MODELLING SPEECH QUALITY FOR NARROWBAND AND WIDEBAND SILK CODEC FOR VOIP APPLICATIONS

Mohammad Goudarzi, Lingfen Sun, Emmanuel Ifeachor
University of Plymouth
School of Computing and Mathematics
Overview

- Introduction
- SILK Codec
- Overview of speech quality models
- Testbed setup
- SILK quality modelling based on E-model
- Model’s performance
- Subjective evaluation of the model
- Summary
Introduction

- VoIP networks, Skype
- Factors affect speech quality
- Proprietary software, open-source codec
- How to measure speech quality
SILK Codec

- Speech codec and compression format developed by Skype
  - Supported by Android and iPhone
  - Expected to be incorporated into softphones, cordless phones, and mobile devices
  - Freely available for third party users

- Highly scalable codec
  - Network bandwidth
  - Send bitrate
  - Complexity

<table>
<thead>
<tr>
<th>Mode</th>
<th>Fs (Hz)</th>
<th>BR(kbps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Narrowband</td>
<td>8000</td>
<td>6 – 20</td>
</tr>
<tr>
<td>Mediumband</td>
<td>12000</td>
<td>7 – 25</td>
</tr>
<tr>
<td>Wideband</td>
<td>16000</td>
<td>8 – 30</td>
</tr>
<tr>
<td>Super Wideband</td>
<td>24000</td>
<td>12 – 40</td>
</tr>
</tbody>
</table>
Overview of measurement methods

- Subjective methods
- Objective models
  - Computational models
    - E-model
    - Regression based models
    - Machine learning
  - Signal-based models
    - PESQ (ITU-T P862) and PESQ-WB
    - POLQA (ITU-T P.863)
Speech quality measurement models

- Measure the end-to-end quality (MOS)
- Impact of network conditions (i.e., packet loss, delay, etc.)
- Maintain a good service quality during the call

- Application level adaptation
  - Bitrate
  - Sampling frequency
  - Packetization
  - Inband FEC

- Online monitoring and control
Aims

- Low complexity, reliable measurement model to quantify the speech quality for SILK codec
- The model can be used in adaptation, FEC tuning and user satisfaction measurement
- Use both application level and network level parameters to predict the speech quality of VoIP calls.
- Can have direct applications in monitoring and adaptation of VoIP applications
Testbed setup

- Packet loss (0% to 30% with 5% intervals)
- 16 British English speech samples (ITU-T P.50)
  - 8 male and 8 female
- 15 different initial seeds
Modelling based on E-model

_E-model (ITU-T G.107):_

\[ R = R_o - Id - Ie + A \]

- \( R_o \): basic signal-to-noise ratio (SNR)
  - 93.2 for NB
  - 129 for WB
- \( Id \): is the impairments caused by delay factors
- \( Ie \): includes distortions by codec and packet loss
- \( A \): Advantage factor.

\[ R = R_o - Ie(\text{bitrate, loss}) \]
Modelling $I_{e\text{.eff}}$

$$R = 3.026MOS^3 - 25.314MOS^2 + 87.060MOS - 57.336$$

$$I_{e\text{.eff}} = Ro - R$$

$$I_{e\text{.eff}} = I_e + (R_0 - I_e) \frac{Ppl}{Ppl + Bpl}$$
Modelling $Ie$

$Ie$ values are obtained by taking the $Ie. eff$ values in 0% packet loss. Each bitrate level, has a fixed $Ie$ value.
Modelling $Bpl$

- Non-linear regression analysis of the model with our training set
- $le.eff$ values obtained in the previous step

\[ Bpl = \alpha_0 + \alpha_1 B + \alpha_2 B^2 + \alpha_3 B^3 \]

<table>
<thead>
<tr>
<th>Coeff</th>
<th>NB</th>
<th>WB</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_0$</td>
<td>36.836</td>
<td>122.617</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>$-4.893$</td>
<td>16.391</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.287</td>
<td>0.011</td>
</tr>
<tr>
<td>$\alpha_3$</td>
<td>$-0.006$</td>
<td>0</td>
</tr>
<tr>
<td>$r^2$</td>
<td>99.94%</td>
<td>99.8%</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.9922</td>
<td>0.9871</td>
</tr>
</tbody>
</table>
Model vs PESQ results

Wideband

PESQ MOS

Predicted MOS Score

WB mode : 98%

Narrowband

PESQ MOS

Predicted MOS Score

NB mode : 99%
Subjective evaluation of the model

- Two sets of subjective tests were carried out for NB and WB samples according to ITU-T P.800 recommendation.
- Sessions were conducted in 2 separate days
  - 60 Samples per session divided into 4 groups
  - Samples were randomly assigned for each user
Subjective results

<table>
<thead>
<tr>
<th>Measure</th>
<th>WB</th>
<th>NB</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r^2$</td>
<td>97.83%</td>
<td>91.86%</td>
</tr>
<tr>
<td>Max error</td>
<td>0.59</td>
<td>0.901</td>
</tr>
<tr>
<td>Min error</td>
<td>$-0.190$</td>
<td>$-0.486$</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.1362</td>
<td>0.2272</td>
</tr>
</tbody>
</table>
Conclusion

- Developed a regression-based model, that can measure the speech quality of VoIP calls for Skype’s SILK codec for VoIP applications
- The model is a reference-free, regression-based model based on sender bitrate and packet loss parameters
- Validated our developed model with subjective test results
- Correlation of 91% for narrowband and 97% for wideband.
- Can be implemented in mobile devices or softphones
  - Monitoring, control, adaptation and tuning for optimal user satisfaction in real time
- Aspects of this work are generic and can be adopted to other multi-rate codecs
Thank you

- Q & A

- Contact information
  - WEBSITE: [http://www.tech.plym.ac.uk/spmc/](http://www.tech.plym.ac.uk/spmc/)
  - EMAIL:
    - Mohammad Goudarzi: mgoudarzi@plymouth.ac.uk
    - Dr Lingfen Sun: lsun@plymouth.ac.uk
    - Prof. E Ifeachor: eifeachor@plymouth.ac.uk