Session III: New ETSI Model on Wideband Speech and Noise Transmission Quality – Phase I

IP transmission simulation

ETSI Workshop on Speech and Noise in Wideband Communication

Isabel Ordás
(Telefónica I+D)
© ETSI 2007. All rights reserved
Index

- Performance parameters
  - Delay, jitter, packet loss
  - Parameter interaction and dependences

- Wideband codecs
  - Overview
  - G.722
  - AMR-WB

- Background noise transmission simulation
  - Steps
    - Step 1: Speech sequences
    - Step 2: Noisy conditions
    - Step 3: Noisy signal processing
    - Step 4: Network simulation

- Database description
  - Speech samples with background noise
  - Noise reduction, coding and network transmission conditions
World Class Standards

Performance parameters
Delay, jitter, packet loss

Lots of conditions and parameters that can influence on speech quality

- **Delay**: amount of time it takes for a signal to reach a destination
  - Very direct impact on user satisfaction
  - ITU: <150ms (preferred); 400 ms (limit)
  - Codec delay + packetization delay + output queuing delay + serialization delay + network delay + network switching delay + propagation delay + de-jitter delay

- **Jitter**: variation of delay
  - Services intolerant of delay variation take solutions to reduce it by means of buffering (de-jitter buffers) increases delay
  - ITU: <1ms (audio applications after de-jitter buffer); <30 ms (no buffer)

- **Packet loss**: percentage of data packets which are lost
  - Very direct effect
  - ITU: < 3% (audio applications)
Performance parameters
Parameter interaction and dependences

- These parameters are not independent one another
Wideband codecs
Overview

- Wider band of frequency (50 Hz to 7000 Hz) compared to traditional Narrowband speech (200 Hz to 3400 Hz)
- Increase intelligibility and naturalness of speech
  - 50 Hz – 200 Hz: increased naturalness, presence and comfort
  - 3400 Hz – 7000 Hz: fricative differentiation and higher intelligibility
- Digitalised at 16 kHz
  - 16-bit integer → 256 kbps
- Speech compression becomes of significant importance
World Class Standards

Wideband codecs
G.722

- ITU-T Recommendation
- SB-ADPCM (Sub Band – Adaptive Differential Pulse Code Modulation)
- 3 modes of operation: 64 kbps; 56 kbps (auxiliary data channel 8kbps) and 48 kbps (auxiliary data channel 16 kbps)
- Encoder
World Class Standards

Wideband codecs
AMR-WB

- 3GPP /ETSI
- Recommendation G.722.2 ITU-T
- ACELP (Algebraic Code Excited Linear Prediction Coder)
- Adaptive codec capable of operating at 9 modes of operation: 6.6 kbps, 8.85 kbps, 12.65 kbps, 14.25 kbps, 15.85 kbps, 18.25 kbps, 19.85 kbps, 23.05 kbps and 23.85 kbps

- Encoder

\* 3 bits, indicating whether information bits are available and if they are speech or SID information
World Class Standards

Background noise transmission simulation
Steps

1. Clean speech sequences recording
   - Noise-free speech samples (8 samples)

2. Noisy conditions adding
   - Speech samples with background noise conditions (80 samples)

3. Wideband terminal simulation
   - Speech samples with noise and terminal conditions (720 samples)

4. Network simulation
   - Speech sample database (4320 samples)

Database of noise type/wideband terminal/network impairment combinations
World Class Standards

Background noise transmission simulation
Step 1: Speech sequences

- Recording a representative number of speech sequences without background noise

- Conditions
  - 48 kHz (16 bit) sampling rate
  - Wave format
  - Active speech level equalized to -26 dBov

- Number of samples
  - 4 speakers (2 male, 2 female), 8 sentences each
  - 2 languages: Czech, French
  - Length of recordings between 24s and 73s
    - Neutral sentences of 2s to 3s separated by pauses
    - Speech activity factor between 30% and 60%
Background noise transmission simulation
Step 2: Noisy conditions

- Different background noises need to be recorded for each speech file
  - Cafeteria noise
  - Office room noise
  - Road noise
  - Crossroads
  - Car noise (car hands-free at 130 km/h)

- Two microphone-loudspeaker positions
  - Typical handset microphone position (with loudness ratings adjusted to 7dB)
  - Hands-free microphone position (with loudness ratings adjusted to 11dB)
Background noise transmission simulation
Step 3: Noisy signal processing

- The noisy signal must be processed to take into account the influence of the terminal
  - Convolution with impulse response of WideBand (WB) terminals
  - Application of WideBand (WB) Noise-Suppression Algorithm (NSA)

- Signal processing implemented for STF 294
  - Signal speech+noise down-sampled (from 48 kHz to 16kHz) and filtered out using band-pass filters
  - Noise reduction algorithms with the following parameters
    - Parameter 1: with/without noise estimation using VAD
    - Parameter 2: smooth/sharp noise reduction filter
    - Parameter 3: noise reduction level of 9dB/18dB
Background noise transmission simulation

Step 4: Noisy signal processing

- The noisy signal must be processed to take into account the influence of the terminal
  - Convolution with impulse response of WideBand (WB) terminals
  - Application of WideBand (WB) Noise-Suppression Algorithm (NSA)

- Signal processing implemented for STF 294
  - Signal speech+noise down-sampled (from 48 kHz to 16kHz) and filtered out using band-pass filters
  - Noise reduction algorithms with the following parameters
    - Parameter 1: with/without noise estimation using VAD
    - Parameter 2: smooth/sharp noise reduction filter
    - Parameter 3: noise reduction level of 9dB/18dB
Background noise transmission simulation
Step 4: Network simulation (I)

- Noisy speech samples are simulated being transmitted over a network, adding delay, jitter and packet loss
- Real-time network emulator: NIST Net
- Procedure for simulation
  1. The call generators establish a call
  2. WAV files are encoded into the proper format (WB codec) by the sender
  3. The transport module produces RTP/UDP/IP packets to be transmitted over the packet network
  4. The source call generator sends the IP packets to NIST Net emulator through IP address 1
  5. NIST Net applies the selected network conditions (delay, jitter, and packet loss)
  6. NIST Net sends the result of the emulation to the receiver through IP address 2
  7. The receiver obtains the packet load
  8. The WB information is decoded and recorded into WAV format
World Class Standards

Background noise transmission simulation
Step 4: Network simulation (II)

- Parameters which have been varied for the purpose of STF 294
  - Packet loss
  - Delay
  - Jitter

- ITU-T Recommendations
  - One-way speech delay <150 ms (400 ms as an absolute limit)
  - Packet loss <3% for audio communications
  - Jitter should not be more than 20 ms to 50 ms (1ms after de-jigger buffering)

- Conditions emulated

<table>
<thead>
<tr>
<th></th>
<th>End-to-end delay (ms)</th>
<th>Jitter (ms)</th>
<th>Packet loss (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>150</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>400</td>
<td>20</td>
<td>3</td>
</tr>
</tbody>
</table>

Delay/Jitter distribution rule: “heavy-tail”
Packet loss distribution rule: random
# Database description

Speech samples with background noise

<table>
<thead>
<tr>
<th>Condition description</th>
<th>Number of conditions</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Languages</td>
<td>French, Czech</td>
<td>2</td>
</tr>
<tr>
<td>Speakers</td>
<td>2 males, 2 females</td>
<td>4</td>
</tr>
<tr>
<td>Noisy background</td>
<td>Cafeteria noise, Office room noise, Road noise, Crossroads, Car noise</td>
<td>5</td>
</tr>
<tr>
<td>Microphone-loudspeaker positions</td>
<td>Typical handset microphone position (with loudness ratings adjusted to 7 dB), Hands-free microphone position (with loudness ratings adjusted to 11 dB)</td>
<td>2</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td></td>
<td>80 (2<em>4</em>5*2)</td>
</tr>
</tbody>
</table>
**Database description**

Noise reduction, coding and network transmission conditions

<table>
<thead>
<tr>
<th>Condition description</th>
<th>Number of conditions</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noise reduction (Flt 135 filter)</td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>Noise estimation using VAD</td>
<td>Smooth noise reduction filter</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Noise reduction level of 9 dB</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Noise reduction level of 18 dB</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Smooth noise reduction filter</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Noise reduction level of 9 dB</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Noise reduction level of 18 dB</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sharp noise reduction filter</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Noise reduction level of 9 dB</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Noise reduction level of 18 dB</td>
<td></td>
</tr>
<tr>
<td>Continuous noise estimation (no VAD)</td>
<td>Smooth noise reduction filter</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Noise reduction level of 9 dB</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Noise reduction level of 18 dB</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sharp noise reduction filter</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Noise reduction level of 9 dB</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Noise reduction level of 18 dB</td>
<td></td>
</tr>
<tr>
<td>Coding</td>
<td>G.722 AMR-WB</td>
<td>2</td>
</tr>
<tr>
<td>Network impairments</td>
<td>Delay 0ms ; Jitter 0ms ; Loss 0% (No impairments)</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Delay 150ms ; Jitter 10ms ; Loss 1%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Delay 400ms ; Jitter 20ms ; Loss 3%</td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td>54 (9<em>2</em>3)</td>
</tr>
</tbody>
</table>
Questions?
Thank you!