ON PHASE IMPORTANCE IN PARAMETER ESTIMATION IN SINGLE-CHANNEL SPEECH ENHANCEMENT

Pejman Mowlaee[†] and Rahim Saeidi[‡]

[†]Signal Processing and Speech Communication Laboratory, Graz University of Technology, Graz, Austria
[‡] Centre for Language and Speech Technology, Radboud University Nijmegen, The Netherlands
`pejman.mowlaee@tugraz.at rahim.saeidi@let.ru.nl`

ABSTRACT

In this paper, we study the impact of exploiting the spectral phase information to further improve the speech quality of the single-channel speech enhancement algorithms. In particular, we focus on the two required steps in a typical single-channel speech enhancement system, namely: parameter estimation solved by a minimum mean square error (MMSE) estimator of the speech spectral amplitude, followed by signal reconstruction stage, where the observed noisy phase is often used. For the parameter estimation stage, in contrast to conventional Wiener filter, a new MMSE estimator is derived which takes into account the clean phase information as a prior information. In our experiments, we show that by including the phase information in the two steps, it is possible to improve the perceived signal quality of the enhanced signal significantly with respect to the methods that do not employ the phase information.

Index Terms— Single-channel speech enhancement, phase prior, Wiener filter.

1. INTRODUCTION

Enhancing a target speech signal from its noise corrupted observation had long been addressed as a challenging topic with applications in hearing aids, mobile telephony and robust automatic speech recognition systems in adverse noise scenarios. While previous methods have already shown successful approaches in speech enhancement, ongoing research is dedicated to further push the limits for the achievable performance especially in adverse non-stationary noise conditions.

Figure 1 shows the two steps required in a single-channel speech enhancement algorithm: i) parameter estimation (magnitude estimation) where the goal is to estimate the speech spectral amplitude DFT coefficients given its noisy observation, and ii) signal reconstruction, where given the modulus of the enhanced speech spectrum, the aim is to recover the time-domain representation of the enhanced signal.

For parameter estimation stage, Wiener filter has been extensively used as the gain function in different domain representation for speech signal, as some examples, the short-time Fourier transformation (STFT) [1, 2], log STFT [3], Karhunen Loeve transform [4], autoregressive modeling [5, 6], cepstral [7] and
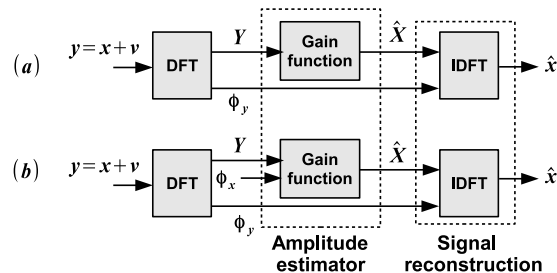
Fig. 1. Showing the two stages required in single-channel speech enhancement methods: 1) magnitude estimation, and 2) signal reconstruction. Top panel shows conventional methods using noisy phase and estimating speech amplitude without any prior on clean phase, and bottom panel shows the proposed approach where speech amplitude is estimated using prior on clean phase.

sinusoidal parameters [8]. All these methods have in common that they aim to estimate the clean speech amplitude without exploiting any prior about the phase of the speech signal (Fig. 1a).

In the signal reconstruction stage, given an estimate of the speech amplitude spectrum, noisy phase is often directly passed to reconstruct the enhanced signal. The STFT phase estimation is more difficult than STFT magnitude estimation due to the difficulty of characterizing the phase spectrum at low-energy regions as well as the fact that the noise reduction algorithms mostly only use second-order statistics in power spectrum estimation for noise reduction. Previous studies shows that accurate spectral amplitude estimation plays a more important role than improving the phase spectrum in the sense of the perceived signal quality [9]. Furthermore, [10] showed that phase becomes important when phase deviation is larger than around $\frac{\pi}{4}$, i.e. when the local signal-to-noise ratio is lower than 6 decibels, where roughness is perceived in the reconstructed signal. More recently, [11, 12], showed that using the phase spectrum leads to increase the intelligibility of the noisy speech signal. Assuming circular symmetric joint probability density function (pdf) for amplitude and phase spectra, phase is uniformly distributed and independent of the amplitude spectrum. Under such assumption, the MMSE optimal estimate of the clean speech phase is equal to the observed noisy phase [1]. This is justified from the histograms plotted in [13, 14]. However, the histograms were calculated based on taking the coefficients with equal SNRs around 20 decibels, ignoring the temporal and spectral information. Taking the temporal dynamics into account, in [15] the deviation phase group delay was shown to follow the spectral amplitude behavior, hence phase

and amplitude spectra are not generally independent. It is shown than one can estimate the phase spectrum of the original signal by iteratively performing DFT and inverse DFT, given the spectral amplitudes only [16].

It is important to emphasize that the gain functions either binary or soft given by the MMSE amplitude estimators, are derived based on the fact that the magnitude and phase spectra are independent and the posterior of the desired magnitude spectrum is derived without any dependence on the phase of the clean phase (Fig. 1a). Here, we derive the optimal estimator which takes the clean phase prior into account (Fig. 1b). In our study we assume that the clean phase of the desired speech signal is available. We acknowledge that the pre-knowledge of clean phase spectrum is unrealistic and needs to be replaced by an estimate of the clean speech phase using phase estimation algorithms, e.g. [17, 18, 19]. In our future work, we aim at presenting the general solution for parameter estimation and phase estimation where we have no prior assumption for clean phase spectrum. However, in this work, given the clean phase spectrum, we focus on presenting the proof of concept that the phase information leads to improved speech quality in single-channel speech enhancement problem. This justifies how much phase of the desired signal brings impact in parameter estimation already. We show that the clean speech phase spectrum brings additional information that can be successfully used to achieve improvement in the parameter estimation stage. For this, we derive a new MMSE estimator for amplitude spectrum given the clean phase spectrum. Through experiments, we show how much improvement is achieved by replacing the mixture phase with clean phase for oracle and estimated spectral amplitudes of speech and noise.

2. GAIN FUNCTION FOR MMSE AMPLITUDE SPEECH ESTIMATION

Let $x(n)$ and $v(n)$ as time domain signals for speech and noise. Then the noisy signal at the lth frame is $y_l(n) = x_l(n) + v_l(n)$ where $n \in [0, N-1]$ and N is the window length. The observed noisy speech STFT is given by

$$Y^c(l, k) = X^c(l, k) + V^c(l, k), \quad (1)$$

where ^{c} super index indicates the complex representation for STFT spectra and k is the frequency bin index and l is the frame index. In the complex spectrum domain, we have:

$$Y(l, k) = \sqrt{X^2(l, k) + V^2(l, k) + 2X(l, k)V(l, k)\cos\theta(l, k)}, \quad (2)$$

where Y, X and V are the magnitude spectra of noisy, clean and noise signal spectra, respectively. It is clear that the phase difference between the underlying individual signal spectra $\theta(l, k) = \phi_x(l, k) - \phi_v(l, k)$ plays an important role in the estimation of $Y(l, k)$. Because of the dependency of the noisy amplitude spectrum to the phase difference θ, the commonly used Wiener power spectrum estimation of $Y^2(l, k)$ is not the optimal estimate for the complex DFT domain. For example in [20], we recently justified the importance of this phase spectrum impact in parameter estimation stage of single-channel speech separation. The results showed that significant improvement is achievable when an estimation of the phase difference is included in the parameter estimation.
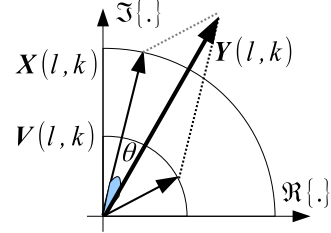

Fig. 2. Showing the speech, noise and the resulting noisy complex spectra denoted by $\mathbf{X}(l, k)$, $\mathbf{V}(l, k)$, and $\mathbf{Y}(l, k)$. The phase difference between the clean speech and noise complex spectra is marked as $\theta = \phi_X(k, l) - \phi_Y(k, l)$ which plays role in the derived mask in Eq. (10)

2.1. Wiener gain function without phase prior

According to Fig. 1a, assuming an oracle estimate of the magnitude spectra of individual signals ($\hat{X} = X$), to reconstruct the enhanced signal we need to associate the magnitude and phase spectrum and take the inverse discrete Fourier transformation (DFT). To this end, previous methods apply a mask either binary or soft (Wiener mask) on the mixture magnitude spectrum, and directly use the mixture phase unaltered for the inverse transformation. The Wiener filter has been widely used as the softmask gain function and is given by:

$$G(l, k) = \frac{X^2(l, k)}{X^2(l, k) + V^2(l, k)} \quad (3)$$

and the time-domain signal estimate for the ith signal is given by $\hat{x}(n) = \text{DFT}^{-1}\{G(l, k)Y(l, k)e^{j\angle\phi_y(l, k)}\}$.

2.2. Proposed gain function given prior clean phase spectrum

For the rest of this paper we remove the time and frequency subindices for simplicity. The vector representation of the target and masker speech spectra in the complex domain is illustrated in Fig. 2. Power spectral subtraction and Wiener filter speech estimates are derived by assuming speech and noise as independent Gaussian random processes. Here, similar to [21], we assume that speech is represented by a deterministic waveform of unknown amplitude and phase spectral coefficients denoted by $X^c = Xe^{j\phi_x}$ where X denotes the speech envelope and ϕ_x as its phase. The pdf for the noisy speech spectral amplitude given the clean speech amplitude and phase spectra at each kth frequency channel is given by:

$$p(Y|X, \phi_x) = \frac{1}{\pi\sigma_v^2}\exp\left[-\frac{Y^2 - 2X\text{Re}\{Y^c e^{-j\phi_x}\} + X^2}{\sigma_v^2}\right]. \quad (4)$$

Previous speech amplitude estimators eliminated the speech phase dependency by averaging the likelihood function for the spectral envelope, i.e., $E\{Y|X\} = \int_0^{2\pi} p(Y|X, \theta)p(\theta)d\theta$ assuming a uniform probability density function for $p(\theta)$ on $(0, 2\pi)$. The conditional probability in (4) simplifies to:

$$p(Y|X, \Phi_x) = \frac{1}{\pi\sigma_v^2}exp\left(-\frac{Y^2 + X^2 - 2XY\cos\theta}{\sigma_v^2}\right), \quad (5)$$

where $E\{V_v^2\} = \sigma_v^2$ is the PSD for the interfering signal, which denotes the complex Gaussian distribution for the interference as

$V = Y - X$, for the joint distribution for Y and Φ_Y. From Bayes rule, we have

$$p(X|Y) = \frac{p(Y|X)p(X)}{p(Y)} = \frac{p(Y|X)P(X)}{\int_0^\infty p(Y|X)P(X)dX}. \quad (6)$$

Assuming the independence between the speech amplitude and phase $p(X, \phi_x) = p(X)p(\phi_x)$, we extend (6) to the MMSE estimator of the magnitude of the clean target speech DFT coefficient given the clean target phase and mixture magnitude and phase spectra given by

$$E(X|Y, \phi_x) = \frac{\int_0^\infty Xp(Y|X, \phi_x)p(X, \phi_x)dX}{\int_0^\infty p(Y|X, \Phi_x)p(X, \phi_x)dX}. \quad (7)$$

A uniform distribution is considered for phase, i.e., $p(\phi_x) = \frac{1}{2\pi}$. To further model the speech spectral amplitudes, we utilize zero-mean Gaussian distribution for the spectral amplitude i.e. $X \sim N(0, \sigma_x^2)$

$$p(X) = \frac{1}{2\pi\sigma_x^2} \exp\left(-\frac{X^2}{2\sigma_x^2}\right), \quad (8)$$

where $\sigma_x^2 = E\{X^2\}$. Plugging Eq.(8) and Eq.(5) in Eq.(7), the numerator of (7) is given by:

$$\text{Num} = \frac{1}{4\pi^2\sigma_v^2} \int_0^\infty X^{v-1} e^{\beta_1 X^2} e^{-\gamma X} dX, \quad (9)$$

where $\beta_1 = (\frac{1}{\sigma_v^2} + \frac{1}{2\sigma_x^2})$ and $\gamma = -\frac{2Y\cos(\phi_y - \phi_x)}{\sigma_v^2}$. The denominator in Eq.(7) is given by replacing $v = 1$ in (7) and we get:

$$\text{Den} = \frac{1}{4\pi^2\sigma_v^2} \int_0^\infty e^{\beta_1 X^2} e^{-\gamma X} dX. \quad (10)$$

Plugging Equations (9),(10) into (7), and using the following integral equation (Eq. 3.462-1, page 365) from table of integrals [22]:

$$\int_0^\infty c^{v-1} e^{-\beta c^2 - \gamma c} dc = (2\beta)^{-\frac{b}{2}} \Gamma(v) \exp(\frac{\gamma^2}{8\beta}) D_{-v}(\frac{v}{\sqrt{2\beta}}), \quad (11)$$

if $\text{Re}\{\beta\} > 0$ and $\text{Re}\{v\} > 0$. For the MMSE estimator we obtain

$$\hat{X} = E(X|Y, \phi_x) = \frac{\Gamma(2) \exp(\frac{\gamma^2}{8\beta_1}) D_{-v}(\frac{\gamma}{\sqrt{2\beta_1}})}{\Gamma(1) \exp(\frac{\gamma^2}{8\beta_1}) D_{-1}(\frac{\gamma}{\sqrt{2\beta_1}})}, \quad (12)$$

where $D_{-v}(\cdot)$ is a special function, called the parabolic cylinder function of order v [23]. Calculating the expectation of the amplitude, $E(X|Y, \phi_x)$, in Eq.(7), speech amplitude spectrum estimate is given by

$$\hat{X} = 2\frac{D_{-v}(z)}{D_{-1}(z)}, \quad (13)$$

using $\beta_1 = \frac{1}{\sigma_v^2} + \frac{1}{2\sigma_x^2}$ the argument z is given by:

$$z = -\frac{2Y\cos(\phi_y - \phi_x)}{\sqrt{2\beta_1}\sigma_v^2}$$
$$= -\sqrt{\frac{2\text{SNR}_{\text{post}}\text{SNR}_{\text{priori}}}{\text{SNR}_{\text{priori}} + 0.5}} \cos(\phi_y - \phi_x), \quad (14)$$

where $\text{SNR}_{\text{priori}} = \frac{\sigma_x^2}{\sigma_v^2}$ is the *a prior SNR* and $\text{SNR}_{\text{post}} = \frac{Y^2}{\sigma_v^2}$ is the *a posterior SNR*. Using the estimated amplitude for the desired speaker used together with mixture phase, we reconstruct the time domain signal as below:

$$\hat{x}(n) = \text{DFT}^{-1}\{\hat{X}(l,k)e^{j\angle\phi_y(l,k)}\}, \quad (15)$$

where $\phi_{y,k}$ is the phase spectrum of the mixed signal. It is clear that the new mask is phase-aware and depends on the phase difference between the clean speech and mixed speech complex spectra, marked in gray color in Fig. 2.

3. EXPERIMENTS

3.1. Experimental setup

In this section, we demonstrate the impact of the proposed phase-aware MMSE amplitude estimator in a single-channel speech enhancement in a realistic scenario. For this, we select 50 utterances from randomly selected speakers among 34 speakers in the GRID corpus [24]. We decrease the sampling rate to 8 kHz. The framing is set to 20 ms as analysis window with a frame shift of 8 ms. The number of DFT points is equal as the window length. As noise signal, we select the real life noise recorded in CHiME challenge dataset [25]. The utterances have a command like structure and are of length around 1.8 second. The speech and noise signals are mixed together at signal-to-noise ratios ranging in [0 10] dB with 1 decibel of step. To quantify the amount of improvement in the single-channel speech enhancement performance, we use PESQ (as implemented in [26]) as an estimate of the perceived signal quality. The results are averaged on 50 speech utterances at 10 SNR levels.

3.2. Spectrogram analysis

Figure 3 compares the spectrograms for the enhanced speech signals using Wiener and the proposed approach in this work which ignore and takes the phase prior, respectively. The spectrograms of noisy and clean speech are also shown for reference. The proposed MMSE estimator recovers more speech and attenuate more noise components.

3.3. Phase importance in MMSE amplitude estimation

We consider the following possible scenarios
Scenario I: Both speech and noise are known (oracle): For the first scenario, Figure 4 shows the PESQ scores obtained by different MMSE amplitude estimators discussed in this work. To study the impact of the phase used in reconstruction of the enhanced signal, we also include the result obtained by replacing the mixture phase with clean phase (right panel). The following observations are made:

- The phase-aware MMSE spectral amplitude estimator consistently improves the PESQ score compared to the Wiener method. The improvement is well pronounced at signal-to-noise ratios lower than 6 decibels (zoomed panel).

- When clean phase is used for reconstruction, the improvement obtained by the proposed MMSE estimator is marginal, but still considerable for signal-to-noise ratios lower than 3 decibels.

- Replacing the mixture phase with the clean phase, results in large improvement for Wiener mask, and less for the proposed method.

Scenario II: estimated noise and oracle speech: We relax the ideal speech spectral amplitude, by replacing it with the speech amplitude spectrum estimated by improved minima controlled recursive averaging (IMCRA) presented in [27]. The PESQ results
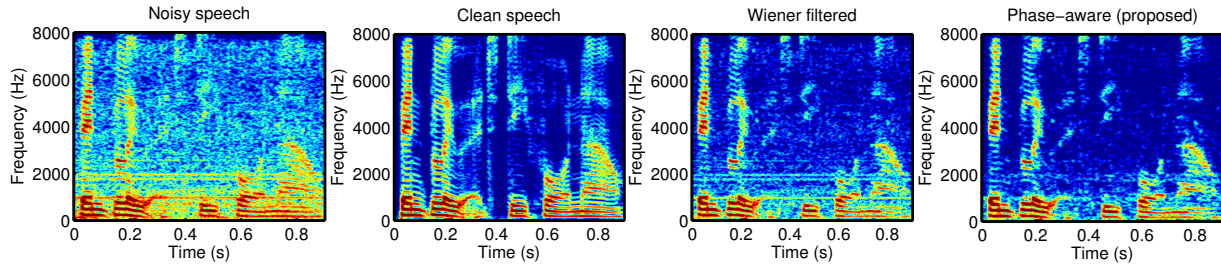
Fig. 3. Comparing the spectrograms of the enhanced speech using Wiener mask and the proposed one. The noisy and clean speech spectrograms are also shown for reference. The speech signal is a male speaker (speaker 10 from GRID corpus) saying: place green in t five again corrupted with noise at SNR = 0 decibels for scenario II: speech estimating but noise is priori known. The higher accuracy obtained by the proposed method versus Wiener filter is highlighted using black boxes. The proposed approach achieves PESQ of 2.35 compared to 2.03 of standard Wiener filter.

	Noisy phase			Clean phase		
SNR-level (dB)	0	5	10	0	5	10
Noisy	1.95±0.1	2.24±0.07	2.80±0.06	2.16±0.08	2.45±0.06	2.80±0.06
Scenario II						
Wiener	2.43±0.07	2.87±0.06	2.4±0.08	2.7±0.08	3.15±0.09	2.59±0.09
proposed	2.66±0.08	3.13±0.07	2.73±0.1	2.72±0.09	3.20±0.08	2.74±0.1
Scenario III						
Wiener	2.12±0.08	2.56±0.06	3.16±0.12	2.33±0.09	2.74±0.08	3.23±0.1
proposed	2.32±0.09	2.74±0.08	3.25±0.09	2.37±0.09	2.78±0.08	3.28±0.09

Table 1. Showing the PESQ scores comparing the proposed amplitude MMSE estimator versus typical Wiener filter for three scenarios: 1) oracle spectral speech and noise amplitudes, 2) estimated speech spectral amplitude using IMCRA and oracle noise spectral amplitude, and 3) estimated speech and noise spectral amplitudes using IMCRA [27].
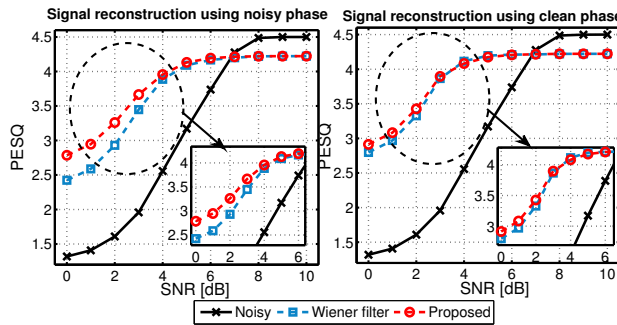

Fig. 4. Comparing the PESQ results obtained by speech amplitude estimators with and without phase prior for two scenarios: (left) noisy phase signal reconstruction and (right) when clean phase is used for signal reconstruction.

are shown in Table 1.

Scenario III: estimated speech and noise: As our last scenario, we relax any prior knowledge about speech or noise signals, and replace them from the amplitude and noise estimated provided by IMCRA. From Table 1, the following observations are made

- For the second and third scenarios, significant improvement in PESQ is obtained when using the proposed phase-aware spectral amplitude estimator rather than typical wiener gain function.

- The PESQ improvement between Wiener and the proposed method becomes less when we replace the noisy phase with the oracle clean phase.

4. CONCLUSION

Previous techniques for single-channel speech enhancement neglected the phase information in estimating the speech spectral amplitude, and used the observed noisy phase unaltered directly in their signal reconstruction stage. In this work, we demonstrated that the phase information is important in both steps. To justify this, we conducted experiments on realistic noisy speech signals at different signal-to-noise ratios, and compared the perceptual scores obtained by the phase-aware versus previous phase-independent Wiener masks. The results reveal that it is possible to achieve significant improvement in the perceived signal quality.

In this work, we assumed that we have access to the clean phase information as prior knowledge, and derived the corresponding MMSE mask function for speech spectral amplitude. In general, however, we are required to estimate the phase spectrum of the clean signal. Future work should be dedicated to estimate the phase of the target speech signal from the noisy signal (for example by using the recent phase estimation algorithms [17, 18, 19], and feed it back to combine with the proposed phase-aware MMSE amplitude estimator presented here.

---

[1]At the time that this paper was under review, we found out that a similar concept was investigated in [28].

5. REFERENCES

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

[2] J. S. Erkelens, R. C. Hendriks, R. Heusdens, and J. Jensen, "Minimum mean-square error estimation of discrete fourier coefficients with generalized gamma priors," *Trans. Audio, Speech and Lang. Proc.*, vol. 15, no. 6, pp. 1741–1752, 2007.

[3] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean square error log-spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 443–445, 1985.

[4] Y. Hu and P.C. Loizou, "A generalized subspace approach for enhancing speech corrupted by colored noise," in *IEEE Trans. Speech Audio Proc.*, 2003, pp. 334–341.

[5] S. Srinivasan, J. Samuelsson, and W. B. Kleijn, "Codebook driven short-term predictor parameter estimation for speech enhancement," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 14, no. 1, pp. 163–176, Jan. 2006.

[6] H. Sameti, H. Sheikhzadeh, L. Deng, and R. L. Brennan, "HMM-based strategies for enhancement of speech signals embedded in nonstationary noise," *IEEE Trans. Speech Audio Process.*, vol. 6, no. 5, pp. 445–455, Sept. 1998.

[7] N. Madhu, C. Breithaupt, and R. Martin, "Temporal smoothing of spectral masks in the cepstral domain for speech separation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 2008, pp. 45–48.

[8] J. Jensen and J.H.L. Hansen, "Speech enhancement using a constrained iterative sinusoidal model," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 7, pp. 731–740, 2001.

[9] D. Wang and J. Lim, "The unimportance of phase in speech enhancement," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 30, no. 4, pp. 679–681, 1982.

[10] P. Vary, "Noise suppression by spectral magnitude estimation –mechanism and theoretical limits–," *Signal Processing*, vol. 8, no. 4, pp. 387 – 400, 1985.

[11] L.D. Alsteris and K.K. Paliwal, "Short-time phase spectrum in speech processing: A review and some experimental results," *Elsevier Signal Processing*, vol. 17, no. 3, pp. 578–616, 2007.

[12] K. Paliwal, K. Wojcicki, and B. Shannon, "The importance of phase in speech enhancement," *Speech Communication*, vol. 53, no. 4, pp. 465 – 494, 2011.

[13] R. C. Hendriks and R. Martin, "MAP estimators for speech enhancement under normal and rayleigh inverse Gaussian distributions," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 15, no. 3, pp. 918–927, 2007.

[14] 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 Transactions on Audio, Speech & Language Processing*, vol. 15, no. 6, pp. 1741–1752, 2007.

[15] A. P. Stark and K. K. Paliwal, "Group-delay-deviation based spectral analysis of speech," in *INTERSPEECH*, 2009, pp. 1083–1086.

[16] D. Griffin and J. Lim, "Signal estimation from modified short-time fourier transform," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 32, no. 6, pp. 236–243, Feb. 1984.

[17] D. Gunawan and D. Sen, "Iterative phase estimation for the synthesis of separated sources from single-channel mixtures," *IEEE Signal Process. Lett.*, vol. 17, no. 6, pp. 421–424, May 2010.

[18] P. Mowlaee, R. Saiedi, and R. Martin, "Phase estimation for signal reconstruction in single-channel speech separation," in *Proceedings of the International Conference on Spoken Language Processing*, 2012.

[19] N. Sturmel and L. Daudet, "Iterative phase reconstruction of Wiener filtered signals," in *ICASSP*, 2012, pp. 101–104.

[20] P. Mowlaee and R. Martin, "On phase importance in parameter estimation for single-channel source separation," in *The International Workshop on Acoustic Signal Enhancement (IWAENC)*, 2012, pp. 1–4.

[21] R. Mcaulay and M. Malpass, "Speech enhancement using a soft-decision noise suppression filter," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 28, no. 2, pp. 137–145, 1980.

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

[23] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover Publications, New York, 1964.

[24] M. Cooke, J. R. Hershey, and S. J. Rennie, "Monaural speech separation and recognition challenge," *Elsevier Computer Speech and Language*, vol. 24, no. 1, pp. 1–15, 2010.

[25] H. Christensen, J. Barker, N. Ma, and P. Green, "The CHiME corpus: a resource and a challenge for computational hearing in multisource environments," in *Proc. Interspeech*, 2010, pp. 1918–1921.

[26] P. Loizou, *Speech Enhancement: Theory and Practice*, CRC Press, Boca Raton, 2007.

[27] I. Cohen, "Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging," *Speech and Audio Processing, IEEE Transactions on*, vol. 11, no. 5, pp. 466 – 475, Sept. 2003.

[28] T. Gerkmann and M. Krawczyk, "Mmse-optimal spectral amplitude estimation given the stft-phase," *Signal Processing Letters, IEEE*, vol. 20, no. 2, pp. 129 –132, feb. 2013.