Exploiting Temporal Correlation of Speech and Noise Magnitudes Using a Modified Kalman Filter for Speech Enhancement

Thomas Esch, Peter Vary

Institute of Communication Systems and Data Processing (ind), RWTH Aachen University, 52056 Aachen, Germany
E-Mail: {esch,vary}@ind.rwth-aachen.de
Web: www.ind.rwth-aachen.de

Abstract

A new speech enhancement algorithm using a modified Kalman filter in the frequency domain is proposed. The new approach consists of two steps. In the first step, the temporal trajectories of the speech and noise magnitudes are modeled by low order autoregressive (AR) processes, i.e., the current coefficients are propagated in time based on information taken from previous, enhanced coefficients, followed by a subsequent phase estimation. In the second step, the first estimation is updated. Therefore, two statistical estimators are utilized. The performance of the proposed method is shown to be considerably better than purely statistical estimators.

1 Introduction

Speech quality and intelligibility may significantly deteriorate in the presence of background noise, e.g., engine noise or street noise. The problem of enhancing speech that is degraded by additive noise has been widely studied in the past and is still an active field of research. Speech enhancement has many applications in voice communications, speech recognition and hearing aids.

The design of many speech enhancement systems is based on modeling the noisy input coefficients in the short-time Fourier transform (STFT) domain by applying individual adaptive gains to each frequency bin. Most of the rules proposed in literature have been derived under certain assumptions about the statistics of the speech and noise signal. Considering a Gaussian speech and noise model, this enables to deduce useful *minimum mean-squared error* (MMSE) estimators, such as the well-known Wiener filter [1] or the *short-term spectral amplitude* (STSA) estimator [2]. Martin [3] proposed the use of a Gamma speech model and derived an MMSE estimator for the complex speech coefficients under the assumption of Gaussian and Laplacian noise models. Lotter [4] derived a *maximum a posteriori* (MAP) estimator using a super-Gaussian speech and Gaussian noise model. All of these estimators only utilize statistical characteristics of speech and noise, correlation in time is explicitly not taken into account.

Paliwal and Basu [5] were the first to propose the use of a Kalman filter for the purpose of speech enhancement. In order to reduce complexity, the authors in [6] derived a Kalman filtering system in the sub-band domain. Puder [7] further investigated the application of a Kalman filter in sub-bands and increased the performance compared to the full-band time domain approach. In addition to the exploitation of intra-frame correlation, model-based approaches that consider the correlation of successive speech frames can be found, e.g, in [8] and [9].

In this paper, the Kalman filter approach of [9] is modified and extended. Instead of using a complex predictor to exploit the temporal correlation of successive spectral coefficients, only the real-valued magnitudes are propagated in time, followed by an additional phase estimation term. Furthermore, the propagation step is not only applied to the speech signal, but also extended to the noise signal. The resulting prediction errors are estimated in a second step by utilizing different statistical estimators. The remainder of this paper is organized as follows: In Sec. 2, a brief overview about the proposed system is given. Secs. 3 and 4 comprise the procedure of propagation and update step in detail.

This work was supported by Nokia, Tampere, Finland.

Experimental results are shown in Sec. 5 and conclusions are drawn in Sec. 6.

2 System Overview

A simplified block diagram of the proposed system is depicted in Fig. 1. It is assumed that the noisy input signal $y(k)$ consists of the clean speech signal $s(k)$ which is degraded by an additive noise signal $n(k)$, i.e.,

$$y(k) = s(k)+n(k), \tag{1}$$

where k is the discrete time index. To decompose speech and noise signal, the noisy signal is transformed into the frequency domain. Therefore, $y(k)$ is segmented into overlapping frames of length L_F. After windowing, the fast Fourier transform (FFT) is applied to these frames. Hence, the spectral coefficient of the noisy input signal at frequency bin μ and frame λ is given by:

$$\begin{aligned} Y(\lambda,\mu) &= S(\lambda,\mu)+N(\lambda,\mu) && (2)\\ &= R(\lambda,\mu)\mathrm{e}^{j\vartheta(\lambda,\mu)} && (3)\\ &= A(\lambda,\mu)\mathrm{e}^{j\alpha(\lambda,\mu)}+B(\lambda,\mu)\mathrm{e}^{j\beta(\lambda,\mu)}, && (4)\end{aligned}$$

where $S(\lambda,\mu)$ and $N(\lambda,\mu)$ represent the spectral coefficients of speech and noise. Moreover, $R(\lambda,\mu)$, $A(\lambda,\mu)$ and $B(\lambda,\mu)$ denote the magnitudes of the noisy, speech, and noise signal and $\vartheta(\lambda,\mu)$, $\alpha(\lambda,\mu)$, $\beta(\lambda,\mu)$ are the corresponding phases respectively.

The investigated system is based on a *Kalman filter* structure that consists of two steps, namely *propagation* and *update step*. Both are briefly explained in the following. In the propagation step, temporal correlation (a priori information of higher order) of successive frames is exploited. The current speech and noise magnitudes are predicted based on information taken from previous, enhanced coefficients. In contrast to [9], possible correlation of the noise magnitudes is also taken into account. Additionally, magnitude and phase estimations are performed consecutively. The resulting estimates

$$\begin{aligned} \hat{S}_\mathrm{prop}(\lambda,\mu) &= \hat{A}_\mathrm{prop}(\lambda,\mu)\mathrm{e}^{j\hat{\alpha}(\lambda,\mu)} \quad \text{and} && (5)\\ \hat{N}_\mathrm{prop}(\lambda,\mu) &= \hat{B}_\mathrm{prop}(\lambda,\mu)\mathrm{e}^{j\hat{\beta}(\lambda,\mu)} && (6)\end{aligned}$$

are combined to get an estimation of the current noisy coefficient

$$\hat{Y}_\mathrm{prop}(\lambda,\mu) = \hat{S}_\mathrm{prop}(\lambda,\mu)+\hat{N}_\mathrm{prop}(\lambda,\mu). \tag{7}$$

In general, the prediction in the propagation step is erroneous and the prediction errors

$$\begin{aligned} \hat{E}_\mathrm{S}(\lambda,\mu) &= S(\lambda,\mu)-\hat{S}_\mathrm{prop}(\lambda,\mu) \quad \text{and} && (8)\\ \hat{E}_\mathrm{N}(\lambda,\mu) &= N(\lambda,\mu)-\hat{N}_\mathrm{prop}(\lambda,\mu) && (9)\end{aligned}$$

occur for the speech and noise signal. Considering the differential signal

$$D(\lambda,\mu) = Y(\lambda,\mu)-\hat{Y}_\mathrm{prop}(\lambda,\mu), \tag{10}$$

the update step estimates these prediction errors based on a conventional statistical estimator, utilizing a priori information of zeroth order. This estimator is adapted to the statistics of speech and

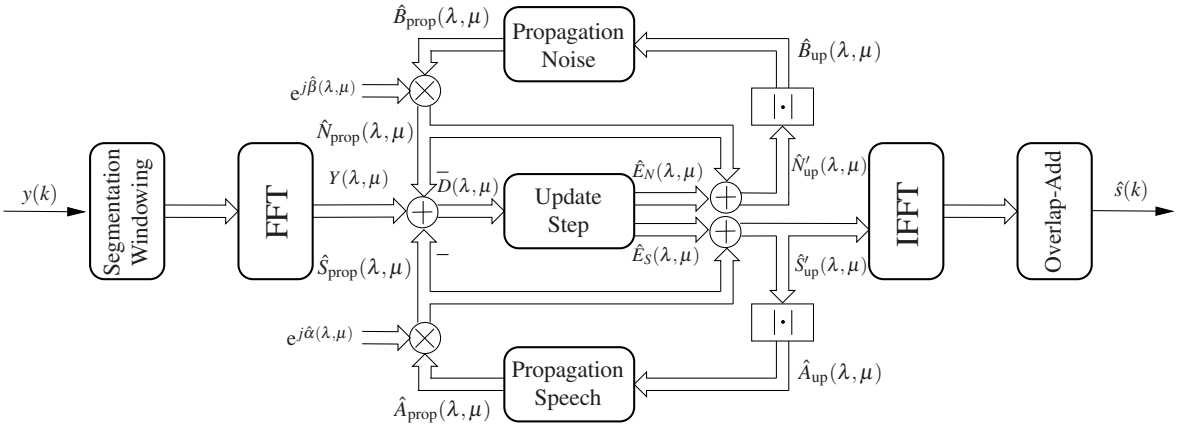

Figure 1: System block diagram

noise and performs a spectral weighting of the differential signal by multiplying the Kalman gain $K(\lambda,\mu)$:

$$\hat{E}_{\text{S}}(\lambda,\mu) = K(\lambda,\mu)D(\lambda,\mu) \quad (11)$$

$$\hat{E}_{\text{N}}(\lambda,\mu) = (1-K(\lambda,\mu))\,D(\lambda,\mu). \quad (12)$$

To obtain the enhanced speech and noise coefficients $\hat{S}'_{\text{up}}(\lambda,\mu)$ and $\hat{N}'_{\text{up}}(\lambda,\mu)$, the initial predictions of the propagation step are updated:

$$\hat{S}'_{\text{up}}(\lambda,\mu) = \hat{S}_{\text{prop}}(\lambda,\mu)+\hat{E}_{\text{S}}(\lambda,\mu) \quad (13)$$

$$\hat{N}'_{\text{up}}(\lambda,\mu) = \hat{N}_{\text{prop}}(\lambda,\mu)+\hat{E}_{\text{N}}(\lambda,\mu). \quad (14)$$

It turned out that speech distortions can be reduced by omitting the phase of $\hat{S}'_{\text{up}}(\lambda,\mu)$ and using the short-time phase of the noisy input coefficient instead for reconstruction. The enhanced speech coefficient is therefore given by:

$$\hat{S}_{\text{up}}(\lambda,\mu) = \left|\hat{S}'_{\text{up}}(\lambda,\mu)\right| e^{j\vartheta(\lambda,\mu)}. \quad (15)$$

In order to obtain the enhanced signal $\hat{s}(k)$ in the time domain, an inverse fast Fourier transform (IFFT) and the overlap-add method are applied.

3 Propagation Step

In this section, further details about the propagation step are given. The magnitudes $A(\lambda,\mu)$ and $B(\lambda,\mu)$ of the speech and the noise signal are both modeled as two independent autoregressive (AR) processes. Based on these predictions, the phases $\alpha(\lambda,\mu)$ and $\beta(\lambda,\mu)$ are estimated.

3.1 Magnitude Estimation

Within the modified Kalman filter, the AR model is used to exploit temporal correlation of the speech and noise magnitudes. In [9], a complex AR model was used to directly predict the spectral coefficient $\hat{S}_{\text{prop}}(\lambda,\mu)$. It was shown that this kind of complex estimator achieves the highest prediction gain

$$G_{\text{P}} = \frac{\mathscr{E}\left\{|S(\lambda,\mu)|^2\right\}}{\mathscr{E}\left\{\left|S(\lambda,\mu)-\hat{S}_{\text{prop}}(\lambda,\mu)\right|^2\right\}}, \quad (16)$$

compared to estimators that predict either magnitude and phase or real and imaginary part separately. The aim in this contribution is not to utilize the AR model for a complex but for a real-valued magnitude prediction. This is motivated by the fact that most part of the temporal correlation of the spectral coefficients can be found in successive magnitudes and only marginally in the phase samples. In addition, the magnitude predictions $\hat{A}_{\text{prop}}(\lambda,\mu)$ and $\hat{B}_{\text{prop}}(\lambda,\mu)$ are used here in a second step to estimate the phases α and β, as described in Sec. 3.2.

The magnitudes $\hat{A}_{\text{prop}}(\lambda,\mu)$ and $\hat{B}_{\text{prop}}(\lambda,\mu)$ for speech and noise can be stated as:

$$\hat{A}_{\text{prop}}(\lambda,\mu) = \sum_{i=1}^{N_{\text{K}}} \hat{a}_i(\lambda,\mu)\hat{A}_{\text{up}}(\lambda-i,\mu) \quad \text{and} \quad (17)$$

$$\hat{B}_{\text{prop}}(\lambda,\mu) = \sum_{i=1}^{M_{\text{K}}} \hat{b}_i(\lambda,\mu)\hat{B}_{\text{up}}(\lambda-i,\mu), \quad (18)$$

where N_{K} and M_{K} represent the orders of the speech and the noise model respectively. The AR coefficients $\hat{a}_i(\lambda,\mu)$ and $\hat{b}_i(\lambda,\mu)$ are estimated in advance by minimizing the prediction error energies. This optimization criterion leads to the well-known Yule-Walker equations [10]. The required autocorrelation vector and matrix are calculated from the previous L_{AC} enhanced magnitudes of either speech or noise.

3.2 Phase Estimation

In contrast to [9], magnitude and phase are estimated consecutively in this contribution. As there is almost no correlation in successive phase samples, linear prediction is explicitly applied to the speech and noise magnitudes in order to exploit the maximum temporal correlation within adjacent magnitudes. If $\hat{A}_{\text{prop}}(\lambda,\mu)$ and $\hat{B}_{\text{prop}}(\lambda,\mu)$ are available for the current frame λ, the phases $\hat{\alpha}(\lambda,\mu)$ and $\hat{\beta}(\lambda,\mu)$ are estimated according to Fig. 2. The aim in this phase estimation process is to ensure that the resulting phase of $\hat{Y}_{\text{prop}}(\lambda,\mu)$ equals the noisy input phase $\vartheta(\lambda,\mu)$. Therefore, the following procedure is applied. Note that the frame index λ and the frequency index μ are omitted in the following for simplicity.

1. At first, a random phase $\hat{\alpha}$ out of the range

$$\vartheta - \arcsin\frac{\hat{B}_{\text{prop}}}{\hat{A}_{\text{prop}}} \leq \hat{\alpha} \leq \vartheta + \arcsin\frac{\hat{B}_{\text{prop}}}{\hat{A}_{\text{prop}}} \quad (19)$$

is selected and applied to $\hat{A}_{\text{prop}}$ in order to obtain $\hat{S}_{\text{prop}}$ (cf. Eq. 5). The limitation in Eq. 19 ensures that in the following step at least one phase $\hat{\beta}$ can be found which satisfies $\angle\{\hat{Y}_{\text{prop}}\} = \vartheta$, where $\angle\{\cdot\}$ represents the phase operator. Obviously, the range of the phase limitation is dependent on the estimated input signal-to-noise Ratio (SNR), where the worst case is given below 0 dB. Assuming that the predictions $\hat{A}_{\text{prop}}$ and $\hat{B}_{\text{prop}}$ are adequate, the phase estimation of α gets more precise with an increasing input SNR.

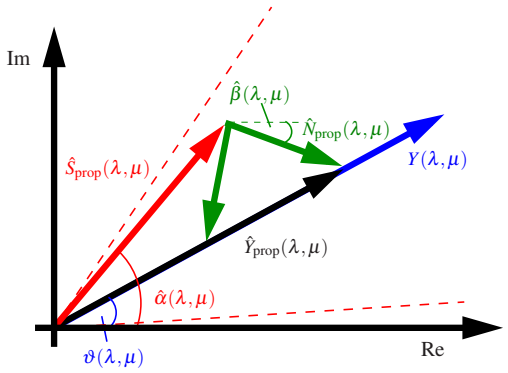

Figure 2: Phase estimation

2. The phase $\hat{\beta}$ that satisfies the following equation:

$$\angle\left\{\hat{A}_{\text{prop}}e^{j\hat{\alpha}}+\hat{B}_{\text{prop}}e^{j\hat{\beta}}\right\}=\vartheta \tag{20}$$

is selected as estimate for the noise phase. In general, there are two solutions of Eq. 20, as can be seen from Fig. 2. In this case, $\hat{\beta}$ is chosen such that the distance $Y(\lambda,\mu)-\hat{Y}_{\text{prop}}(\lambda,\mu)$ is minimized. Hence, $\hat{N}_{\text{prop}}$ and $\hat{Y}_{\text{prop}}$ are calculated according to Eqs. 6 and 7 respectively.

4 Update Step

While in the propagation step, the temporal correlation of successive speech and noise magnitudes is exploited, the update step makes use of the statistical characteristics of both signals. The objective in this step is to estimate the prediction errors $E_{\text{S}}(\lambda,\mu)$ and $E_{\text{N}}(\lambda,\mu)$, caused in the propagation step. By reorganizing Eq. 10, it can be shown that the differential signal $D(\lambda,\mu)$ consists of the speech prediction error $E_{\text{S}}(\lambda,\mu)$ that is degraded by the noise prediction error $E_{\text{N}}(\lambda,\mu)$:

$$\begin{aligned} D(\lambda,\mu) &= Y(\lambda,\mu)-\hat{Y}_{\text{prop}}(\lambda,\mu) \\ &= S(\lambda,\mu)-\hat{S}_{\text{prop}}(\lambda,\mu)+N(\lambda,\mu)-\hat{N}_{\text{prop}}(\lambda,\mu) \\ &= E_{\text{S}}(\lambda,\mu)+E_{\text{N}}(\lambda,\mu). \end{aligned} \tag{21}$$

The estimation problem in the update step reduces to a classical noise reduction problem: The target coefficient $E_{\text{S}}(\lambda,\mu)$ is assumed to be degraded by the additive 'noise' coefficient $E_{\text{N}}(\lambda,\mu)$ to produce the noisy coefficient $D(\lambda,\mu)$. Thus, a conventional statistical estimator can be applied which is adapted to the statistics of the prediction errors.

Based on the assumption that the coefficients $E_{\text{S}}(\lambda,\mu)$ and $E_{\text{N}}(\lambda,\mu)$ are statistically independent, two estimators are considered in the following for the update step, namely an MMSE estimator [1] and a recently published super-Gaussian joint MAP estimator [4]. Both estimators rely on a Gaussian model for the noise signal. Indeed, even if the initial speech signal $s(k)$ is degraded by a colored noise $n(k)$, the propagation step has the effect of a prewhitening filter as it reduces possible temporal correlation. In addition, both estimators require the *a posteriori* SNR $\gamma(\lambda,\mu)$ and the *a priori* SNR $\xi(\lambda,\mu)$, which are defined here as follows:

$$\gamma(\lambda,\mu)=\frac{|D(\lambda,\mu)|^2}{\mathcal{E}\{|E_{\text{N}}(\lambda,\mu)|^2\}} \quad \text{and} \quad \xi(\lambda,\mu)=\frac{\mathcal{E}\{|E_{\text{S}}(\lambda,\mu)|^2\}}{\mathcal{E}\{|E_{\text{N}}(\lambda,\mu)|^2\}}. \tag{22}$$

The two estimators are briefly described in the following:

1. *Gaussian MMSE Estimator/Wiener Filter*
This Gaussian MMSE estimator corresponds to the well-known Wiener filter solution and is derived from the optimal filter theory [1]. This linear estimator minimizes the mean square error between clean and enhanced coefficient. Applied to the update step, the enhanced coefficient $\hat{E}_{\text{S}}(\lambda,\mu)$ can be stated as:

$$\begin{aligned} \hat{E}_{\text{S}}(\lambda,\mu) &= \mathcal{E}\{E_{\text{S}}(\lambda,\mu)|D(\lambda,\mu)\} && (23)\\ &= \underbrace{\frac{\xi(\lambda,\mu)}{\xi(\lambda,\mu)+1}}_{K_{\text{G}}(\lambda,\mu)}D(\lambda,\mu), && (24) \end{aligned}$$

Note that this MMSE estimator in the update step equals the conventional Kalman filter gain as it arises from the same assumption that the prediction errors for speech and noise are Gaussian distributed [6].

2. *Super-Gaussian Joint MAP Estimator*
Applied to the update step, this generalized super-Gaussian estimator [4] uses the following parametric function to approximate the *probability density function* (PDF) of the spectral magnitude $|E_{\text{S}}|$:

$$p(|E_{\text{S}}|) = \frac{\delta^{\eta+1}}{\Gamma(\eta+1)}\frac{|E_{\text{S}}|^{\eta}}{\sigma_{E_{\text{S}}}^{\eta+1}}\exp\left\{-\delta\frac{|E_{\text{S}}|}{\sigma_{E_{\text{S}}}}\right\}, \tag{25}$$

where $\Gamma(\cdot)$ states the Gamma function and $\sigma_{E_{\text{S}}}$ the standard deviation of the speech prediction error. The parameters δ and η can be selected in order to obtain the optimal approximation. Therefore, the Kullback-Leibler distance between measured and modeled PDF is used [4]. In combination with a Gaussian noise model, this MAP estimator jointly maximizes the a posteriori PDF of amplitude *and* phase of the prediction error E_{S}, given the noisy sample D:

$$|\hat{E}_{\text{S}}| = \arg\max_{|E_{\text{S}}|} p\left(|E_{\text{S}}|,\angle\{E_{\text{S}}\}|D\right) \tag{26}$$

$$\angle\{\hat{E}_{\text{S}}\} = \arg\max_{\angle\{E_{\text{S}}\}} p\left(|E_{\text{S}}|,\angle\{E_{\text{S}}\}|D\right), \tag{27}$$

resulting in the following weighting rule:

$$\hat{E}_{\text{S}}(\lambda,\mu)=\underbrace{\left(u(\lambda,\mu)+\sqrt{u^2(\lambda,\mu)+\frac{\eta}{2\gamma(\lambda,\mu)}}\right)}_{K_{\text{S}}(\lambda,\mu)}D(\lambda,\mu), \tag{28}$$

where $u(\lambda,\mu)=\frac{1}{2}-\frac{\delta}{4\sqrt{\gamma(\lambda,\mu)\xi(\lambda,\mu)}}$.

Based on the calculation of either K_{G} or K_{S}, the noise prediction error can be estimated according to Eq. 12.

5 Results

For the evaluation of the proposed noise reduction scheme, five speech signals from the NTT speech database were each degraded by six different noise types (f16, babble, car, factory1, factory2, white), taken from the NOISEX-92 database. Among the five speech signals, there were three sqeuences from a male and two from a female speaker, each with a length of 8 seconds. The input SNR was varied between -10 dB and 35 dB (step size: 5 dB). For the analysis and synthesis structure, 75% overlapping Hann windows with a length of 20 ms and a 256-FFT (including zero-padding) were used. It turned out that good results were achieved by the following parameters applied to the modified Kalman filter: $L_{\text{AC}}=6$, $N_{\text{K}}=3$ and $M_{\text{K}}=2$ (sampling frequency f_{s}=8 kHz). While the power of the noise prediction error $\mathcal{E}\{|E_{\text{N}}(\lambda,\mu)|^2\}$ was estimated by using [11], the *decision-directed* approach [2] was utilized for the estimation of the a priori SNR.

A total of six different noise suppression techniques were investigated. Among them were the purely statistical weighting rules: Wiener filter [1] and super-Gaussian joint MAP (JMAP) estimator [4]. They were compared with the modified Kalman filter in [9] (Kalman filter *S*) and the new approach (Kalman filter

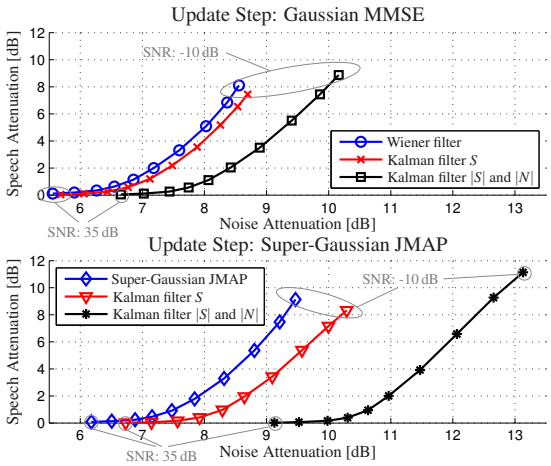

Figure 3: Speech attenuation vs. noise attenuation

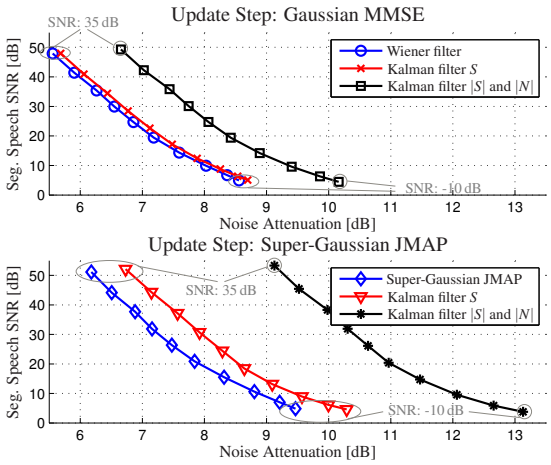

Figure 4: Segmental speech SNR vs. noise attenuation

$|S|$ and $|N|$) that is proposed in this paper. For each Kalman filter, the above mentioned weighting rules (cf. Sec. 4) were applied in the update step respectively. For the evaluation, three different kinds of instrumental measurements were used, namely the segmental noise attenuation (NA), the segmental speech attenuation (SA) and the segmental speech SNR (SegSNR) (e.g., [12]).

Figs. 3 and 4 illustrate the averaged results for SA and SegSNR, respectively, both plotted over NA with the input SNR as control variable. This procedure makes a fair comparison between noise attenuation and speech distortion possible. In Fig. 3, a low SA and a high NA is desirable, in Fig. 4 a high SegSNR and a high NA. In the upper plots of Figs. 3 and 4, the Gaussian MMSE estimator was used in the update step of the Kalman filters, in the lower plots the super-Gaussian JMAP estimator respectively. A lower a priori SNR threshold was applied to all estimators in a way that they yield nearly the same SA at 5 dB input SNR (cf. Fig. 3).

The results show that both types of Kalman filters achieve better results than the corresponding purely statistical estimator. In addition, the new Kalman filter based on consecutive magnitude and phase estimation in the propagation step outperforms the approach in [9]. The results show a considerable enhancement by the new estimator, e.g., if keeping the SA and SegSNR constant, the new approach increases the NA by a maximum of 2 dB in case the super-Gaussian JMAP estimator is applied in the update step. Furthermore, it can be seen that the utilization of the super-Gaussian JMAP estimator, i.e., the adaptation to the PDF of the prediction error signal, leads to better results than the application of the Gaussian MMSE estimator. The instrumental measurements were confirmed by informal listening tests.

6 Conclusions

A new method for single channel speech enhancement is presented in this paper which relies on a Kalman filter structure. In the first step, this model-based approach exploits the temporal correlation of successive speech and noise magnitudes. Based on these predictions, the phase samples are estimated subsequently. In the second step, the statistics of the differential signal are utilized to estimate the prediction errors by applying two different statistical estimators. Although the complexity is moderately increased by the proposed technique, the instrumental measurements in terms of segmental speech SNR, speech and noise attenuation clearly show the better performance compared to the Wiener filter, the super-Gaussian JMAP estimator and another recently published Kalman filter approach.

References

[1] J. S. Lim and A. V. Oppenheim, "Enhancement and Bandwidth Compression of Noisy Speech," *Proceedings of the IEEE*, vol. 67, no. 12, pp. 1586–1604, 1979.

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

[3] R. Martin, "Speech Enhancement Using MMSE Short Time Spectral Estimation with Gamma Distributed Speech Priors," in *Proc. of ICASSP*, Orlando, USA, 2002.

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

[5] K. Paliwal and A. Basu, "A Speech Enhancement Method Based on Kalman Filtering," in *Proc. of ICASSP*, Dallas, USA, 1987.

[6] W.-R. Wu and P.-C. Chen, "Subband Kalman Filtering for Speech Enhancement," *IEEE Transactions on Circuits and Systems II*, vol. 45, no. 8, pp. 1072–1083, Aug. 1998.

[7] H. Puder, "Kalman-Filters in Subbands for Noise Reduction with Enhanced Pitch-Adaptive Speech Model Estimation," *European Transactions on Telecommunications*, vol. 13, no. 2, pp. 139–148, 2002.

[8] E. Zavarehei and S. Vaseghi, "Speech Enhancement in Temporal DFT Trajectories Using Kalman Filters," in *Proc. of INTERSPEECH*, Lisbon, Portugal, 2005.

[9] T. Esch and P. Vary, "Speech Enhancement Using a Modified Kalman Filter Based on Complex Linear Prediction and Supergaussian Priors," in *Proc. of ICASSP*, Las Vegas, USA, 2008.

[10] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing: Principles, Algorithms, and Applications*, Prentice Hall, Upper Saddle River, New Jersey, 1996.

[11] R. Martin, "Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 5, pp. 501–512, 2001.

[12] S. Gustafsson, R. Martin, P. Jax, and P. Vary, "A Psychoacoustic Approach to Combined Acoustic Echo Cancellation and Noise Reduction," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 5, pp. 245–256, July 2002.