MMSE-OPTIMAL ENHANCEMENT OF COMPLEX SPEECH COEFFICIENTS WITH UNCERTAIN PRIOR KNOWLEDGE OF THE CLEAN SPEECH PHASE

Timo Gerkmann

Speech Signal Processing Group, Department of Medical Physics and Acoustics
Cluster of Excellence *Hearing4all*, Universität Oldenburg, Germany
`timo.gerkmann@uni-oldenburg.de`

ABSTRACT

In most STFT-based speech enhancement algorithms only the STFT amplitude of speech is processed, while the STFT phase of the noisy signal is neither modified nor employed to improve amplitude estimation. This is also, because modifying the spectral phase often yields undesired artifacts and unnatural sounding speech.

In this paper, we first obtain a clean speech phase estimate using a recent phase reconstruction algorithm. Then, we propose to treat this reconstructed phase as uncertain a priori knowledge when deriving a joint MMSE estimate of the clean speech amplitude and phase. The resulting MMSE-estimator yields a compromise between the phase of the noisy signal and the prior phase estimate. Instrumental measures and informal listening show that the proposed estimator reduces undesired artifacts and results in an improved speech quality.

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

1. INTRODUCTION

In many situations where speech signals are captured by speech communication devices like hearing aids or cell phones, the speech signals are disturbed by additive noise. In this paper we address the enhancement of such noisy signals in the short time discrete Fourier transform (STFT) domain. In most commercially available noise reduction systems only the clean speech magnitude is changed, while the noisy phase is kept unchanged [1]. However, recently it has been shown that speech enhancement algorithms may be improved if the clean speech phase is known [2] leading to an increased interest in phase processing [3]–[9]. In principle, the estimation of the clean speech phase is possible for instance by iteratively synthesizing and reanalyzing the clean speech amplitudes [10]. However, these phase reconstruction algorithms require the clean speech amplitudes to be known. When only estimates are available, the resulting speech quality may degrade greatly. See also [3, 11, 7, 9] for more information and recent advances in iterative phase estimation.

Recently, we proposed to estimate the clean speech phase in voiced speech based on an estimate of the speech fundamental frequency [4]. We could show that this phase estimate can be employed as an additional input to improve the estimation of the clean speech spectral amplitudes [6]. However, when replacing the noisy phase by the estimated clean speech phase for signal reconstruction, artifacts may occur [3, 11, 4]. These artifacts occur whenever the phase estimate is erroneous. Using [4], this may for instance happen when reconstructing the phase at higher signal harmonics, as small errors in the fundamental frequency estimate are multiplied by the harmonic number.

In this paper, we assume we have an estimate of the clean speech phase available and treat it as uncertain a prior knowledge when deriving a joint minimum mean square error (MMSE)-estimate of the clean speech phase and amplitude (Section 3). Without loss of generality, this prior phase information is obtained using the sinusoidal model based phase reconstruction algorithm [4]. We show that taking the uncertainty about the prior phase estimate into account allows us to increase the instrumentally predicted perceptual quality of speech (Section 4), which is also confirmed by informal listening. Section 5 concludes this paper.

2. SIGNAL MODEL AND NOTATION

In this work, we assume we observe a noisy speech signal $Y_k(\ell)$ given by the additive superposition of a speech signal $S_k(\ell)$ and a noise signal $N_k(\ell)$. Thus, in the STFT domain, the noisy observation is denoted by

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

The segment index ℓ and the frequency index k, are omitted in the sequel unless needed. The complex coefficients can be represented by their amplitudes and phases denoted as

$$Y = Re^{j\Phi_Y}; \qquad S = Ae^{j\Phi_S}; \qquad N = Ve^{j\Phi_N}. \tag{2}$$

We denote random variables by capital letters, e.g. S, A, Φ_S, and their realizations by the corresponding lower case letters, e.g. s, a, ϕ_S.

3. DERIVATION OF THE PROPOSED ESTIMATOR

The goal of this paper is to find the MMSE-optimal estimator of the (C)omplex speech coefficients given (U)ncertain (P)hase information (CUP). For this we have to solve $\mathrm{E}(S \mid Y, \widetilde{\phi}_\mathrm{S})$, where $\widetilde{\phi}_\mathrm{S}$ denotes a priori knowledge on the clean speech phase. Without loss of generality, this prior phase information can be obtained using the phase reconstruction algorithm proposed in [4]. In [12] it has been argued that estimating logarithmically compressed spectral amplitudes is perceptually advantageous. You et al. generalized the log-artihmic amplitude compression [12] by employing a compression parameter β [13]. As in [14, 6] we adopt this idea and incorporate the compression parameter β in our estimator. Given prior knowledge on the phase, we thus have to solve

$$
\widehat{S}^{(\beta)} = \mathrm{E}\left(A^\beta \mathrm{e}^{\mathrm{j}\Phi_\mathrm{S}} \mid y, \widetilde{\phi}_\mathrm{S}\right)
$$
$$
= \int_0^\infty \int_0^{2\pi} a^\beta \mathrm{e}^{\mathrm{j}\phi_\mathrm{S}} \, p_{A,\Phi_\mathrm{S}|y,\widetilde{\phi}_\mathrm{S}}\left(a, \phi_\mathrm{S} \mid y, \widetilde{\phi}_\mathrm{S}\right) \mathrm{d}\phi_\mathrm{S} \mathrm{d}a. \quad (3)
$$

In order to solve (3), we need a model for the posterior function $p_{A,\Phi_\mathrm{S}|y,\widetilde{\phi}_\mathrm{S}}$. With Bayes' theorem we can write

$$
p_{A,\Phi_\mathrm{S}|y,\widetilde{\phi}_\mathrm{S}}\left(a, \phi_\mathrm{S} \mid y, \widetilde{\phi}_\mathrm{S}\right) = \frac{p_{Y,A,\Phi_\mathrm{S},\widetilde{\Phi}_\mathrm{S}}\left(y, a, \phi_\mathrm{S}, \widetilde{\phi}_\mathrm{S}\right)}{p_{Y,\widetilde{\Phi}_\mathrm{S}}\left(y, \widetilde{\phi}_\mathrm{S}\right)}
$$
$$
= \frac{p_{Y|s,\widetilde{\phi}_\mathrm{S}}\left(y \mid a, \phi_\mathrm{S}, \widetilde{\phi}_\mathrm{S}\right) p_{A,\Phi_\mathrm{S},\widetilde{\Phi}_\mathrm{S}}\left(a, \phi_\mathrm{S}, \widetilde{\phi}_\mathrm{S}\right)}{\int\int p_{Y|s,\widetilde{\phi}_\mathrm{S}}\left(y \mid a, \phi_\mathrm{S}, \widetilde{\phi}_\mathrm{S}\right) p_{A,\Phi_\mathrm{S},\widetilde{\Phi}_\mathrm{S}}\left(a, \phi_\mathrm{S}, \widetilde{\phi}_\mathrm{S}\right) \mathrm{d}a\mathrm{d}\phi_\mathrm{S}}
$$
$$(4)$$

meaning that we now need models for $p_{Y|s,\widetilde{\phi}_\mathrm{S}}$ and $p_{A,\Phi_\mathrm{S},\widetilde{\Phi}_\mathrm{S}}$. To find a model for $p_{Y|s,\widetilde{\phi}_\mathrm{S}}$ we assume that if the clean speech realization $s = a\mathrm{e}^{\mathrm{j}\phi_\mathrm{S}}$ is known, the estimated phase $\widetilde{\phi}_\mathrm{S}$ gives no further information on Y, i.e.

$$
p_{Y|s,\widetilde{\phi}_\mathrm{S}} = p_{Y|s}. \quad (5)
$$

For complex Gaussian distributed noise spectral coefficients, the probability density function (PDF) $p_{Y|s}$ is also Gaussian and with (5) we have

$$
p_{Y|s,\widetilde{\phi}_\mathrm{S}}(y \mid a, \phi_\mathrm{S}) = \frac{1}{\pi\sigma_\mathrm{N}^2}\exp\left(-\frac{|y - a\mathrm{e}^{\mathrm{j}\phi_\mathrm{S}}|^2}{\sigma_\mathrm{N}^2}\right). \quad (6)
$$

To solve (4), we still need a model for the joint PDF of the clean speech amplitude, phase, and the phase estimate $p_{A,\Phi_\mathrm{S},\widetilde{\Phi}_\mathrm{S}}$. As empirically shown in [15], amplitudes and phases can be assumed to be mutually independent. Then, the joint PDF $p_{A,\Phi_\mathrm{S},\widetilde{\Phi}_\mathrm{S}}$ can be rewritten as

$$
p_{A,\Phi_\mathrm{S},\widetilde{\Phi}_\mathrm{S}}\left(a, \phi_\mathrm{S}, \widetilde{\phi}_\mathrm{S}\right) = p_A(a) \, p_{\Phi_\mathrm{S},\widetilde{\Phi}_\mathrm{S}}\left(\phi_\mathrm{S}, \widetilde{\phi}_\mathrm{S}\right)
$$
$$
= p_A(a) \, p_{\Phi_\mathrm{S}|\widetilde{\phi}_\mathrm{S}}\left(\phi_\mathrm{S} \mid \widetilde{\phi}_\mathrm{S}\right) p_{\widetilde{\Phi}_\mathrm{S}}\left(\widetilde{\phi}_\mathrm{S}\right). \quad (7)
$$

Using (5) and (7) in (4), the posterior results in

$$
p_{A,\Phi_\mathrm{S}|y,\widetilde{\phi}_\mathrm{S}}\left(a, \phi_\mathrm{S} \mid y, \widetilde{\phi}_\mathrm{S}\right)
$$
$$
= \frac{p_{Y|s}(y \mid a, \phi_\mathrm{S}) \, p_A(a) \, p_{\Phi_\mathrm{S}|\widetilde{\phi}_\mathrm{S}}\left(\phi_\mathrm{S} \mid \widetilde{\phi}_\mathrm{S}\right)}{\int\int p_{Y|s}(y \mid a, \phi_\mathrm{S}) \, p_A(a) \, p_{\Phi_\mathrm{S}|\widetilde{\phi}_\mathrm{S}}\left(\phi_\mathrm{S} \mid \widetilde{\phi}_\mathrm{S}\right) \mathrm{d}a\mathrm{d}\phi_\mathrm{S}} \quad (8)
$$

where we can cancel $p_{\widetilde{\Phi}_\mathrm{S}}$, as it is not part of the integral. To model the posterior we still need models for the speech amplitude prior p_A and for the uncertainty of the prior phase estimate, i.e. the conditional PDF $p_{\Phi_\mathrm{S}|\widetilde{\phi}_\mathrm{S}}$.

As in [14] we propose to model the speech spectral amplitudes using the χ-distribution

$$
p_A(a) = \frac{2}{\Gamma(\mu)}\left(\frac{\mu}{\sigma_\mathrm{s}^2}\right)^\mu a^{2\mu-1}\exp\left(-\frac{\mu}{\sigma_\mathrm{s}^2}a^2\right), \quad (9)
$$

with the Gamma function $\Gamma(\cdot)$ [16, Eq. (8.31)], and shape parameter μ. The parameter μ allows to model different shapes of the PDF of the speech amplitudes. Assuming speech coefficients S are complex Gaussian distributed, the speech amplitudes can be modeled by a χ-distribution with shape parameter $\mu = 1$. However, in the speech enhancement context, speech has been shown to follow a heavy-tailed (also known as super-Gaussian) distribution [17], which can be modeled by setting $\mu < 1$ [18, 14].

The novelty of this paper comes with employing an explicit model for the error between the prior phase estimate $\widetilde{\phi}_\mathrm{S}$ and the true phase ϕ_S in (8). For this we propose to employ the *von Mises* distribution with concentration parameter κ

$$
p_{\Phi_\mathrm{S}|\widetilde{\phi}_\mathrm{S}}\left(\phi_\mathrm{S} \mid \widetilde{\phi}_\mathrm{S}\right) = \exp\left(\kappa\cos(\phi_\mathrm{S} - \widetilde{\phi}_\mathrm{S})\right)/2\pi I_0(\kappa). \quad (10)
$$

The circular variance of the *von Mises* distribution is $\mathrm{var}(\Phi_\mathrm{S} \mid a, y) = 1 - I_1(\kappa)/I_0(\kappa)$ [19], from which it follows that the variance decreases for an increasing κ. Here, large values for κ mean that we are very certain about the prior phase estimate $\widetilde{\phi}_\mathrm{S}$, while low values for κ reflect a large degree of uncertainty.

With (6), (9), and (10), we have all models at hand to determine the posterior (8). This posterior enables us to formulate the MMSE estimator of the compressed speech coefficients with uncertain prior knowledge of the clean speech phase (3). Similar to [6], with [16, Eq. (3.462.1)], the integral over the amplitude can be solved and we get the CUP:

$$
\widehat{S}^{(\beta)} = \mathrm{E}\left(A^\beta \mathrm{e}^{\mathrm{j}\Phi_\mathrm{S}} \mid y, \widetilde{\phi}_\mathrm{S}\right) =
$$
$$
\left(\sqrt{\frac{1}{2}\frac{\xi}{\mu+\xi}\sigma_\mathrm{N}^2}\right)^\beta \frac{\Gamma(2\mu+\beta)}{\Gamma(2\mu)} \times
$$
$$
\frac{\int_0^{2\pi} \mathrm{e}^{\mathrm{j}\phi_\mathrm{S}}\exp(\nu^2/4)\,\mathrm{D}_{-(2\mu+\beta)}(\nu)\,p_{\Phi_\mathrm{S}|\widetilde{\phi}_\mathrm{S}}\mathrm{d}\phi_\mathrm{S}}{\int_0^{2\pi}\exp(\nu^2/4)\,\mathrm{D}_{-(2\mu)}(\nu)\,p_{\Phi_\mathrm{S}|\widetilde{\phi}_\mathrm{S}}\mathrm{d}\phi_\mathrm{S}}
$$
$$(11)$$

where ν contains the phase difference between the phases of the clean speech and noisy speech signals $\phi_Y - \phi_S$ and is defined as,

$$\nu = -\frac{r}{\sigma_N} \sqrt{2\frac{\xi}{\mu + \xi}} \cos(\phi_S - \phi_Y), \qquad (12)$$

and $D_.(\nu)$ is the parabolic cylinder function [16, Eq. (9.24)], $\xi = \sigma_S^2/\sigma_N^2 = E(|S|^2)/E(|N|^2)$ is the *a priori* SNR. The speech estimate is then obtained as

$$\widehat{S} = |\widehat{S}^{(\beta)}|^{1/\beta} \frac{\widehat{S}^{(\beta)}}{|\widehat{S}^{(\beta)}|} = \widehat{A} e^{j\widehat{\Phi}_S}. \qquad (13)$$

3.1. Implementation of the proposed CUP estimator

Solving the integral over the speech spectral phase ϕ_S in (11) is quite difficult, as it involves the integration over the parabolic cylinder function $D_.(\nu)$. However, as the phase has a limited span between $0 \leq \phi_S < 2\pi$, the intergral in (11) can be solved numerically with high precision. Furthermore, in practice, speech enhancement gain functions that involve computationally complex special functions are often precomputed and tabulated. Thus, we propose to solve the integral in (11) numerically and store the result in a table. For a given shape parameter μ and compression parameter β, this table has four dimensions, the *a priori* SNR ξ, the *a posteriori* SNR r^2/σ_N^2, the concentration parameter κ, and the phase difference $\phi_S - \phi_Y$. During runtime, the computational complexity is thus very low and given by a table look-up.

3.2. Interpretation of the proposed CUP estimator

If we assume that the prior phase estimate $\widetilde{\phi}_S$ obtained using [4] perfectly reflects the true phase, i.e. $\phi_S = \widetilde{\phi}_S$, then the phase is known and deterministic. This can be modeled by setting $\kappa \to \infty$. Then the PDF $p_{\Phi_S|\widetilde{\phi}_S}$ is given by a delta at $\widetilde{\phi}_S$. As a consequence, for $\kappa \to \infty$ the estimate of the amplitude obtained by (11) resembles the amplitude estimator in [6] while the estimate of the phase equals the prior phase information, i.e. $\widehat{\phi}_S = \widetilde{\phi}_S$. Thus, the estimator in [6] is a special case of the proposed estimator (11). The other extreme-case is obtained by setting $\kappa = 0$. Then, the PDF $p_{\Phi_S|\widetilde{\phi}_S}$ is uniform and for $\mu = \beta = 1$ the behavior of the proposed estimator approaches the linear behavior of a Wiener filter. The resulting phase estimate for $\kappa = 0$ resembles the noisy phase ϕ_Y. However, for any $0 < \kappa < \infty$, the proposed estimator will be a compromise between the two cases $\kappa = 0$ and $\kappa \to \infty$, both in terms of the amplitude attenuation as well as in terms of the phase estimate. This behavior is illustrated in Figure 1 for $\widetilde{\phi}_S - \phi_Y = \pi/4$ and $\widetilde{\phi}_S - \phi_Y = \pi/2$, where we set $\mu = \beta = 1$ and $\xi = 0.2$. It can be seen that for large a posteriori SNRs r^2/σ_N^2, the estimated phase will be close to the observed phase $\widehat{\phi}_S \to \phi_Y$, while for low a posteriori SNRs the estimated phase is closer to the model phase $\widehat{\phi}_S \to \widetilde{\phi}_S$.
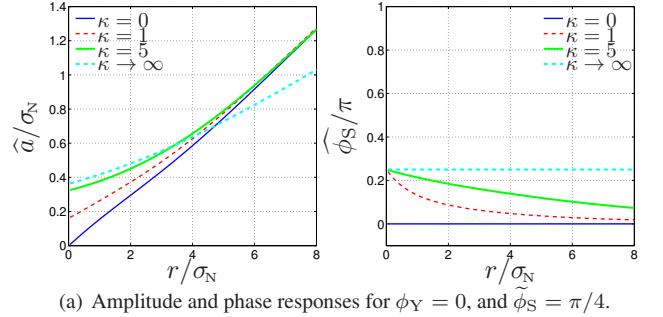
(a) Amplitude and phase responses for $\phi_Y = 0$, and $\widetilde{\phi}_S = \pi/4$.
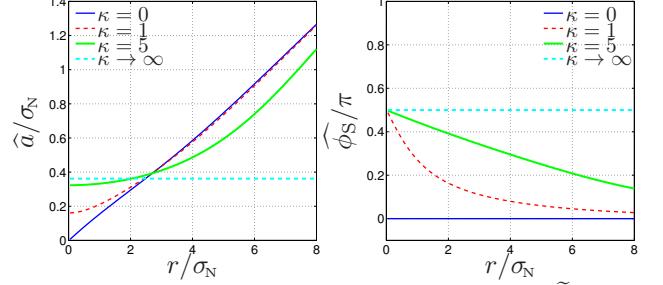
(b) Amplitude and phase responses for $\xi = 0.2$, $\phi_Y = 0$, and $\widetilde{\phi}_S = \pi/2$

Fig. 1. Amplitude and phase responses for $\mu = \beta = 1$ and $\xi = 0.2$ for different values of κ in (10). For $\kappa = 0$ the amplitude estimate approaches the behavior of a Wiener filter (left) and the phase estimate results in $\widehat{\phi}_S = \phi_Y$ (right). For $\kappa \to \infty$ the amplitude estimate approaches the result in [6] (left) and the phase estimate results in $\widehat{\phi}_S = \widetilde{\phi}_S$ (right).

4. EXPERIMENTAL RESULTS

In this section, we apply the proposed estimator in a speech enhancement task. For the estimation of the parameters, we employed the phase reconstruction along frequency proposed in [4] to obtain an estimate of the prior phase estimate $\widetilde{\phi}_S$ in (10). This algorithm requires an estimate of the speech fundamental frequency which is obtained using PEFAC [20] which also yields the probability of a signal segment being voiced, which we denote by $P_{H_V}(\ell)$. The noise power spectral density σ_N^2 is estimated using the a posteriori speech presence probability with fixed priors [21], while the *a priori* SNR is estimated using the decision-directed approach [22] with smoothing constant $\alpha_{DD} = 0.96$. All applied gain functions are limited to be larger than -15 dB.

The parameter κ can now be adjusted to reflect the certainty of the prior phase estimate $\widetilde{\phi}_S$. As in unvoiced speech the sinusoidal model based approach [4] does not yield reasonable phase estimates, we control the value of κ by the probability that a signal frame contains voiced speech $P_{H_V}(\ell)$. Furthermore, in [4] the phase estimates at all frequencies are based on multiples of the estimate of the fundamental frequency. This model is well fulfilled at lower frequencies and less well fulfilled at the higher frequencies. In the proposed algorithm this can be considered by using

lower values for κ at high frequencies. With these considerations, we set κ to

$$\kappa(k, \ell) = \begin{cases} 4P_{H_V}(\ell) & , kf_s/N < 4000 \, \text{Hz} \\ 2P_{H_V}(\ell) & , kf_s/N \geq 4000 \, \text{Hz} \end{cases}, \quad (14)$$

where $N = 512$ is the length of the discrete Fourier transform and $f_s = 16 \, \text{kHz}$ is the sampling rate. The values of 4 and 2 are chosen based on informal listening.

The proposed CUP estimator is compared to three estimators, denoted as "phase-sensitive [23]", "phase-sensitive [23] $\times \exp(j\widetilde{\phi}_S)$" and "phase insensitive [14]". In all four algorithms we set $\mu = \beta = 0.5$ to incorporate the super-Gaussian character of speech and the perceptually motivated compression of spectral amplitudes. The "phase insensitive [14]" estimator does not modify the phase of the noisy speech in the STFT domain, but only modifies the clean speech amplitudes without taking the phase into account. Also the "phase-sensitive [23]" approach does not modify the noisy phase. However, in contrast to [14], in this approach the prior phase estimate $\widetilde{\phi}_S$ is employed for an improved estimation of the speech amplitudes using the phase sensitive amplitude estimators proposed in [6, 23]. In the approach "phase-sensitive [23] $\times \exp(j\widetilde{\phi}_S)$" we use the same phase sensitive amplitude estimator [6, 23] but replace the phase of the noisy signal by the prior phase estimate $\widetilde{\phi}_S$. Note, that this corresponds to setting $\kappa \to \infty$ in the proposed CUP estimator, i.e. treating the prior phase estimate as deterministic.

The performance of the four algorithms are compared using the (P)erceptual (E)valuation of (S)peech (Q)uality measure, as provided by (L)oizou [24] (PESQL). The PESQL score is computed over the entire speech signals, i.e. without separating voiced and unvoiced speech sounds. The results are given in Figure 2. It can be seen that using the "phase-sensitive [23]" approach outperforms the "phase insensitive [14]" estimator, especially in babble noise. Using the prior phase information $\widetilde{\phi}_S$ to replace the noisy phase, in low signal to noise ratios (SNRs) the "phase-sensitive [23] $\times \exp(j\widetilde{\phi}_S)$" approach can improve the performance even more. However, in high SNRs, errors in the phase estimate decrease the performance. In contrast, the proposed CUP estimator, which treats the phase estimate $\widetilde{\phi}_S$ as *uncertain* prior knowledge, results in a consistently larger PESQL-score for the wide range of considered SNRs. Informal listening reveals that replacing the noisy phase by the prior phase estimate $\widetilde{\phi}_S$ in "phase-sensitive [23] $\times \exp(j\widetilde{\phi}_S)$" may result in unnatural artifacts and unnatural sounding speech (note that these artifacts do not occur when the phase estimate is only employed for an improved amplitude estimation as in "phase-sensitive [23]"). Furthermore, informal listening confirmed that these undesired artifacts are reduced by taking the uncertainty of the prior phase estimate into account by using the proposed CUP estimator.
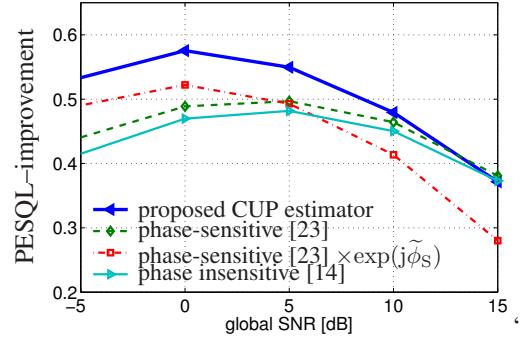

(a) Nonstationary factory noise
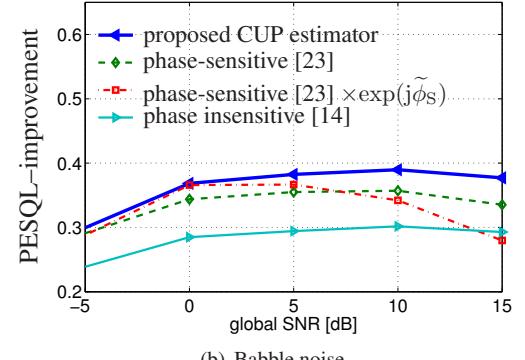
(b) Babble noise

Fig. 2. Instrumental evaluation of the perceptual speech quality for different input SNRs and noise types averaged over 200 sentences (100 spoken by female speakers and 100 spoken by male speakers) from the TIMIT database. We show the improvement over the unprocessed noisy speech.

5. CONCLUSIONS

While in most STFT-based single channel speech enhancement algorithms the phase of the noisy signal is not explicitly modified, more recently attention is drawn towards the importance of phase in speech enhancement [2]. While phase estimates can be robustly employed for improving spectral amplitude estimation [6, 23], replacing the phase of the noisy signal by a phase estimate may result in undesired artifacts in the enhanced speech [3, 11, 4].

In this paper we proposed to employ an estimated clean speech phase, obtained e.g. from the sinusoidal model based phase reconstruction algorithm [4], as uncertain prior knowledge when finding a joint MMSE-estimate of the clean speech amplitude and phase. The resulting phase estimate is a compromise between the phase of the noisy signal and the prior phase estimate. The estimated amplitude is a compromise between the outputs of the phase sensitive amplitude estimator [6] and the output when assuming a uniformly distributed phase. We showed that the proposed approach increases the speech quality further and informal listening confirmed that artifacts are reduced.

6. REFERENCES

[1] R. C. Hendriks, T. Gerkmann, and J. Jensen, *DFT-Domain Based Single-Microphone Noise Reduction for Speech Enhancement: A Survey of the State-of-the-art.* Colorado, USA: Morgan & Claypool, Feb. 2013.

[2] 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.

[3] N. Sturmel and L. Daudet, "Signal reconstruction from STFT magnitude: a state of the art," in *International Conference on Digital Audio Effects (DAFx)*, Paris, France, Sep. 2011, pp. 375–386.

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

[5] P. Mowlaee, R. Saeidi, and R. Martin, "Phase estimation for signal reconstruction in single-channel speech separation," in *ISCA Interspeech*, Portland, OR, USA, Sep. 2012.

[6] T. Gerkmann and M. Krawczyk, "MMSE-optimal spectral amplitude estimation given the STFT-phase," *IEEE Signal Process. Lett.*, vol. 20, no. 2, pp. 129–132, Feb. 2013.

[7] J. Le Roux and E. Vincent, "Consistent Wiener filtering for audio source separation," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 217–220, Mar. 2013.

[8] A. Sugiyama and R. Miyahara, "Phase randomization – A new paradigm for single-channel signal enhancement," in *IEEE ICASSP*, Vancouver, Canada, May 2013, pp. 7487–7491.

[9] P. Mowlaee and R. Saeidi, "Iterative closed-loop phase-aware single-channel speech enhancement," *IEEE Signal Process. Lett.*, vol. 20, no. 12, pp. 1235–1239, Dec. 2013.

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

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

[12] Y. Ephraim and D. Malah, "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.

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

[14] C. Breithaupt, M. Krawczyk, and R. Martin, "Parameterized MMSE spectral magnitude estimation for the enhancement of noisy speech," in *IEEE ICASSP*, Las Vegas, NV, USA, Apr. 2008, pp. 4037–4040.

[15] 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 Trans. Audio, Speech, Language Process.*, vol. 15, no. 6, pp. 1741–1752, Aug. 2007.

[16] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals Series and Products*, 7th ed. San Diego, CA, USA: Academic Press, Feb. 2007.

[17] R. Martin, "Speech enhancement based on minimum mean-square error estimation and supergaussian priors," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 5, pp. 845–856, Sep. 2005.

[18] I. Andrianakis and P. R. White, "MMSE speech spectral amplitude estimators with Chi and Gamma speech priors," in *IEEE ICASSP*, Toulouse, France, May 2006, pp. 1068–1071.

[19] M. Evans, N. Hastings, and B. Peacock, "von Mises distribution," in *Statistical Distributions*, 4th ed. New York: John Wiley & Sons, 2000, ch. 45, pp. 191–192.

[20] S. Gonzalez and M. Brookes, "PEFAC – a pitch estimation algorithm robust to high levels of noise," *IEEE Trans. Audio, Speech, Language Process.*, vol. 22, no. 2, pp. 518–530, Feb. 2014.

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

[22] 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.

[23] M. Krawczyk, R. Rehr, and T. Gerkmann, "Phase-sensitive real-time capable speech enhancement under voiced-unvoiced uncertainty," in *EUSIPCO*, Marrakech, Morocco, Sep. 2013.

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