A Priori SNR Estimation Using Weibull Mixture Model

12. ITG Fachtagung Sprachkommunikation

Aleksej Chinaev, Jens Heitkaemper, Reinhold Haeb-Umbach

Department of Communications Engineering
Paderborn University

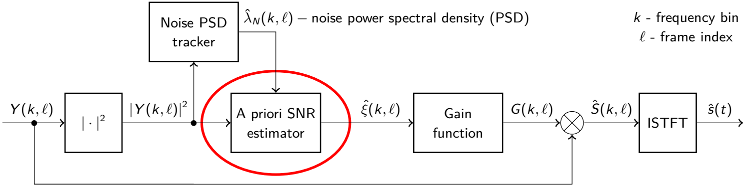
7. Oktober 2016

Table of contents

Problem formulation and motivation

- Single-channel clean speech $s(t)$ contaminated by an additive noise $n(t)$:

$$y(t) = s(t) + n(t) \quad \overset{STFT}{\circ\!\!-\!\!\bullet} \quad Y(k,\ell) = S(k,\ell) + N(k,\ell)$$



- A priori SNR $\xi(k,\ell) = \frac{\lambda_S(k,\ell)}{\lambda_N(k,\ell)}$ – a key component in enhancement system

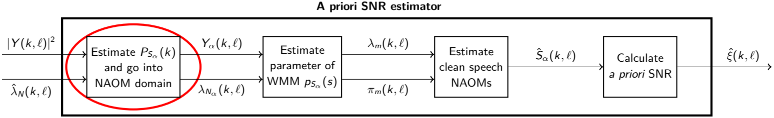
$\lambda_S(k,\ell) = \mathrm{E}\left[|S(k,\ell)|^2\right]$ - clean speech PSD, $\lambda_N(k,\ell) = \mathrm{E}\left[|N(k,\ell)|^2\right]$ - noise PSD

- Motivated by a generalized spectral subtraction (GSS) denoising $|Y(k,\ell)|^\alpha$ for $\alpha \in \mathbb{R}_{>0}$ not restricted to $(\alpha = 1)$ or $(\alpha = 2)$ with assumption

$$|Y(k,\ell)|^\alpha = |S(k,\ell)|^\alpha + |N(k,\ell)|^\alpha$$

Table of contents

Normalized α-order magnitude (NAOM) domain



- Normalize $|Y(k,\ell)|^\alpha$ to a root of an averaged power $P_{S_\alpha}(k)$ of $|S(k,\ell)|^\alpha$

$$Y_\alpha(k,\ell) = \frac{|Y(k,\ell)|^\alpha}{\sqrt{P_{S_\alpha}(k)}} = S_\alpha(k,\ell) + N_\alpha(k,\ell) \quad \text{with} \quad P_{S_\alpha}(k) = \frac{1}{L}\sum_{\ell=1}^{L} |S(k,\ell)|^{2\alpha}$$

  - Statistical models independent of speaker loudness
  - Normalized energy of clean speech NAOMs $E[S_\alpha^2(k)] = 1$

- $S_\alpha(k,\ell)$ & $N_\alpha(k,\ell)$ – realizations of random variables $S_\alpha(k)$ & $N_\alpha(k)$

Estimate $S_\alpha(k,\ell)$ from $Y_\alpha(k,\ell)$ given models for $S_\alpha(k)$ & $N_\alpha(k)$

Modeling of noise NAOM coefficients $N_\alpha(k,\ell)$

- $N(k,\ell) \sim \mathcal{N}_c(n; 0, \lambda_N(k,\ell))$
- $N_\alpha(k,\ell)$ – Weibull distributed

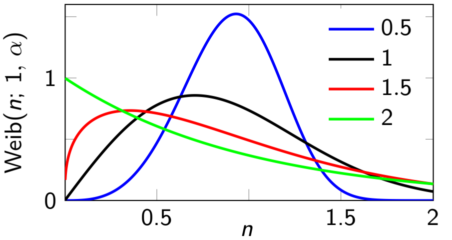
  $p_{N_\alpha(k,\ell)}(n) = \mathrm{Weib}(n; \lambda_{N_\alpha}(k,\ell), \alpha)$
  - Shape parameter $\alpha \in \mathbb{R}_{>0}$
  - Scale parameter

  $$\lambda_{N_\alpha}(k,\ell) = \frac{\lambda_N(k,\ell)}{\sqrt[\alpha]{P_{S_\alpha}(k)}} \in \mathbb{R}_{>0}$$
- Model $N_\alpha(k)$ with Weibull PDF

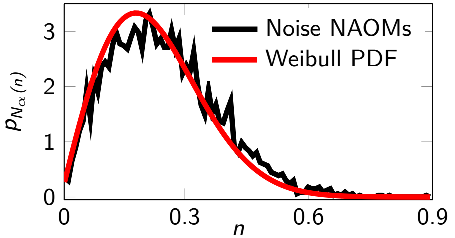
  $$p_{N_\alpha(k)}(n) = \mathrm{Weib}(n; \lambda_{N_\alpha}(k), \alpha)$$

  with $\lambda_{N_\alpha}(k) = \frac{1}{L}\sum_{\ell=1}^{L} \lambda_{N_\alpha}(k,\ell)$
- NAOM coefficients of white noise signal and estimated $p_{N_\alpha(k)}(n)$

Weibull PDF for $\lambda = 1$ and different α



Histogram and Weibull PDF for $\alpha = 0.7$



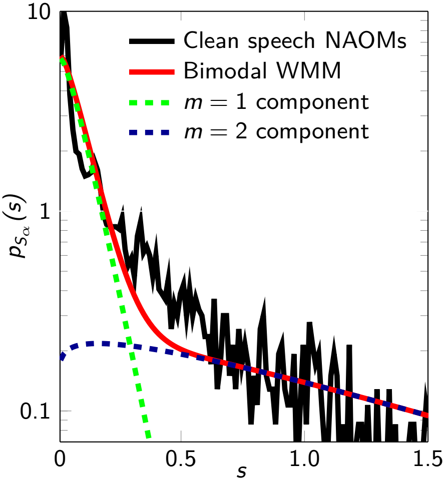
Modeling of NAOM coefficients of clean speech $S_\alpha(k,\ell)$

- $S(k,\ell) \sim \mathcal{N}_c(n; 0, \lambda_S(k,\ell))$
- Bimodal Weibull mixture model (WMM) to model $S_\alpha(k)$

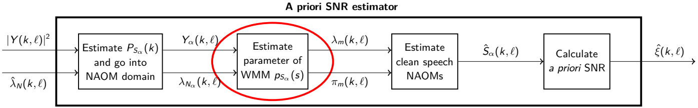
$$p_{S_\alpha(k)}(s) = \sum_{m=1}^{2} \pi_m(k) \cdot \mathrm{Weib}(s;\ \lambda_m(k),\ \beta)$$

  - $m = 1$: silence
  - $m = 2$: activity
  - $\pi_m(k) \in [0,\ 1]$: weights
  - $\lambda_m(k)$: scale parameters
  - β: shape parameter
- $\beta \neq \alpha$: additional degree of freedom in the model
- Clean speech NAOMs & estimated WMM ($\alpha = 0.7$; $\beta = 2.5$)

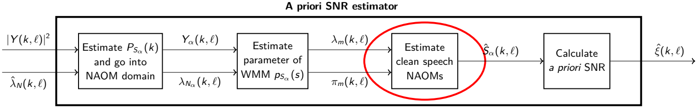
Histogram and estimated WMM



Estimation of WMM parameters and clean speech NAOMs



- Set $\lambda_1(k)$ acc. to ξ_{min} usually used in *a priori* SNR estimation [Cappe 94]
- Expectation Maximization algorithm to estimate $\lambda_2(k)$, $\pi_m(k)$
  - After EM, weights $\pi_m(k)$ are corrected with the constraint $E[S_\alpha^2(k)] = 1$

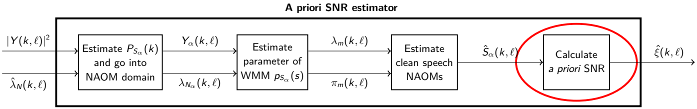


- Maximum a posteriori (MAP) estimation:

$$\hat{S}_\alpha^{\text{MAP}}(k,\ell) = \underset{s}{\operatorname{argmax}}\; p_{S_\alpha(k)\,|\,Y_\alpha(k,\ell)}(s|y)$$

  - $Y_\alpha(k,\ell)$ is a realisation of random variable $Y_\alpha(k) = S_\alpha(k) + N_\alpha(k)$
  - Approximative computationally efficient solution for $\beta = \alpha = 1$

Calculation of *a priori* SNR and causal implementation



- Go back into domain of power spectral density by calculating

$$\hat{\xi}(k,\ell) = \max\left(\frac{\left[\hat{S}_\alpha(k,\ell)\cdot\sqrt{P_{S_\alpha}(k)}\right]^{\frac{2}{\alpha}}}{\lambda_N(k,\ell)}, \xi_{\min}\right)$$

Causal implementation of WMM-based *a priori* SNR estimators

- Calculate $P_{S_\alpha}(k)$ and $\lambda_{N_\alpha}(k)$ in a causal way
- Causal EM for $\lambda_2(k)$ and $\pi_2(k)$ with one EM-iteration per time frame

- Note, parameters α and β have to be set appropriately $\rightarrow$ optimization

Table of contents

Experimental evaluation

Data and setup

- Clean speech: *Wall Street Journal* database 16 kHz (male and female)
- 7 different noise types of *Noisex92* database: *white*, *pink*, *f16*, *hfchannel*, *factory-1*, *factory-2*, *babble*
- Input global SNR from -5 dB up to 25 dB in 5 dB steps

Spectral speech enhancement framework

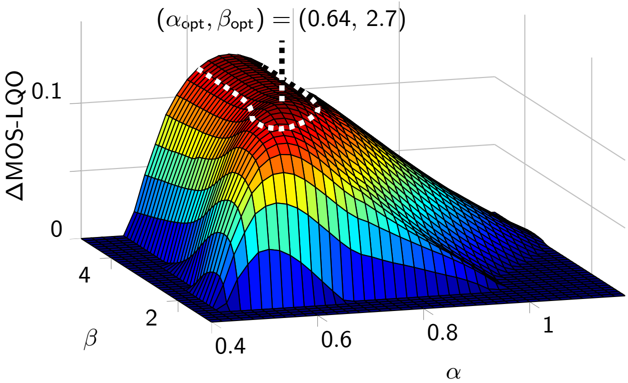
- Noise PSD tracking using *Minimum statistics* approach [Martin 01]
- A priori SNR estimation with $\xi_{\text{min}} = -18$ dB [Cappe 94]
  - Proposed WMM-based approach with Wiener filter
  - Reference approach: *Decision Directed* [Ephraim 84]

Optimization of α and β

- Speech quality maximization in terms of wide-band mean opinion score listening quality objective (MOS-LQO) with

$$\Delta\text{MOS-LQO} = \max(\,\text{MOS-LQO}_{\text{WMM}} - \text{MOS-LQO}_{\text{DD}}\,,\ 0\,)$$

- Averaging over genders, noise types and input global SNR values



Final experimental results

- Clean speech: WSJ database signals other than used for optimization
- Estimation error – Itakura-Saito distance (ISD) and estimator's variance – logarithmic error variance (LEV): the smaller the better

Resulting ISD, LEV and MOS-LQO values averaged over noise types

SNR, dB		−5	0	5	10	15	20	25	AVG
ISD	DD	48.8	44.0	39.6	34.9	30.2	24.5	19.1	34.4
	WMM	**42.6**	**38.1**	**34.1**	**30.4**	**27.3**	**23.0**	**18.9**	**30.6**
LEV	DD	53.1	49.0	46.4	45.1	45.5	47.4	50.5	48.1
	WMM	**45.6**	**43.9**	**42.6**	**41.1**	**39.0**	**37.0**	**35.9**	**40.7**
MOS-LQO	DD	1.11	1.30	1.63	2.09	2.57	3.00	3.39	2.16
	WMM	**1.18**	**1.46**	**1.77**	**2.13**	**2.62**	**3.16**	**3.61**	**2.28**

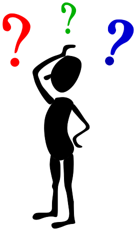
Conclusions and outlook

Conclusions

- Novel causal *a priori* SNR estimator based on a bimodal Weibull mixture model for the normalized α-order spectral magnitudes (NAOMs)
- Optimization of the proposed approach by maximization of speech quality
  - Power exponent $\alpha_{\text{opt}} = 0.64$ smaller than 1 (spectral magnitudes)
  - Shape factor $\beta_{\text{opt}} = 2.7$ – a heavier tailed Weibull distribution
- Compared to the wide-spread *Decision Directed* approach:
  - Reduced error and variance of the WMM-based *a priori* SNR estimator
  - Improvement of speech quality of the enhanced signals
  - Higher computational effort

Outlook

- Reduction of computational effort – fixed speaker-independent models
- Development of model-based spectral enhancement using generalized (arbitrary) power exponent in the spirit of generalized spectral subtraction



Thank you for your attention!

Questions?

Paderborn University

Department of
Communications Engineering

Web: nt.upb.de

Resulting WMM parameter and audio samples

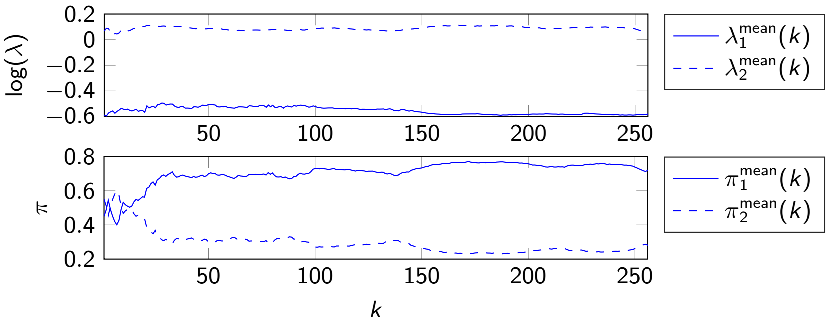


Figure : Resulting WMM parameter over frequency bins

- Exemplarily speech samples: Noisy DD WMM