A COMFORT NOISE ADDITION POST-PROCESSOR FOR ENHANCING LOW BIT-RATE SPEECH CODING IN NOISY ENVIRONMENTS

Guillaume Fuchs, Anthony Lombard, Emmanuel Ravelli and Martin Dietz
Fraunhofer Institut für Integrierte Schaltungen (IIS), Germany

Fraunhofer IIS

OUTLINE

- Motivations
- Comfort Noise Addition
  - Principle
  - Noise estimation
  - Noise injection
- Listening test results
- Conclusions

Motivations

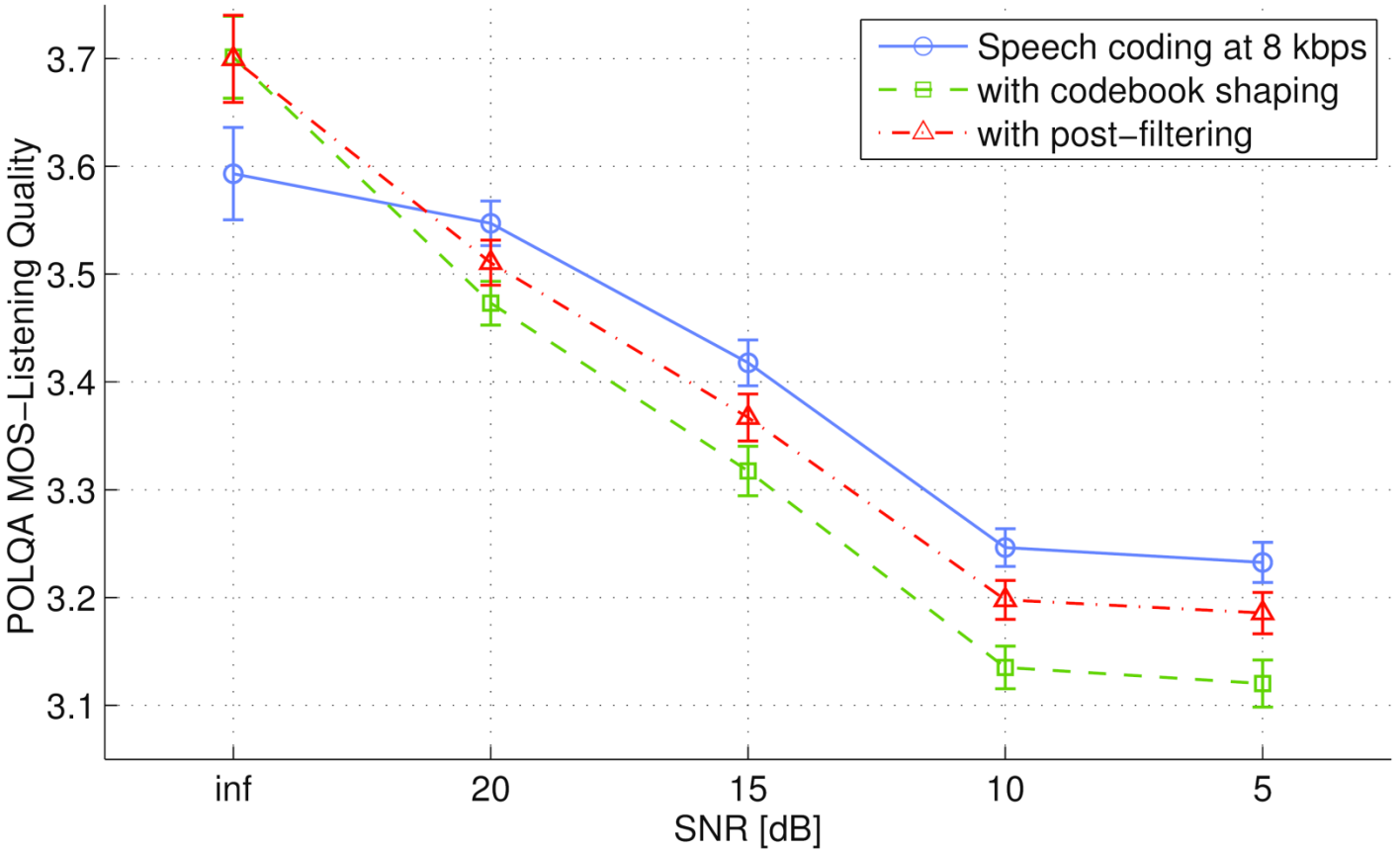
Degradation of quality at low bit-rates under noisy conditions

Degradation comes from:

- Source modeling of speech coders inappropriate for background noise
  - Loss of energy in background noise
  - Fluctuating/unstable background noise
- Switch between different coding schemes
  - Discontinuities and inconsistent rendering in background noise
- Speech enhancement techniques
  - Create audible musical noise on noisy speech

Motivations

Formant enhancement in low bit-rate speech coding



Fraunhofer IIS

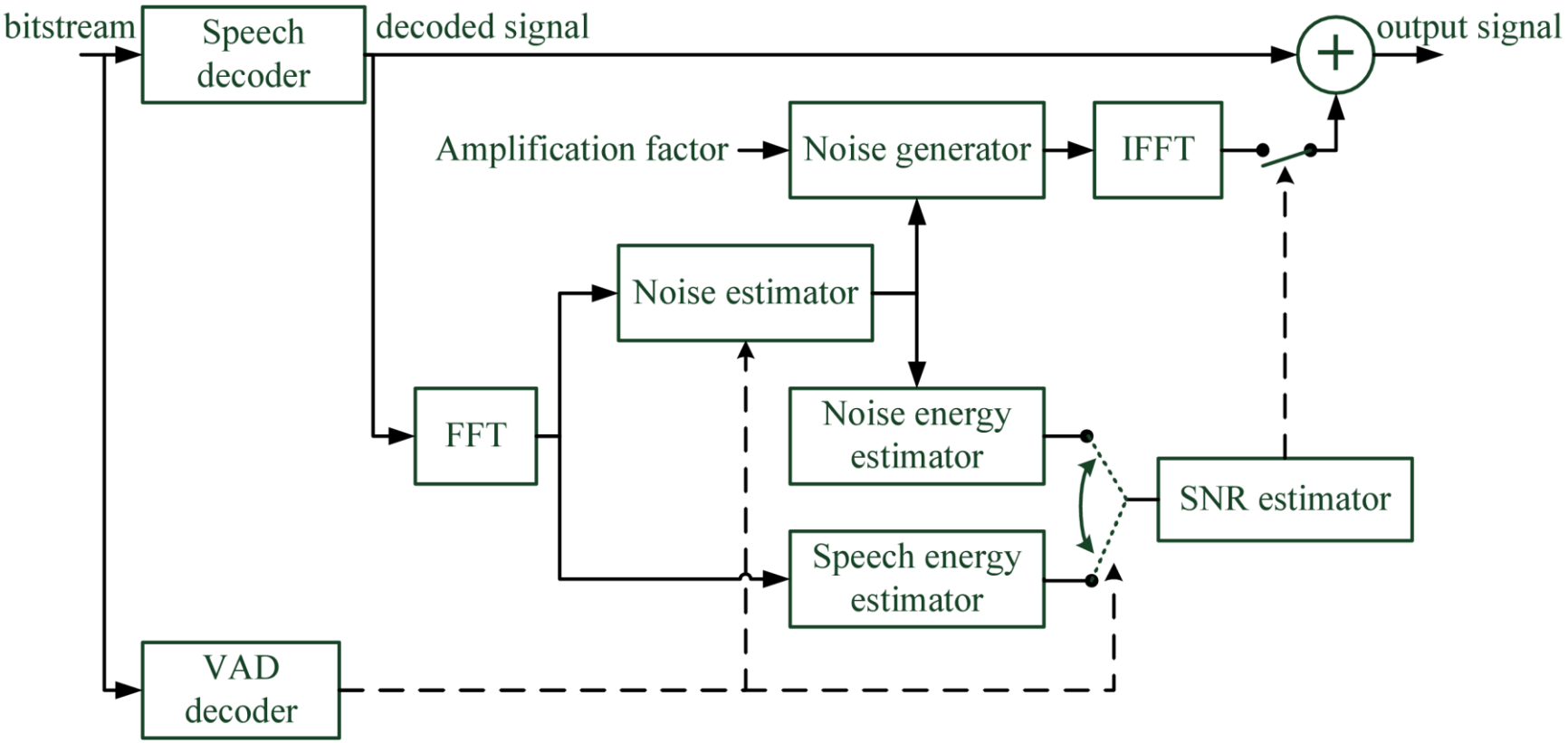
Comfort Noise Addition
Principle

- Characteristics
  - Works as a post-processor
  - Does not need any extra side-information
  - Is only used when estimated SNR is low (<28 dB)

- Aims
  - Mask coding artefacts and discontinuities
  - Compensate the loss of energy in the background noise
  - Can be used in DTX on or off
  - Can be combined with speech enhancement techniques

Comfort Noise Addition

Block diagram



Comfort Noise Addition

Noise estimation

- Based on **Minimum Statistics**, and relies on two assumptions
  - Speech and noise are statistically independent
  - Power of the noisy signal often decays to the power level of the noise signal (e.g., in speech pauses)

$$E_n(n,k) = B(n,k) \cdot \min_{i=n-D+1\ldots n}(P(i,k))$$

where

  - $B(n,k)$ is a bias compensation factor (overestimation factor)
  - $P(i,k)$ is the recursively smoothed periodogram of the decoded signal
  - D analysis window size of about 1.5s

- **DTX off**: noise estimated only on **inactive** frames
- **DTX on**: non-artificial noise only present and estimated on speech pauses of **active** frames. Estimation complements the CNG of DTX.

Comfort Noise Addition

Noise injection

- Artificial noise is injected in both **active** and **inactive** frames.
- Added noise $w(k)$ is generated by a Gaussian noise generator.
- $w(k)$ is shaped by the noise estimate and scaled to reach a targeted gain of the background noise energy:

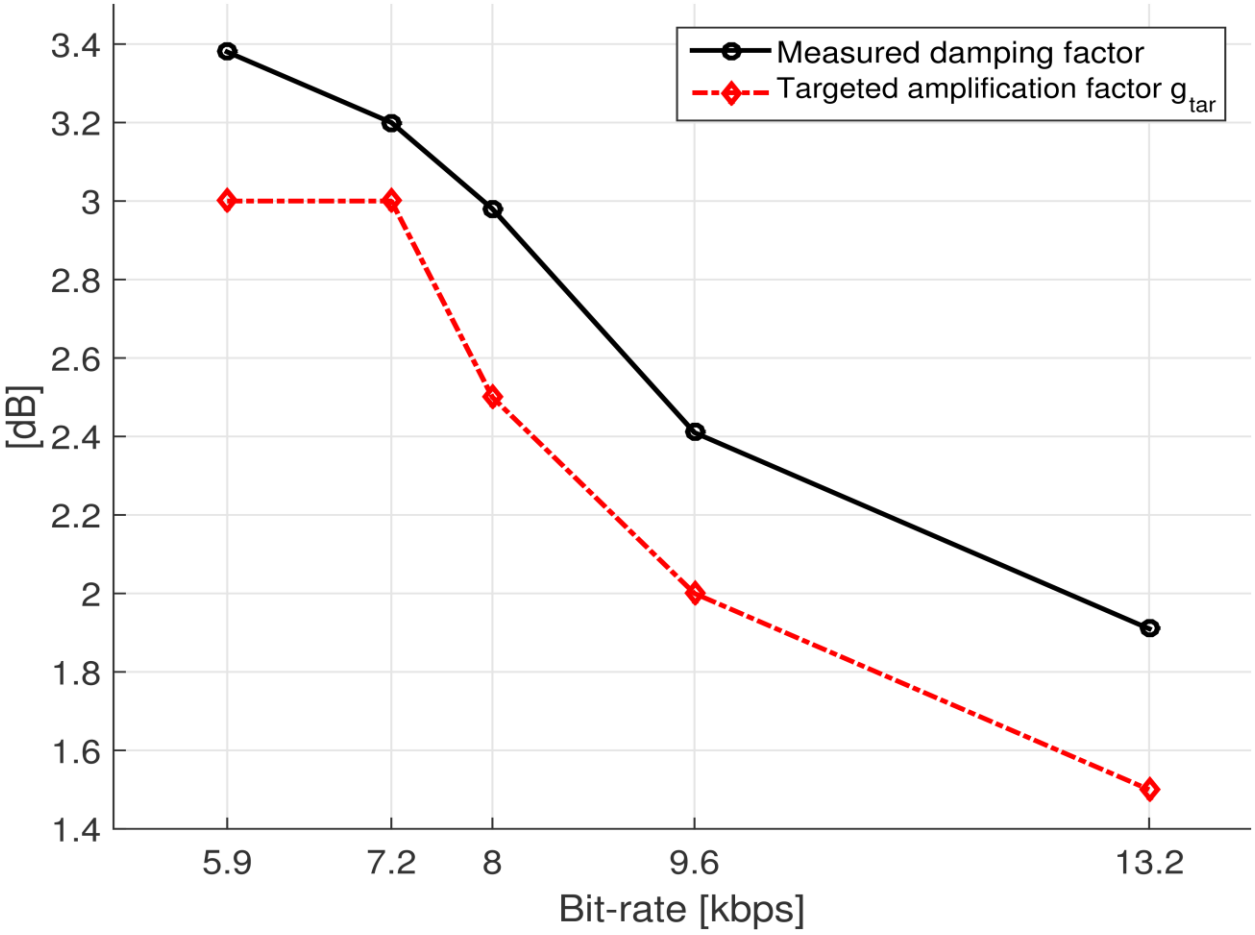
$$E_w(n,k) = \max((g_{tar} - 1) \cdot E_n(n,k), 0)$$

, where

- g_{tar} is the targeted amplification factor for the background noise.

Comfort Noise Addition

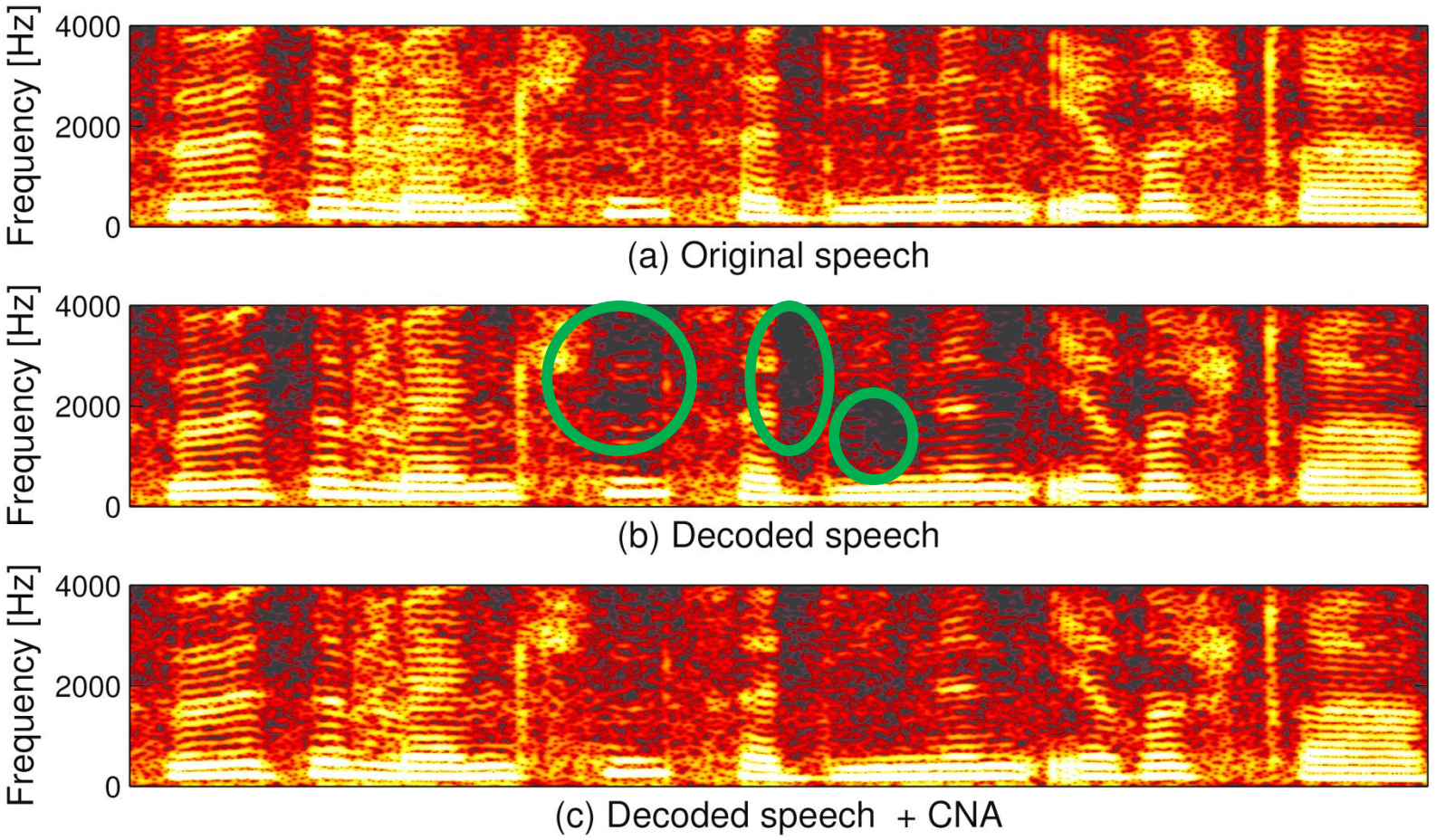
Noise injection – targeted amplification factor



Fraunhofer IIS

Comfort Noise Addition

Illustration of the effect



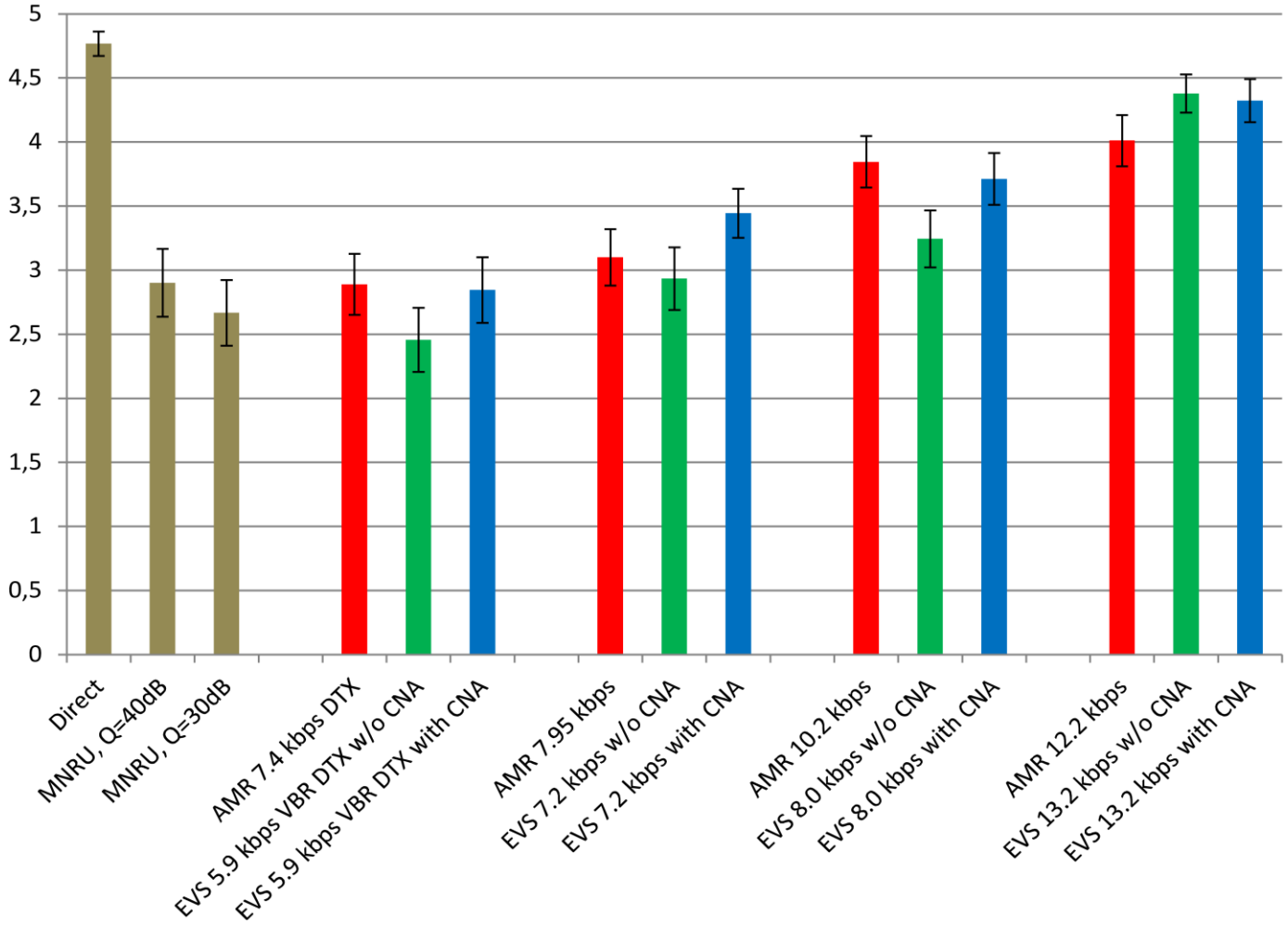
Evaluation

Subjective listening tests

- Noisy speech NB & WB
- P.800 DCR
  - Naive listeners (15 for NB, 18 for WB)
- DTX on and off
- Conditions:
  - Direct
  - MNRUs
  - AMR-WB
  - EVS with CNA
  - EVS without CNA

Evaluation : P.800 results

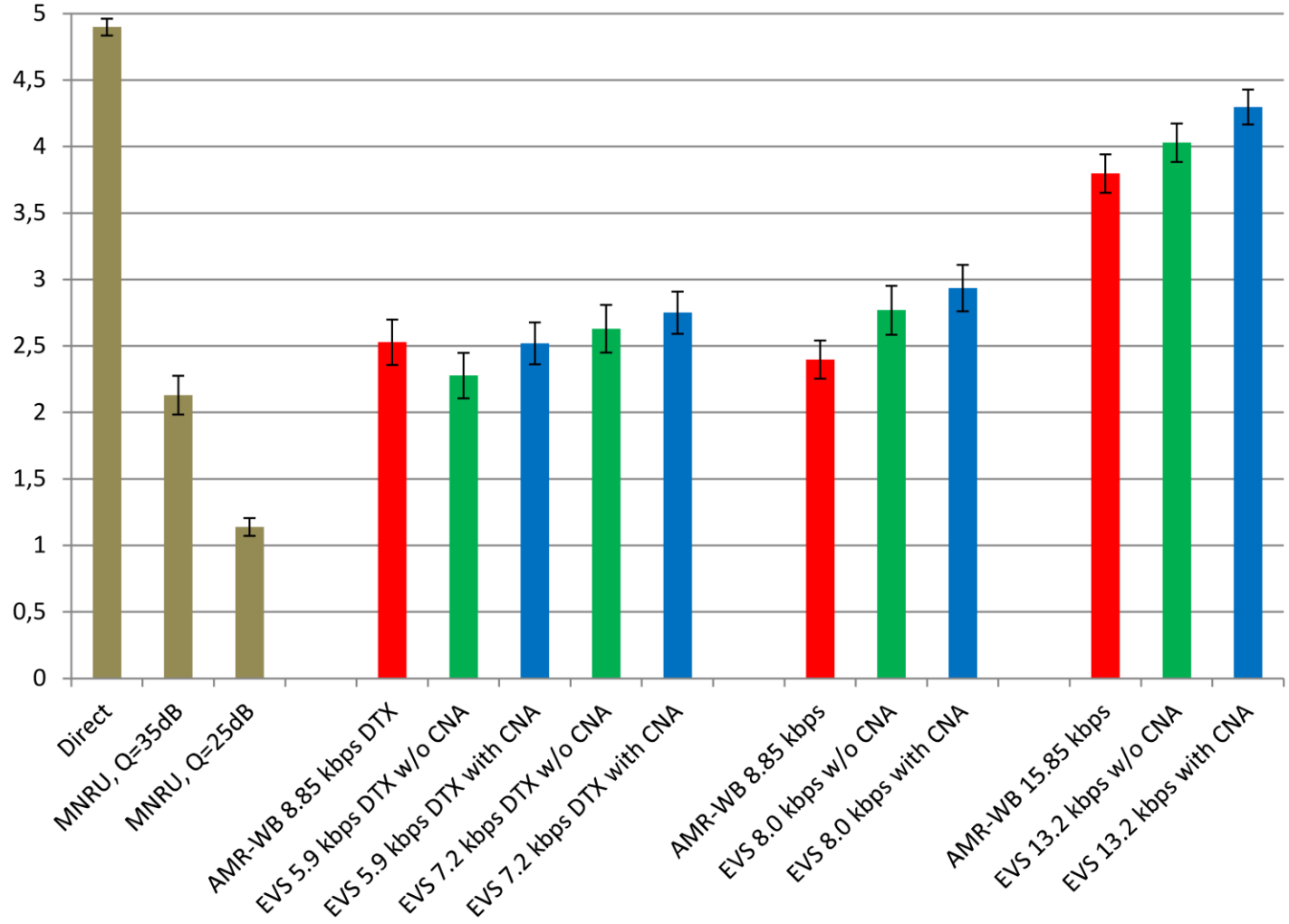
NB speech over 15 dB SNR car noise



Fraunhofer IIS

Evaluation : P.800 results

WB speech over 20 dB SNR street noise



Fraunhofer IIS

Conclusions

- CNA is a new speech coding post-processor for noisy speech:
  - Is a good complementary tool to speech enhancement techniques
  - Works in DTX on or off
  - Works over conventional speech coders
  - Improves the perceptual quality of noisy speech coded at low bit-rates
  - Is adopted in EVS NB, WB and AMR-WB IO for rates lower or equal to 13.2 kbps

Thank you for your attention!

Questions?