MMSE SPEECH SPECTRAL AMPLITUDE ESTIMATION ASSUMING NON-GAUSSIAN NOISE

Balázs Fodor and Tim Fingscheidt

Institute for Communications Technology, Technische Universität Braunschweig
Schleinitzstr. 22, 38106 Braunschweig, Germany
{b.fodor, t.fingscheidt}@tu-bs.de

ABSTRACT

In many applications non-Gaussian noises, such as babble noise, can be observed. In this paper we present a minimum mean square error (MMSE) estimation of the speech spectral amplitude. It principally allows for arbitrary speech spectral amplitude probability density function (pdf) models (Rayleigh, Chi, ...), while the pdf of the noise DFT coefficients is modeled by a Gaussian mixture (GMM). Applying for both approaches an idealized a priori SNR estimator that works well in babble noise, we can show clear improvements compared to the MMSE spectral amplitude estimator with Gaussian noise assumption.

1. INTRODUCTION

In speech enhancement, a low level of speech distortion can be achieved by employing an appropriate speech model. Classically, the discrete Fourier transform (DFT) coefficients of the speech signal are commonly modeled by a Gaussian distribution [1, 2]. However, the actual speech content in a noisy signal can be better preserved by applying different speech amplitude priors, such as super-Gaussian [3], or more general, generalized gamma [4].

In analogy, further reduction of the residual noise level can be obtained by a proper selection of the noise model. In [3] it was shown that the histogram of the noise DFT coefficients is closer to a Gaussian distribution then it is the case for the speech DFT coefficients. However, this is not true for special noises, such as babble or fan noise. Therefore, instead of a Gaussian distribution [2, 3], in [4, 5] a Gaussian mixture model (GMM) of the noise DFT coefficients were employed in the context of a minimum mean sqare error (MMSE) estimator.

The system proposed in [5] has the disadvantage, that it does not have the "classical" speech enhancement structure with the noise power and SNR estimation steps. Therefore, it is quite inflexible against varying speech and/or noise power. The estimator proposed in [4] was designed for multichannel speech enhancement. It was shown that using the GMM noise model and a generalized gamma model for the speech, the proposed system cannot be separated into a beamformer and a postfilter. However, the performance of the new estimator was not evaluated.

Modeling the noise DFT coefficients by means of Gaussian mixtures has the advantage that—with an infinite number of mixtures—any distribution can be approximated perfectly. Such a model is suitable for environments with non-stationary non-Gaussian noise, such as interfering talkers (babble noise).

In this contribution we develop an MMSE estimator for speech enhancement under a non-Gaussian noise assumption which will be reflected by a GMM. For ease of comparison to [1], a Gaussian model of the speech DFT coefficients is employed.

The paper is organized as follows: Section 2 will address the problem of the noise spectral statistics. In Section 3, a short review of a reference MMSE estimator with Gaussian noise model will be given. Section 4 will present the new MMSE estimator based on a noise GMM, followed by the evaluation of the proposed weighting rule in Section 5. Finally, Section 6 will give some concluding remarks.

2. NOISE MODELING

Most spectral weighting rules for noise reduction assume that the probability density function (pdf) of the complex-valued noise DFT coefficients is Gaussian. Only few investigations have shown that this assumption does not hold for some noise types, such as fan or babble noise. Although according to the central limit theorem, the Gaussian noise model is adequate for a wide range of stationary noises, Lotter pointed out in [3] that there is no overall distribution, which fits for all possible noise histograms.

2.1 Histogram Measurement

In order to investigate the properties of non-Gaussian noises, we first measured the histogram of the DFT coefficients of babble noises (acquired, e. g., in shops, amusement parks, exhibitions, etc.) taken from the NTT Ambient Noise Database [6]. After typical signal analysis as used for speech enhancement[1], the real and imaginary parts of the noise DFT coefficients $N \in \mathbb{C}$ were subsequently processed separately. In order to make the data independent from the noise level, all (real and imaginary) DFT coefficients of a database file of 3 minutes were scaled to unit variance. After processing the whole data set of 24 files, histograms were computed as illustrated in Figure 1. Please note the significant deviation of the pdf of the babble noise DFT coefficients from the solid line Gaussian pdf with the same (unit) variance.

In order to further demonstrate the deviation of the measured pdf's from the Gaussian model, the kurtosis values [7]

$$\kappa = \frac{\mu_4}{\sigma^4} \qquad (1)$$

of the observations were computed, with μ_4 being the fourth moment about the mean, and σ being the standard deviation

[1] We worked at the sample rate of 8kHz with an analysis frame length of 256 samples, a Hann window and an analysis frame shift of 128 samples. Note that in the majority of the paper we omit frame index ℓ and frequency bin index k.

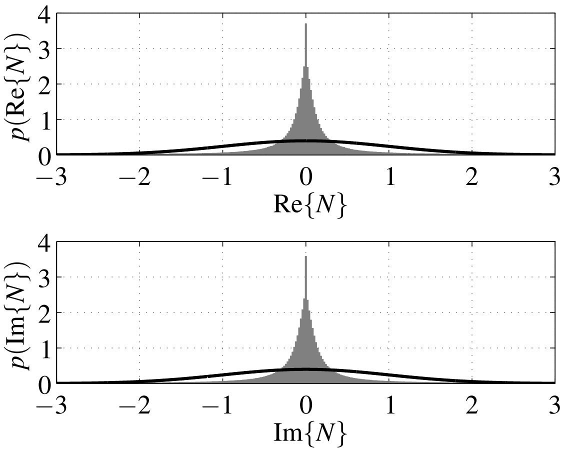

Figure 1: Histogram of babble noise DFT coefficients (black curve: Gaussian model, grey curve: histogram of the observation)

$\kappa_{\mathrm{Re}\{N\}}$	$\kappa_{\mathrm{Im}\{N\}}$	$\kappa_{\mathrm{Gaussian}}$
305.95	165.86	3

Table 1: Kurtosis of the real and imaginary parts of the observed babble noise DFT coefficients with unity variance

of $\mathrm{Re}\{N\}$ and $\mathrm{Im}\{N\}$, respectively. $\mathrm{Re}\{\cdot\}$ and $\mathrm{Im}\{\cdot\}$ denote the real and imaginary part operator. The results are summarized in Table 1. Kurtosis values greater than three (Gaussian case) are called super-Gaussian. The indication, that the histograms illustrated in Figure 1 are distributed highly non-Gaussian, is confirmed.

2.2 Gaussian Mixture Model

In order to allow a flexible pdf model in the DFT domain, we employed a Gaussian mixture model (GMM). Therefore, the pdf of the noise DFT coefficients in a 2-dimensional notation $\boldsymbol{N} = [\mathrm{Re}\{N\}\ \mathrm{Im}\{N\}]^T \in \mathbb{R}^2$ turns out to be:

$$p(\boldsymbol{N}) = \sum_{m=1}^{M} c_m \mathcal{N}(\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m) \quad (2)$$
$$= \sum_{m=1}^{M} \frac{c_m}{2\pi|\boldsymbol{\Sigma}_m|^{1/2}} e^{-\frac{1}{2}(\boldsymbol{N}-\boldsymbol{\mu}_m)^H \boldsymbol{\Sigma}_m^{-1}(\boldsymbol{N}-\boldsymbol{\mu}_m)},$$

with c_m, $\boldsymbol{\mu}_m$ and $\boldsymbol{\Sigma}_m$ representing the weight, the mean vector and the 2×2 covariance matrix of the m-th of M Gaussian modes, respectively.
It can be seen in Figure 1 that the histograms of the real and imaginary parts are very similar. Moreover, the pdf is zero-mean and the peak can be approximated by a sum of several zero-mean low-variance Gaussians.
GMM parameters according to (2) can be computed by the expectation maximization (EM) algorithm. As training data we used two thirds of the babble noise taken from the NTT Ambient Noise Database [6]. The EM training results showed that the covariance matrices are almost diagonal, the variance of the real and imaginary noise DFT coefficients were almost equal and the means were almost zero. Because of the peakedness of the measured histogram, the resulting GMM had many low-variance Gaussian modes. Therefore, the real and imaginary parts of the noise DFT coefficients can sufficiently well be described as independent and identically distributed (i.i.d.) non-Gaussian. Accordingly, the pdf of the complex-valued noise DFT coefficients

m	c_m	σ_m^2	$c_m\sigma_m^2/\sigma_{\mathrm{GMM}}^2$ [%]
1	0.05	14.84	71.98%
2	0.18	1.3	23.57%
3	0.26	0.13	3.53%
4	0.21	0.03	0.69%
5	0.31	0.01	0.23%

Table 2: Parameters of the noise GMM with $M = 5$. The third column represents the so-called variance contribution of the m-th Gaussian mode of the GMM.

M	D_{KL}^{Re}	D_{KL}^{Im}
1	427.4	414.74
2	54.95	47.99
5	5.52	2.24
10	3.65	1.66
20	2.02	1.15

Table 3: Kullback-Leibler divergence of the Gaussian mixture model with order M.

$N = \mathrm{Re}\{N\} + j\mathrm{Im}\{N\}$ can be reduced to

$$p(N) = \sum_{m=1}^{M} c_m \mathcal{N}(0, \sigma_m^2) = \sum_{m=1}^{M} \frac{c_m}{\pi\sigma_m^2} e^{-\frac{1}{\sigma_m^2}|N|^2}, \quad (3)$$

with σ_m^2 being the variance of the m-th complex-valued Gaussian. A summary of the GMM training results for $M = 5$ can be found in Table 2. The last column of Table 2 shows the variance contribution of each Gaussian mode to the total variance of the GMM.
The performance of the mixture model can be evaluated by the Kullback-Leibler divergence D_{KL} which is defined as

$$D_{KL} = \int_{\mathbb{C}} p_{\mathrm{test}}(x) \log \frac{p_{\mathrm{test}}(x)}{p_{\mathrm{ref}}(x)} dx, \quad (4)$$

with $p_{\mathrm{test}}(x)$ and $p_{\mathrm{ref}}(x)$ being the test and reference pdfs in the complex plane $x \in \mathbb{C}$, respectively. The Kullback-Leibler divergence values for different M's were calculated commonly for all frequency bins by means of the normalized histograms numerically and are given in Table 3. As it can be seen, the Kullback-Leibler divergence decreases roughly exponentially with increasing M and approximately five GMM modes ensure a satisfying fit w.r.t. the Kullback-Leibler divergence.
The kurtosis test and the trained GMM parameters proved that the Gaussian mixture model for babble noise should be preferred against the Gaussian model.

3. REFERENCE ESTIMATOR WITH GAUSSIAN NOISE MODEL

This section briefly summarizes the derivation and the used assumptions of Ephraim-Malah's MMSE short-time spectral amplitude estimator of speech DFT coefficients based on a Gaussian speech and noise model [1].
The input signal $y(n)$ of a speech enhancement system is assumed to consist of the clean speech signal $s(n)$ and the additive noise signal $n(n)$. After segmentation, windowing, and a DFT transform, the input signal can be rewritten as $Y(\ell,k) = S(\ell,k) + N(\ell,k)$ with ℓ being the analysis frame index, k being the frequency bin index. Using

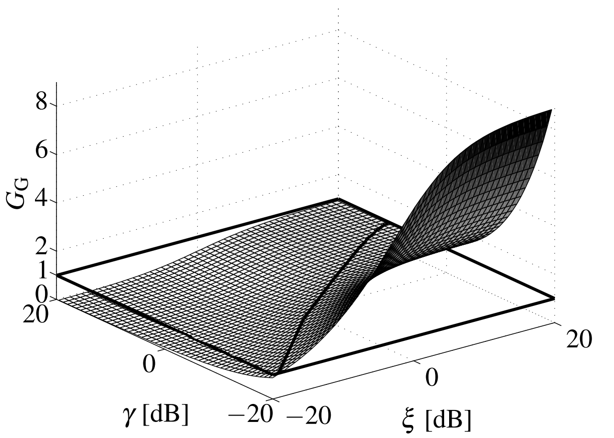

Figure 2: Weighting rule G_G of the MMSE estimator using a Gaussian speech and noise model [1]. $G_\mathrm{G} = 1$ is highlighted by a solid black curve.

polar coordinates, the input signal can be reformulated as $R(\ell,k)e^{j\Theta(\ell,k)} = A(\ell,k)e^{j\alpha(\ell,k)} + B(\ell,k)e^{j\beta(\ell,k)}$ where R, A, B, (Θ, α, β) are the magnitudes (phases) of the short-time spectra Y, S, and N, respectively.
The real and imaginary parts of the noise DFT coefficients N are assumed to be Gaussian i.i.d., therefore, $p(Y|A,\alpha)$ turns out to be (again subscript G for Gaussian noise model)

$$p_\mathrm{G}(Y|A,\alpha) = \frac{1}{\pi\sigma_N^2} e^{-\frac{1}{\sigma_N^2}|Y-Ae^{j\alpha}|^2}, \tag{5}$$

with σ_N^2 being the noise variance in the DFT domain.
Applying a Gaussian model also for speech DFT coefficients, Ephraim and Malah derived the following weighting rule [1] (subscript G for the Gaussian noise model):

$$G_\mathrm{G} = \Gamma(^3/_2)\frac{\sqrt{v}}{\gamma}M(-^1/_2;1;-v), \tag{6}$$
$$\hat{S} = G_\mathrm{G} \cdot Y,$$

where $\Gamma(\cdot)$ is the gamma function, $v = \gamma\xi/(1+\xi)$, $M(\cdot)$ is the confluent hypergeometric function, $\gamma = R^2/\sigma_N^2$ is the *a posteriori* signal-to-noise ratio (SNR), $\xi = \sigma_S^2/\sigma_N^2$ is the *a priori* SNR, and $\hat{S}$ is the estimated speech DFT coefficient.
A plot of the weighting rule G_G can be seen in Figure 2.

4. NEW MMSE ESTIMATOR WITH GMM NOISE MODEL

As outlined in Section 2.2, the pdf of the noise DFT coefficients can be modeled by a Gaussian mixture. Therefore, the conditional pdf of the noisy speech spectrum given the speech spectrum can now be formulated as (subscript GMM for the GMM noise model)

$$p_\mathrm{GMM}(Y|A,\alpha) = \sum_{m=1}^{M} \frac{c_m}{\pi\sigma_m^2} e^{-\frac{1}{\sigma_m^2}|Y-Ae^{j\alpha}|^2}. \tag{7}$$

Assuming again a Gaussian speech model and a GMM noise model, we obtain the following expression (subscript GMM for the GMM noise model):

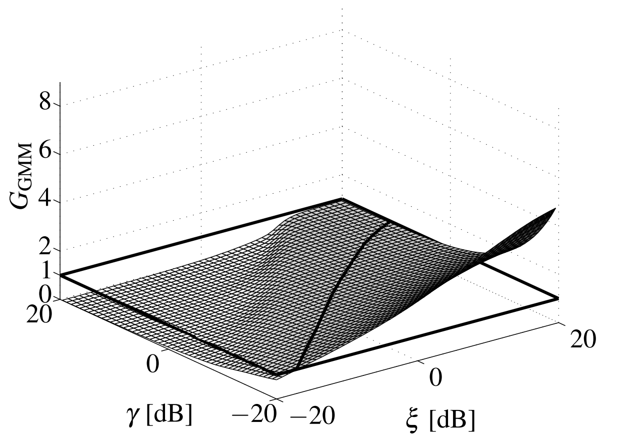

Figure 3: Weighting rule G_GMM of the MMSE estimator using a Gaussian speech model and a noise GMM with five Gaussian mixtures according to Table 2. $G_\mathrm{G} = 1$ is highlighted by a solid black curve.

$$G_\mathrm{GMM} = \Gamma\left(^3/_2\right) \frac{\sum_{m=1}^{M} \frac{c_m}{\sigma_m^2}\left(\frac{\sqrt{v_m}}{\gamma_m}\right)^3 \exp\left(-\frac{v_m}{\xi_m}\right) M(-^1/_2;1;-v_m)}{\sum_{m=1}^{M} \frac{c_m}{\sigma_m^2}\left(\frac{\sqrt{v_m}}{\gamma_m}\right)^2 \exp\left(-\frac{v_m}{\xi_m}\right)}, \tag{8}$$

with $v_m = \gamma_m\xi_m/(1+\xi_m)$, $\gamma_m = R^2/\sigma_m^2$, $\xi_m = \sigma_S^2/\sigma_m^2$ and $\exp(\cdot)$ being the exponential function. The derivation is summarized in Appendix A. Please note, that if $M = 1$ than $\gamma_m = \gamma$ and $\xi_m = \xi$ and therefore, $G_\mathrm{GMM} = G_\mathrm{G}$. To illustrate the resulting weighting rule G_GMM we assume now the noise pdf being modeled by a GMM as given in Table 2. Please note, that this GMM has unit variance, with fixed σ_N^2/σ_m^2 ratios. Using the estimated noise power $\widehat{\sigma_N^2}$, mode variances σ_m^2 are scaled, respectively. In Fig. 3, the resulting weighting rule G_GMM is plotted in the range $(\gamma,\xi) = [-20, 20]$ dB. $G_\mathrm{GMM} = 1$ is highlighted by a solid curve.
In the $G_\mathrm{GMM} > 1$ region (right from the $G_\mathrm{GMM} = 1$ solid curve in Fig. 3), the proposed weighting rule behaves more conservative than the reference in Fig. 2, as it can also be seen on the left of Fig. 4. Then a plateau region follows, where the proposed weighting rule again is more conservative, this time by not suppressing that much, which ensures a good speech preservation performance.
Towards large γ's and small ξ's, where the additive noise dominates, lower weights are needed in order to attenuate the disturbing noise. In this area, the proposed weighting rule behaves more aggressive than the reference, which is reflected by smaller weights. Therefore, a greater amount of noise reduction can be obtained.

5. EVALUATION

In order to precisely show the merit of the proposed MMSE estimator, the evaluation was performed with the test data set of the babble noise signals from the NTT Ambient Noise Database [6]. As clean speech, we employed 96 speech signals (four male and four female speakers) taken from the NTT Multi-Lingual Speech Database [8], downsampled to 8 kHz sampling rate. The active speech level was set to −26dB below the clipping level and the noise signal level adjusted to the desired SNR, according to ITU-T Recommendation P.56 [9]. Next, both signals were superimposed, in order to obtain

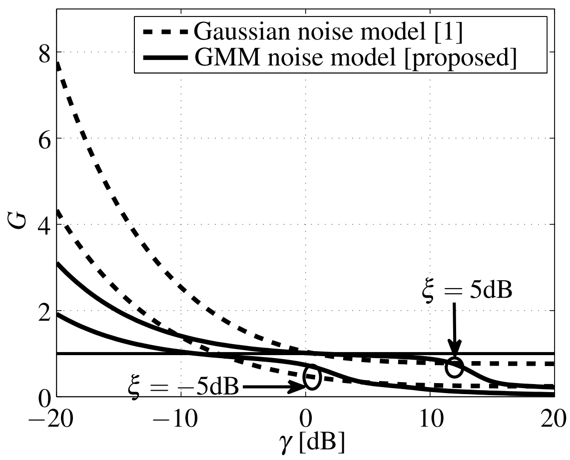

Figure 4: Weighting rules G_{G} (dashed line) and G_{GMM} (solid line) for $\xi = -5$dB and $\xi = 5$dB of the MMSE estimator using a Gaussian speech model.

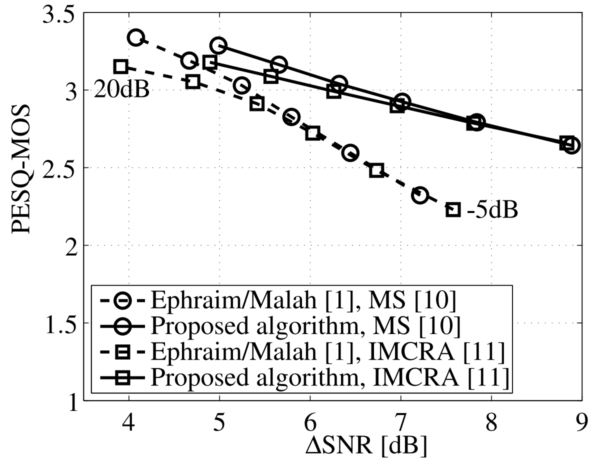

Figure 5: Performance of the new MMSE estimator compared to Ephraim and Malah's spectral amplitude estimator.

Noise power estimator	SNR [dB]					
	-5	0	5	10	15	20
MS	5.16	5.17	5.20	5.27	5.40	5.64
IMCRA	5.01	5.16	5.39	5.78	6.41	7.35

Table 4: Noise tracking performance (NTP) measure of the noise power estimators MS [10] and IMCRA [11] in dB for different SNR values.

the desired SNR ratio. The signal analysis was the same as used for the histogram measurement above: At a sampling frequency of 8 kHz, the segmentation was done with a Hann window, the analysis frame length was 256 samples, the analysis frame shift contained 128 samples.

Since most of the state-of-the-art noise power estimators $\widehat{\sigma_N^2}(\ell,k)$ such as minimum statistics (MS) [10] are designed for tracking the spectral noise power of stationary interferers, the spontaneous fluctuations of the babble noise power cannot be estimated well with these algorithms. Beside minimum statistics, we also employed the improved minima controlled recursive averaging (IMCRA) algorithm [11], which is better able to track the noise power of non-stationary noises. In order to prove the effectiveness of the estimators, we measured their so-called noise tracking performance (NTP) measure [12], which is defined as

$$\mathrm{NTP} = \frac{1}{KN}\sum_{k=1}^{K}\sum_{\ell=1}^{L}\left|10\log_{10}\left[\frac{\sigma_N^2(\ell,k)}{\widehat{\sigma_N^2}(\ell,k)}\right]\right|. \quad (9)$$

The results at different SNR values for MS and IMCRA can be seen in Table 4. Interestingly, in the presence of non-stationary babble noise, MS achieves better (i. e., lower) NTP values at higher SNRs, while IMCRA performs better at very low SNR conditions.

The widely employed decision-directed *a priori* SNR estimation [1] is calculated as follows:

$$\widehat{\xi}(\ell,k) = (1-\alpha)P\{\gamma(\ell,k)-1\} + \alpha\frac{|\hat{S}(\ell-1,k)|^2}{\widehat{\sigma_N^2}(\ell-1,k)}, \quad (10)$$

with $P\{x\} = \max\{0,x\}$. This approach does not work well for non-stationary noise, therefore, we employed the clean speech power spectral density (PSD) $|S(\ell-1,k)|^2$ as numerator in (10). Employing such an idealized a priori SNR estimator has a few advantages: On the one hand, it yields independence from the estimated speech $\hat{S}$. Therefore, the use of (10) with ideal nominator ensures, that both weighting rules are addressed with exactly the same (γ, ξ) pairs which allows precise comparison of the estimators G_{G} and G_{GMM}. On the other hand, the use of an (idealized) well-working SNR estimator makes it even more difficult for a new weighting rule to yield substantial advantages.

The weighting rule was then computed by using the proposed MMSE estimator (8) with the noise GMM as described in Table 2 with $M = 5$, realized by a table lookup of $G_{\mathrm{GMM}} = f(\gamma,\xi)$ with both γ and ξ varying from $-20...+20$dB in 0.4dB steps. As a reference, we also employed the MMSE estimator with a Gaussian noise model (6). Since the use of the confluent hypergeometric function $M(\cdot)$ is computationally intensive, we also realized G_{G} as a table lookup, in the same fashion as G_{GMM}.

We evaluated the performance of the proposed weighting rules w.r.t. speech preservation and noise attenuation performance. Given a noisy speech signal $y(n)$ we employed the respective clean speech $\tilde{s}(n)$ and noise component $\tilde{n}(n)$ of the enhanced signal $\hat{s}(n) = \tilde{s}(n) + \tilde{n}(n)$. Through the clean speech signal $s(n)$ and its processed replica $\tilde{s}(n)$, the speech preservation performance was represented by the PESQ-MOS score [13].

The SNR improvement ΔSNR was measured as the difference between the output and the input SNR values. The input (output) SNR is determined as the ratio of the power of $s(n)$ ($\tilde{s}(n)$) to that of $n(n)$ ($\tilde{n}(n)$). All SNR measurements were performed using [9] for active speech and noise level measurements.

In Figure 5, a comparison of the proposed and the reference weighting rule can be seen, based on the ITU-T measures ΔSNR and PESQ-MOS for different SNR values. The optimum in Figure 5 resides in the right top corner, which means a high SNR improvement and a good quality of the clean speech component, simultaneously. It can generally be said that at higher SNRs more speech preservation can be observed, while at very low SNRs greater ΔSNR improvement and can be attained. At very low SNRs (-5 and 0 dB) the proposed weighting rule achieves approximately 1.5dB enhancement in terms of ΔSNR and 0.5 PESQ-MOS score improvement. Towards greater SNRs, this gap is getting smaller. At 15 and 20dB SNR, the ΔSNR improvement is approximately 1dB, the PESQ-MOS scores are quite comparable. We can summarize that if a well working a priori SNR estimator for non-stationary noises were available, our proposed GMM-based weighting rule would allow for further substantial improvements in low SNR conditions.

6. CONCLUSIONS

In this paper, we analyze the statistics of noise DFT coefficients in the context of non-Gaussian non-stationary noises. We employ a Gaussian mixture, modeling the non-Gaussian pdf of the noise DFT coefficients and present an MMSE speech spectral amplitude estimator. It turns out that in non-stationary non-Gaussian babble noise the proposed approach outperforms the known Gaussian noise model-based estimator. Comparison tests have been made with the clean speech PSD being made perfectly known to the decision-directed *a priori* SNR estimator for all algorithms. This leads us to the conclusion that there is potential in the area of non-Gaussian noise modeling, but simultaneously, there is a need for proper noise power and particularly SNR estimators for non-stationary noises.

A. APPENDIX: DERIVATION OF THE NEW MMSE ESTIMATOR WITH GMM NOISE MODEL

Using the minimum mean square error (MMSE) criterion $\min_{\hat{A}} E\{(\hat{A}-A)^2|Y\}$, Ephraim und Malah have shown that the estimated magnitude $\hat{A}$ of the clean speech spectrum can be computed as [1]

$$\hat{A} = E\{A|Y\} = \frac{\int_0^\infty \int_0^{2\pi} A p(Y|A,\alpha) p(A,\alpha) d\alpha dA}{\int_0^\infty \int_0^{2\pi} p(Y|A,\alpha) p(A,\alpha) d\alpha dA}, \qquad (11)$$

where $E\{\cdot\}$ and $p(\cdot)$ denote the expectation operator and the corresponding pdf, respectively.
Let us assume, that the real and imaginary parts of the speech DFT coefficients are Gaussian i.i.d., A is statistically independent from α, and α being uniformly distributed. A Gaussian speech model leads then to a Rayleigh pdf of the speech amplitude yielding

$$p(A,\alpha) = p(\alpha)\cdot p(A) = \frac{1}{2\pi}\cdot 2\frac{A}{\sigma_S^2} e^{-\frac{A^2}{\sigma_S^2}}, \qquad (12)$$

with σ_S^2 being the speech variance in the DFT domain.
Applying a GMM as a noise model introduced in 2.2, substituting (7) and (12) into (11), we obtain the following expression:

$$\hat{A} = \frac{\sum_{m=1}^{M} \frac{c_m}{\sigma_m^2} e^{-\frac{R^2}{\sigma_m^2}} \int_0^\infty A^2 e^{-\frac{A^2}{\lambda_m}} \frac{1}{2\pi}\int_0^{2\pi} e^{2\frac{AR}{\sigma_m^2}\cos(\Theta-\alpha)} d\alpha dA}{\sum_{m=1}^{M} \frac{c_m}{\sigma_m^2} e^{-\frac{R^2}{\sigma_m^2}} \int_0^\infty A e^{-\frac{A^2}{\lambda_m}} \frac{1}{2\pi}\int_0^{2\pi} e^{2\frac{AR}{\sigma_m^2}\cos(\Theta-\alpha)} d\alpha dA}, \qquad (13)$$

with $\frac{1}{\lambda_m} = \frac{1}{\sigma_m^2} + \frac{1}{\sigma_S^2}$. Using [14, (8.431)], (13) turns out to be

$$\hat{A} = \frac{\sum_{m=1}^{M} \frac{c_m}{\sigma_m^2} e^{-\gamma_m} \int_0^\infty A^2 e^{-\frac{A^2}{\lambda_m}} I_0(2a\sqrt{\frac{v_m}{\lambda_m}}) dA}{\sum_{m=1}^{M} \frac{c_m}{\sigma_m^2} e^{-\gamma_m} \int_0^\infty A e^{-\frac{A^2}{\lambda_m}} I_0(2a\sqrt{\frac{v_m}{\lambda_m}}) dA}, \qquad (14)$$

with I_0 being the modified Bessel function of the first kind and the zeroth order.
Applying [14, (6.631)] with $\mu = 2$ and $\mu = 1$ for the numerator and the denominator, respectively, as well as $\nu = 0$, $\alpha = \lambda_m^{-1}$ and $\beta = j2\sqrt{v_m/\gamma_m}$ for both of numerator and denominator, (14) can be rewritten as

$$\hat{A} = \frac{\sum_{m=1}^{M} \frac{c_m}{\sigma_m^2} e^{-\gamma_m} \Gamma(^3/_2) \lambda_m^{^3/_2} M(^3/_2;1;v_m)}{\sum_{m=1}^{M} \frac{c_m}{\sigma_m^2} e^{-\gamma_m} \lambda_m M(1;1;v_m)}. \qquad (15)$$

According to [15, (A.1.17a), (A.1.19a)], (15) turns out to be

$$\hat{A} = \Gamma(^3/_2) \frac{\sum_{m=1}^{M} \frac{c_m}{\sigma_m^2} e^{-\gamma_m+v_m} \lambda_m^{^3/_2} M(-^1/_2;1;-v_m)}{\sum_{m=1}^{M} \frac{c_m}{\sigma_m^2} e^{-\gamma_m+v_m} \lambda_m}. \qquad (16)$$

Finally, using that $-\gamma_m + v_m = -v_m/\xi_m$, $\lambda_m = (v_m/\gamma_m^2)R^2$ and $G = \hat{A}/R$, (16) can be reformulated as (8).

REFERENCES

[1] Ephraim, Y.; Malah, D., "Speech Enhancement Using a Minimum-Mean Square Error Short-Time Spectral Amplitude Estimator," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 32, no. 6, pp. 1109–1121, Dec. 1984.

[2] Wolfe, P. J.; Godsill, S. J., "Efficient Alternatives to the Ephraim and Malah Suppression Rule for Audio Signal Enhancement," *EURASIP Journal on Applied Signal Processing*, vol. 2003, no. 10, pp. 1043–1051, 2003.

[3] Lotter, T.; Vary, P., "Speech Enhancement by MAP Spectral Amplitude Estimation Using a Super-Gaussian Speech Model," *EURASIP Journal on Applied Signal Processing*, vol. 7, pp. 1110–1126, 2005.

[4] Hendriks, R. C.; Heusdens, R.; Kjems, U.; Jensen, J., "On Optimal Multichannel Mean-Squared Error Estimators for Speech Enhancement," *IEEE Signal Processing Letters*, vol. 16, no. 10, pp. 885–888, Oct. 2009.

[5] Potamitis, I.; Fakotakis, N.; Kokkinakis, G., "A Trainable Speech Enhancement Technique Based on Mixture Models for Speech and Noise," in *EUROSPEECH 2003*, Geneva, Switzerland, Sep. 2003.

[6] "Ambient Noise Database for Telephonometry," NTT Advanced Technology Corporation (NTT-AT), 1996.

[7] Abramowitz, M.; Stegun, I. A., Ed., *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover Publications, New York, USA, 1972.

[8] "Multi-Lingual Speech Database for Telephonometry," NTT Advanced Technology Corporation (NTT-AT), 1994.

[9] "Objective Measurement of Active Speech Level," ITU-T P.56, Mar. 1993.

[10] Martin, R., "Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 5, pp. 504–512, July 2001.

[11] Cohen, I., "Noise Spectrum Estimation in Adverse Environments: Improved Minima Controlled Recursive Averaging," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 5, pp. 466–475, Sept. 2003.

[12] Hendriks, R. C.; Heusdens, R. and Jensen, J., "MMSE Based Noise PSD Tracking with Low Complexity," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Dallas, TX, USA, March 2010, pp. 4266–4269.

[13] "Perceptual Evaluation of Speech Quality (PESQ)," ITU-T P.862, Feb. 2001.

[14] Gradshteyn, I. S.; Ryzhik, I. M., *Table of Integral, Series, and Products*, Academic Press, 4th edition, 1965.

[15] David Middleton, *An Introduction to Statistical Communication Theory*, Wiley-IEEE Press, 1996.