A generalized log-spectral amplitude estimator for single-channel speech enhancement

Aleksej Chinaev, Reinhold Haeb-Umbach

Department of Communications Engineering
Paderborn University

March 8, 2017

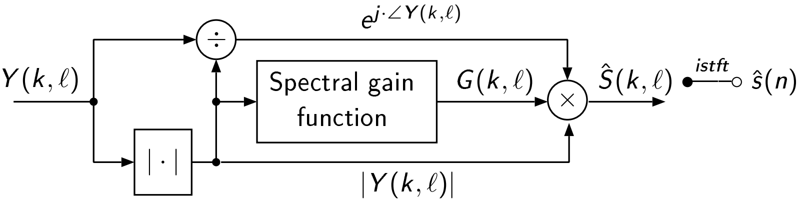
Table of contents

① Generalized single-channel spectral speech enhancement

- Clean speech $s(n)$ contaminated by an additive noise $d(n)$

$$y(n) = s(n) + d(n) \quad \overset{stft}{\circ\!\!-\!\!\bullet} \quad Y(k,\ell) = S(k,\ell) + D(k,\ell)$$

- Conventional single-channel spectral speech enhancement system



- 2 types of gain functions $G(k,\ell) \in \mathbb{R}_{>0}$ for generalized estimators
  a) Generalized spectral amplitude (GSA) $|S(k,\ell)|^{\alpha}$ for $\alpha \in \mathbb{R}_{>0}$
  b) Generalized probability density function (PDF) $p_S(s)$ for $\alpha = 1$

Generalized model-based gain functions

	a) GSA $\lvert S\rvert^{\alpha}$	b) Gen. PDF	Gen. Gamma	Gamma	Chi	Weibull	MMSE	MAP	log-estimator
1. [Sim, 1998]	✓						✓		
2. [You, 2003]	✓						✓		
3. [Dat, 2005]		✓	✓					✓	
4. [Lotter, 2005]		✓		✓				✓	
5. [Andrianakis, 2006]		✓		✓	✓		✓	✓	
6. [Erkelens, 2007]		✓	✓	✓	✓		✓		
7. [Breithaupt, 2008]	✓	✓			✓		✓		
8. [Borgstrom, 2011]		✓		✓	✓		✓		✓
9. [Zhao, 2012]		✓		✓			✓		✓
10. Proposed	✓	✓				✓		✓	✓

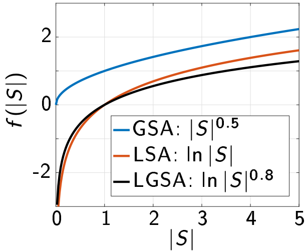
A novel nonlinearity in a MAP-based estimator

- Merging two nonlinearities $f(\cdot)$ of MMSE spectral estimators

$$f(|\hat{S}|) = \mathbb{E}[\, f(|S|) \,|\, Y]$$

- $f_1(|S|) = |S|^\alpha$ e.g. with $\alpha = 0.5$ from super Gaussian amplitude root estimator [Breithaupt, 2008]

- $f_2(|S|) = \ln |S|$ from perceptually motivated log-spectral amplitude (LSA) estimator [Ephraim, 1985]

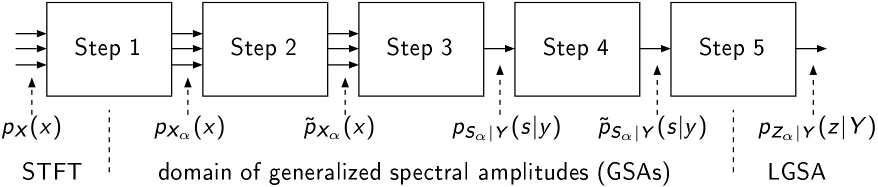


- A novel nonlinearity $\boxed{Z_\alpha = f_3(|S|) = \ln |S|^\alpha}$ resulting in a logarithmic GSA (LGSA) e.g. with $\alpha = 0.8$ for a MAP estimator

$$f_3(|\hat{S}|) = \arg\max_z \; p_{Z_\alpha|Y}(z|y)$$

② Logarithmic generalized spectral amplitude (LGSA) estimator

- 5 steps from PDFs $p_X(x)$ for $X \in \{S, D, Y\}$ to PDF $p_{Z_\alpha|Y}(z|y)$



- Step 1: Transformation to GSAs $X_\alpha = |X|^\alpha \Rightarrow p_{X_\alpha}(x)$
- Step 2: Approximation by consistent Gaussians $\Rightarrow \tilde{p}_{X_\alpha}(x)$
- Step 3: Calculation of $p_{S_\alpha|Y}(s|y)$ under assumption $Y_\alpha = S_\alpha + D_\alpha$
- Step 4: Zero-truncation for log-transformation $\Rightarrow \tilde{p}_{S_\alpha|Y}(s|y)$
- Step 5: Transformation to logarithmic GSA $Z_\alpha = \ln S_\alpha$

Statistical modelling of generalized spectral amplitudes

- Step 1: Transformation to GSAs $X_\alpha = |X|^\alpha$ for $X \in \{S, D, Y\}$

$$p_X(x) = \mathcal{N}_C\left(x; 0, \lambda_X\right) \quad \Rightarrow \quad p_{X_\alpha}(x) = \mathrm{Weib}(x;\, \lambda_X,\, \alpha)$$

with power spectral densities $\lambda_X(k,\ell) = \mathbb{E}[|X(k,\ell)|^2]$.

- Step 2: Approximation by consistent Gaussians $\Rightarrow \tilde{p}_{X_\alpha}(x)$

$$\mathrm{Weib}(x;\, \lambda_X,\, \alpha) \approx \mathcal{N}\left(x; \mu_X,\, \sigma_X^2\right)$$

with μ_X and $\sigma_X^2 = g(\alpha) \cdot \mu_X^2$ using statistical moment matching.

- Step 3: Calculation of $p_{S_\alpha|Y}(s|y)$ under assumption $Y_\alpha = S_\alpha + D_\alpha$

$$p_{S_\alpha|Y}(s|y) = \mathcal{N}(s; \mu_{S|Y},\, \sigma^2_{S|Y})$$

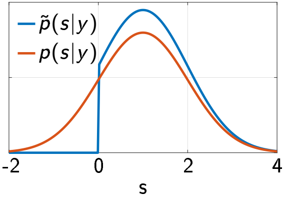
with MMSE-GSS estimator $\hat{S}_\alpha^{\mathrm{GSS}} = \mu_{S|Y} = G_\alpha^{\mathrm{GSS}} \cdot Y_\alpha$ [Sim, 1998].

Zero-truncation and logarithmic transformation

- Step 4: Zero-truncation of $p_{S_\alpha|Y}(s|y)$ for log-transformation

$$p_{S_\alpha|Y}(s|y) \approx \tilde{p}_{S_\alpha|Y}(s|y)$$

to get a conditional PDF only for positive GSA values $S_\alpha > 0$.



- Step 5: Transformation to logarithmic GSA $Z_\alpha = \ln S_\alpha$

$$p_{Z_\alpha|Y}(z|y) = \frac{e^z}{Q\left(-\frac{\mu_{S|Y}}{\sigma_{S|Y}}\right)} \cdot \mathcal{N}(e^z; \mu_{S|Y}, \sigma^2_{S|Y})$$

with the complementary cumulative distribution of the standard normal density $Q(x)$.

MAP-based LGSA estimator

- Computationally efficient MAP-based LGSA estimator

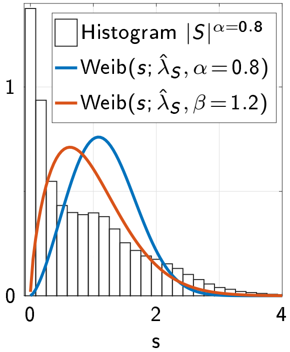
$$\hat{S}_{\alpha}^{\mathrm{LGSA}} = \frac{\mu_{S|Y}}{2} + \sqrt{\left(\frac{\mu_{S|Y}}{2}\right)^2 + \sigma_{S|Y}^2} = G_{\alpha}^{\mathrm{LGSA}} \cdot Y_{\alpha}$$

- Additional modeling freedom
  - Decouple compression factor α from a Weibull shape parameter β

$$p_{X_{\alpha}(k,\ell)}(x) = \mathrm{Weib}(x;\, \lambda_X(k,\ell),\, \beta)$$

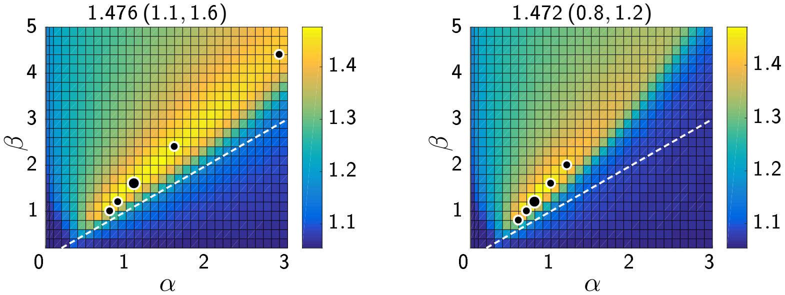
  - Resulting in a spectral gain dependent only on β

$$\boxed{\hat{S}_{\alpha}^{\mathrm{LGSA}} = G_{\beta}^{\mathrm{LGSA}} \cdot Y_{\alpha}}$$



③ Parameter optimization of GSS and LGSA gain functions

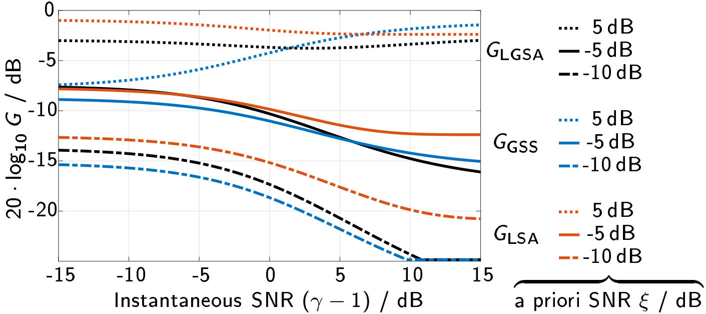
- A loosely coupled version of the MMSE-based generalized spectral subtraction (GSS) estimator $\hat{S}_{\alpha}^{\mathrm{GSS}} = G_{\beta}^{\mathrm{GSS}} \cdot Y_{\alpha}$ [Sim, 1998]



(a) MMSE-GSS (b) MAP-LGSA

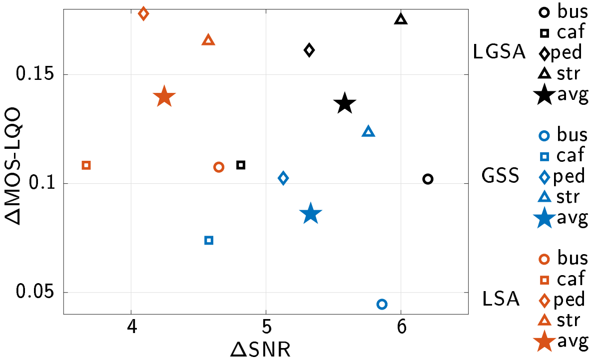
- Quality of denoised speech titled by $\text{MOS-LQO}_{\text{opt}}(\alpha_{\text{opt}}, \beta_{\text{opt}})$
  - $\text{SNR}_{\text{IN}} = 5\,\text{dB}$: The same quality in operating point $\beta_{\text{opt}} > \alpha_{\text{opt}}$
  - $\text{SNR}_{\text{IN}} = \{\text{-5} : 5 : 15\}\text{dB}$: GSS optima scatter more than of LGSA

Discussion of resulting gain curves after optimization



- Reducing musical tones: decreasing of curves with growing γ
  - LSA: Price to pay for high quality - weak noise suppression
  - GSS: Good noise suppression but poor speech quality
  - LGSA: LSA behavior for higher ξ with better noise suppression

④ Experimental results on CHiME-3 database



- Speech quality improvement over noise suppression gain
  - LSA: Good overall speech quality and poor noise suppression
  - GSS: Better noise suppression on cost of speech quality
  - LGSA: The best noise suppression w.o. loss of speech quality

⑤ Conclusions

- Generalized model-based spectral speech enhancement
- Modified version of MMSE-based GSS estimator from [Sim, 1998]
- Novel logarithmic generalized spectral amplitude (LGSA) estimator
  - New perceptually motivated nonlinearity $f(|S|) = \ln |S|^{\alpha}$
  - Computationally efficient MAP-based gain function
- Optimization of proposed estimators for specific global input SNR
- Better tradeoff speech quality/noise suppression for LGSA estimator

Thank you very much for your attention!

Sound examples

Example 227

	clean	noisy	MMSE-LSA	MMSE-GSS	MAP-LGSA
MOS-LQO		1.208	**1.377**	1.318	**1.372**
SNR/dB		2.29	9.84	**12.16**	**12.18**

Example 649

	clean	noisy	MMSE-LSA	MMSE-GSS	MAP-LGSA
MOS-LQO		1.346	**1.454**	1.417	**1.468**
SNR/dB		9.40	12.62	**13.42**	**13.64**

Example 1280

	clean	noisy	MMSE-LSA	MMSE-GSS	MAP-LGSA
MOS-LQO		1.219	1.611	1.605	**1.664**
SNR/dB		4.40	11.70	**13.89**	**13.83**

Subfamilies of Generalized Gamma distribution

- Generalized Gamma distribution with three parameters

$$\mathrm{GenGam}(x;\, \alpha, \tau, \lambda) = \frac{2}{\alpha \cdot \lambda \cdot \Gamma(\tau)} \cdot x^{\frac{2\tau}{\alpha} - 1} \cdot \exp\left(-\frac{x^{2/\alpha}}{\lambda^{1/\tau}}\right)$$

- Subfamilies of $\mathrm{GenGam}(x)$: Gamma, Chi and Weibull distributions

Probability distribution	Degree of freedom 1	Degree of freedom 2	Scale parameter
Generalized Gamma	τ	α	λ
Gamma	τ	2	λ
Chi	τ	1	λ
Weibull	1	α	λ