonant-vowel logatomes
  - "ama"
  - "aba"
  - "afa"
  - "ana"
  - "apa"
  - "asa"
  - "awa"
  - "ascha"
Speech material
## Speech material

<table>
<thead>
<tr>
<th>Bandwidth</th>
<th>Codec</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct (fs=8 kHz)</td>
<td>No</td>
</tr>
<tr>
<td>NB</td>
<td>G.712 filter</td>
</tr>
<tr>
<td>NB</td>
<td>G711 A-law (64)</td>
</tr>
<tr>
<td>NB</td>
<td>G711 A-law (64) + IRS</td>
</tr>
<tr>
<td>NB</td>
<td>G723.1 (6.3)</td>
</tr>
<tr>
<td>NB</td>
<td>GSMEFR (12.2)</td>
</tr>
<tr>
<td>NB</td>
<td>AMRNB (4.75)</td>
</tr>
<tr>
<td>NB</td>
<td>AMRNB (12.2)</td>
</tr>
<tr>
<td>NB</td>
<td>Speex (2.15)</td>
</tr>
<tr>
<td>NB</td>
<td>Speex (11)</td>
</tr>
<tr>
<td>NB</td>
<td>Speex (24.6)</td>
</tr>
<tr>
<td>Direct (fs=16 kHz)</td>
<td>No</td>
</tr>
<tr>
<td>WB</td>
<td>P.341 filter</td>
</tr>
<tr>
<td>WB</td>
<td>G722 (64)</td>
</tr>
<tr>
<td>WB</td>
<td>AMRWB (12.65)</td>
</tr>
<tr>
<td>WB</td>
<td>AMRWB (23.05)</td>
</tr>
<tr>
<td>WB</td>
<td>Speex (3.95)</td>
</tr>
<tr>
<td>WB</td>
<td>Speex (23.8)</td>
</tr>
<tr>
<td>WB</td>
<td>Speex (42.2)</td>
</tr>
<tr>
<td>Direct (fs=32 kHz)</td>
<td>No</td>
</tr>
<tr>
<td>SWB</td>
<td>14KBP filter</td>
</tr>
<tr>
<td>SWB</td>
<td>G722.1C (24)</td>
</tr>
<tr>
<td>SWB</td>
<td>G722.1C (48)</td>
</tr>
<tr>
<td>SWB</td>
<td>EVS (24.4)</td>
</tr>
<tr>
<td>SWB</td>
<td>EVS (48)</td>
</tr>
<tr>
<td>SWB</td>
<td>Opus (160)</td>
</tr>
<tr>
<td>Direct (fs=48 kHz)</td>
<td>No</td>
</tr>
</tbody>
</table>
Outline

• Motivation
• Speech material
• Intelligibility test
• Subjective vs. objective intelligibility
• Conclusions
Intelligibility test

- 30 listeners (15m, 15f), normal-hearing, German
- 54m² acoustically-treated room
- Shure SHR240 headphones
  - freq.: 20-20,000 Hz
  - diotic listening
- 864 stimuli
- Duration: one hour
Intelligibility test
Intelligibility test

[Graph showing logatome recognition performance across different channel conditions]
Intelligibility test

- Better performance with higher bandwidth and bitrate

- WB significant over NB?
  - only for WB codecs at high bitrate and NB codecs at low bitrate
  - Speex-WB@42.2, G.722@64, P.341 filter better than NB@ ≤11

- SWB significant over WB?
  - No, except for Speex-WB@3.95

- SWB significant over NB?
  - Not always
  - G.722.1C@24 and @48 only better than Speex-NB@2.15
  - 14 KPB filter same as G.711@64 and G.712 filter
Intelligibility test

• Logatome confusions
  • “aba”—“awa” -> prevail in SWB and in “Direct” with fs ≥ 16 kHz
  • “awa”—“aba” -> reduced from NB to WB
  • “afa”—“asa” -> reduced from NB to WB
Outline

• Motivation
• Speech material
• Intelligibility test
• Subjective vs. objective intelligibility
• Conclusions
Objective intelligibility

- POLQA-intelligibility model (V1490intellV2)
- POLQA standard (V2.4.1)

- 2nd-order curves can be fitted to subjective vs. objective intelligibility
  - slightly better fit with POLQA-intelligibility as predictor
2nd-order fits

R2 = 0.870, RMSE = 2.10

R2 = 0.858, RMSE = 2.20
Outline

• Motivation
• Speech material
• Intelligibility test
• Subjective vs. objective intelligibility
• Conclusions
Conclusions

• Intelligibility test
  • closed-response with 8 VCV logatomes
  • 23 channel degradations: NB, WB, SWB
    • + “Direct” with fs = 8, 16, 32, and 48 kHz

• Greater gain in intelligibility for the transition NB → WB
• No statistical differences between WB and SWB (except for Speex-WB@3.95)

• Strong quadratic correspondences between subjective and objective intelligibility
Thank you for your attention!

Laura Fernández Gallardo, Ph.D.
laura.fernandezgallardo@tu-berlin.de
Questions?

Laura Fernández Gallardo, Ph.D.
laura.fernandezgallardo@tu-berlin.de