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

Kimitaka TSUTSUMI[1], Kei Kikuiri[1], Jérémie Lecomte[2]

[1]NTT DOCOMO Inc., [2]Fraunhofer IIS

Agenda

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

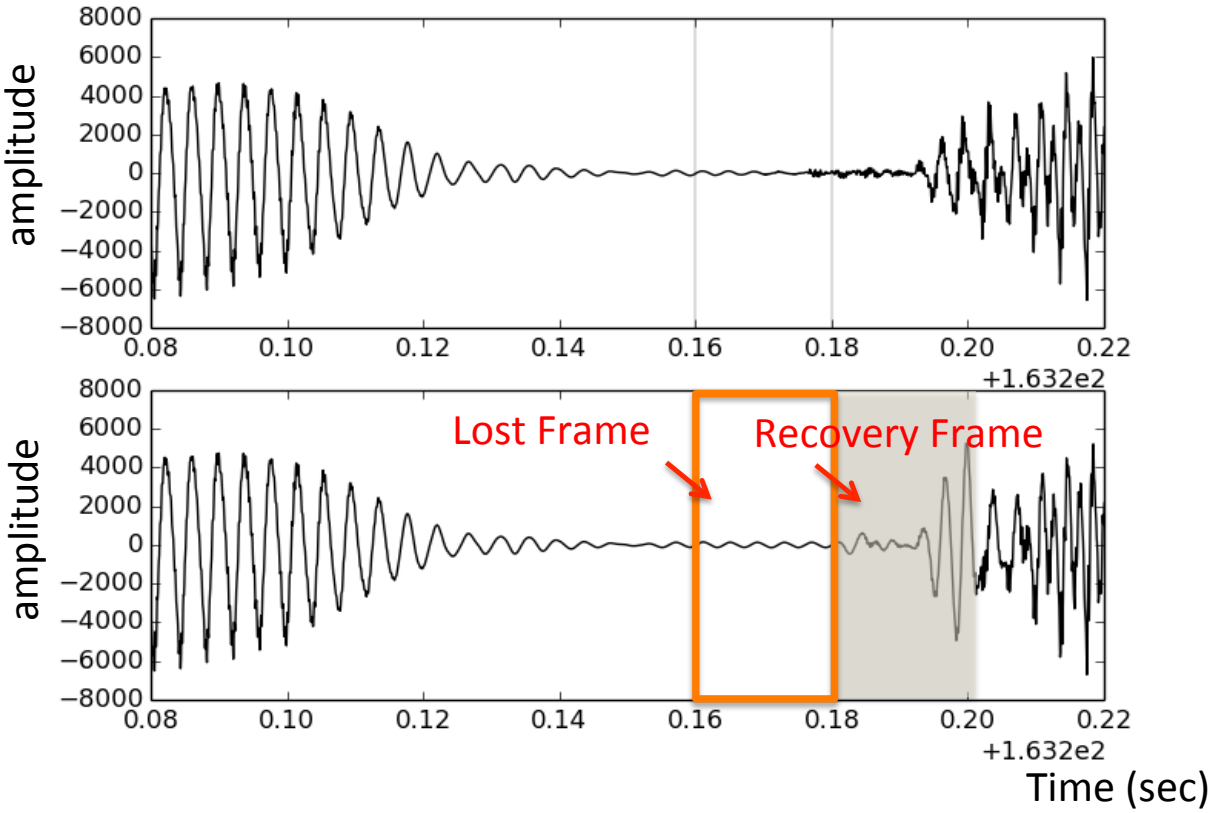
NTT

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

Innovative R&D by NTT

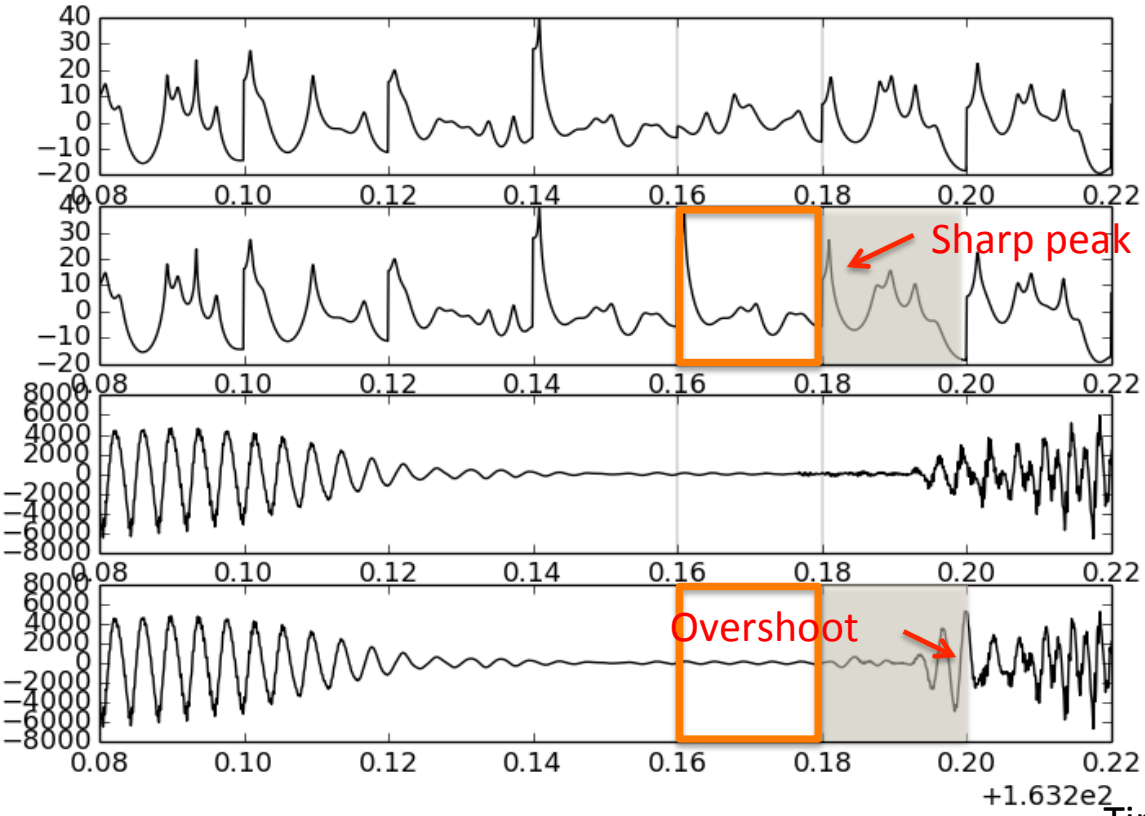
Sharp peak in LPC spectrum observed at the recovery frame suffering from speech overshoot

LPC Spectrum under Clean channel

LPC Spectrum under Erroneous channel

Waveform under Clean channel

Waveform under Erroneous channel

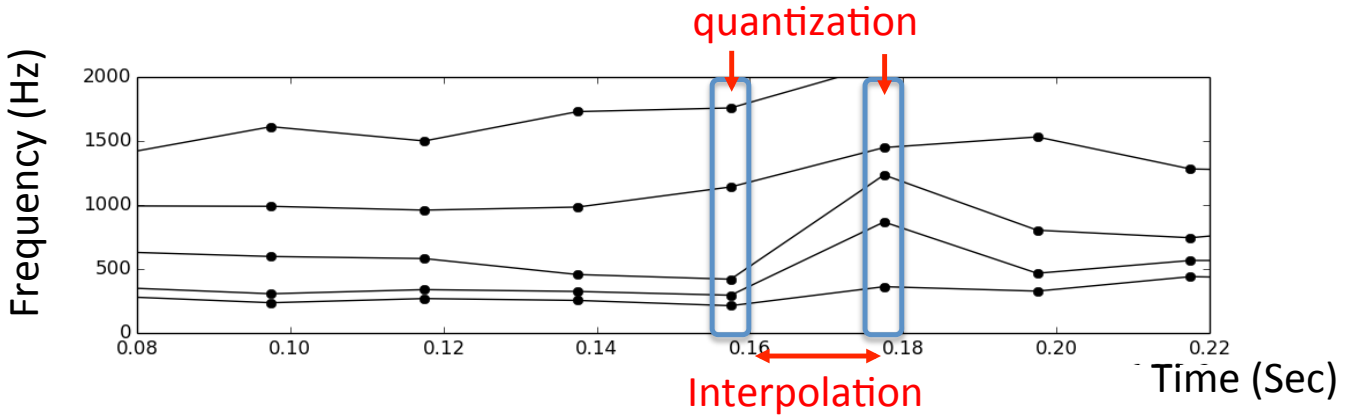


NTT

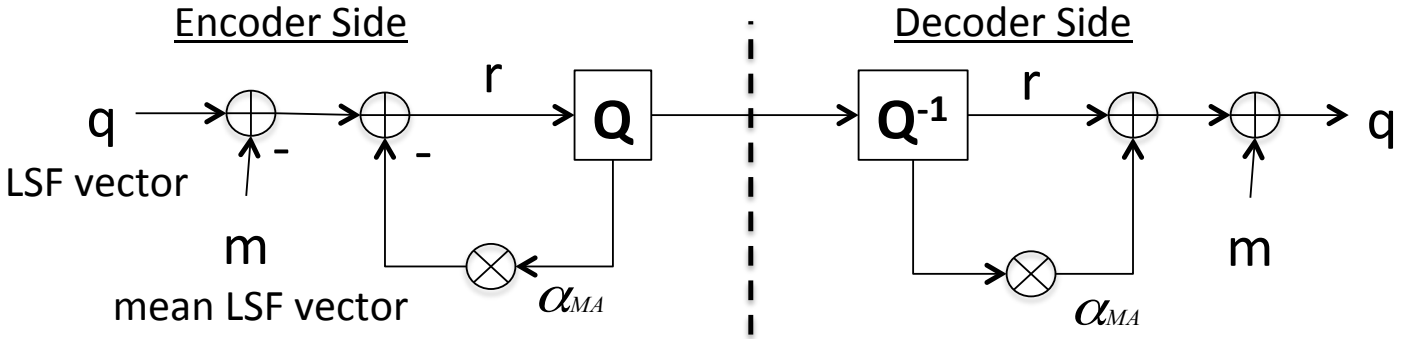
LSF quantization in EVS codec

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

LSF under clean channel



- MA prediction for quantization at speech onset



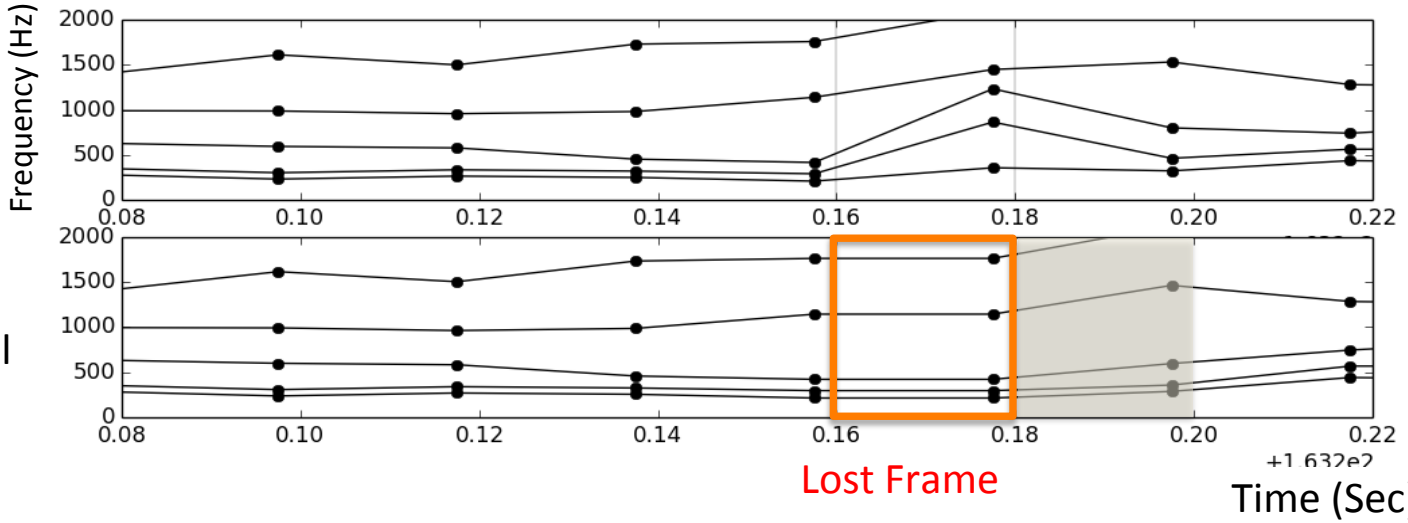
Recovery of LSF in EVS codec

- 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



Analysis on speech overshoot

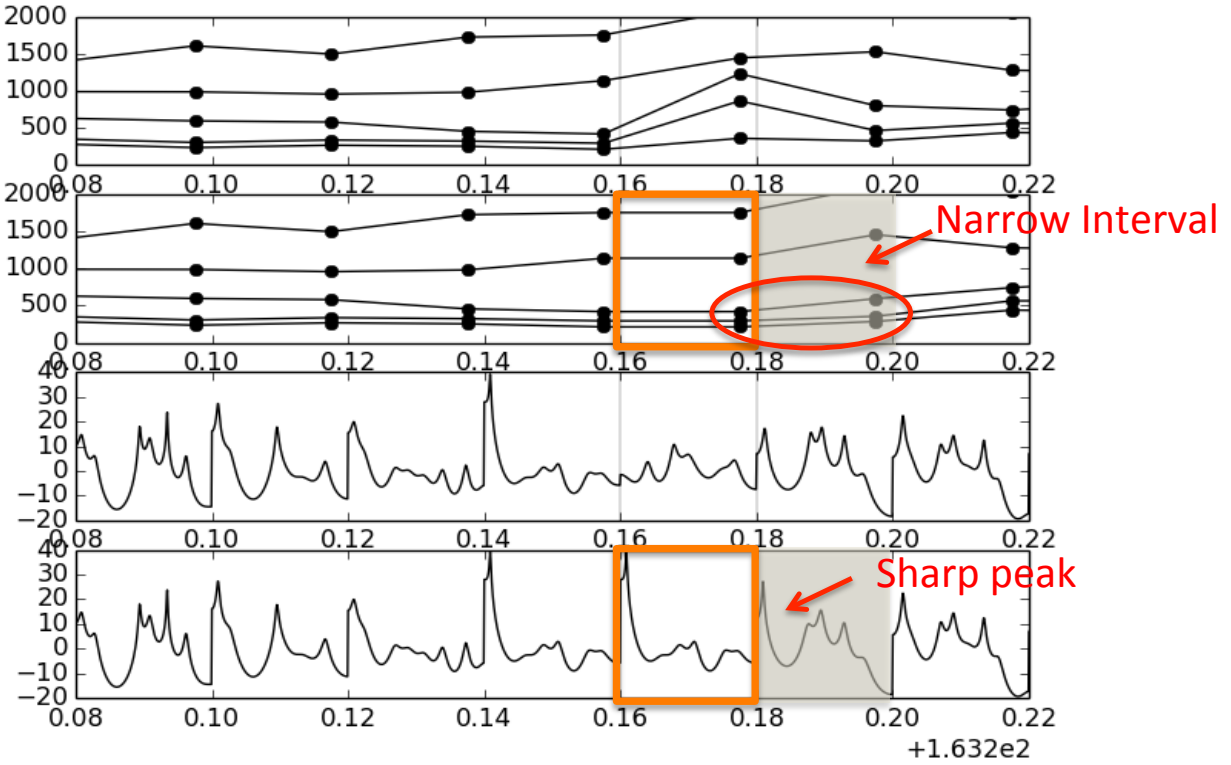
- Sharp peak in LPC spectrum caused by narrow interval between LSF parameters

LSF under Clean channel

LSF under Erroneous channel

LPC Spectrum under Clean channel

LPC Spectrum under Erroneous channel



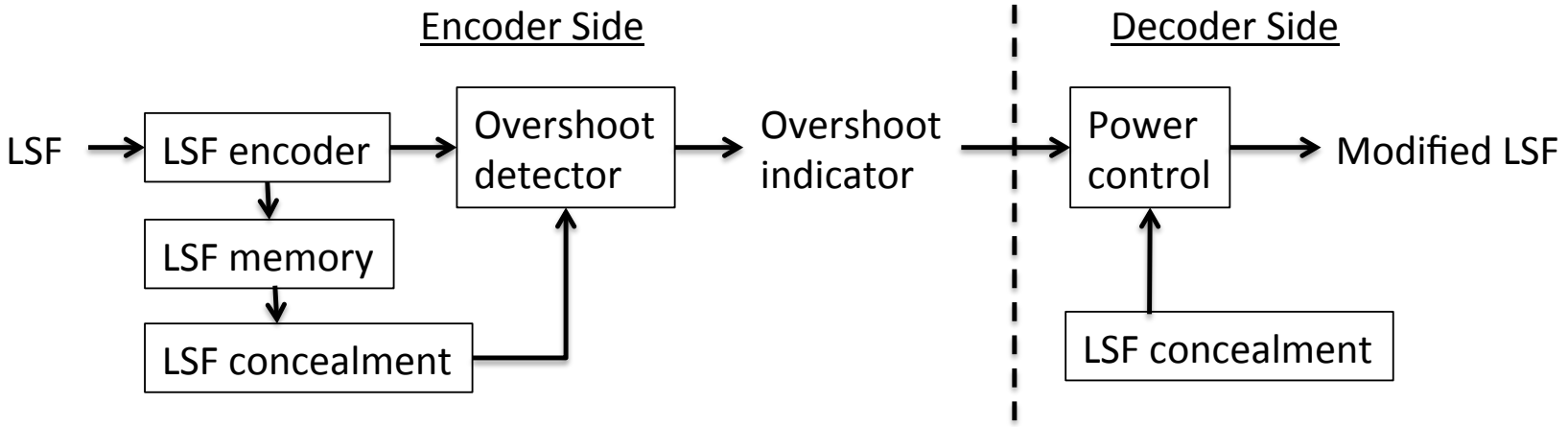
Speech overshoot detection

- 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

Frame class of overshoot frame	The number of frames
Unvoiced	183
Unvoiced transition	0
Voiced transition	1
Voiced	65
Onset	7
Total	256

Proposed method: overview

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



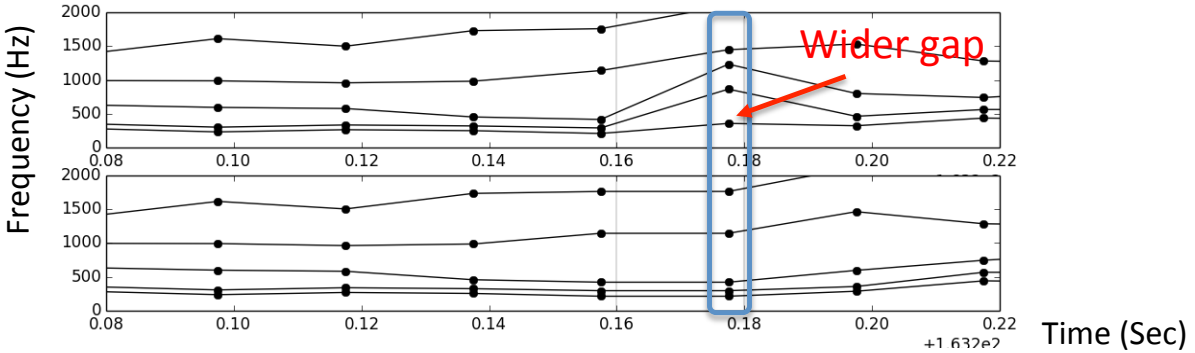
Block-diagram of proposed technology

Proposed method: power control

- Decoder side tool
- Narrow interval compared to clean channel

LSF under Clean channel

LSF under Erroneous channel

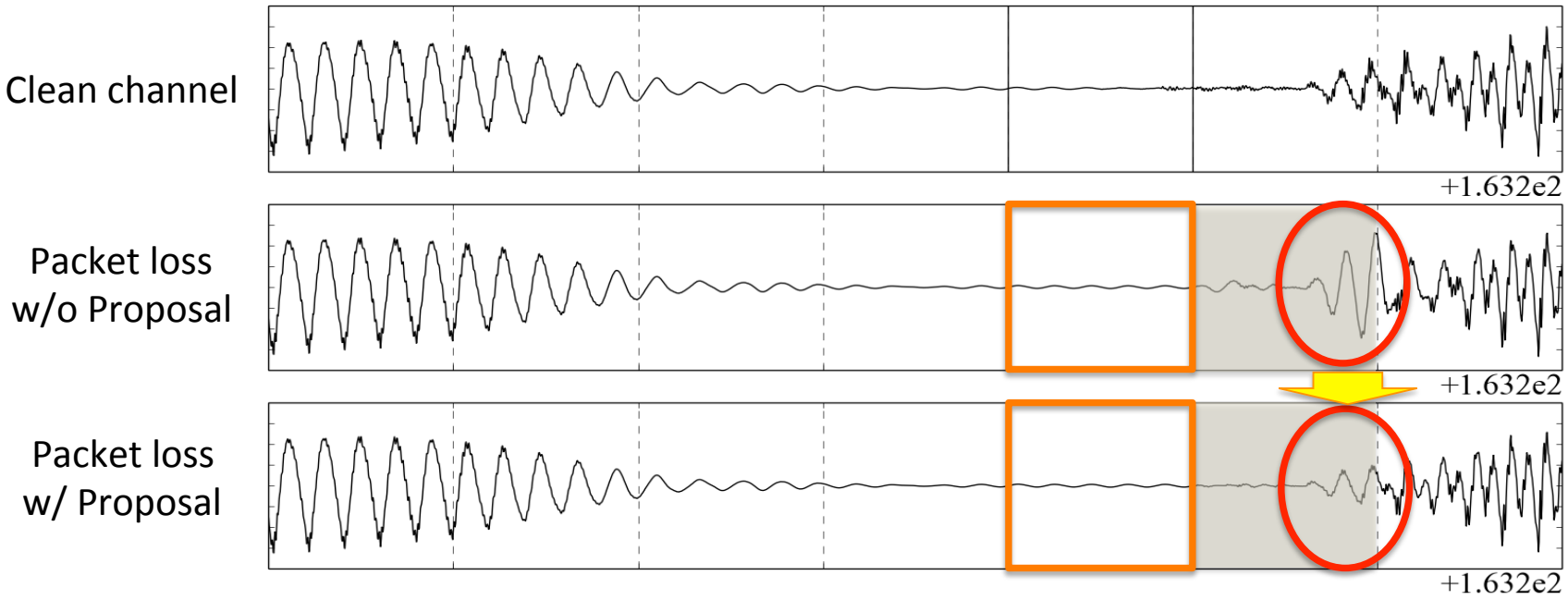


- Increase interval of the output vector of LSF concealment

$$\tilde{\omega}_j^{[-1]} = \begin{cases} j \cdot \delta & (1 \le j < idx) \\ \dot{\omega}_j^{[-1]} & (idx \le j < 16) \end{cases} \qquad \delta = \omega_{idx-1}^{[-1]} / (idx - 1)$$

Proposed method: power control

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



Proposed method: overshoot detector

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

concealed LSF

Modified LSF based on concealed LSF

LSF under clean channel

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

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

$$\left|\dot{\omega}_j^{[-1]}-\omega_j^{[-1]}\right|^2 > \left|\tilde{\omega}_j^{[-1]}-\omega_j^{[-1]}\right|^2$$

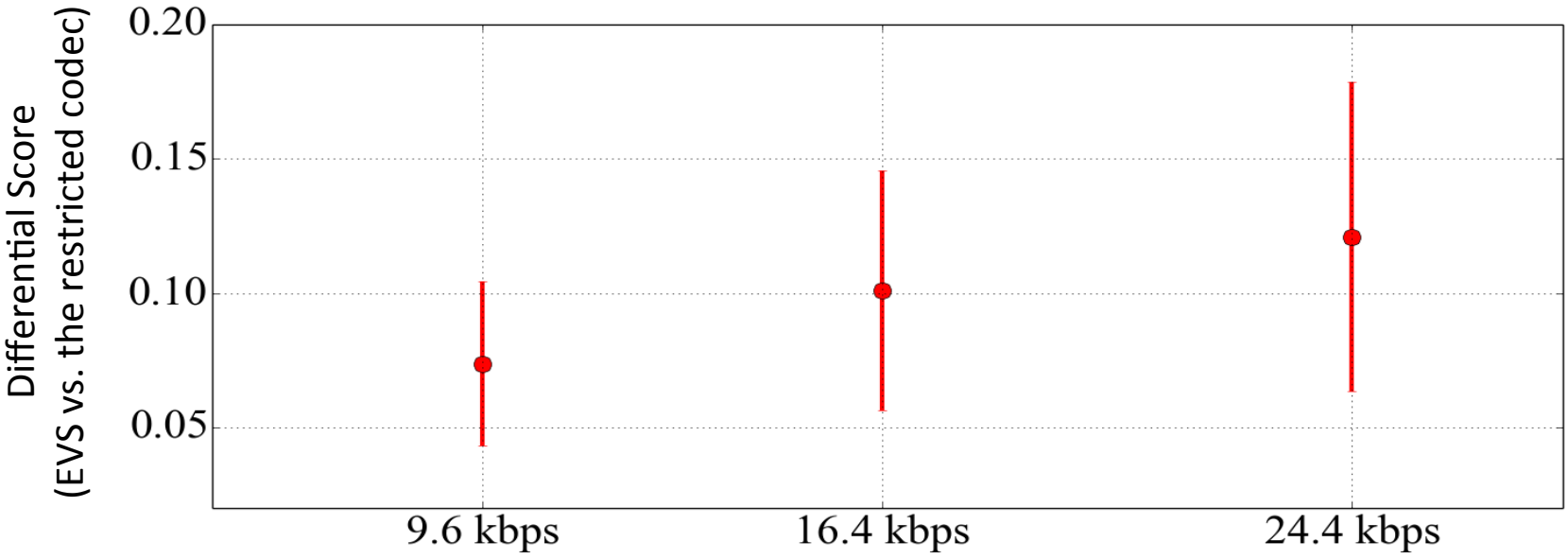
- Limit activation only at onset

Minimum value of Algebraic CB gain → $g_{\min} > g_{th}$ ← Predetermined threshold

NTT

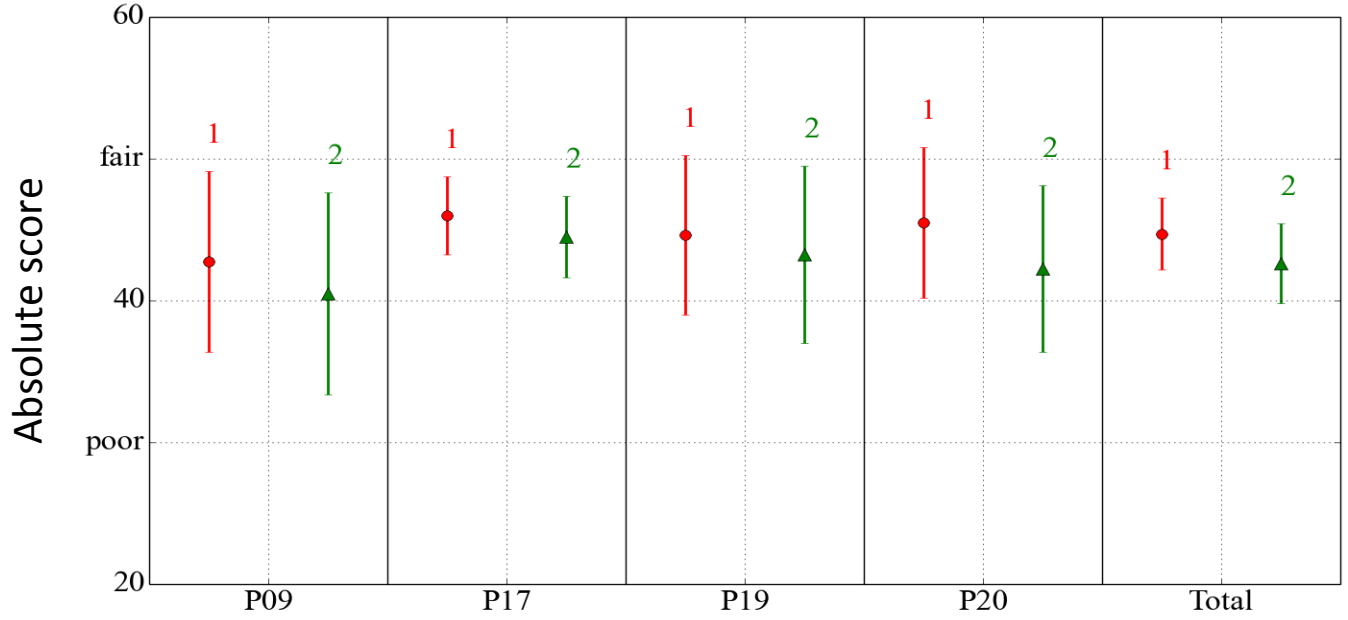
Experimental result: objective quality

- 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

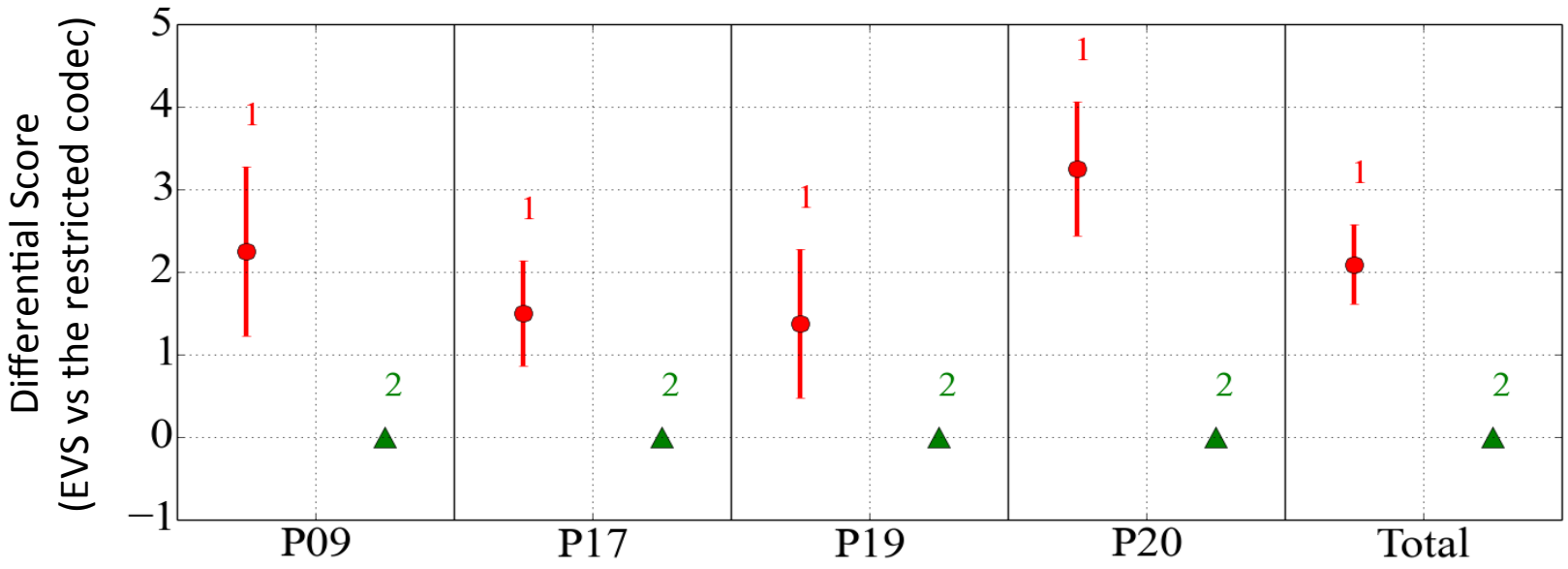


1. EVS, 2. the restricted codec (EVS codec with the proposed method deactivated)

Experimental result: additional analysis

Innovative R&D by NTT

Observed significant improvement with proposed method



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

NTT

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