Speech Quality and Speech Quality Assessments Methods

Dr. Peter Počta
{pocta@fel.uniza.sk}

Department of Telecommunications and Multimedia
Faculty of Electrical Engineering
University of Žilina, Slovakia
Outline

- Speech Quality Definition
- Speech Quality Assessment Methods
- Subjective Testing
- Objective Methods
- Parametric Methods
- Performance Assessment of Objective and Paramatric Models
Speech Quality Definition

Quality is the:

- Result of the judgement of the perceived composition of an entity with respect to its desired composition [Jekosch 2005, pp.15]

- Perceived composition: Totality of features of an entity. Signal for the identity of the entity to visible to the perceiver.
- Entity: Material or immaterial object under observation
- Desired Composition: Totality of features of individual expectations and/or relevant demands and/or social requirements.
- Feature: Recognizable and nameable characteristic of an entity
Speech Quality Assessment Methods (Listening-only)

- **Subjective Testing**
  - Test subjects (group of people)
  - Higher validity and reliability of results
  - Time-consuming and costly

- **Objective Methods**
  - Algorithms
  - Good correlation with subjective tests

- **Parametric Methods**
  - Parametric or computational models (based on equations)
  - Mainly, weaker correlation with subjective test than objective methods
Subjective Testing

- Described in ITU-T Recommendation P.800 and related recommendations
- Anechoic room usage
- Female and male talkers (recordings) employed
- Two to five independent, short, meaningful and simple sentences usage (from newspapers, not technical literature)
Subjective Testing

- Overall samples duration: below 10 seconds
- Samples are presented to 24 to 32 naïve subjects
- Subjects vote on the quality of each sample, most frequently using five-point absolute category rating (ACR) listening quality (LQ) scale (see in Table 1).
Subjective Testing

Table 1: Opinion Scales
(MOS values)
(adopted by Raake)

<table>
<thead>
<tr>
<th>(a) Listening quality</th>
<th>(b) Impairment</th>
</tr>
</thead>
<tbody>
<tr>
<td>5. Excellent</td>
<td>5.0 Imperceptible</td>
</tr>
<tr>
<td>4. Good</td>
<td>4.0 Perceptible, but not annoying</td>
</tr>
<tr>
<td>3. Fair</td>
<td>3.0 Slightly annoying</td>
</tr>
<tr>
<td>2. Poor</td>
<td>2.0 Annoying</td>
</tr>
<tr>
<td>1. Bad</td>
<td>1.0 Very annoying</td>
</tr>
</tbody>
</table>
Subjective Testing

Subjective Testing Methods:

- **Absolute Category Rating (ACR)**
  - Based only on degraded samples
  - 5-point ACR Scale (see in Table 1)

- **Degradation Category Rating (DCR)**
  - Enables a more fine-grained resolution of small quality differences than ACR method
  - Original and degraded samples usage
  - Each stimulus is preceded by clean reference stimulus representing top-line quality
  - Subjects are asked to rate the degradation of test stimulus relative to the clean reference
Subjective Testing

- **Comparison Category Rating (CCR)**
  - Original and degraded samples usage
  - Employs pairs of stimuli: the quality of the second stimulus is rated relative to the first
  - Both stimuli are randomly selected from the set of all test stimuli
  - Both (CCR and DCR) use similar category rating to 5-point ACR Scale (see in Table 1)
Objective Methods

- To reduce the necessity for time-consuming and costly perception tests to measure the quality of networks or systems

**Objective Methods (Signal-based Methods):**

- **Intrusive:**
  - original and degraded samples usage
  - correlation with subjective test around 0.93 (PESQ)

- **Nonintrusive:**
  - only degraded sample usage
  - correlation with subj. test around 0.77 (3SQM)
Intrusive Objective Models

Fig. 1: Principle of intrusive signal-based models (adopted by Raake)
Intrusive Objective Models

- **PSQM (Beerends, standardized as ITU-T P.861)**
  - Very good cognitive model
  - Problems related to time-alignment and time clipped passages (for instance: lost packets)

- **PAMS (Rix and Hollier, British Telecom)**
  - Very good time-alignment model

- **PSQM+ (modified version of PSQM)**
  - Problems pointed out above, partially resolved
Intrusive Objective Models

- **PESQ (Rix, standardized as ITU-T P.862)**
  - Combinations of good properties in case of PSQM+ and PAMS models
  - Good correlation with subjective tests (0.93)
  - Mostly employed, at this time

- **P.OLQA**
  - Currently under development in ITU-T/SG12 working group)
Intrusive Objective Models

Fig. 2: The structure of PESQ algorithm (adopted by Opticom)
Nonintrusive Objective Models

Fig. 3: Principle of single-ended (nonintrusive) signal-based models (adopted by Raake)
Noninvasive Objective Models

- **ANIQUE**
  - Peripheral and central levels of auditory signal processing are modeled to extract the perceptual modulation spectrum.
  - Modulation spectrum is then related to the mechanical limitations of speech production systems to quantify the degree of naturalness in speech signals.
Nonintrusive Objective Models

- **SEAM (3SQM, standardized as ITU-T P.563)**
  - Based on three different models (Gray, Beerends and Hekstra)
  - Set of key parameters are extracted for the analysis of:
    1. Vocal tract and unnaturalness of speech
    2. Strong additive noise
    3. Interruptions, mutes and time clipping
  - Based on those parameters, the intermediate speech quality is estimated for each distortion class
  - Overall quality is obtained by linear combination of distortion class qualities
Parametric Methods

- Mainly used for planning purposes
- E-model → typical representative of this model group
- The primary output of E-model → quality rating factor R (on 0-100 scale)
- R factor can be transformed to MOS by:

\[
MoS = \begin{cases} 
1 & ; \quad R < 0 \\
1 + 0.03R + R(R - 60)(100 - R) \times 10^{-6} & ; \quad 0 < R < 100 \\
4.5 & ; \quad R > 100
\end{cases}
\]
Parametric Methods

**E-model principle:** \( R = R_0 - I_S - I_D - I_E + A \)

- \( R_0 \) represents the basic signal-to-noise ratio
- \( I_S \) is a combination of all impairments which occur more or less simultaneously with the voice signal
- \( I_D \) represents the impairments caused by delay
- \( I_E \) represents impairments caused by low bit-rate codecs and packet losses and other nonlinear effects
- \( A \) is advantage factor, which allows for compensation of impairment factors when there are other advantages of access to the user
Performance Assessment of Models

- Objective and parametric models designed to used in place of subjective tests
- Accuracy evaluated by comparison to subjective data

For this purpose, ITU-T P.800.1 defines terminology to assist this:

- **MOS-LQS** – subjective MOS derived using ACR LQ subjective test
- **MOS-LQO** – objective assessment of MOS-LQS, typically from an intrusive or signal-based nonintrusive models
- **MOS-LQE** – parametric estimate of MOS-LQS, typically from E-model
References

Thank you for your attention!

Questions?