A Packet Loss Recovery Technique with Line Spectral Frequency Modification in 3GPP EVS Codec

Kimitaka TSUTSUMI\textsuperscript{1}, Kei Kikuiri\textsuperscript{1}, Jérémie Lecomte\textsuperscript{2}

\textsuperscript{1}NTT DOCOMO Inc., \textsuperscript{2}Fraunhofer IIS
Agenda

• Problem statement
• LSF quantization in EVS codec
• Analysis on speech overshoot
• Proposed method
• Experimental result
• Conclusion
Problem: Speech overshoot

Observed power surge at the first frame recovered from packet loss at a frame before speech onset

Waveform under Clean channel

Waveform under Erroneous channel
Problem: LPC spectrum

Sharp peak in LPC spectrum observed at the recovery frame suffering from speech overshoot
LSF quantization in EVS codec

- Quantization only at selected sub-frames
- Interpolation from available LSF vectors

**Encoder Side**
- \( q \) \rightarrow \text{LSF vector} \rightarrow \text{r} \rightarrow \text{Q} \rightarrow \text{Decoder Side}
- \( \alpha_{MA} \times \text{mean LSF vector} \)

**Decoder Side**
- \( Q^{-1} \) \rightarrow \text{r} \rightarrow \text{q} \rightarrow \text{m} \)
- \( \alpha_{MA} \times m \)}
Recovery of LSF in EVS codec

- Two frames required for complete recovery
- The first Recovery frame: affected by $r^{-1}$
- The Second recovery frame: affected by interpolation

$$z[0] = r[0] + \alpha_{MA} r^{-1}$$

LSF under clean channel

LSF under erroneous channel

Lost Frame
Analysis on speech overshoot

- Sharp peak in LPC spectrum caused by narrow interval between LSF parameters

![Graph showing LSF and LPC spectra under clean and erroneous channels.](image)
Speech overshoot detection

• Difficult to detect at the decoder side
  ✓ Due to packet loss, or power surge of original?

• Frame class at overshoot not classified as “Onset”
  ✓ Waveform at overshoot looks like speech onset

<table>
<thead>
<tr>
<th>Frame class of overshoot frame</th>
<th>The number of frames</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unvoiced</td>
<td>183</td>
</tr>
<tr>
<td>Unvoiced transition</td>
<td>0</td>
</tr>
<tr>
<td>Voiced transition</td>
<td>1</td>
</tr>
<tr>
<td>Voiced</td>
<td>65</td>
</tr>
<tr>
<td>Onset</td>
<td>7</td>
</tr>
<tr>
<td>Total</td>
<td>256</td>
</tr>
</tbody>
</table>
Proposed method: overview

- Speech overshoot detection based on PLC simulation at encoder side
- Transmit the simulation result to the decoder

Block-diagram of proposed technology
Proposed method: power control

- Decoder side tool
- Narrow interval compared to clean channel
  - LSF under Clean channel
  - LSF under Erroneous channel
- Increase interval of the output vector of LSF concealment

\[
\tilde{\omega}_j^{[-1]} = \begin{cases} 
  j \cdot \delta & (1 \leq j < idx) \\
  \hat{\omega}_j^{[-1]} & (idx \leq j < 16)
\end{cases} \\
\delta = \omega_{idx-1}^{[-1]}/(idx - 1)
\]
Proposed method: power control

- Sharp peak in LPC spectrum prevented
- Power surge at the recovery frame suppressed

Clean channel

Packet loss w/o Proposal

Packet loss w/ Proposal
Proposed method: overshoot detector

- Encoder side tool
- Test if modified LSF vector closer to LSF under clean channel than concealed LSF

\[
\sum_{i=1}^{16} |\tilde{\omega}_j[-1] - \omega_j[-1]|^2 > \alpha \cdot \sum_{i=1}^{16} |\tilde{\omega}_j[-1] - \omega_j[-1]|^2
\]

- Test if each of the elements in modified LSF vector closer to LSF under clean channel than concealed LSF

\[
|\tilde{\omega}_j[-1] - \omega_j[-1]|^2 > |\tilde{\omega}_j[-1] - \omega_j[-1]|^2
\]

- Limit activation only at onset

Minimum value of Algebraic CB gain \( g_{\text{min}} > g_{\text{th}} \) Predetermined threshold
Experimental result: objective quality

- Wideband PESQ for artificially created error pattern
- Speech materials (wideband) of 170 seconds in total
- Observed significant improvement for all the operating bit-rate
- Comparison with the restricted codec (EVS codec with the proposed method deactivated)
Experimental result: subjective quality

- Mushra test for 8 expert listeners
- 4 speech items with 6% random error pattern
- Comparison between EVS codec and the restricted EVS codec
- Observed improvement in average score for all the items

1. EVS, 2. the restricted codec (EVS codec with the proposed method deactivated)
Experimental result: additional analysis

Observed significant improvement with proposed method

Differential Score (EVS vs the restricted codec)

1. EVS, 2. the restricted codec (EVS codec with the proposed method deactivated)
Conclusion

• Examined a behavior of speech overshoot at the recovery frame from packet loss around speech onset
• Power control method, and speech overshoot detection method developed
• Observed quality improvement by experimental results