A KALMAN FILTER BASED NOISE SUPPRESSION ALGORITHM USING SPEECH AND NOISE MODELS DERIVED FROM SPATIAL INFORMATION

Ramun Schmid, Guido M. Schuster

University of Applied Sciences of Eastern Switzerland in Rapperswil, Switzerland

ABSTRACT

In this paper, a novel Kalman filter based noise suppression algorithm for hearing aids, using spatial information for estimating the required noise and speech models, is proposed. The main assumption of the scheme is that the target (usually the speech signal) is directly in front of the hearing aid user while the interference (usually the noise signal) comes from the back hemisphere. While in an earlier paper [1], a related approach based on instantaneous Wiener filters using a Weighted Overlap Add (WOLA) decomposition has been presented, this paper focuses on a time domain approach employing a time varying Kalman filter. Clearly, with the proper noise and speech models, one would expect a better performance of a time varying Kalman filter than of a WOLA Wiener filter. Hearing tests as well as objective performance measures show the excellent performance of the Kalman filter based noise suppression algorithm.

1. INTRODUCTION

The proposed algorithm is based on the LOCO (LOw COmplexity) idea, which was originally published in [1]. Based on an adaptive Elko-beamformer, LOCO describes a new way to estimate the statistical properties of the signal and noise in the beamformed signal. While in traditional approaches, the single beamformed signal is used to drive the statistical estimators which attempt to estimate the power spectral density (PSD) of the noise and the PSD of the speech, LOCO makes use of the spatial information. Based on our main assumption that the target is directly in front of the hearing aid user while the noise comes from the back hemisphere, LOCO uses the front- and back-cardioids of the Elko-beamformer for the estimation of the speech and the noise properties respectively (Fig. 3).

This idea can be implemented in different ways. In [1], it was used to estimate the PSDs as the squared magnitude of a frame based Fast Fourier Transform (FFT). Based on these PSDs, a corresponding instantaneous Wiener filter was calculated and applied to the beamformed signal, which resulted in excellent acoustic properties. Implemented in a WOLA framework (WOLA-LOCO), this scheme results in a data expansion. To avoid this, alternative schemes based on wavelets (Wavelet LOCO) have been proposed [1]. These Wavelet LOCO algorithms showed similar acoustic properties to the WOLA-LOCO, but as they don't result in a data expansion, they are computationally more efficient.

While applying a Wiener filter, the WOLA-LOCO as well as the Wavelet LOCO algorithm treat the beamformed signal as if it were stationary. Since this assumption is incorrect for a natural speech signal, the applied Wiener filter has to be changed from frame to frame. Since a Kalman filter is the nonstationary equivalent of a causal Wiener filter, the step away from the instantaneous Wiener filter towards a time varying Kalman filter should result in a smaller mean squared error and hence in improved performance. Clearly, while for the Wiener filter the estimation of the speech and noise PSDs is the critical part, for a Kalman filter, the estimation of the speech and noise model parameters is of fundamental importance.

Note that this paper presents the most important results of a larger thesis (in German), which can be found in [7]. Furthermore, the source code required to generate the results presented in this paper can also be found in [7].

The paper is organized as follows. In section 2, the notation and the performance measures used throughout this paper are introduced. In section 3, the main ideas behind the proposed scheme and the underlying speech/noise models are discussed and one selected implementation based on LPC analysis is presented in detail. Finally, the experimental results are shown in section 5, where the results of the proposed scheme are compared to the well-known Elko-beamformer [8] and the WOLA-LOCO [1] scheme.

2. NOTATION AND PERFORMANCE MEASURES

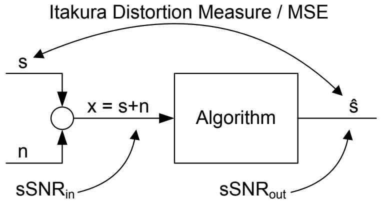

Figure 1: Notation

To measure the final speech signal quality, different quality measures were used. For good comparability with [1] the same two representative objective speech quality measures will be used here: the segmental Signal to Noise Ratio (sSNR) and the Itakura Distortion Measure (ID). And finally, since the Kalman filter is built to optimize the Mean Squared Error (MSE), we will use this measure as well.

2.1 Mean Squared Error

It is well known that the MSE is not an adequate measure for speech quality. It is nevertheless important for this paper, since the Kalman filter minimizes the MSE:

$$\mathrm{MSE} = \frac{1}{N}\sum_{n=1}^{N}\left(s(n) - \hat{s}(n)\right)^2 \tag{1}$$

where N is the length of the speech signal in samples.

2.2 Segmental SNR

The segmental SNR is a simple and effective speech quality measure which allows for good comparability:

$$\mathrm{sSNR_{dB}} = 10 \cdot \frac{1}{M} \cdot \sum_{m=0}^{M-1} \log\left(\frac{\sum\limits_{n=N\cdot m}^{N\cdot m+N-1} s^2[n]}{\sum\limits_{n=N\cdot m}^{N\cdot m+N-1} n^2[n]}\right) \tag{2}$$

where N denotes the segment width in samples. During the project, various segment sizes suggested in the literature were evaluated and $20\,ms = 410$ Samples (at a sampling rate of 20480 Samples/s) resulted in the best performance. To calculate the instantaneous signal and noise output powers, the algorithm is fed with the $x = s$ and $x = n$ signals separately. However, all internal parameters are adapted as in the $x = s+n$ case. In other words, this allows the calculation of the output sSNR because the response of the system to the noise *only* as well as to the signal *only* can be measured.

2.3 Itakura Distance Measure

The well known Itakura Distance Measure, which is also called the Log Likelihood Ratio, is selected as the second representative objective speech quality measure. The Itakura Distance Measure is defined as follows:

$$d_{\mathrm{ID}}\left(S_m(k), \hat{S}_m(k)\right) = \ln\left(\frac{\boldsymbol{b}^T\,\boldsymbol{R}_{SS}\,\boldsymbol{b}}{\boldsymbol{a}^T\,\boldsymbol{R}_{SS}\,\boldsymbol{a}}\right) \qquad (3)$$

where $k = n \in [N\cdot m\,,\ N\cdot m + N - 1]$, $\boldsymbol{R}_{SS}$ is the correlation matrix of the clean signal and $\boldsymbol{a}$ and $\boldsymbol{b}$ are the LPC coefficient vectors of the approximated (output) signal and the clean signal, respectively. Again, segments of $20\,ms$ and LPC order of 14 showed good results. In the end, all segmental values are arithmetically averaged. Even though objective quality measures are important, the final judgment of the speech quality is reserved for human listeners. For this purpose, the original and processed sound files can be found in [7].

2.4 Scenarios

During this project, carefully recorded sound files using a KEMAR were used to test the algorithm. The KEMAR manikin was equipped with two behind the ear (BTE) hearing aids. Each hearing aid contained two microphones in end-fire configuration that were connected to a digital audio recording system. For the results reported in this paper, the recording was done in an anechoic chamber.

Furthermore, several acoustic scenarios were used, the four most common ones being shown here as examples. The desired speech signal always comes from the front (0°), but the direction of the interfering signal differs. This different direction of the interfering signal exhibits itself in a time delay between the front microphone signal and the back microphone signal.

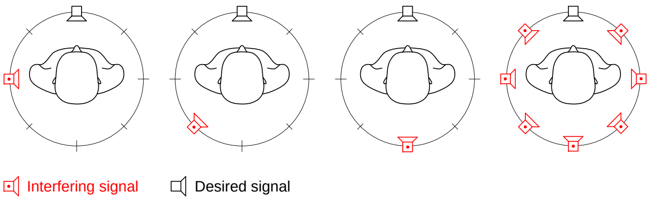

Figure 2: Acoustic scenarios

In the above figures, the interference is a female speech signal, while the desired signal (the signal at 0°) is a male speech signal and the listener stands in the middle of the circle. The three leftmost scenarios show the interference at 90°, 135° and 180°, while the rightmost scenario shows the so called cocktail-party situation, where there are multiple interferences (male and female) from all around the listener at 45°, 90°, 135°, 180°, 225°, 270° and 315°.

2.5 LOCO

The proposed algorithm, which has been named LOCO, is based on an Elko-beamformer (see Fig. 3). Since we expect the desired speech signal to come from the front and define everything from the back as noise, we can use the front and back cardioid signals (which are already available from the Elko-beamformer) as estimators of the speech and noise signals (Fig. 3).

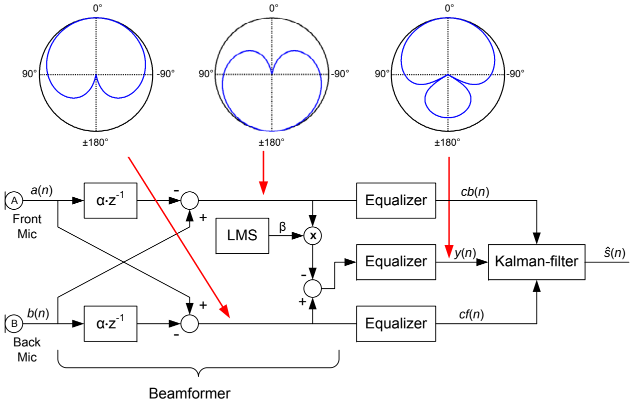

Figure 3: LOCO Algorithm

The front and back cardioid signals as well as the beamformed signal show highpass characteristics ($1 - z^{-2}$ for signals from the front with $\alpha = 1$). The beamformed signal can be equalized very efficiently with an IIR filter which has the inverse transfer function

$$H(z) = \frac{1}{1 - \beta\cdot(1-\alpha) - \alpha\cdot z^{-2}} \qquad (4)$$

where β is the adaptive parameter which determines the directivity of the Elko-beamformer. The cardioid signals used for the speech/noise model parameter estimates can be equalized with the following filter:

$$H(\mathrm{z}) = \frac{1}{1 - \alpha\,\mathrm{z}^{-2}} \qquad (5)$$

Choosing $\alpha < 1$ ensures the stability of these equalizers.

3. SPEECH MODEL AND KALMAN FILTER

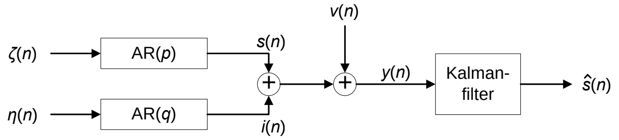

Figure 4: The speech / interference model

Using a Kalman filter for speech enhancement asks for a state space model. An appropriate one that is often used (e.g. in [3]), assumes that the speech signal $s(n)$ as well as the interference signal $i(n)$ can be adequately modeled by Autoregressive (AR) processes of order p and q respectively:

$$s(n) = -\sum_{k=1}^{p} a_k(n)\, s(n-k) + \zeta(n) \qquad (6)$$

$$i(n) = -\sum_{k=1}^{q} b_k(n)\, i(n-k) + \eta(n) \qquad (7)$$

The excitation signals $\zeta(n)$ and $\eta(n)$ are assumed to be independent zero mean white Gaussian noise with variance

$\sigma_\zeta^2(n)$ and $\sigma_\eta^2(n)$ respectively.
A corresponding state space model with the state vector $\boldsymbol{x}(n) = [s(n-p+1) \quad \cdots \quad s(n) \quad i(n-q+1) \quad \cdots \quad i(n)]^T$ can be given as

$$\boldsymbol{x}(n) = \boldsymbol{A}(n-1)\,\boldsymbol{x}(n-1) + \boldsymbol{B}\,\boldsymbol{u}(n) \tag{8}$$

$$y(n) = \boldsymbol{C}\,\boldsymbol{x}(n) + \nu(n) \tag{9}$$

where $\nu(n)$ is the white, Gaussian measurement error with variance σ_ν^2 and the input $u(n) = [\zeta(n) \quad \eta(n)]^T$. The transition matrix $\boldsymbol{A}(n)$, the input matrix $\boldsymbol{B}$ and the output matrix $\boldsymbol{C}$ are defined as follows:

$$\boldsymbol{A}(n) = \left[\begin{array}{cccc|cccc} \multicolumn{4}{c|}{\overbrace{\hphantom{-a_p \quad \cdots \quad \cdots \quad -a_1}}^{p}} & & & & \\ 0 & 1 & & & & & & \\ & & \ddots & & & \boldsymbol{0}_{p,q} & & \\ & & & 1 & & & & \\ -a_p & \cdots & \cdots & -a_1 & & & & \\ \hline & & & & 0 & 1 & & \\ & \boldsymbol{0}_{q,p} & & & & & \ddots & \\ & & & & & & & 1 \\ & & & & -b_q & \cdots & \cdots & -b_1 \\ & & & & \multicolumn{4}{c}{\underbrace{\hphantom{-b_q \quad \cdots \quad \cdots \quad -b_1}}_{q}} \end{array}\right] \tag{10}$$

$$\boldsymbol{B} = \left[\begin{array}{cccc|cccc} \multicolumn{4}{c|}{\overbrace{\hphantom{0 \cdots 0 \; 1}}^{p}} & \multicolumn{4}{c}{\overbrace{\hphantom{0 \cdots 0 \; 1}}^{q}} \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 \end{array}\right]^T \tag{11}$$

$$\boldsymbol{C} = \left[\begin{array}{cccc|cccc} \multicolumn{4}{c|}{\overbrace{\hphantom{0 \cdots 0 \; 1}}^{p}} & \multicolumn{4}{c}{\overbrace{\hphantom{0 \cdots 0 \; 1}}^{q}} \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 1 \end{array}\right] \tag{12}$$

Note that for simplicity $\boldsymbol{A}(n)$ in (10) is shown as time invariant, while in fact, the parameters a_k and b_k may change at every time step n (as shown in (6) and (7)). This ability of the Kalman filter to deal with a time variant signal and speech model is essential for the use of the Kalman filter instead of a Wiener filter.

Since in real world applications the input $u(n)$ is unknown, one will consider it to be zero. Based on this simplification one will have an uncertainty in the state vector $\boldsymbol{x}(n)$. The covariance matrix $\boldsymbol{Q}_{\boldsymbol{w}}(n)$ of the corresponding state error can be calculated as follows:

$$\boldsymbol{Q}_{\boldsymbol{w}}(n) = \boldsymbol{B}\,E\left\{\boldsymbol{u}(n)\,\boldsymbol{u}^T(n)\right\}\boldsymbol{B}^T = \boldsymbol{B}\begin{bmatrix}\sigma_\zeta^2(n) & 0 \\ 0 & \sigma_\eta^2(n)\end{bmatrix}\boldsymbol{B}^T \tag{13}$$

Based on this state space model, a Kalman filter can be used to estimate the state vector $\boldsymbol{x}(n)$ based on the noisy measurements $y(k)$ (k up to n). This estimate $\hat{\boldsymbol{x}}(n)$ is given as follows [2].

$$\begin{aligned} \hat{\boldsymbol{x}}(n|n-1) &= \boldsymbol{A}(n-1)\hat{\boldsymbol{x}}(n-1|n-1) \\ \hat{\boldsymbol{P}}(n|n-1) &= \boldsymbol{A}(n-1)\hat{\boldsymbol{P}}(n-1|n-1)\boldsymbol{A}^T(n-1) + \boldsymbol{Q}_{\boldsymbol{w}}(n) \\ \boldsymbol{K}(n) &= \frac{\hat{\boldsymbol{P}}(n|n-1)\boldsymbol{C}^T}{\boldsymbol{C}\hat{\boldsymbol{P}}(n|n-1)\boldsymbol{C}^T + \boldsymbol{Q}_{\boldsymbol{v}}(n)} \qquad (14) \\ \hat{\boldsymbol{x}}(n|n) &= \hat{\boldsymbol{x}}(n|n-1) + \boldsymbol{K}(n)[\boldsymbol{y}(n) - \boldsymbol{C}(n)\hat{\boldsymbol{x}}(n|n-1)] \\ \hat{\boldsymbol{P}}(n|n) &= [\boldsymbol{I} - \boldsymbol{K}(n)\boldsymbol{C}(n)]\,\hat{\boldsymbol{P}}(n|n-1) \end{aligned}$$

Where $\hat{\boldsymbol{P}}(n|n-1)$ and $\hat{\boldsymbol{P}}(n|n)$ are the *a priori* and the *a posteriori* error covariance matrices respectively. $\boldsymbol{K}(n)$ is the Kalman gain vector and $\boldsymbol{I}$ the identity matrix of order $p+q$. The estimated speech signal $\hat{s}(n)$ can be found at the pth position of the estimated state vector $\hat{\boldsymbol{x}}(n|n)$.

Note that because of the special structure of the vector $\boldsymbol{x}(n)$, one will estimate not only $s(n)$ but also $s(n-1)\cdots s(n-p+1)$. Since these estimates are all based on measurements $y(k)$ with k up to n, they correspond to fixed-lag estimates $\hat{s}(n-1|n)\cdots\hat{s}(n-p+1|n)$ [6]. As shown in [4], fixed-lag smoothers can give better results because they lead to better suppression in spectral valleys.

While working with artificial speech signals, with a growing fixed-lag, the improvement was evident [7]. However, with real-world speech signals fixed-lag smoothing did not result in the expected better performance.

4. DIFFERENT SCHEMES

The way the proposed speech and noise models and the corresponding Kalman filter are employed for the purpose of noise suppression is not unique. One can think of several different approaches, to estimate the parameters and to run the Kalman filter. Most meaningful combinations have been implemented in [7]. In our tests, the three schemes shown in Fig. 5 resulted in the best performances.

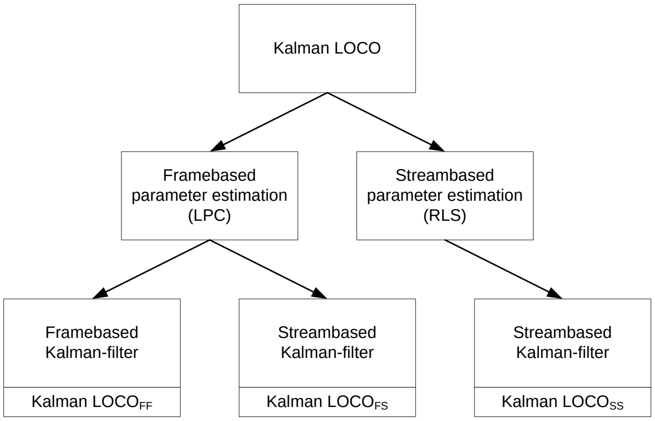
Kalman LOCO
Framebased parameter estimation (LPC)
Streambased parameter estimation (RLS)
Framebased Kalman-filter
Streambased Kalman-filter
Streambased Kalman-filter
Kalman LOCO$_{\text{FF}}$
Kalman LOCO$_{\text{FS}}$
Kalman LOCO$_{\text{SS}}$

Figure 5: Different Kalman LOCO Schemes

Because of the brevity of this paper, we will focus on the Kalman LOCO$_{\text{FS}}$ scheme. From all the tested schemes, this has shown to be the most successful one. For more information on the other schemes, refer to [7]. Figure 6 gives an overview of the proposed Kalman LOCO$_{\text{FS}}$ scheme.

4.1 Parameter estimation

The estimation of the target (speech) and interference (noise) parameters $\boldsymbol{a}_k$, $\boldsymbol{b}_k$, σ_ζ^2 and σ_η^2 proved to be one of the key points in the proposed scheme. Based on the LOCO idea, several algorithms have been implemented and tested on real-world speech signals. Estimating the parameters by Linear Predictive Coding (LPC) analysis showed the best overall performance. The frames of length 128 samples used for the LPC analysis are windowed (with a Hann window) and overlapped by 75%.
To ensure stability of the estimated systems, we use the *autocorrelation method* for the LPC analysis [5]. With the autocorrelation vector $\boldsymbol{r}_{\boldsymbol{x}} = [r_x(1) \quad \cdots \quad r_x(p)]^T$ of the windowed frame, the corresponding LPC parameters $\boldsymbol{a} = [a_1 \cdots a_k]^T$ can be calculated with the following formula:

$$\boldsymbol{R}_{\boldsymbol{x}}\boldsymbol{a} = -\boldsymbol{r}_{\boldsymbol{x}} \tag{15}$$

where $\boldsymbol{R}_{\boldsymbol{x}}$ is the $p \times p$ Toeplitz autocorrelation matrix. The power σ^2 of the corresponding excitation signal can be cal-

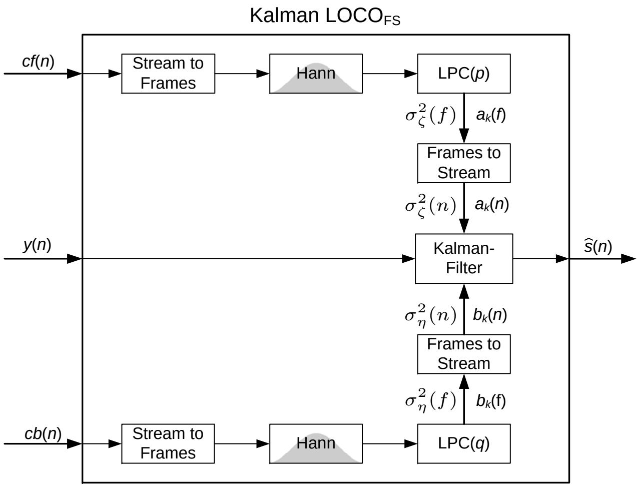

Figure 6: Kalman LOCO$_{\mathrm{FS}}$

culated as follows:

$$\sigma^2 = \frac{r_x(0) + \sum_{k=1}^{p} a_k\, r_x(k)}{N} \qquad (16)$$

where $N = 128$ is the number of samples per frame.

To suppress possible artifacts, the estimated parameters need to be smoothed. To avoid the risk of producing unstable systems, the parameters are not smoothed directly. Instead, the autocorrelation vector $\boldsymbol{r_x}$ used in the LPC analysis is smoothed with a simple first order IIR lowpass filter with a time constant of $\tau = (-32/20480)/\ln(0.95) \approx 30\,ms$.

For the estimation of the variance σ_ν^2 of the measurement error, we assume that it is a property of the measurement equipment and stays constant over time. Based on this assumption, it is straight forward to measure this variance offline during a speech pause. It is then implemented as a constant in the algorithm.

4.2 Usage of the estimated parameters

Since the Kalman filter works sample-by-sample (stream) based, the estimated parameters $\boldsymbol{a}$ and σ^2, which are estimated per frame f, need to be structured into a stream as well (the Kalman filter expects one parameter set for every time instant n). We assume that the estimated parameters are the most accurate in the middle of a frame. This assumption leads to using the parameters for the middle 32 samples of the corresponding frame (Fig. 7).

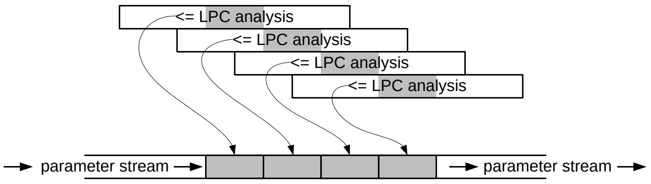

Figure 7: Usage of the estimated parameters

4.3 Model order

Two obvious parameters which have to be optimized are the model orders p and q. As we expect the target as well as the interference to be a speech signal, these orders have always been considered equal. Generally speaking, one can say that with higher orders, the speech and interference signal can be modeled more precisely and therefore the results become better. This behavior can be seen in Figs. 8 to 10.

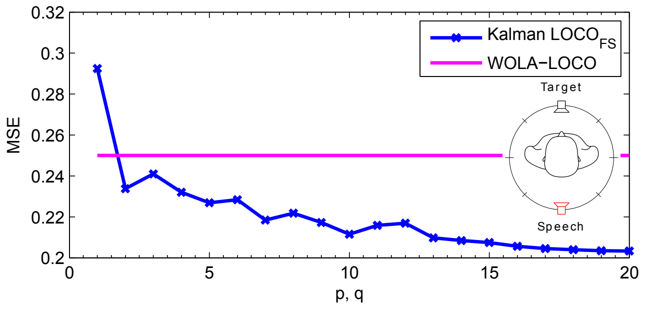

Figure 8: The MSE vs. the model order

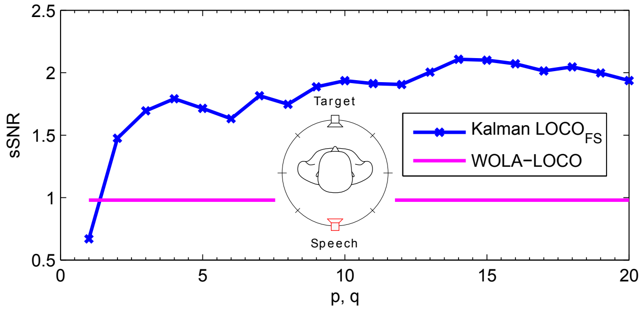

Figure 9: The sSNR vs. the model order

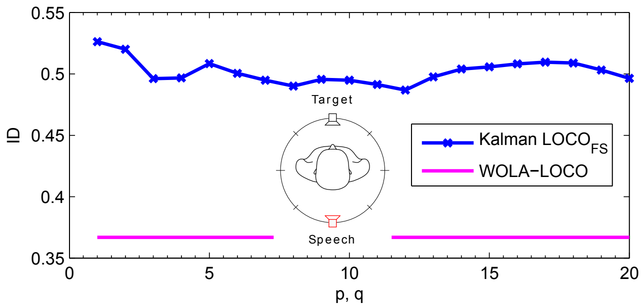

Figure 10: The ID vs. the model order

However, with higher order models, the matrices used for LPC analysis as well as for the Kalman filter become bigger. From a computational point of view, this is clearly undesirable. Considering these two effects of higher model orders, a reasonable tradeoff are model orders $p = q = 10$.

5. EXPERIMENTAL RESULTS

Tables 1-4 compare the proposed Kalman LOCO$_{\mathrm{FS}}$ with the WOLA-LOCO and the simple Elko-Beamformer. The Kalman LOCO$_{\mathrm{FS}}$ is implemented with model orders $p = q = 10$. The different tables show the results for the four acoustic scenarios shown in Fig. 2. Note that for the scenarios where the interference comes from the side, only the results from the left channel are shown, since the right channel is in the acoustic shadow of the head and hence the interference is not really a problem on that side.

As expected, the Kalman LOCO$_{\mathrm{FS}}$ achieves in every scenario a considerably better MSE than the WOLA-LOCO.

Also, the sSNR achieved by the Kalman LOCO_{FS} is always better than the one achieved by WOLA-LOCO. While the better MSE shows that the Kalman filter can handle the nonstationary speech signals better than the Wiener filter, the improved sSNR shows that the Kalman filter can also improve the perceptual quality of the speech signal. However, the WOLA-LOCO shows the better Ithakura Distance (ID).

On average, the objective results of our algorithm are better than the ones of WOLA-LOCO. But even more important than these objective measures are listening tests, since for noise suppression human listeners must be the ultimate judges of the quality. Therefore, the original and processed sound files as well as the MATLAB code can be found in [7]. These files show that our new algorithm performs acoustically comparable to WOLA-LOCO. The suppression of the interference is very similar and it is hard to tell which algorithm sounds better, since both of them produce slight, but different, acoustic artifacts.

Method	MSE	sSNR	ID
Only Elko-beamformer	1.936	-0.190	0.597
WOLA-LOCO	1.095	-0.229	0.563
Kalman LOCO_{FS}	0.714	0.277	0.654

Table 1: Interference at 90°, left channel only

Method	MSE	sSNR	ID
Only Elko-beamformer	0.651	2.39	0.494
WOLA-LOCO	0.269	2.88	0.409
Kalman LOCO_{FS}	0.212	3.63	0.472

Table 2: Interference at 135°, left channel only

Method	MSE	sSNR	ID
Only Elko-beamformer	0.735	1.04	0.438
WOLA-LOCO	0.250	0.98	0.367
Kalman LOCO_{FS}	0.212	1.94	0.495

Table 3: Interference at 180°, average of the left and the right channels

6. SUMMARY AND CONCLUSION

The proposed Kalman filter based noise suppression algorithm shows that the advantage of a Kalman filter over a Wiener filter can successfully be exploited. While considering the MSE, the proposed scheme outperforms the WOLA-LOCO algorithm significantly, as it is able to track the nonstationery speech signals better than the WOLA-LOCO. Tests with artificial speech signals (generated with time varying all-pole models) showed that using a Kalman smoother instead of a Kalman filter can further improve the performance of the proposed scheme significantly. The fact that with real world speech signals a Kalman smoother does not lead to better performance, suggests that an improvement of the speech/noise models and/or the estimation of the speech/noise model parameters could lead to even better results.

The segmental SNR and the Ithakura Distance as well as subjective listening tests show, that the Kalman-LOCO performs similarly to the WOLA-LOCO if one considers the human perception. In contrasts to WOLA-LOCO, the Kalman-LOCO has the advantage that one has a model of the speech/noise processes. As shown in [4], such models could be used for deemphasizing and emphasizing filters. With such filters, the perceptional quality can be improved at the price of a higher MSE. The optimal use of these models for such pre- and post processing is currently one of our research efforts.

Method	MSE	sSNR	ID
Only Elko-beamformer	1.157	2.16	0.495
WOLA-LOCO	0.405	2.15	0.473
Kalman LOCO_{FS}	0.361	2.81	0.632

Table 4: Cocktail-party noise at 45°, 90°, 135°, 180°, 225°, 270° and 315°, average of the left and the right channels

As the matrices involved in the LPC analysis and the Kalman filter are quite large, the computational effort of the proposed scheme is higher than the one of the WOLA-LOCO. There are many possibilities to reduce this computational effort (e.g. subsampling of the Kalman filter or making use the sparse $\boldsymbol{A}$ matrix). However, as the computational effort was not the main topic of this work, it has not been further investigated for this paper. Now that the potential of this approach has been shown, we are currently investigating efficient computational approaches.

In summary, the novel scheme presented in this paper, Kalman-LOCO, shows that a Kalman filter based approach to noise suppression is quite competitive with the existing schemes such as WOLA-LOCO. From a perceptual point of view, Kalman-LOCO and WOLA-LOCO sound similar, both with their distinct artifacts. From an objective measure point of view, Kalman-LOCO outperforms WOLA-LOCO. On the other hand, currently Kalman-LOCO consumes more computational resources than WOLA-LOCO. One great advantage of Kalman-LOCO which has not yet been exploited is the ability to use the noise and the speech model necessary for the Kalman filter for a pre- and/or post- filter. These filters are known to further improve the perceptional quality at the expense of the MSE and are currently subject to further research.

REFERENCES

[1] H. Lang, R. Hegner, G. Schuster, "A high performance low complexity noise suppression algorithm" in *Proceedings of the European Signal Processing Conference*, 2009

[2] Monson H. Hayes, "Statistical Digital Signal Processing and Modeling", John Wiley & Sons, 1996

[3] P. Sorqvast, P. Handel, B. Ottersten, "Kalman filtering for low distortion speech enhancement in mobile communication" in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1997

[4] V. Grancharov, J. Samuelsson, B. Kleijn, "On Causal Algorithms for Speech Enhancement" in *IEEE Transactions on Audio, Speech, and Language Processing*, 2006

[5] John Makhoul, "Linear Prediction: A Tutorial Review" in *Proceedings of the IEEE, Vol. 63*, 1975

[6] John B. Moore, "Discrete-Time Fixed-Lag Smoothing Algorithms", Pergamon Press, 1973

[7] Ramun Schmid, "Kalman LOCO", HSR Hochschule fuer Technik Rapperswil, 2010 `http://www.medialab.ch/EUSIPCO2010/`.

[8] G.W. Elko and Anh-Tho Nguyen Pong, "A simple adaptive first order differential microphone" in *IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 169–172, Oct 1995.