A Priori SNR Estimation Using a Generalized Decision Directed Approach

Aleksej Chinaev and Reinhold Haeb-Umbach, University of Paderborn, Germany, {chinaev, haeb}@nt.upb.de, http://nt.upb.de

PADERBORN UNIVERSITY The University for the Information Society

Computer Science, Electrical Engineering and Mathematics

NT Communications Engineering Prof. Dr.-Ing. Reinhold Häb-Umbach

Introduction

- Generalized spectral subtraction has been shown to be superior.
- Here we generalize the Decision Directed method for an improved estimation of a priori SNR $\xi(k,\ell)$.
- Generalization: Move from the PSD domain to a generalized ρ-domain
  - Generalized a priori SNR

$$\xi_\rho(k,\ell) \triangleq \frac{E[|S(k,\ell)|^{2\rho}]}{E[|D(k,\ell)|^{2\rho}]} \qquad (1)$$

  with clean speech STFT $S(k,\ell)$, noise STFT $D(k,\ell)$ and an arbitrary $\rho \in \mathbb{R}_{>0}$
  - Generalized a posteriori SNR

$$\gamma_\rho(k,\ell) \triangleq \frac{|Y(k,\ell)|^{2\rho}}{E[|D(k,\ell)|^{2\rho}]}$$

  with noisy STFT $Y(k,\ell) = S(k,\ell) + D(k,\ell)$

Decision Directed (DD)

- Assumption: Uncorrelated, complex normal distributed $S(k,\ell)$ and $D(k,\ell)$
  - $\gamma(k,\ell)$ is exponentially distributed
- Most popular a priori SNR estimator:

$$\hat{\xi}^{\text{DD}}(k,\ell) = \alpha \cdot \tilde{\xi}(k,\ell-1) + (1-\alpha) \cdot \hat{\xi}^{\text{ML}}(k,\ell)$$

  - propagated a priori SNR of previous processing step

$$\tilde{\xi}(k,\ell-1) = G^2(k,\ell-1) \cdot \hat{\gamma}(k,\ell-1)$$

  - maximum likelihood (ML) estimate based on the current observation

$$\hat{\xi}^{\text{ML}}(k,\ell) = \max(\hat{\gamma}(k,\ell)-1\,,\,0)$$

- Drawback: Slow response to an abrupt change in the instantaneous SNR

Generalized DD (GDD)

- Generalized a posteriori SNR

$$\gamma_\rho(k,\ell) = \gamma^\rho(k,\ell)/\Gamma(\rho+1) \qquad (2)$$

  is Weibull distributed

$$p_{\gamma_\rho(k,\ell)}(\gamma_\rho) = \text{Weib}(\gamma_\rho;\ \xi_\rho(k,\ell),\ \rho)$$

  with $\xi_\rho(k,\ell)$ as a parameter.
- Estimation of $\xi_\rho(k,\ell)$ using DD approach:

$$\hat{\xi}_\rho^{\text{DD}}(k,\ell) = \alpha_\rho \cdot \tilde{\xi}_\rho(k,\ell-1) + (1-\alpha_\rho) \cdot \hat{\xi}_\rho^{\text{ML}}(k,\ell)$$

  - propagated a priori SNR of previous processing step

$$\tilde{\xi}_\rho(k,\ell-1) = G^{2\rho}(k,\ell-1) \cdot \hat{\gamma}_\rho(k,\ell-1)$$

  - ML estimate of $\xi_\rho(k,\ell)$ from the current observation

$$\hat{\xi}_\rho^{\text{ML}}(k,\ell) = [\max(\hat{\gamma}(k,\ell)-1\,,\,0)]^\rho$$

- Two ways to get a priori SNR estimate
  - Conventional way: Based on definition (1)

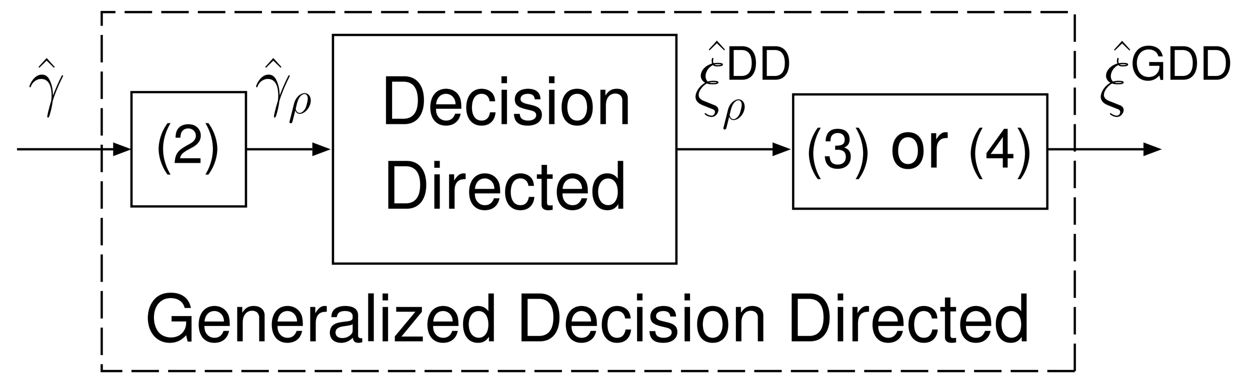
$$\hat{\xi}^{\text{GDD}}(k,\ell) = \left[\hat{\xi}_\rho^{\text{DD}}(k,\ell)\right]^{\frac{1}{\rho}} \qquad (3)$$

  - Alternative way: Considering $\hat{\xi}_\rho^{\text{DD}}(k,\ell)$ as a realization of a random variable

$$\xi_\rho^{\text{DD}}(k,\ell) \triangleq \frac{|S(k,\ell)|^{2\rho}}{E[|D(k,\ell)|^{2\rho}]}$$

  one has an alternative GDD estimate

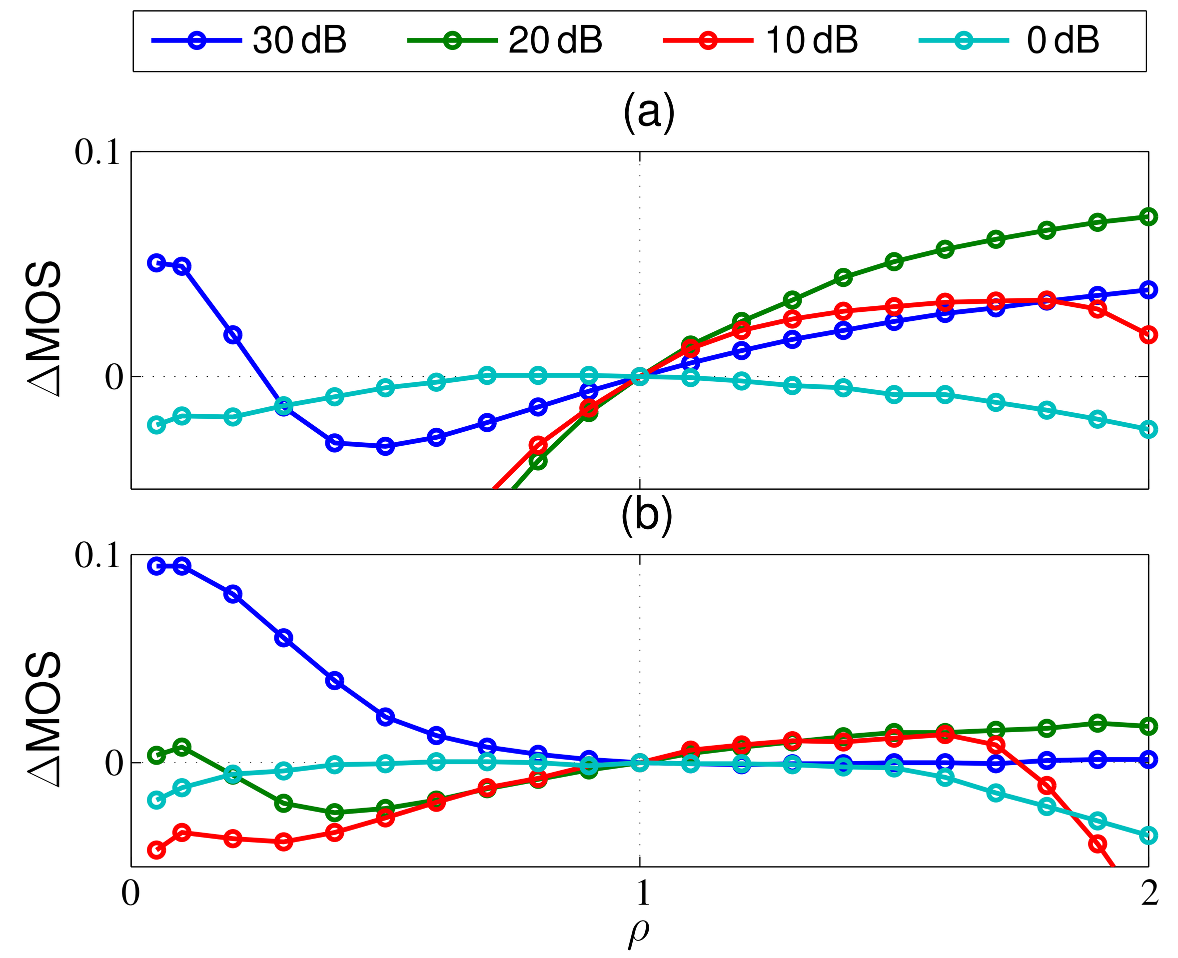
$$\hat{\xi}^{\text{GDD}}(k,\ell) = \left[\hat{\xi}_\rho^{\text{DD}}(k,\ell) \cdot \Gamma(\rho+1)\right]^{\frac{1}{\rho}}. \qquad (4)$$



- GDD simplifies to DD method for $\rho = 1$.

Parameterization of GDD

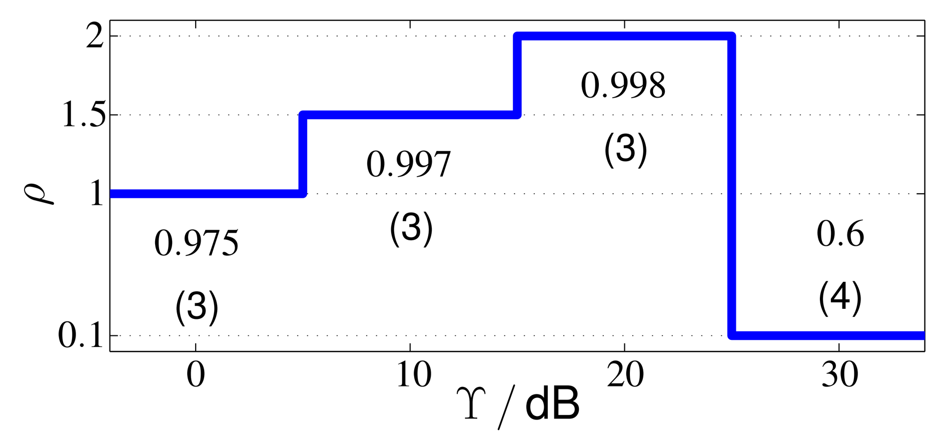
- Set parameters ρ and α_ρ in experiments with white noise (Noisex-92)



Gain of mean opinion score (ΔMOS) of the GDD over the DD method using α_ρ^{opt} for global input SNR $\{0, 10, 20, 30\text{ dB}\}$: (a) for (3), (b) for (4).

  - Low input SNR: DD method is an optimal choice for a priori SNR estimation.
  - Medium to high input SNR: The GDD approach outperforms the DD method.
  - Optimal choice of ρ and α_ρ depends on the global input SNR.

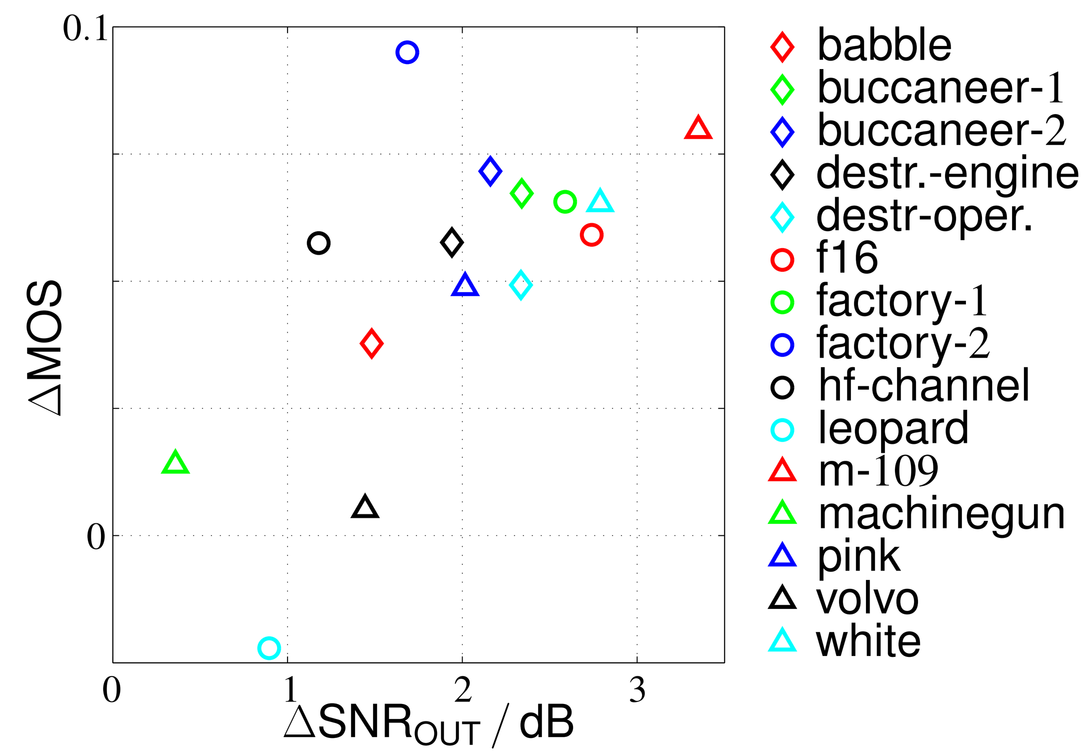
- Estimate a global input SNR Υ and apply an adaptation function $\rho(\Upsilon)$:



$\rho(\Upsilon)$ with corresponding values of α_ρ^{opt} and applied final equations.

Experimental results

- Performance gain of the GDD method with $\rho(\Upsilon)$ over the DD approach



Gain in terms of ΔMOS and a global output SNR ($\Delta\text{SNR}_{\text{OUT}}$) for different noise types (Noisex-92) averaged over results for $\Upsilon \in \{0, 10, 20, 30\text{ dB}\}$.

- Averaged over all noise types:

$$\Delta\text{MOS} \approx 0.05 \text{ and } \Delta\text{SNR}_{\text{OUT}} \approx 2\,\text{dB}$$

  - $\Upsilon_{\text{AVG}} = 15$ dB: from 18 dB up to 20 dB

Υ / dB	0	10	20	30
ΔMOS	0	0.01	0.04	0.15
$\Delta\text{SNR}_{\text{OUT}}$ / dB	0.5	1.1	1.2	5

- CHiME-3 database: Without loss of speech quality $\Delta\text{SNR}_{\text{OUT}} = 0.9$ dB
  - $\Upsilon_{\text{AVG}} = 5.8$ dB: from 9.6 dB up to 10.5 dB

Conclusions

- Extension of the Decision Directed approach for the use in the generalized SNR domain
- Notable increase in noise suppression with a moderate gain in speech quality