MMSE-optimal spectral amplitude estimation given the STFT-phase

Timo Gerkmann*, *Member, IEEE,* and Martin Krawczyk

***Abstract*—In this letter, we derive a minimum mean squared error (MMSE) optimal estimator for clean speech spectral amplitudes, which we apply in single channel speech enhancement. As opposed to state-of-the-art estimators, the optimal estimator is derived for a given clean speech spectral phase. We show that the phase contains additional information that can be exploited to distinguish outliers in the noise from the target signal. With the proposed technique, incorporating the phase can potentially improve the PESQ-MOS by 0.5 in babble noise as compared to state-of-the-art amplitude estimators. In a blind setup we achieve a PESQ improvement of around 0.25 in voiced speech.**

***Index Terms*—Speech enhancement, Phase estimation, noise reduction, signal reconstruction.**

I. Introduction

SINGLE channel speech enhancement describes the problem of estimating a clean speech signal from a noisy recording with only one microphone. Typical applications can be found in the area of speech communications, such as hearing aids, telephony and automatic speech recognition in man-machine interfaces. Research in this area has been going on for decades – with quite some success. Nowadays, speech enhancement algorithms are implemented on mobile devices such as smart phones and hearing aids. As these devices are often used in noisy environments, recent research addresses the robustness of the algorithms in nonstationary noise, e.g. babble, and low signal-to-noise ratios.

For the improvement of noisy speech and the estimation of parameters required for speech enhancement, in the vast majority of proposals, the noisy speech signal is transformed to some spectral domain, as it allows for a better separation of speech and noise. In this paper we will focus on short time discrete Fourier transform (STFT)-based speech enhancement. Already in the early eighties researchers have investigated in what direction research in speech enhancement can be expected to be fruitful. For instance, Wang and Lim [1] have done experiments in which they investigated the importance of phase estimation in speech enhancement. For this, they synthesized noisy speech by taking the amplitude and phase from signals with different signal to noise ratios (SNRs). They observed that improving the noisy spectral amplitude is more important for the signal quality than improving the noisy spectral phase. Following this observation, it was concluded that clean speech phase estimation is unimportant in speech enhancement, and until today most research aims at estimating the clean speech amplitude only, while keeping the noisy phase unaltered. Among the most prominent amplitude estimators are the proposals by Ephraim and Malah. For instance in [2] the spectral amplitudes are estimated, while in [3] it is argued that estimating logarithmically compressed amplitudes is perceptually more meaningful. In [4] a more flexible estimator with a parameterized compression function is derived that generalizes the estimators of [3] and [2]. While in [2]–[4] it is assumed that the speech spectral coefficients are complex Gaussian distributed, more recent work focuses on deriving Bayesian estimators for more heavy-tailed distributions [5]–[9]. The most general estimators have parameters to control both the degree of heavy-tailedness and the degree of compression [9]. All estimators [2]–[9] have in common that only the speech spectral amplitude is altered, while the noisy phase is left unchanged. Also in other speech enhancement techniques, like sinusoidal modeling, it has been proposed to combine improved spectral amplitudes with the noisy STFT phase [10].

Despite the general trend of neglecting STFT phase estimation, Paliwal et al. argue that, potentially, the role of the phase in speech enhancement has been underestimated in the past [11]. They showed that if the segment overlap and the length of the Fourier transform are increased, the impact of the clean speech phase is larger than observed by Wang and Lim [1]. Thus, Paliwal et al. concluded that estimating the clean speech phase can indeed be beneficial. While they proposed some methods for speech enhancement that involve a modification of complex spectral coefficients [11], the direct estimation of the clean spectral phase is considered a difficult task and only few proposals exist. For instance, Griffin and Lim proposed to estimate the spectral phase by iteratively analyzing and synthesizing the signal starting from only the spectral amplitudes [12]. However, their approach is computationally quite demanding and requires knowledge of the clean spectral amplitudes.

While the noisy phase has been shown to be the optimal Bayesian estimator if the clean speech phase is uniformly distributed [2], [6], in [13] we showed that with a given fundamental frequency it is possible to reconstruct the clean speech phase both on and between speech spectral harmonics in voiced segments directly in the STFT domain. In a speech enhancement framework, this reconstructed clean speech STFT phase increases the Perceptual Evaluation of Speech Quality (PESQ) mean opinion score (MOS) by up to 0.1 as compared to using the noisy phase [14].

In this letter, we argue that the clean speech phase pro-

The authors are with the Speech Signal Processing Group, Department for Medical Physics and Acoustics, Universität Oldenburg, 26111 Oldenburg, Germany, e-mail: {timo.gerkmann, martin.krawczyk}@uni-oldenburg.de, web: www.speech.uni-oldenburg.de.

vides additional information that can also be exploited for an improved estimation of the clean speech spectral amplitudes. For this, we derive a novel MMSE optimal estimator for the clean speech spectral amplitude when the spectral phase is given.

II. Signal Model and Notation

We observe noisy speech in the STFT domain, where the noisy speech $Y_k(\ell)$ is an additive superposition of speech $S_k(\ell)$ and noise $N_k(\ell)$,

$$Y_k(\ell) = S_k(\ell) + N_k(\ell)\,. \tag{1}$$

Here, k is the frequency index and ℓ is the segment index. In the sequel, we will omit the indices k and ℓ for brevity. The complex spectral coefficients can be written in terms of their amplitude and phase, denoted as

$$Y = R\mathrm{e}^{\mathrm{j}\Phi_{\mathrm{Y}}}; \quad S = A\mathrm{e}^{\mathrm{j}\Phi_{\mathrm{S}}}; \quad N = D\mathrm{e}^{\mathrm{j}\Phi_{\mathrm{N}}}. \tag{2}$$

Further, we will denote random variables by capital letters, e.g. S, A, Φ_{S}, while their realizations are denoted by the corresponding lower case letters, e.g. s, a, ϕ_{S}. Estimated quantities are marked by a hat symbol, e.g. $\widehat{\phi_{\mathrm{S}}}$ is an estimate of ϕ_{S}.

III. Amplitude Estimation Given Phase

In this section we estimate the clean speech amplitudes provided that we know the clean speech phase ϕ_{S}, as well as the clean speech power spectral density (PSD) $\sigma_{\mathrm{S}}^2 = \mathrm{E}\big(A^2\big)$ and the noise PSD $\sigma_{\mathrm{N}}^2 = \mathrm{E}\big(D^2\big)$. Similar to [4], [9] we want to minimize the mean squared error between the compressed clean speech amplitudes A^β and the estimator for compressed amplitudes $\widehat{A}^\beta$. A compression factor $\beta < 1$ allows us to emphasize estimation errors of low amplitudes, and for $\beta \to 0$ a logarithmic spectral amplitude estimator is approximated [4], [9]. To obtain our novel estimator we have to solve

$$\begin{aligned}\widehat{A^\beta} &= \mathrm{E}\big(A^\beta \mid r, \phi_{\mathrm{Y}}, \phi_{\mathrm{S}}\big) \\ &= \int_{-\infty}^{\infty} a^\beta \, p_{A|R,\Phi_{\mathrm{Y}},\Phi_{\mathrm{S}}}(a \mid r, \phi_{\mathrm{Y}}, \phi_{\mathrm{S}})\,\mathrm{d}a.\end{aligned} \tag{3}$$

Using Bayes' theorem we obtain

$$\widehat{A^\beta} = \frac{\int_{-\infty}^{\infty} a^\beta \, p_{R,\Phi_{\mathrm{Y}}|A,\Phi_{\mathrm{S}}}(r, \phi_{\mathrm{Y}} \mid a, \phi_{\mathrm{S}}) \; p_{A,\Phi_{\mathrm{S}}}(a, \phi_{\mathrm{S}})\,\mathrm{d}a}{\int_{-\infty}^{\infty} p_{R,\Phi_{\mathrm{Y}}|A,\Phi_{\mathrm{S}}}(r, \phi_{\mathrm{Y}} \mid a, \phi_{\mathrm{S}}) \; p_{A,\Phi_{\mathrm{S}}}(a, \phi_{\mathrm{S}})\,\mathrm{d}a}.$$

With the assumption that the clean speech amplitude is independent of the clean speech phase, we can write

$$\widehat{A^\beta} \;=\; \frac{\int_{-\infty}^{\infty} a^\beta \, p_{R,\Phi_{\mathrm{Y}}|A,\Phi_{\mathrm{S}}}(r, \phi_{\mathrm{Y}} \mid a, \phi_{\mathrm{S}}) \; p_A(a)\,\mathrm{d}a}{\int_{-\infty}^{\infty} p_{R,\Phi_{\mathrm{Y}}|A,\Phi_{\mathrm{S}}}(r, \phi_{\mathrm{Y}} \mid a, \phi_{\mathrm{S}}) \; p_A(a)\,\mathrm{d}a}. \tag{4}$$

As in [6]–[9], we assume that the real and imaginary parts of the complex noise spectral coefficients are independent and Gaussian distributed. Thus, if the speech coefficients are given, after polar transformation we obtain the conditioned probability density function (PDF) of the noisy coefficients as

$$\begin{aligned}&p_{R,\Phi_{\mathrm{Y}}|A,\Phi_{\mathrm{S}}}(r, \phi_{\mathrm{Y}} \mid a, \phi_{\mathrm{S}}) = \\ &\qquad \frac{r}{\pi\sigma_{\mathrm{N}}^2}\exp\left(-\frac{r^2 + a^2 - 2ar\cos(\phi_{\mathrm{Y}} - \phi_{\mathrm{S}})}{\sigma_{\mathrm{N}}^2}\right).\end{aligned} \tag{5}$$

To model the PDF of the speech spectral amplitudes, as in [7], [9], we employ the χ-distribution with shape parameter μ. Note that the χ-distribution is a special case of the generalized Gamma-distribution [8] when the parameters in [8, Eq. (1)] are set to $\gamma^{\,[8]} = 2$ and $\beta^{\,[8]} = \mu/\sigma_{\mathrm{S}}^2$. The χ-distribution is defined as

$$p_A(a) = \frac{2}{\Gamma(\mu)}\left(\frac{\mu}{\sigma_{\mathrm{S}}^2}\right)^{\mu} a^{2\mu-1}\exp\left(-\frac{\mu}{\sigma_{\mathrm{S}}^2}a^2\right), \tag{6}$$

with the Gamma function $\Gamma(\cdot)$ [15, Eq. (8.31)].

If the spectral coefficients S are complex Gaussian distributed, the resulting spectral amplitudes $A = |S|$ are χ-distributed with $\mu = 1$. More heavy-tailed (super-Gaussian) priors can be modeled by setting $\mu < 1$. The solutions to the integrals resulting from inserting (6) and (5) into (4) are listed in [15, Eq. (3.462.1)], and yield our proposed estimator

$$\boxed{\begin{aligned}\widehat{A} &= \big(\mathrm{E}\big(A^\beta \mid r, \phi_{\mathrm{Y}}, \phi_{\mathrm{S}}\big)\big)^{\frac{1}{\beta}} \\ &= \sqrt{\frac{1}{2}\frac{\xi}{\mu+\xi}\sigma_{\mathrm{N}}^2}\left(\frac{\Gamma(2\mu+\beta)}{\Gamma(2\mu)}\frac{\mathrm{D}_{-(2\mu+\beta)}(\nu)}{\mathrm{D}_{-(2\mu)}(\nu)}\right)^{\frac{1}{\beta}}\end{aligned}} \tag{7}$$

where $\mathrm{D}_{\cdot}(\nu)$ is the parabolic cylinder function [15, Eq. (9.24)], $\xi = \sigma_{\mathrm{S}}^2/\sigma_{\mathrm{N}}^2$ is the *a priori* SNR, and the argument

$$\nu = -\frac{r}{\sigma_{\mathrm{N}}}\sqrt{2\frac{\xi}{\mu+\xi}}\cos(\underbrace{\phi_{\mathrm{Y}} - \phi_{\mathrm{S}}}_{\Delta\phi}) \tag{8}$$

contains the phase difference $\phi_{\mathrm{Y}} - \phi_{\mathrm{S}} = \Delta\phi$.

IV. Benefits of the proposed estimator

To understand the benefits of the proposed estimator, the input-output curve of (7) is given in Fig. 1 and Fig. 2 for an a priori SNR of $\xi = 0.2$. To draw conclusions independent of an absolute signal-scaling, we normalize the input R and the output $\widehat{A}$ by σ_{N}. In Fig. 1 we set the shape parameter to $\mu = 1$ and the compression parameter to $\beta = 0.001$. Thus, without incorporating the phase we would approximate the log-spectral amplitude estimator (LSA) proposed in [3]. For reference, we include the input-output-curves of the LSA and the Wiener filter in Fig. 1.

The phase information employed by the proposed estimator can help to distinguish if large amplitudes $R/\sigma_{\mathrm{N}} \gg 1$ originate from speech or represent outliers in the noise. For this distinction, state-of-the-art estimators only have the *a priori* SNR ξ and R/σ_{N} available. Taking the phase into account, we now have additional information for an improved separation of noise outliers from speech: if R/σ_{N} is large due to a contribution from speech, then the phase of the noisy speech will be close to the clean speech phase [16], i.e. $|\Delta\phi| \to 0$. Consequently, if $\Delta\phi = 0$, the proposed estimator (top solid line) applies less attenuation and thus less speech distortions than the LSA. However, if R/σ_{N} is large because of noise outliers, the phase difference $|\Delta\phi|$ is likely to be larger than zero. Employing this larger $|\Delta\phi|$ in (8) and (7) results in an efficient attenuation of noise outliers that is not possible without taking the phase into account. This larger attenuation can be seen in the second, third, and fourth solid line that represent $|\Delta\phi| = \pi/4, \pi/2, \pi$ (top to bottom).

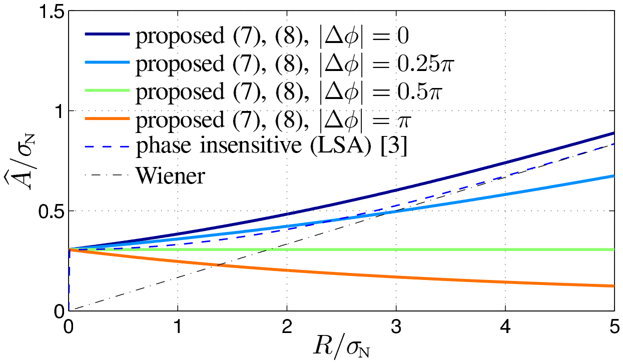

Fig. 1. Input-output curve of the proposed estimator (7) for $\mu = 1$, $\beta = 0.001$ and $\xi = 0.2$. The solid lines are the results for different values for $|\Delta\phi| = |\phi_{\mathrm{Y}} - \phi_{\mathrm{S}}|$, namely $0, \pi/4, \pi/2, \pi$ (top to bottom). For reference we also include the phase insensitive LSA [3] (dashed) and the Wiener filter (dash-dotted).

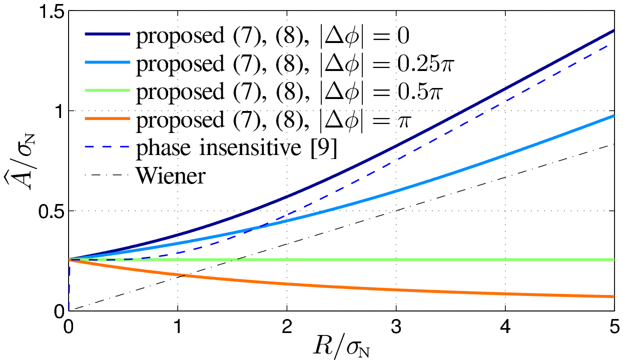

Fig. 2. As in Fig. 1 we compare the output of the derived estimator for different $|\Delta\phi|$ but now set $\mu = \beta = 0.5$. For comparison we also include the phase insensitive amplitude estimator from [9] with $\mu = \beta = 0.5$ (dashed).

From these considerations we see that incorporating the phase provides a novel mechanism to distinguish noise from speech. The proposed estimator can efficiently reduce undesired noise outliers while preserving the speech signal.

Similar to [7]–[9], we can also model heavy-tailed speech distributions by setting $\mu < 1$. In Fig. 2 we plot the results for $\beta = 0.5$ and $\mu = 0.5$. It can be seen that this estimator results in lower outputs for low input values as compared to Fig. 1. However, for large input values and $|\Delta\phi| \leq \pi/2$, less attenuation is applied as compared to $\beta = 0.001$ and $\mu = 1$ in Fig. 1. The reason is that for $\mu < 1$ the speech distribution is modeled as being more heavy-tailed than the noise distribution, meaning that, as compared to $\mu = 1$, large noisy amplitudes are assumed to originate more likely from speech rather than noise.

V. Evaluation

In this section, we employ the proposed estimator in a speech enhancement framework. For this we use a randomly chosen subset of 10 female and 10 male speakers from the TIMIT database [17] and additive babble noise at various input SNRs. For the estimation of the noise PSD we use the unbiased MMSE estimator [18]. Subsequently, the *a priori* SNR is estimated using the *decision-directed* approach [2]. The sampling rate is set to $8\,\mathrm{kHz}$ and the segment length is $32\,\mathrm{ms}$. In figures 3–5 we evaluate the proposed estimator using the PESQ MOS as implemented in [19]. While PESQ has been initially developed for assessing the perceived quality of coded speech, it also shows good correlation with speech quality in the speech enhancement context [19]. We provide the results for the phase insensitive amplitude estimator [9] and compare it to the proposed estimator (7) with $\beta = \mu = 0.5$. To quantify the achievable gain when the clean speech phase is given, in Fig. 3 we employ the true clean speech phase ϕ_{S} in (8). We see that employing the clean phase as extra information in amplitude enhancement results in an *additional* PESQ improvement of almost 0.35 PESQ-MOS. Informal listening reveals that the proposed method is capable of reducing annoying outliers in the residual noise. If we additionally use the clean speech phase instead of the noisy phase for signal reconstruction as $\widehat{S} = \widehat{A}\exp(\mathrm{j}\Phi_{\mathrm{S}})$, an overall PESQ improvement of more than 0.5 PESQ-MOS can be achieved. It is interesting to note that, in contrast to [11], the performance gain using phase information is achieved without zero-padding and with a segment-overlap of only 50% in the spectral analysis.

In practice, we have to estimate the clean speech phase. Thus, in the next experiment we employ the phase estimation method proposed in [13]. As [13] only provides an estimate of the phase change $\Delta\phi_{\mathrm{S}}(\ell) = \phi_{\mathrm{S}}(\ell) - \phi_{\mathrm{S}}(\ell - 1)$ but not the absolute phase $\phi_{\mathrm{S}}(\ell)$, we use the phase changes $\Delta\phi_{\mathrm{S}}$, $\Delta\phi_{\mathrm{Y}}$ instead of the absolute phases $\phi_{\mathrm{S}}, \phi_{\mathrm{Y}}$ in (8). To facilitate the estimation of $\Delta\phi_{\mathrm{S}}$, we use a segment overlap of 87.5% in this experiment.

In Fig. 4 the results for an estimated phase are given. The fundamental frequency, which is needed for phase estimation via [13], has been obtained from the clean speech signal using PEFAC as proposed in [20] with the voiced/unvoiced decision employed in [21]. As in [13] the phase is only estimated in voiced speech, we also evaluate the performance only in voiced speech as indicated by using PEFAC on the clean signal. It can be seen that a robust estimate of the fundamental frequency is sufficient to obtain a clean phase estimate that results in large improvements of PESQ. Using the clean phase estimate also for reconstruction in moderate and low SNRs improves results further (upper dash-dotted line in Fig. 4). However, also some artifacts are introduced and the performance gain vanishes for large SNR. This is not the case when the estimated phase is only used to improve amplitudes. Thus, we conclude that employing an estimated phase for amplitude estimation is more robust as compared to a direct employment of a clean speech phase estimate.

In the final experiment, we investigate the performance gain in a blind setup, i.e. when the fundamental frequency is estimated on noisy speech. In Fig. 5 it can be seen that even in this blind setup a performance gain of around 0.25 PESQ-MOS is achieved.

VI. Conclusions

In this letter we showed how knowledge of the clean speech spectral phase can be employed for a more robust amplitude estimation. For this, we have derived an MMSE optimal

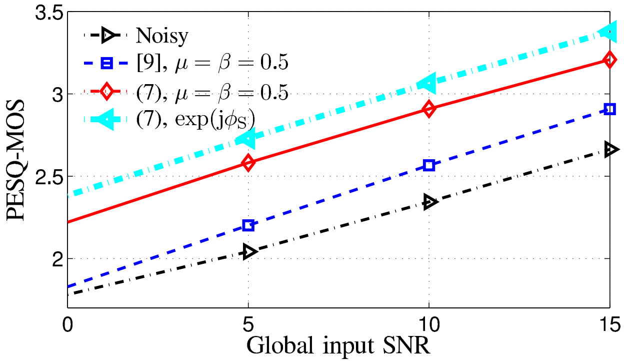

Fig. 3. PESQ-MOS for the MMSE-optimal amplitude estimator [9] with $\mu = \beta = 0.5$ and the proposed estimator (7). Here, the true clean speech spectral phase is used in (7)(8) and the segment overlap is 50%. We also plot the result when the clean phase is used for reconstruction, as $\widehat{S} = \widehat{A}\exp(\mathrm{j}\phi_\mathrm{S})$ (upper dash-dotted line).

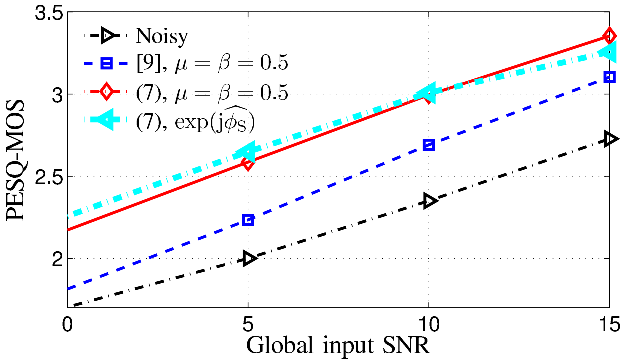

Fig. 4. As Fig. 3, but the clean speech spectral phase is estimated using [13]. The fundamental frequency required in [13] is estimated on the clean speech signal using [20]. The segment overlap is 87.5%.

estimator for the clean speech spectral amplitudes when, besides the speech and noise power spectral densities, also the clean speech phase is known. The proposed estimator improves single channel speech enhancement further, as the additional information provided by the phase helps to distinguish outliers in the noise from speech. We showed that incorporating the phase can potentially improve the PESQ-MOS by 0.5 in babble noise as compared to state-of-the-art amplitude estimators. In a blind setup we achieve a PESQ improvement of 0.25 in voiced speech. Results demonstrate that clean speech phase estimation is an interesting field of research that can push the limits of single channel speech enhancement algorithms further.

Acknowledgments

The authors would like to thank Robert Rehr for the implementation of PEFAC, and the associate editor and reviewers for their valuable comments that helped improving this letter.

References

[1] D. L. Wang and J. S. Lim, "The unimportance of phase in speech enhancement," *IEEE Trans. Acoust., Speech, Signal Process.*, no. 4, pp. 679–681, 1982.

[2] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 32, no. 6, pp. 1109–1121, Dec. 1984.

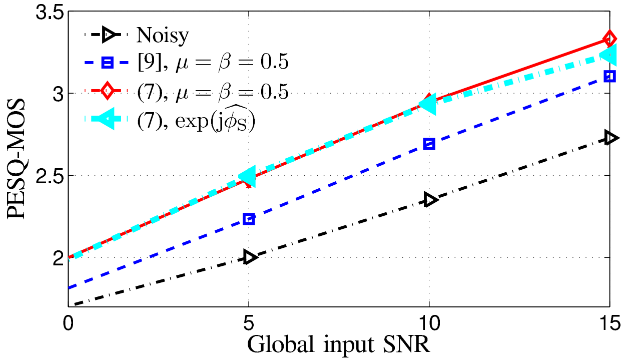

Fig. 5. Blind phase estimation: as Fig. 4, but the fundamental frequency required in [13] is estimated on the noisy speech signal using [20].

[3] ——, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 2, pp. 443–445, Apr. 1985.

[4] C. H. You, S. N. Koh, and S. Rahardja, "β-order MMSE spectral amplitude estimation for speech enhancement," *IEEE Speech Audio Process.*, vol. 13, no. 4, pp. 475–486, Jul. 2005.

[5] R. Martin, "Speech enhancement using MMSE short time spectral estimation with Gamma distributed speech priors," in *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 2002, pp. 253–256.

[6] T. Lotter and P. Vary, "Speech enhancement by MAP spectral amplitude estimation using a super-Gaussian speech model," *EURASIP J. Applied Signal Process.*, vol. 2005, no. 7, pp. 1110–1126, Jan. 2005.

[7] I. Andrianakis and P. R. White, "MMSE speech spectral amplitude estimators with Chi and Gamma speech priors," in *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Toulouse, France, May 2006, pp. 1068–1071.

[8] J. S. Erkelens, R. C. Hendriks, R. Heusdens, and J. Jensen, "Minimum mean-square error estimation of discrete Fourier coefficients with generalized Gamma priors," *IEEE Audio, Speech, Language Process.*, vol. 15, no. 6, pp. 1741–1752, Aug. 2007.

[9] C. Breithaupt, M. Krawczyk, and R. Martin, "Parameterized MMSE spectral magnitude estimation for the enhancement of noisy speech," in *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Apr. 2008, pp. 4037–4040.

[10] J. Jensen and J. H. Hansen, "Speech enhancement using a constrained iterative sinusoidal model," *IEEE Speech Audio Process.*, vol. 9, no. 7, pp. 731–740, Oct. 2001.

[11] K. Paliwal, K. Wójcicki, and B. Shannon, "The importance of phase in speech enhancement," *ELSEVIER Speech Commun.*, vol. 53, no. 4, pp. 465–494, Apr. 2011.

[12] D. Griffin and J. Lim, "Signal estimation from modified short-time fourier transform," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 32, no. 2, pp. 236–243, Apr. 1984.

[13] M. Krawczyk and T. Gerkmann, "STFT phase improvement for single channel speech enhancement," *Int. Workshop Acoustic Echo, Noise Control (IWAENC)*, Sep. 2012.

[14] T. Gerkmann, M. Krawczyk, and R. Rehr, "Phase estimation in speech enhancement — unimportant, important, or impossible?" in *IEEE Conv. Elect. Electron. Eng. Israel*, Eilat, Israel, Nov. 2012.

[15] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals Series and Products*, 7th ed. San Diego, CA, USA: Academic Press, 2000.

[16] P. Vary, "Noise suppression by spectral magnitude estimation – mechanism and theoretical limits," *ELSEVIER Signal Process.*, vol. 8, pp. 387–400, May 1985.

[17] J. S. Garofolo, "DARPA TIMIT acoustic-phonetic speech database," *National Institute of Standards and Technology (NIST)*, 1988.

[18] T. Gerkmann and R. C. Hendriks, "Unbiased MMSE-based noise power estimation with low complexity and low tracking delay," *IEEE Audio, Speech, Language Process.*, vol. 20, no. 4, pp. 1383–1393, 2012.

[19] P. C. Loizou, *Speech Enhancement - Theory and Practice*. Boca Raton, FL, USA: CRC Press, Taylor & Francis Group, 2007.

[20] S. Gonzalez and M. Brookes, "A pitch estimation filter robust to high levels of noise (pefac)," in *EURASIP Europ. Signal Process. Conf. (EUSIPCO)*, Barcelona, Spain, Sep. 2011, pp. 451–455.

[21] ——, "Voicebox – PEFAC pitch tracker," http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/doc/voicebox/fxpefac.html, Aug. 2012.