NTT

Innovative R&D by NTT

Adaptive selection of lag-window shape for linear predictive analysis in the 3GPP EVS codec

Yutaka Kamamoto, Takehiro Moriya, Noboru Harada

NTT Communication Science Laboratories,
Nippon Telegraph and Telephone Corporation

IEEE GlobalSIP
3rd IEEE Global Conference on Signal & Information Processing
Orlando, Florida, USA December 14-16 2015
IEEE

Outline

- Lag-window has been used for linear predictive analysis to prevent possible instability of the synthesis filter.
- Lag-window shape has the trade-offs between stability and fidelity.
- Adaptive lag-windowing scheme depending on the periodicity is proposed to obtain good compromise.
- The codec with adaptive lag-window shows better quality by the subjective evaluation.

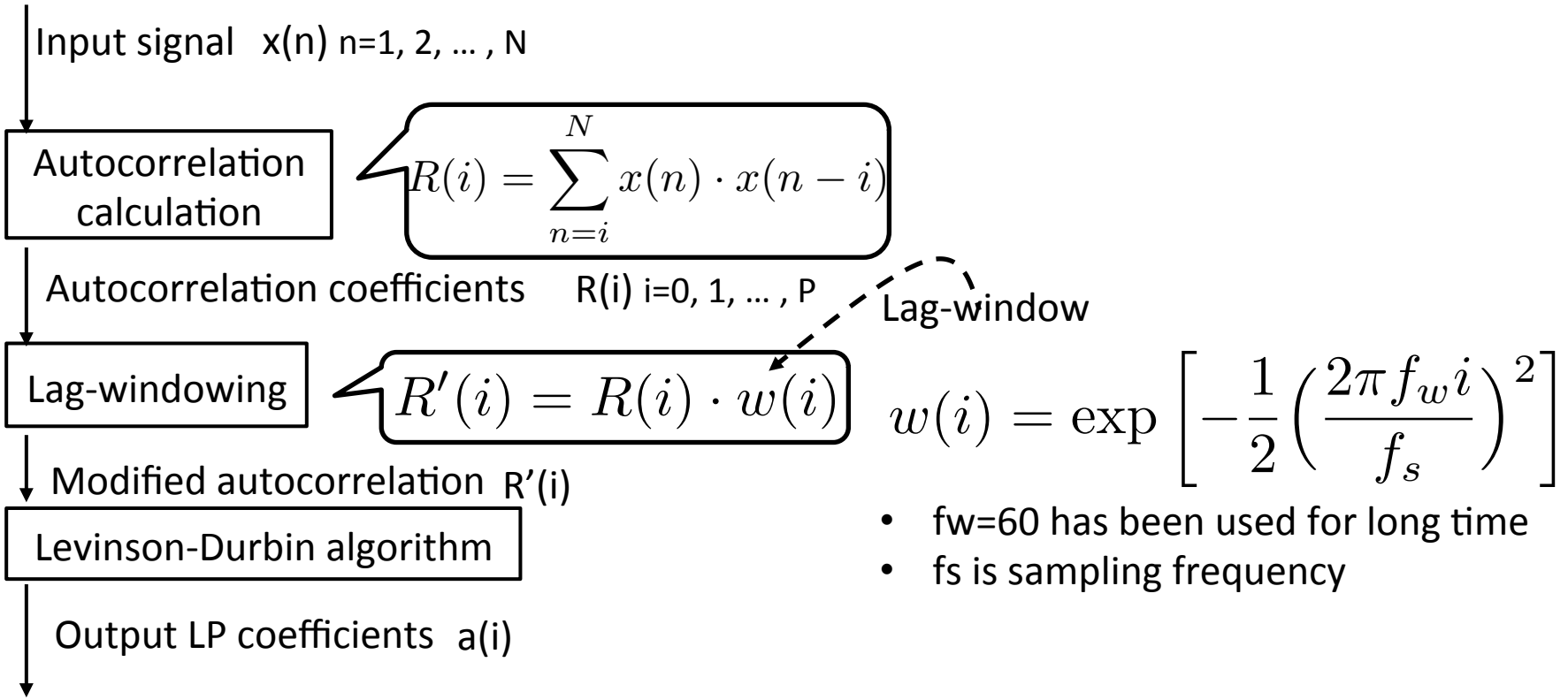
Introduction

- Linear predictive (LP) analysis is widely used
  - Speech coding: code-excited linear prediction (CELP)
    e.g.) G.729, AMR, AMR-WB, G.718
  - Audio coding: spectral envelope in frequency-domain
    e.g.) TwinVQ, USAC, TCX, AMR-WB+
- New 3GPP EVS codec also uses LP analysis
  - More realistic conversation over the mobile network
  - Lower delay is preferable even for music contents
  - Switching strategy between ACELP speech coding and frequency-domain audio coding achieves excellent quality for VoIP/VoLTE

NTT

LP analysis by auto-correlation method

- Spectral envelop is estimated by LP coefficients
- Lag-window is applied to avoid instability of filter



- fw=60 has been used for long time
- fs is sampling frequency

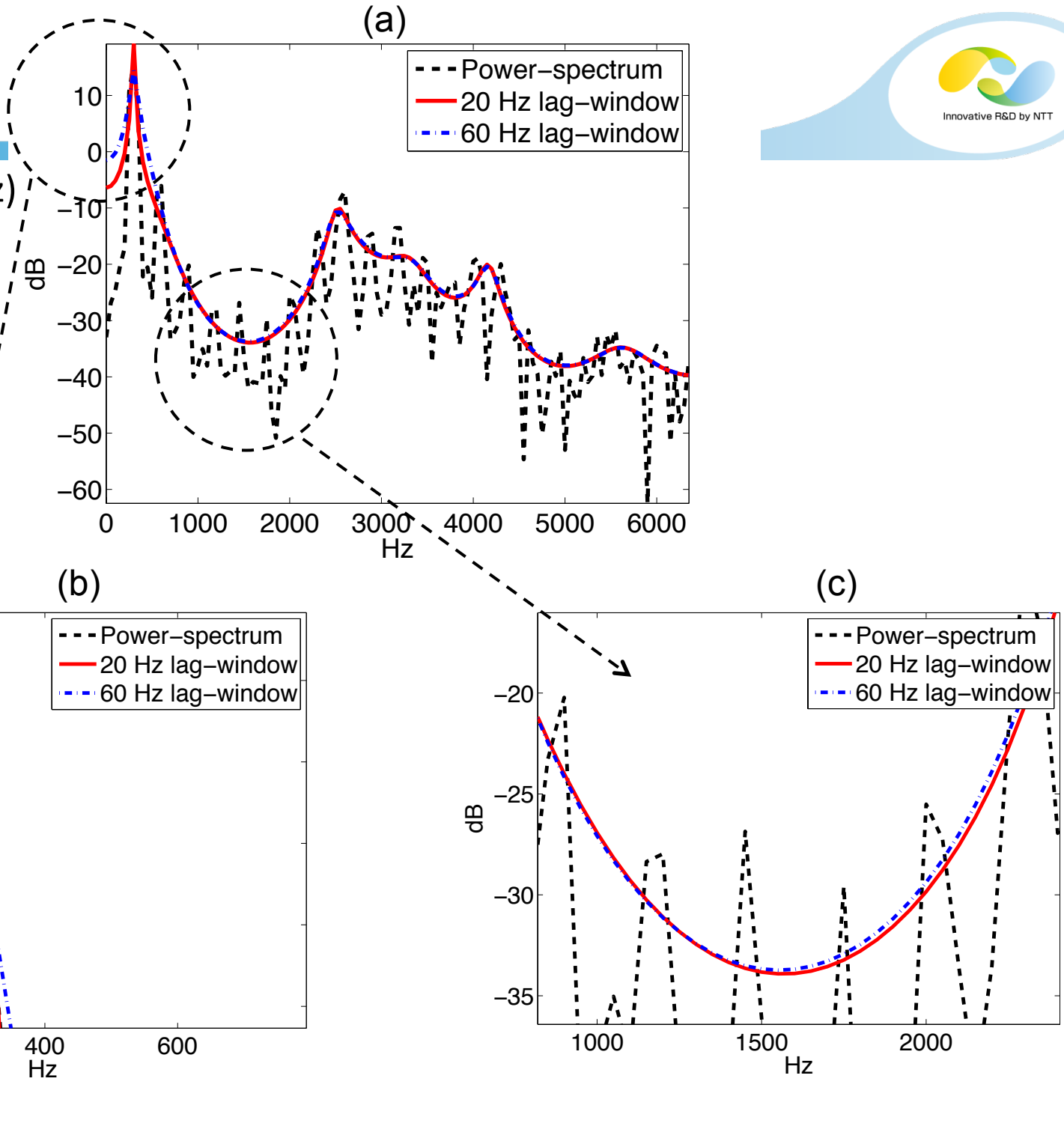
Lag-windowing

- A.k.a. band-width expansion or spectral smoothing
- Preventing instability for LP analysis
  - Simple but efficient, small complexity
- Sacrifice of fidelity of spectral envelop
  - Smaller PARCOR coefficients means whiteness
  - Spectral envelop becomes flatter
- Fixed 60-Hz Gaussian window has been used since early times
- Lag-Window is not needed for some cases
  - Lossless compression without long-term prediction does not use lag-windowing such as G.711.0

Pros: Obtain stable LP coefficients
Cons: Degrade fidelity of spectral envelop

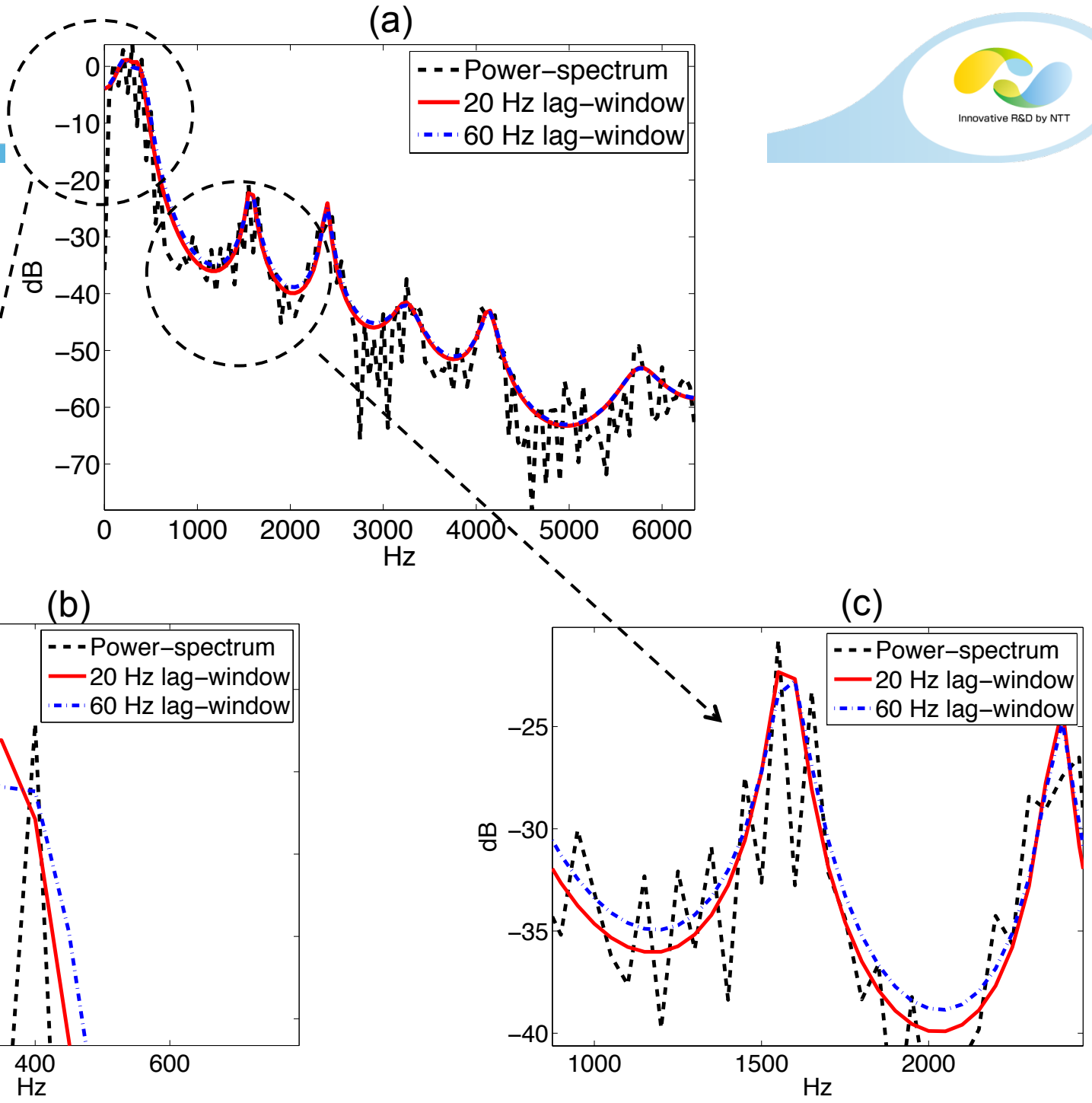
Pros

High-pitch (294 Hz)



Cons

Low-pitch (98 Hz)



Adaptive lag-windowing

- Appropriate control of lag-window shape
- Pilot study
  - Input: Synthesized speech signal to know actual pitch frequency and pitch gain
  - Observed: SNR, segSNR, MOS-LQO by POLQA
  - Results: Higher pitch frequency and pitch gain needs strongly attenuated lag-window, and lower pitch frequency and pitch gain needs weak lag-window
- Lag-window shape should depend on pitch frequency and pitch gain

$$w(i) = \exp\left[-\frac{1}{2}\left(\frac{2\pi(\alpha F_0 + \beta G)i}{f_s}\right)^2\right]$$

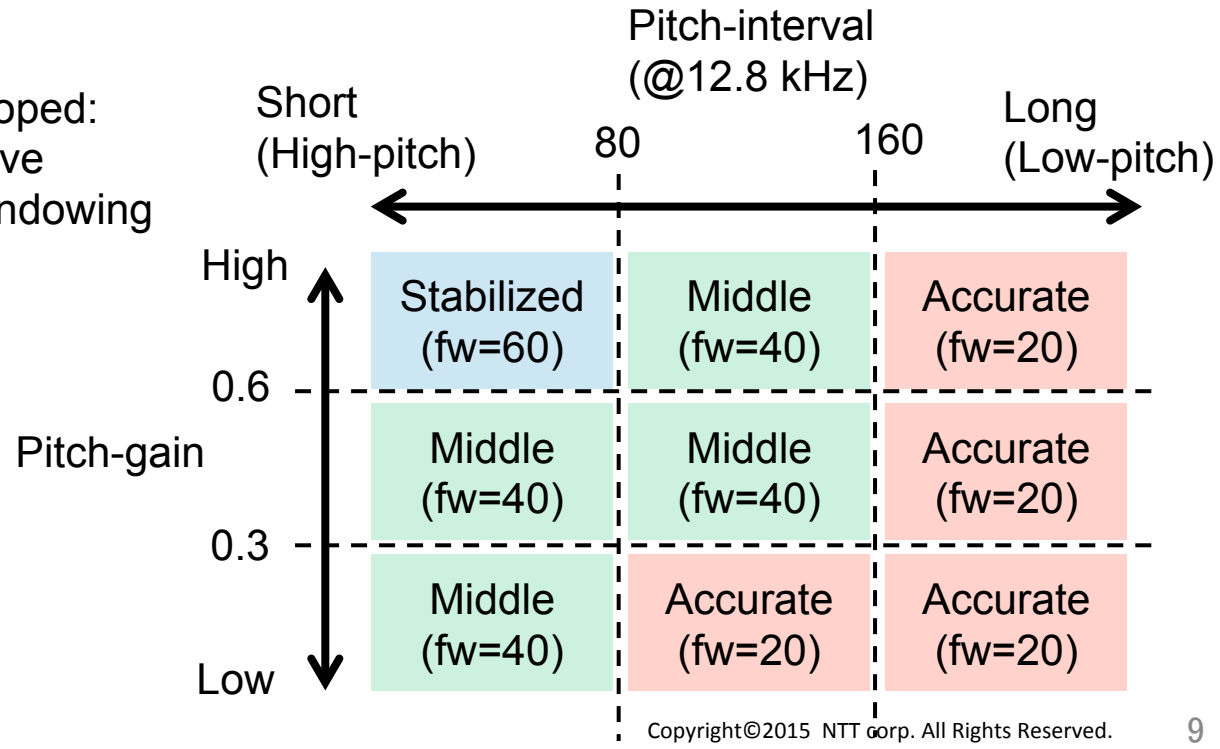
Simplified adaptive lag-windowing

- Selection criterion of lag-window shape
- EVS has enhanced signal analysis tools:
  open-loop pitch analysis, background noise energy estimation, signal activity detection

Conventional:
fixed lag-windowing
(G.729, G.718,
AMR, AMR-WB, etc.)

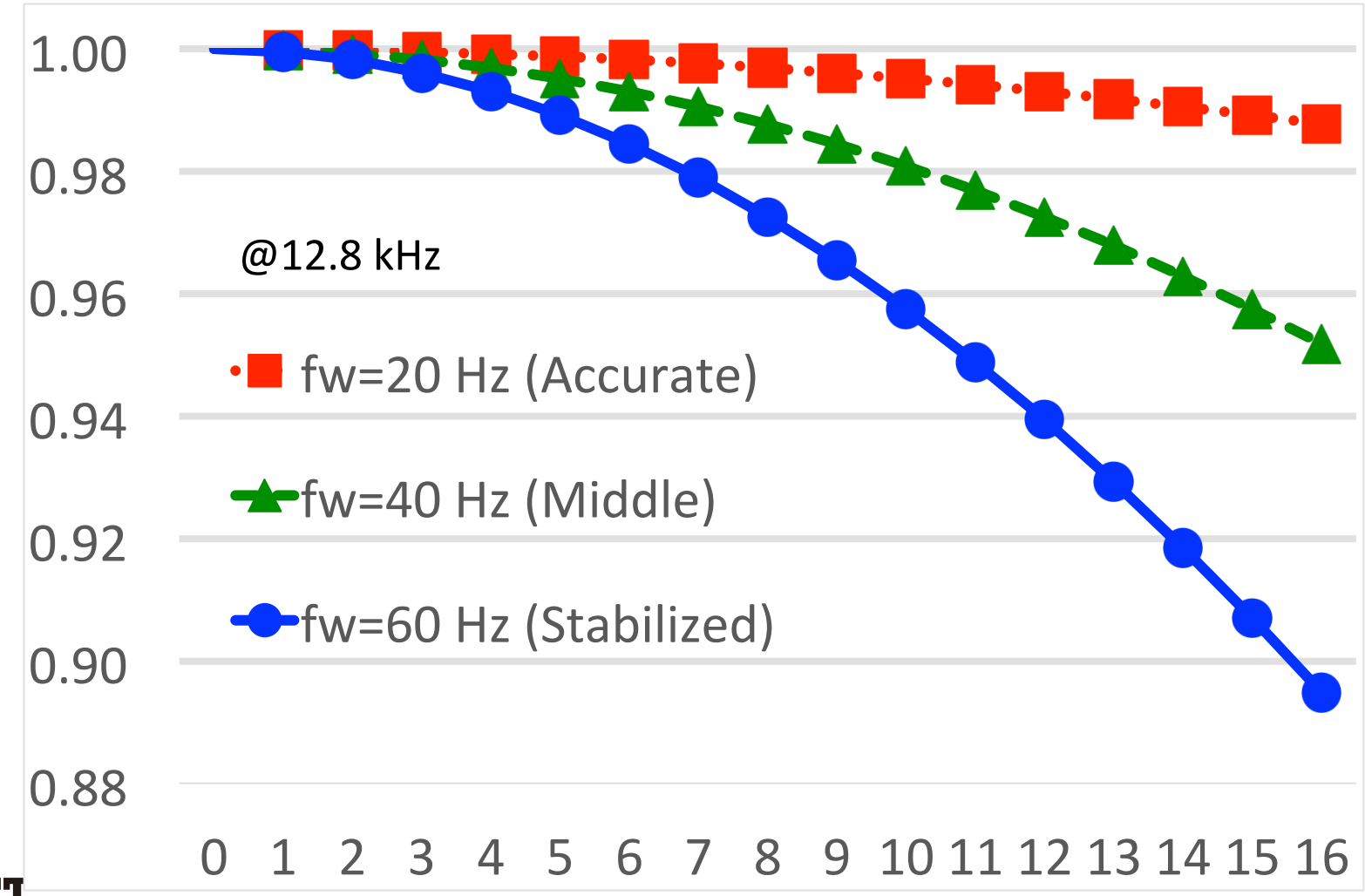
Stabilized
(fw = 60)

Developed:
adaptive
lag-windowing
(EVS)



Shape of lag-window

- Several bytes of ROM, a few if-clause to decide

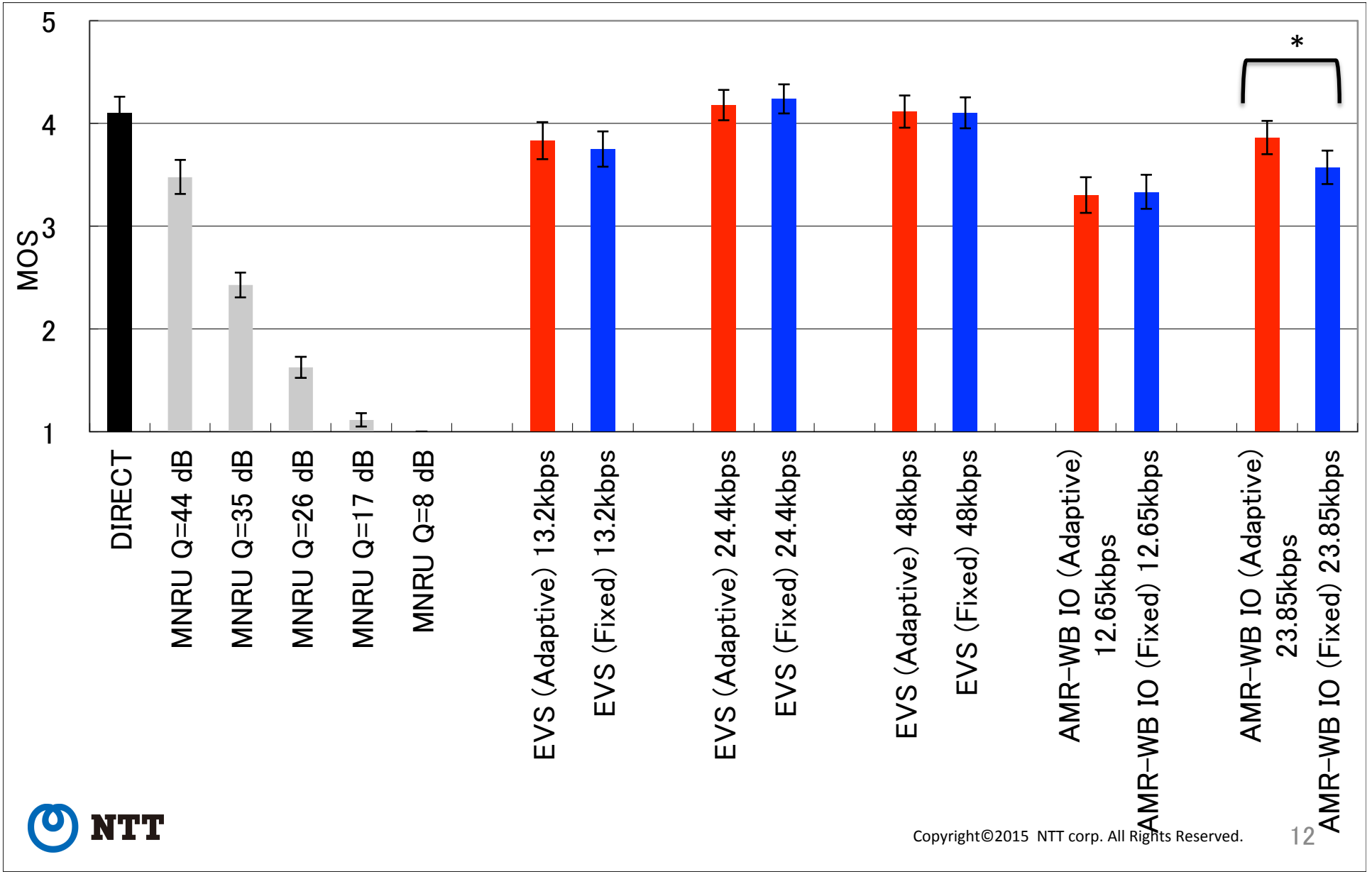


Subjective evaluations

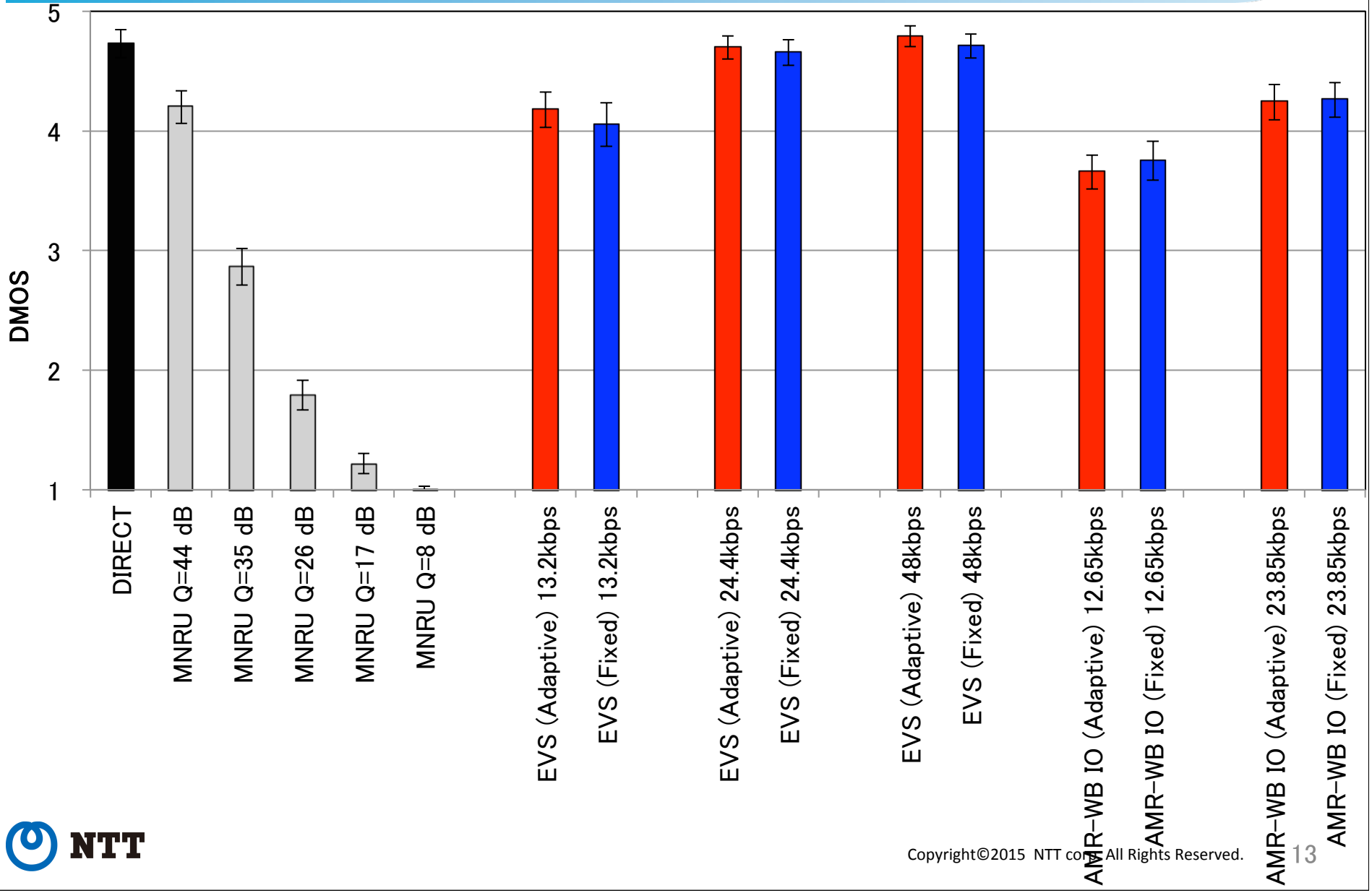
- ITU-T P.800 & EVS test plan
  - Test items: Wideband (16 kHz sampling), -26 dBov, 8 sec., 4 talkers * 6 sentence-pairs
    - Clean speech (ACR)
    - Noisy speech (DCR) – Car noise @ 20 dB
    - Mixed and Music contents (DCR)
  - 24 naïve and native Japanese listeners, (24 listeners * 4 sentence-pairs = 96 votes)
- Codecs
  - Adaptive lag-windowing (developed) vs. Fixed @ 60-Hz lag-windowing (conventional)
  - EVS: 13.2 kbps, 24.4 kbps, & 48 kbps
  - AMR-WB IO: 12.65 kbps & 23.85 kbps

NTT

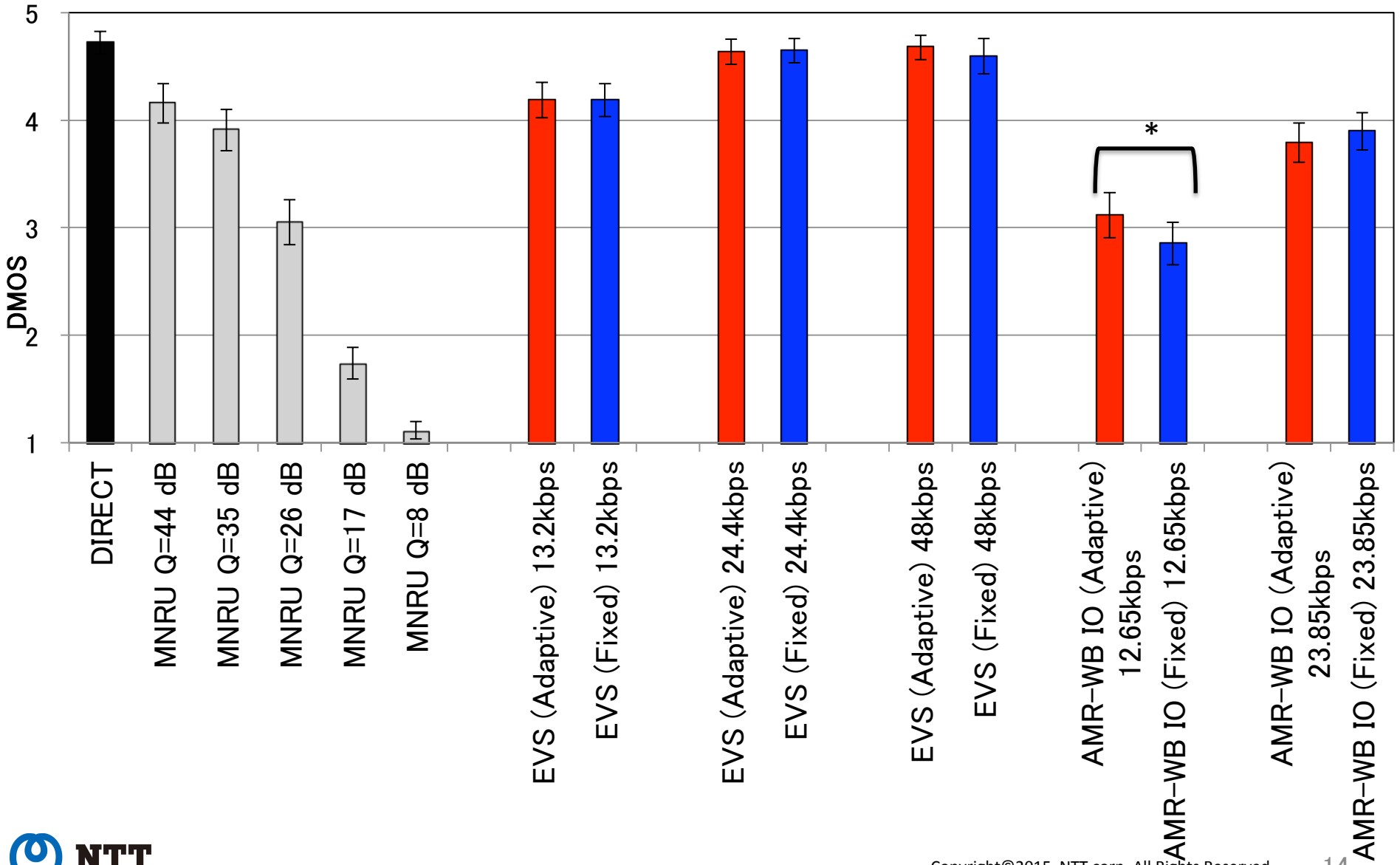
MOS (Clean speech) - ACR



MOS (Noisy speech) - DCR



MOS (Mixed and Music) - DCR



Discussion

- Summary of subjective evaluations
  - No significant degradations
  - Slightly better tendencies
  - Significant improvement for two conditions
- Adaptive lag-windowing enhances the quality of speech and music
  - LP-based codec: IO modes use ACELP, EVS@48 uses TCX
  - Switching LP-based & non-LP-based codecs: EVS@13.2 & EVS@24.4 changes codec strategy frame-by-frame

Adaptive/Fixed	EVS@13.2	EVS@24.4	EVS@48	IO@12.65	IO@23.85
Clean speech	3.8 > 3.7	4.2 = 4.2	4.1 = 4.1	3.3 = 3.3	3.9 >> 3.6
Noisy speech	4.2 > 4.1	4.7 = 4.7	4.8 > 4.7	3.7 < 3.8	4.2 = 4.2
Mixed & Music	4.2 = 4.2	4.6 = 4.6	4.7 > 4.6	3.1 >> 2.9	3.8 < 3.9

Conclusion

- Lag-window has been used for LP analysis to obtain stable coefficients of the synthesis filter.
- Lag-window shape has the trade-offs between stability and fidelity of spectral envelop.
- Adaptive lag-windowing scheme depending on the pitch-lag and the pitch-gain was developed to achieve better compromise.
- The EVS codec with adaptive lag-window showed better quality by the subjective evaluation.
- The adaptive lag-windowing selection method is adopted in 3GPP EVS codec.
- The adaptive lag-windowing scheme may be useful other LP analysis purposes.

NTT

Acknowledgement

- The authors thank all researchers who have contributed to EVS, especially people who kindly checked the performance of the adaptive lag windowing during the standardization phase.