A Study on the Benefits of Phase-Aware Speech Enhancement in Challenging Noise Scenarios

Martin Krawczyk-Becker(✉) and Timo Gerkmann

Universität Hamburg, 20148 Hamburg, Germany
{martin.krawczyk-becker,timo.gerkmann}@uni-hamburg.de

Abstract. In recent years, there has been a renaissance of research on the role of the spectral phase in single-channel speech enhancement. One of the recent proposals is to not only estimate the clean speech phase but also use this phase estimate as an additional source of information to facilitate the estimation of the clean speech magnitude. To assess the potential benefit of such approaches, in this paper we systematically explore in which situations additional information about the clean speech phase is most valuable. For this, we compare the performance of phase-aware and phase-blind clean speech estimators in different noise scenarios, i.e. at different signal to noise ratios (SNRs) and for noise sources with different degrees of stationarity. Interestingly, the results indicate that the greatest benefits can be achieved in situations where conventional magnitude-only speech enhancement is most challenging, namely in highly non-stationary noises at low SNRs.

Keywords: Phase · Speech enhancement · Noise reduction

1 Introduction

The enhancement of speech that is corrupted by noise is a long-standing research topic that has seen many new ideas and improvements over the last decades. In this paper, we focus on single-channel speech enhancement, i.e. approaches that are applied to a single microphone signal or to the output of a multichannel preprocessing stage. Specifically, we consider minimum mean square error (MMSE) optimal Bayesian estimators of the clean speech in the short-time discrete Fourier transform (STFT) domain. Well-known examples of this class of estimators are the Wiener filter and Ephraim and Malah's short-time spectral amplitude estimator (STSA) [2]. Over the years, numerous improvements have been proposed, including the use of super-Gaussian speech priors [4,21] and/or different optimization criteria [1,3,28]. See e.g. [13] for a concise overview. What the vast majority of mainstream approaches have in common is that they are magnitude-centric, meaning that the spectral phase is neither used as a source of information nor is the noise corrupted spectral phase enhanced, which is frequently justified by the statement that the enhancement of the spectral phase

Y. Deville et al. (Eds.): LVA/ICA 2018, LNCS 10891, pp. 407–416, 2018.
https://doi.org/10.1007/978-3-319-93764-9_38

is unimportant [27]. However, contrary to the widespread believe at the time, more recent studies, including [9,24], showed that the spectral phase is indeed important for speech enhancement. These findings sparked a renewed interest in the estimation of the clean speech spectral phase for speech enhancement, e.g. [10,16,22].

With the availability of phase estimates, also the interest in how these phase estimates can best be utilized has risen. A straight forward way is to simply exchange the noisy phase with the phase estimate and combine it with a clean speech magnitude estimate that has been obtained with an existing of-the-shelf estimator. A more elaborate way to utilize the newly available phase estimate is to use it as an additional source of information that facilitates the estimation of the clean speech magnitudes [8,18] or even the complex-valued coefficients [6]. We denote such approaches as being *phase-aware*, while conventional magnitude-centric approaches like the Wiener filter or the STSA are considered *phase-blind*.

Phase-aware approaches have been shown to be capable of generally outperforming their phase-blind counterparts in terms of instrumental measures, e.g. in [8,18,23], and also by means of formal listening experiments [17]. To assess the potential of phase-aware speech enhancement in more detail, in this paper we systematically investigate in which acoustic situations it provides the largest benefits. For this, we directly compare the performance of two phase-aware estimators based on [6,18] to that of their phase-blind counterparts, namely the STSA and the Wiener filter, at different SNRs and for noise sources with different degrees of stationarity. First, we consider pink noise and modulate it with an increasing modulation frequency, which allows us to adjust the amount of non-stationarity in a very controlled way. As a second, practically very relevant example, we use babble noise, where the non-stationarity is adjusted by deliberately changing the number of talkers. The results indicate that the greatest benefits can be achieved in situations where conventional phase-blind speech enhancement is most challenging, i.e. in highly non-stationary noises at low SNRs.

2 Signal Model and Notation

In each time-frequency point of the STFT domain we have a additive superposition of mutually independent speech and noise,

$$Y = S + V = A\mathrm{e}^{\mathrm{j}\Phi^S} + D\mathrm{e}^{\mathrm{j}\Phi^V} = R\mathrm{e}^{\mathrm{j}\Phi^Y}, \qquad (1)$$

where Y, S, and V denote the complex coefficients of the observed noisy speech, the desired clean speech, and the additive noise, respectively. The spectral phases are denoted by Φ^Y, Φ^S, and Φ^V, while the spectral magnitudes are denoted by R, A, and D. Here we make the common assumption that the noise coefficients V follow a circular symmetric zero-mean complex Gaussian distribution with a power spectral density (PSD) of σ_{V}^2, where the circular symmetry implicates a uniformly distributed noise phase Φ^V. The PSD of speech is denoted as σ_{S}^2. We use the hat-symbol to distinguish estimates from their true counterparts, i.e. $\widehat{S}$ is an estimate of S.

3 Conventional Phase-Blind Clean Speech Estimation

Commonly, MMSE estimators of the clean speech S, or more generally any function $f(S)$, are derived by finding the expected value of $f(S)$ given the noisy observation and the PSDs of speech and noise:

$$\widehat{f(S)} = \mathrm{E}\big(f(S) \mid Y, \sigma_{\mathrm{S}}^2, \sigma_{\mathrm{V}}^2\big) = \int_0^{\infty}\int_0^{2\pi} f(S) p\big(A, \Phi^S \mid Y, \sigma_{\mathrm{S}}^2, \sigma_{\mathrm{V}}^2\big)\, \mathrm{d}\Phi^S\, \mathrm{d}A \tag{2}$$

$$= \frac{\int_0^{\infty}\int_0^{2\pi} f(S) p\big(y|A, \Phi^S, \sigma_{\mathrm{V}}^2\big)\, p\big(A \mid \sigma_{\mathrm{S}}^2\big)\, p\big(\Phi^S\big)\, \mathrm{d}\Phi^S \mathrm{d}A}{\int_0^{\infty}\int_0^{2\pi} p(y|A, \Phi^S, \sigma_{\mathrm{V}}^2)\, p(A \mid \sigma_{\mathrm{S}}^2)\, p(\Phi^S)\, \mathrm{d}\Phi^S \mathrm{d}A}, \tag{3}$$

where the second line is obtained by applying Bayes' rule. For complex Gaussian noise, the likelihood is given as $p\big(y|A, \Phi^S, \sigma_{\mathrm{V}}^2\big) = \mathcal{N}\big(S, \sigma_{\mathrm{V}}^2\big)$, see e.g. [2]. For a uniform *phase* prior, i.e. $p\big(\Phi^S\big) = 1/(2\pi)$ for $\Phi^S \in [-\pi, \pi)$, Eq. (3) has been solved analytically for different *magnitude* priors $p\big(A \mid \sigma_{\mathrm{S}}^2\big)$ and functions $f(S)$. Assuming a Rayleigh distribution for $p\big(A \mid \sigma_{\mathrm{S}}^2\big)$, for instance, the Wiener filter is obtained as the MMSE optimal estimator of the complex clean speech coefficients ($f(S) = S$) and Ephraim and Malah's STSA as the MMSE optimal estimators of the clean speech magnitudes ($f(S) = A$). Also more elaborate super-Gaussian clean speech estimators have been derived via (3) by using, e.g., a χ distribution [1] or a generalized gamma distribution [4] for $p\big(A \mid \sigma_{\mathrm{S}}^2\big)$ with different functions $f(S)$. However, in all these approaches the phase prior $p\big(\Phi^S\big)$ is modeled as a uniform distribution, which implies that the complex clean speech coefficients are circularly-complex distributed. Without any prior information about the clean speech spectral phase, the uniform distribution is indeed a reasonable assumption that is supported by long term histogram data [4].

4 Phase-Aware Clean Speech Estimation

In contrast to the conventional phase-blind approaches discussed above, phase-aware estimators such as the ones in [6,8,18] assume that besides the speech and noise PSDs also a prior estimate of the clean speech spectral phase is available. Such a phase estimate can be obtained from the noisy signal for instance based on a harmonic model such as in [16,22] or using an iterative approach similar to Griffin and Lim [12] and its successors [19,25]. To derive MMSE optimal phase-aware estimators, we propose to compute the expected value of $f(S)$ conditioned not only on Y, σ_{S}^2, and σ_{V}^2 as for conventional estimators, but also on the prior phase estimate $\widehat{\Phi^S}$ [6,8,18]:

$$\widehat{f(S)} = \mathrm{E}\Big(f(S) \mid Y, \sigma_{\mathrm{S}}^2, \sigma_{\mathrm{V}}^2, \widehat{\Phi^S}\Big) \tag{4}$$

$$= \frac{\int_0^{\infty}\int_0^{2\pi} f(S) p\big(y|A, \Phi^S, \sigma_{\mathrm{V}}^2\big)\, p\big(A \mid \sigma_{\mathrm{S}}^2\big)\, p\Big(\Phi^S|\widehat{\Phi^S}\Big)\, \mathrm{d}\Phi^S \mathrm{d}A}{\int_0^{\infty}\int_0^{2\pi} p(y|A, \Phi^S, \sigma_{\mathrm{V}}^2)\, p(A \mid \sigma_{\mathrm{S}}^2)\, p\Big(\Phi^S|\widehat{\Phi^S}\Big)\, \mathrm{d}\Phi^S \mathrm{d}A}, \tag{5}$$

where the second line is again obtained using Bayes' rule and making only mild assumptions. Comparing the phase-aware estimator in (5) and the phase-blind estimator in (3), it can be seen that the only difference is the replacement of $p(\Phi^S)$ by $p\left(\Phi^S|\widehat{\Phi^S}\right)$. If the prior phase estimate $\widehat{\Phi^S}$ is informative, the true clean speech phase Φ^S does not follow a uniform distribution anymore. Instead, $p\left(\Phi^S|\widehat{\Phi^S}\right)$ reflects uncertain information about the true clean speech phase. Similar to [6,18] we employ a von Mises distribution with mean direction $\widehat{\Phi^S}$ to model this uncertainty in the prior phase estimate:

$$p\left(\Phi^S|\widehat{\Phi^S}\right) = \exp\left(\varkappa \cos\left(\Phi^S - \widehat{\Phi^S}\right)\right) / \left(2\pi I_0(\varkappa)\right), \tag{6}$$

where $\varkappa$ is the concentration parameter and $I_0(\cdot)$ is the modified Bessel function of the first kind and zeroth-order. Examples for $p\left(\Phi^S|\widehat{\Phi^S}\right)$ for $\widehat{\Phi^S} = 0$ and different concentration parameters $\varkappa$ are presented in Fig. 1. The larger $\varkappa$, the more $p\left(\Phi^S|\widehat{\Phi^S}\right)$ is concentrated around the prior phase estimate $\widehat{\Phi^S}$. Accordingly, $\widehat{\Phi^S}$ is modeled as an increasingly accurate estimate of the true clean speech phase Φ^S. For the extreme case of $\varkappa \to \infty$, the distribution reduces to a single peak at $\widehat{\Phi^S}$, i.e. the prior phase estimate is assumed to be exactly the true clean speech phase Φ^S. On the contrary, the lower $\varkappa$, the wider $p\left(\Phi^S|\widehat{\Phi^S}\right)$, which corresponds to modeling less accurate prior estimates. For $\varkappa = 0$, $p\left(\Phi^S|\widehat{\Phi^S}\right)$ reduces to a uniform distribution and the prior phase estimate $\widehat{\Phi^S}$ does not provide any useful information about the true phase Φ^S, i.e. $p\left(\Phi^S \mid \widehat{\Phi^S}\right) = p(\Phi^S)$. In this special case, the phase-aware estimator in (5) degenerates to a conventional phase-blind estimator similar to (3).

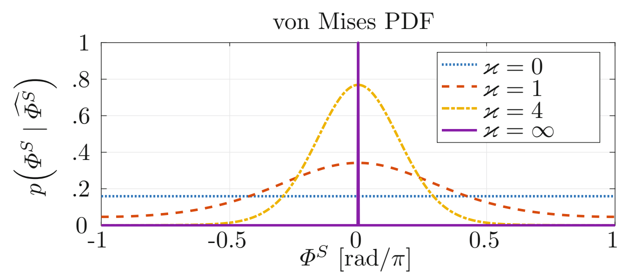

Fig. 1. Von Mises distribution for $p\left(\Phi^S|\widehat{\Phi^S}\right)$ with a mean direction of $\widehat{\Phi^S} = 0$ and an increasing concentration parameter $\varkappa$.

Very general super-Gaussian phase-aware estimators of the clean speech spectral magnitudes $f(S) = A^\beta$ and the complex clean speech coefficients $f(S) = A^\beta e^{j\Phi^S}$ have been derived in [6,18] by solving (5) using a flexible χ distribution for the magnitude prior $p(A \mid \sigma_s^2)$. The magnitude estimator uses the prior phase estimate $\widehat{\Phi^S}$ only to facilitate the estimation of the clean speech magnitude. Similar to the phase-blind estimators, the spectral phase is not modified and the estimated magnitude is combined with the noisy phase Φ^Y to obtain the final estimate. The complex estimator, however, not only enhances the spectral magnitude but also jointly enhances the spectral phase.

For simplicity, in this paper we consider only two special cases of the general phase-aware estimators in [6,18]. Specifically, we set $\beta = 1$, i.e. we estimate $f(S) = Ae^{j\Phi^S} = S$ and $f(S) = A$, and choose the parameter of the χ distribution such that it reduces to a Rayleigh distribution. Both, the simplified estimator of the clean speech coefficients $f(S) = S$ as well as the simplified estimator of the clean speech magnitudes $f(S) = A$ have well-known phase-blind counterparts: If the prior phase estimate is assumed to provide no useful information, i.e. $\varkappa = 0$, it has been shown in [18] that the simplified estimator of S reduces to the Wiener filter, while the simplified magnitude estimators reduces to the STSA [2]. This direct relation between phase-aware and well-known phase-blind estimators allow to investigate the effects of phase-aware speech enhancement in isolation. We denote the simplified phase-aware magnitude estimator ($f(S) = A$) as PAM and the simplified phase-aware complex estimator ($f(S) = S$) as PAC.

5 Evaluation

In this section, we evaluate in which acoustic scenarios phase-aware speech enhancement is most beneficial. For this, the two simplified phase-aware clean speech estimators are compared to their respective phase-blind counterparts. Specifically, we compare Ephraim and Malah's STSA [2] to the PAM and the conventional Wiener filter to the PAC. The evaluation is performed on 128 gender balanced utterances taken from the TIMIT database [5] at a sampling rate of 16 kHz. In the first part, the clean speech utterances are deteriorated by stationary pink noise and pink noise modulated with an increasing modulation frequency, i.e. 0.5 Hz, 1 Hz, and 2 Hz. This allows us to investigate how the performance of phase-aware speech enhancement depends on the non-stationarity of the noise in a very controlled manner. Furthermore, to assess the influence of the SNR on phase-aware speech enhancement, this experiment is conducted for two SNRs, namely 0 dB and 10 dB. We present three measures: global SNR, raw wideband 'Perceptual Evaluation of Speech Quality' (WB-PESQ) scores [15], and raw 'Short-Time Objective Intelligibility Measure' (STOI) values [26].

As a less controlled but practically very relevant example, in the second experiment we deteriorate the clean speech utterances with babble noise, where the amount of non-stationarity is controlled by the number of speakers. The noise is created by randomly superimposing TIMIT sentences that have not been used as clean speech material, with the number of speakers ranging from 40, which represents the most stationary example, to a single competing talker as the most non-stationary noise.

In both experiments, the STFT representation is obtained with a segment length of 32 ms and a segment shift of 8 ms with a square-root Hann window for spectral analysis and synthesis without zero padding. The speech PSD is estimated using the decision-directed approach [2] with a smoothing parameter of 0.96. The noise PSD is estimated via [7]. The maximum attenuation in each time frequency point is set to −15 dB, which is a common way to reduce artifacts in the enhanced signal by introducing a residual noise floor. To assess the full potential of phase-aware speech enhancement without the shortcomings of current phase estimators, here the true clean speech phase Φ^S is provided as the prior phase estimate $\widehat{\Phi^S}$. The concentration parameter in (6) is accordingly set to $\kappa \rightarrow \infty$. Please note that for this specific choice of $\varkappa$, the only difference between PAM and PAC is that PAC combines the magnitude estimate with the noisy phase, while PAC uses the prior phase $\widehat{\Phi^S}$. In practice, the clean speech phase is however not available. Therefore, we finally also present results for the case that the prior phase is blindly estimated via [16] to further confirm the outcome of the oracle experiments.

5.1 Modulated Pink Noise

In Fig. 2, we present global SNR, WB-PESQ, and STOI for pink noise with an increasing amount of non-stationarity. For a better accessibility, we do not present absolute values but rather the improvement of the phase-aware estimator over its conventional phase-blind counterpart. First, it can be seen that the benefit of phase-aware speech enhancement is generally larger at low SNRs (top) than at higher SNRs (bottom). Second, independent of the SNR, the benefit of phase-aware speech enhancement increases with increasing non-stationarity. Generally, in non-stationary noises at low SNRs, speech enhancement is most challenging, specifically because the estimation of the speech PSD σ_{S}^2 and the noise PSD σ_{V}^2 becomes increasingly difficult. For instance, most noise PSD estimators, including minimum statistics [20] and the estimator based on speech presence probability [7] that is employed here, rely on the assumption that noise is more stationary than speech. Such approaches consequently become less accurate for highly non-stationary noise. This is also reflected in an increasing log distortion error [14]

$$\mathrm{LOG{-}Err_{seg}} = \mathrm{mean} \left| 10 \log_{10} \frac{\sigma_{\mathrm{V}}^2}{\widehat{\sigma_{\mathrm{V}}^2}} \right|, \tag{7}$$

where the mean is taken over all time-frequency points. In the noise-only case, $\mathrm{LOG{-}Err_{seg}}$ gradually increases from 1.5 for stationary pink noise to 3.1 for pink noise modulated with 2 Hz. See e.g. [7,14] for a more detailed discussion on this topic.

Since the conventional phase-blind estimators (3) rely solely on the PSD estimates, PSD estimation errors propagate through to the final estimate, leading to noise leakage and/or speech distortions. These artifacts can substantially reduce the speech enhancement performance. The fact that phase-aware speech enhancement provides the most benefit specifically in these situations highlights its relevance and potential. Furthermore, comparing the complex estimator PAC to the magnitude estimator PAM it can be seen that the phase enhancement of PAC leads to an additional improvement in all three measures and all acoustic situations.

The consistently small gains in STOI at 10 dB SNR on the bottom right of Fig. 2 can be explained by the fact that the speech intelligibility, which is predicted by STOI, is already close to 100% even for the noisy signal at high SNRs. Thus there is only little room for improvement.

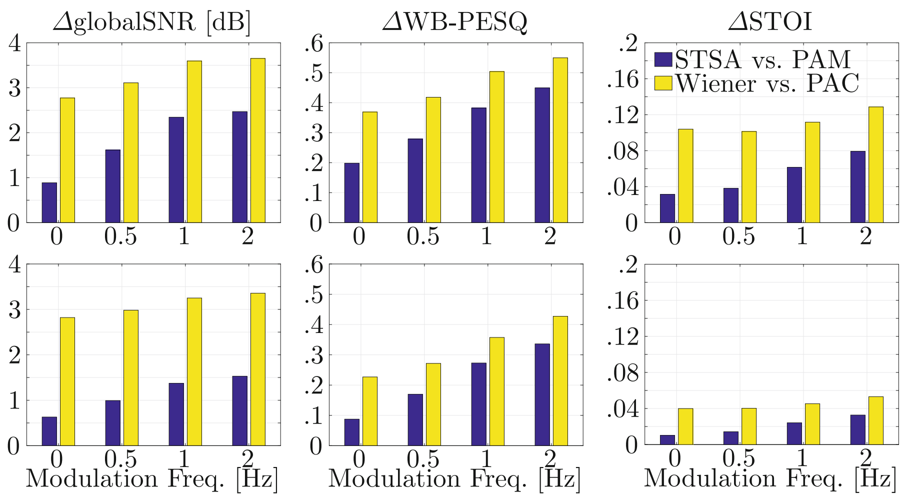

Fig. 2. Improvement of the phase-aware estimators over their respective phase-blind counterparts in SNR, WB-PESQ, and STOI for pink noise modulated with an increasing modulation frequency, representing an increasing degree of non-stationarity. SNR: 0 dB (top) and 10 dB (bottom).

5.2 Babble Noise

In Fig. 3, we present the results for babble noise with a varying number of talkers. The fewer talkers the noise is comprised of, the less stationary it is. Similar to the first experiment in Fig. 2, the benefit of phase-aware speech enhancement is most prominent in highly non-stationary noise at low SNRs. The largest improvements are achieved for 5-talker babble noise, while a for a single interfering talker the improvement in WB-PESQ and STOI is somewhat lower, especially for the complex estimator PAC.

Finally, in Fig. 4, we present results that are achieved when the prior phase $\widehat{\Phi^S}$ is estimated blindly on the noisy signal Y via [16]. No oracle information is used. Although the improvements are substantially smaller than for the oracle

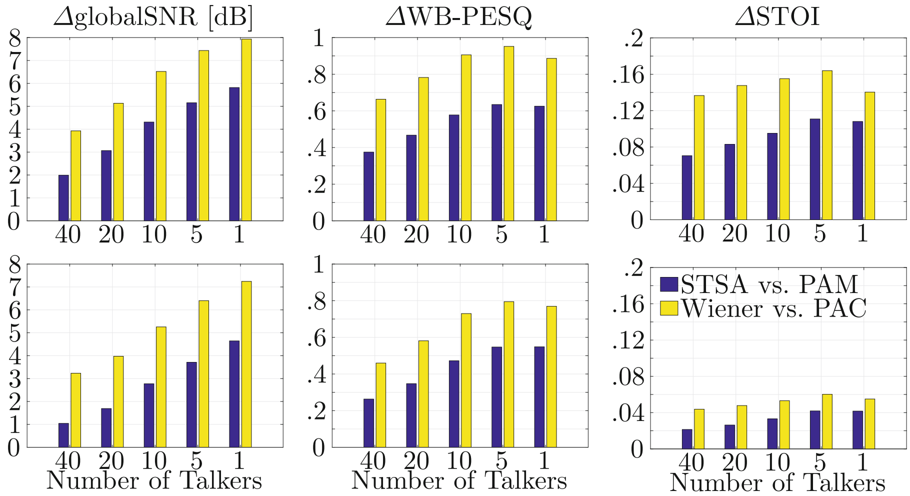

Fig. 3. Improvement of the phase-aware estimators over their respective phase-blind counterparts in SNR, WB-PESQ, and STOI for babble noise with a decreasing number of talkers, representing an increasing degree of non-stationarity. SNR: 0 dB (top) and 10 dB (bottom).

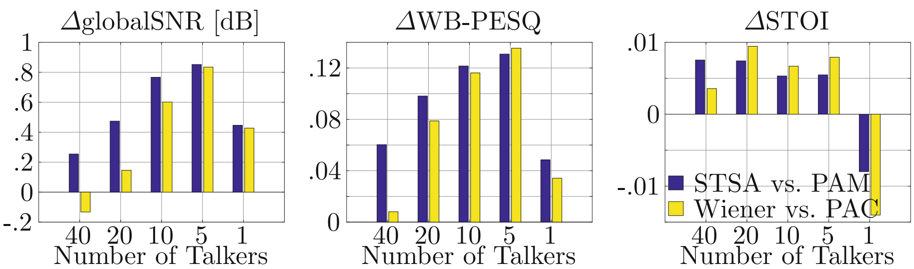

Fig. 4. Improvement of the phase-aware estimators over their respective phase-blind counterparts in SNR, WB-PESQ, and STOI for babble noise with a decreasing number of talkers at 0 dB SNR. The prior phase is blindly estimated via [16].

experiment in Fig. 3, similar trends can be observed. While the improvement in STOI is generally small, SNR and WB-PESQ improvements are again the largest for the highly non-stationary 5-talker babble. The reduced performance for the single competing talker is likely a consequence of the phase estimation process: In [16], the spectral phase of voiced speech is estimated based on a harmonic signal model, which in turn relies on a fundamental frequency estimate obtained via [11]. For a single competing talker at 0 dB SNR, estimating the fundamental frequency only of the desired speaker becomes extremely challenging. Thus the prior phase estimate can strongly deviate from the true clean speech phase in this situation and might even resemble the phase of the competing talker at times, which limits the overall speech enhancement performance.

While the impressive performance gains for the oracle experiments in Figs. 2 and 3 clearly highlight the potential of phase-aware speech enhancement, the current gap between the oracle performance and the one in Fig. 4 makes research into more robust and accurate phase estimation techniques a relevant and promising topic for single-channel speech enhancement.

6 Conclusions

In this paper, we investigated in which situations additional information of the clean speech phase is most valuable. The results show that the greatest benefits can be achieved in situations where conventional magnitude-only speech enhancement is most challenging, namely in highly non-stationary noises at low SNRs. The current gap between the optimal performance of phase-aware speech enhancement and the performance obtained using blindly estimated prior phases highlight the importance of ongoing research into robust and accurate phase estimation techniques.

References

1. Breithaupt, C., Gerkmann, T., Martin, R.: A novel a priori SNR estimation approach based on selective cepstro-temporal smoothing. In: IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Las Vegas, pp. 4897–4900 (2008)
2. Ephraim, Y., Malah, D.: Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. IEEE Trans. Acoust. Speech Signal Process. **32**(6), 1109–1121 (1984)
3. Ephraim, Y., Malah, D.: Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. IEEE Trans. Acoust. Speech Signal Process. **33**(2), 443–445 (1985)
4. Erkelens, J.S., Hendriks, R.C., Heusdens, R., Jensen, J.: Minimum mean-square error estimation of discrete Fourier coefficients with generalized Gamma priors. IEEE Trans. Audio Speech Lang. Process. **15**(6), 1741–1752 (2007)
5. Garofolo, J.S., Lamel, L.F., Fisher, W.M., Fiscus, J.G., Pallett, D.S., Dahlgren, N.L.: DARPA TIMIT acoustic phonetic continuous speech corpus CDROM (1993)
6. Gerkmann, T.: Bayesian estimation of clean speech spectral coefficients given a priori knowledge of the phase. IEEE Trans. Signal Process. **62**(16), 4199–4208 (2014)
7. Gerkmann, T., Hendriks, R.C.: Unbiased MMSE-based noise power estimation with low complexity and low tracking delay. IEEE Trans. Audio Speech Lang. Process. **20**(4), 1383–1393 (2012)
8. Gerkmann, T., Krawczyk, M.: MMSE-optimal spectral amplitude estimation given the STFT-phase. IEEE Signal Process. Lett. **20**(2), 129–132 (2013)
9. Gerkmann, T., Krawczyk, M., Rehr, R.: Phase estimation in speech enhancement – unimportant, important, or impossible? In: IEEE Convention of Electrical and Electronics Engineers in Israel, Eilat, Israel (2012)

10. Gerkmann, T., Krawczyk-Becker, M., Le Roux, J.: Phase processing for single channel speech enhancement: history and recent advances. IEEE Signal Process. Mag. **32**(2), 55–66 (2015)
11. Gonzalez, S., Brookes, M.: PEFAC - a pitch estimation algorithm robust to high levels of noise. IEEE Trans. Audio Speech Lang. Process. **22**(2), 518–530 (2014)
12. Griffin, D.W., Lim, J.S.: Signal estimation from modified short-time Fourier transform. IEEE Trans. Acoust. Speech Signal Process. **32**(2), 236–243 (1984)
13. Hendriks, R.C., Gerkmann, T., Jensen, J.: DFT-Domain Based Single-Microphone Noise Reduction for Speech Enhancement: A Survey of the State-of-the-Art. Morgan & Claypool, Colorado (2013)
14. Hendriks, R.C., Jensen, J., Heusdens, R.: Noise tracking using DFT domain subspace decompositions. IEEE Trans. Audio Speech Lang. Process. **16**(3), 541–553 (2008)
15. ITU-T: Perceptual evaluation of speech quality (PESQ). ITU-T Recommendation P.862 (2001)
16. Krawczyk, M., Gerkmann, T.: STFT phase reconstruction in voiced speech for an improved single-channel speech enhancement. IEEE/ACM Trans. Audio Speech Lang. Process. **22**(12), 1931–1940 (2014)
17. Krawczyk-Becker, M., Gerkmann, T.: An evaluation of the perceptual quality of phase-aware single-channel speech enhancement. J. Acoust. Soc. Am. **140**(4), EL364–EL369 (2016)
18. Krawczyk-Becker, M., Gerkmann, T.: On MMSE-based estimation of spectral speech coefficients under phase-uncertainty. IEEE/ACM Trans. Audio Speech Lang. Process. **24**(12), 2251–2262 (2016)
19. Le Roux, J., Vincent, E.: Consistent Wiener filtering for audio source separation. IEEE Signal Process. Lett. **20**(3), 217–220 (2013)
20. Martin, R.: Noise power spectral density estimation based on optimal smoothing and minimum statistics. IEEE Trans. Speech Audio Process. **9**(5), 504–512 (2001)
21. Martin, R.: Speech enhancement based on minimum mean-square error estimation and supergaussian priors. IEEE Trans. Speech Audio Process. **13**(5), 845–856 (2005)
22. Mowlaee, P., Kulmer, J.: Harmonic phase estimation in single-channel speech enhancement using phase decomposition and SNR information. IEEE/ACM Trans. Audio Speech Lang. Process. **23**(9), 1521–1532 (2015)
23. Mowlaee, P., Saeidi, R.: Iterative closed-loop phase-aware single-channel speech enhancement. IEEE Signal Process. Lett. **20**(12), 1235–1239 (2013)
24. Paliwal, K., Wójcicki, K., Shannon, B.: The importance of phase in speech enhancement. ELSEVIER Speech Commun. **53**(4), 465–494 (2011)
25. Sturmel, N., Daudet, L.: Signal reconstruction from STFT magnitude: a state of the art. In: International Conference on Digital Audio Effects (DAFx), Paris, France, pp. 375–386 (2011)
26. Taal, C.H., Hendriks, R.C., Heusdens, R., Jensen, J.: An algorithm for intelligibility prediction of time-frequency weighted noisy speech. IEEE Trans. Audio Speech Lang. Process. **19**(7), 2125–2136 (2011)
27. Wang, D.L., Lim, J.S.: The unimportance of phase in speech enhancement. IEEE Trans. Acoust. Speech Signal Process. **30**(4), 679–681 (1982)
28. You, C.H., Koh, S.N., Rahardja, S.: β-order MMSE spectral amplitude estimation for speech enhancement. IEEE Trans. Speech Audio Process. **13**(4), 475–486 (2005)