Objective Quality Measure for Artificially Bandwidth-Extended Speech

Johannes Abel and T. Fingscheidt

Institute for Communications Technology, Technische Universität Braunschweig
Speech Quality Assessment...

Subjective Assessment

MOS \(_{LQS}\)

Objective Assessment (reference-based)

MOS \(_{LQO}\)

System Under Test

speech enhancement alg., speech codec, etc...

WB-PESQ/ POLQA

reference speech

degraded speech
- WB-PESQ and POLQA have not been designed for use with ABE speech signals
- WB-PESQ and POLQA are often misused for ABE speech quality estimation

- In several previous studies, WB-PESQ and POLQA were compared to listening test results
  - Results are inconsistent!
Outline

1. Motivation

2. An Objective Measure for ABE-Processed Speech
   - Overview
   - Feature Extraction
   - Training, Evaluation, and Evaluation Metrics
   - Underlying Subjective Listening Test

3. Cross-Validation Experiments
   - Disjoint Languages
   - Disjoint Speakers
   - Disjoint ABE Solutions

4. Summary
Overview

Input Signal Preprocessing
- Delay estimation and compensation
- Voice activity detection (VAD)
- Level adjustment to -26 dBov

Feature Extraction
- Combination of non-perceptual and perceptual features

MOS Predictor
- Powerful Support Vector Regression (SVR)
Feature Extraction and Concatenation

**Basic Features**
- Global Signal-to-Degraded-Speech Ratio
- Seg. Speech-to-Speech Distortion Ratio
- Log Spectral Distance

**Perceptual Processing**
- Sottek’s Hearing Model

**Perceptual Features**
- Spectral Balance Ratio
- Absolute Distortion
- Relative Distortion
- Mod. Normalized Covariance Metric
- 2D-Pearson’s Correlation
Training, Evaluation, and Evaluation Metrics

- **Training**
  - Feature Normalization
  - SVR Training
  - $x(i) \rightarrow \text{MOS}_{LQS}(i)$

- **Evaluation**
  - Feature Normalization
  - SVR
  - $x(j) \rightarrow \text{MOS}_{LQS}(j)$

**MOS Predictor**
- Pretrained Parameters

**Evaluation Metrics**
- Condition-based Pearson’s correlation
- Condition-based Kendal’s rank order

$i, j$ file indices
Underlying Subjective Listening Test

- Absolute category rating listening test following ITU-T P.800
- Conditions under test preprocessed following EVS standardization

- Languages: English (US), German, Chinese (Mandarin), and Korean
- Speakers: Two female (F1, F2) and two male (M1, M2) speakers per language
- Conditions: 6 NB-MNRUs, 6 WB-MNRUs, 1 AMR-NB, 3 AMR-WB, 12 ABE conditions under test processed by

  - Koç University, Turkey
  - McGill University, Canada
  - Microsoft/Aalto University, Finland
  - RWTH Aachen University, Germany
  - Technion, Israel
  - NXP Software/TU Braunschweig, Belgium/Germany

Input for ABE solutions

Different numbers of parameterizations led to 12 ABE conditions
Disjoint Languages

- Disjoint languages explicitly lead to disjoint speakers
- Overlap in conditions

<table>
<thead>
<tr>
<th>Exp.</th>
<th>Languages for Training</th>
<th>Language for Evaluation</th>
<th>Correlation</th>
<th>Rank Order</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Proposed</td>
<td>WB-PESQ</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Proposed</td>
<td>WB-PESQ</td>
</tr>
<tr>
<td>A1</td>
<td>Ch, Ge, Ko</td>
<td>En</td>
<td>0.960</td>
<td>0.888</td>
</tr>
<tr>
<td>A2</td>
<td>En, Ge, Ko</td>
<td>Ch</td>
<td>0.871</td>
<td><strong>0.882</strong></td>
</tr>
<tr>
<td>A3</td>
<td>En, Ch, Ko</td>
<td>Ge</td>
<td><strong>0.972</strong></td>
<td>0.849</td>
</tr>
<tr>
<td>A4</td>
<td>En, Ch, Ge</td>
<td>Ko</td>
<td>0.915</td>
<td>0.772</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td></td>
<td><strong>0.929</strong></td>
<td>0.848</td>
</tr>
</tbody>
</table>

Features are able to reflect relevant information to evaluate ABE solutions!
High and consistent rank order
## Disjoint Speakers

- Gender imbalance in training
- Overlap in conditions

<table>
<thead>
<tr>
<th>Exp.</th>
<th>Speakers for Training</th>
<th>Speaker for Evaluation</th>
<th>Correlation</th>
<th>Rank Order</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Proposed</td>
<td>WB-PESQ</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Proposed</td>
<td>WB-PESQ</td>
</tr>
<tr>
<td>B1</td>
<td>F2, M1, M2</td>
<td>F1</td>
<td>0.979</td>
<td>0.889</td>
</tr>
<tr>
<td>B2</td>
<td>F1, M1, M2</td>
<td>F2</td>
<td>0.926</td>
<td>0.873</td>
</tr>
<tr>
<td>B3</td>
<td>F1, F2, M2</td>
<td>M1</td>
<td>0.955</td>
<td>0.862</td>
</tr>
<tr>
<td>B4</td>
<td>F1, F2, M1</td>
<td>M2</td>
<td>0.960</td>
<td>0.786</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td></td>
<td>0.955</td>
<td>0.852</td>
</tr>
</tbody>
</table>

→ Very high correlation and rank order in each subexperiment!
Disjoint ABE Solutions

Probably the most common use case of the proposed measure!

Nested cross-validation setup to avoid speaker overlap:

![Diagram showing disjoint ABE solutions]

<table>
<thead>
<tr>
<th>Training</th>
<th>C1</th>
<th>C2</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABE01</td>
<td>ABE02</td>
<td></td>
</tr>
<tr>
<td>ABE03</td>
<td>ABE04</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>ABE11</td>
<td>ABE12</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Evaluation</th>
<th>C1</th>
<th>C2</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABE02</td>
<td>ABE01</td>
<td></td>
</tr>
<tr>
<td>ABE04</td>
<td>ABE03</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>ABE12</td>
<td>ABE11</td>
<td></td>
</tr>
</tbody>
</table>
## Disjoint ABE Solutions

<table>
<thead>
<tr>
<th>Exp.</th>
<th>Correlation</th>
<th>Rank Order</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Proposed</td>
<td>WB-PESQ</td>
</tr>
<tr>
<td>C1∩B1</td>
<td>0.921</td>
<td>0.774</td>
</tr>
<tr>
<td>C2∩B1</td>
<td>0.969</td>
<td>0.923</td>
</tr>
<tr>
<td>C1∩B2</td>
<td>0.891</td>
<td>0.882</td>
</tr>
<tr>
<td>C2∩B2</td>
<td>0.924</td>
<td>0.950</td>
</tr>
<tr>
<td>C1∩B3</td>
<td>0.895</td>
<td>0.900</td>
</tr>
<tr>
<td>C2∩B3</td>
<td>0.937</td>
<td>0.882</td>
</tr>
<tr>
<td>C1∩B4</td>
<td>0.914</td>
<td>0.913</td>
</tr>
<tr>
<td>C2∩B4</td>
<td>0.962</td>
<td>0.908</td>
</tr>
<tr>
<td>Mean</td>
<td>0.927</td>
<td>0.892</td>
</tr>
</tbody>
</table>

- High correlation and rank order even though amount of training data was further reduced
- Results proof good generalization capabilities of the proposed measure
Summary

The objective measures WB-PESQ and POLQA have not been designed for use with ABE speech signals and if used anyway for ABE signals, they lead to inconsistent results.

We proposed a reference-based objective measure for ABE speech signals, based on
- an SVR as high-performance regression model
- a multidimensional feature set, designed for ABE speech signal quality estimation
- a carefully designed subjective listening test, covering the ABE solutions of a variety of different institutions and research labs

The proposed measure shows consistency and high performance w.r.t. correlation and rank order in the three cross-validation experiments:
- disjoint languages
- disjoint speakers
- disjoint ABE solutions
Thank you for your attention

Johannes Abel
abel@ifn.ing.tu-bs.de