IMPROVED MEASUREMENT NOISE COVARIANCE ESTIMATION FOR N-CHANNEL FEEDBACK CANCELLATION BASED ON THE FREQUENCY DOMAIN ADAPTIVE KALMAN FILTER

Jan Franzen, Tim Fingscheidt

Institute for Communications Technology, Technische Universität Braunschweig,
Schleinitzstr. 22, 38106 Braunschweig, Germany
{j.franzen, t.fingscheidt}@tu-bs.de

ABSTRACT

Acoustic feedback cancellation has gained a major and steady role in the research fields of signal processing over the past decades, since it is inevitable for numerous applications such as hearing aids or in-car communication systems. In this paper, we investigate measurement noise covariance estimation approaches for feedback cancellation based on the frequency domain adaptive Kalman filter (FDAKF). The capabilities of these estimation methods significantly affect the performance of the FDAKF. We summarize and investigate existing approaches from literature and furthermore provide two new proposals that are explicitly motivated for the use in acoustic feedback cancellation. Experimental validation in the context of an in-car communication system shows that our proposals obtain much better speech quality compared to existing approaches and additionally increase the overall feedback suppression.

***Index Terms*—** feedback cancellation, frequency domain adaptive Kalman filter, measurement noise covariance estimation

1. INTRODUCTION

Applications with a loudspeaker-enclosure-microphone (LEM) system are very common these days: teleconference and hands-free devices, hearing aids and in-car communication systems, only to name a few. They all have in common that the microphone signal includes an undesired echo or feedback component from the loudspeakers. Typically, these systems use an adaptive filter to estimate the LEM impulse response and create an echo estimate which is subtracted from the microphone signal. Thereby an echo- (or feedback-)free enhanced signal is obtained. As for example pointed out in [1], one well-known and approved approach in adaptive filter theory is the Kalman filter. As frequency domain adaptive Kalman filter (FDAKF) it has been succesfully deployed for acoustic echo cancellation (AEC) [2] and further developed in many publications such as [3, 4, 5, 6]. Advantages of the FDAKF by now are its near-optimal stepsize control, robust performance even when 'near-end' and 'far-end' are active simultaneously (double talk), and the possibility to efficiently combine it with a residual echo suppression [2, 7].

In the context of acoustic *feedback* cancellation (AFC), where the deployed approaches are additionally challenged by the closed feedback loop, a wide variety of algorithms is available. These reach from time domain approaches [8] to subband approaches [9, 10], systems based on the normalized least mean squares [11, 12], and many more. Just in the recent few years, a selection of publications accredits the capability of the FDAKF for AFC and proposes systems based on it, as for example [13, 14, 15] or [16].

However, as stated in [17], it is of crucial importance for the adaptive Kalman filter to estimate the covariances of its process noise and measurement noise correctly. Not till then it is possible to obtain an optimal estimation of the true system state, i.e., here the impulse response transfer functions. Though literature provides well-investigated covariance estimation rules for Kalman-based AEC, e.g., [2] or [18], publications with focus on AFC merely adopt these techniques without considering the additional signal correlation in the closed feedback loop. So far, literature covers investigations of covariance learning rules for Kalman-based AFC insufficiently.

In this paper, we investigate measurement noise covariance estimation methods for feedback cancellation based on the frequency domain adaptive Kalman filter. We summarize and investigate existing approaches from the literature and furthermore provide two new proposals that are explicitly motivated for (yet not limited to) the use in acoustic feedback cancellation. If not yet published before, all approaches are extended for the use in an N-channel system, consisting of one microphone and N loudspeakers. The experimental validation is performed for the single-channel as well as multi-channel case in the practical context of an in-car communication (ICC) system.

The remainder of this paper is structured as follows: In Section 2, a system overview is given and the N-channel FDAKF including residual echo suppression (RES) is described. Known approaches for the estimation of the measurement noise covariance are summarized in Section 3, followed by the two new proposals of this paper. In Section 4, the experimental validation of all approaches is given. Section 5 provides conclusions.

2. SYSTEM MODEL AND ALGORITHM

A block diagram of the underlying system model is shown in Figure 1. Starting in the top left of the figure, an optional block indicates the possibility to connect an audio player (or similar) as 'additional signals' to the feedback loop. The respective 1 to N channels $x_j(n)$, $j \in \mathcal{N} = \{1, ..., N\}$, provide the reference signals for feedback cancellation and are connected to the 1 to N loudspeakers in the loudspeaker-enclosure-microphone (LEM) model. The LEM model consists of the impulse responses $h_j(n)$ (for simplicity written as time-invariant) providing the feedback signals $d_j(n)$. Superimposed with the desired signal $s(n)$ and additional ambient noise $n(n)$, the microphone signal $y(n)$ is formed. By subtracting the estimated feedback signals, the FDAKF computes an error signal $e(n)$ which is then subject to a residual feedback suppression, amplified with a gain factor, and finally added to the loudspeaker signals in the feedback loop.

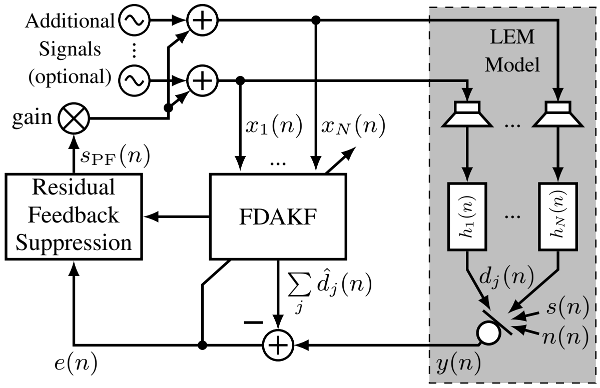

Fig. 1. System model of N-channel feedback cancellation.

The algorithm of the variationally-diagonalized FDAKF is well-known from former publications as for example [3, 19] or [5]. Since the FDAKF and its postfilter for residual echo (or here rather *feedback*) suppression go hand in hand [7], we treat them as one ensemble within this work. On a frame basis, the algorithm is given in compact form as follows.

As preparatory processing reference signals ($j \in \mathcal{N}$) are obtained $\mathbf{x}_j(\ell) = \big[x_j\big((\ell-1)\cdot R+R-K\big), ..., x_j\big((\ell-1)\cdot R+R-1\big)\big]^T$, with frame index $\ell \in \{1,2,...\}$ and frame shift R. Their DFT is computed as $\mathbf{X}_j(\ell) = \mathrm{diag}\{\mathbf{F}_{K\times K}\cdot \mathbf{x}_j(\ell)\}$ with the K-point DFT matrix $\mathbf{F}_{K\times K}$. The corresponding part of the microphone signal $\mathbf{y}(\ell) = \big[\mathbf{0}_{K-R}, y\big((\ell-1)\cdot R\big), ..., y\big((\ell-1)\cdot R+R-1\big)\big]^T$ is as well transformed to the DFT domain $\mathbf{Y}(\ell) = \mathrm{diag}\{\mathbf{F}_{K\times K}\cdot \mathbf{y}(\ell)\}$.

Starting the coefficient adaptation, the system functions are predicted as $\hat{\mathbf{H}}_j^+(\ell) = a\hat{\mathbf{H}}_j(\ell-1)$, initially with $\hat{\mathbf{H}}_j(0) = \mathbf{0}_{K\times K}$. The process noise covariance matrices are computed for the intra-channel case: $\mathbf{\Psi}^{\Delta}_{j,j}(\ell-1) = (1-a^2)\big[\hat{\mathbf{H}}_j(\ell-1)\hat{\mathbf{H}}_j^H(\ell-1) + \mathbf{P}_{j,j}(\ell-1)\big]$, with $\mathbf{P}_{j,j}(0) = \mathbf{I}_{K\times K}$, while the cross-channel terms remain zero: $\mathbf{\Psi}^{\Delta}_{j,i\neq j}(\ell-1) = \mathbf{0}_{K\times K}$. $()^H$ denotes the Hermitian. As next steps the state error covariance matrices are predicted as $\mathbf{P}^+_{j,i}(\ell) = a^2\mathbf{P}_{j,i}(\ell-1) + \mathbf{\Psi}^{\Delta}_{j,i}(\ell-1)$, and a preliminary error signal (see [5]) is computed in the DFT domain $\tilde{\mathbf{E}}(\ell) = \mathbf{Y}(\ell) - \sum_j \mathbf{G}\cdot\big(\mathbf{X}_j(\ell)\cdot\hat{\mathbf{H}}_j^+(\ell)\big)$, with the overlap-save constraint matrix $\mathbf{G} = \mathbf{F}_{K\times K}\mathbf{Q}\mathbf{Q}^T\mathbf{F}^{-1}_{K\times K}$ and $\mathbf{Q} = (\mathbf{0}_{R\times K-R}\;\mathbf{I}_{R\times R})^T$.

Subsequently, the variable of interest for this work, namely the measurement noise covariance matrix $\mathbf{\Psi}^S(\ell)$, is computed. Various methods along with our new proposals will be discussed in detail within the next section.

Now, the respective stepsizes are given by $\boldsymbol{\mu}_{j,i}(\ell) = \frac{R}{K}\mathbf{P}^+_{j,i}(\ell)\big[\frac{R}{K}\big(\sum_j\sum_i \mathbf{X}_j(\ell)\mathbf{P}^+_{j,i}(\ell)\mathbf{X}_i^H(\ell)\big) + \mathbf{\Psi}^S(\ell)\big]^{-1}$, which are in turn used to compute the Kalman gain for each channel as $\mathbf{K}_j(\ell) = \sum_i \boldsymbol{\mu}_{j,i}(\ell)\mathbf{X}_i^H(\ell)$. Using the Kalman gains, the state error covariance matrices are updated to $\mathbf{P}_{j,i}(\ell) = \mathbf{P}^+_{j,i}(\ell) - \frac{R}{K}\mathbf{K}_j(\ell)\big(\sum_j \mathbf{X}_j(\ell)\mathbf{P}^+_{j,i}(\ell)\big)$, and finally the estimated system functions are updated: $\hat{\mathbf{H}}_j(\ell) = \hat{\mathbf{H}}_j^+(\ell) + \mathbf{K}_j(\ell)\cdot\tilde{\mathbf{E}}(\ell)$.

Having obtained the new estimates of the system functions, the actual feedback cancellation can be performed. To accomplish this, the estimated feedback signals $\hat{\mathbf{D}}_j(\ell) = \mathbf{G}\cdot\big(\mathbf{X}_j(\ell)\cdot\hat{\mathbf{H}}_j(\ell)\big)$ are used to obtain the enhanced signal: $\mathbf{E}(\ell) = \mathbf{Y}(\ell) - \sum_j \hat{\mathbf{D}}_j(\ell)$. Using an inverse DFT, the time domain equivalent of the enhanced signal is obtained as $\mathbf{e}(\ell) = \mathbf{F}^{-1}_{K\times K}\cdot\mathbf{E}(\ell)$. Due to overlap-save constraints, only the last R samples of the output $\mathbf{e}(\ell)$ are used, all others are simply discarded.

As derived in [7] (for $\lambda = 1$), an efficient formulation to obtain the Wiener postfilter coefficients for residual feedback suppression directly from the previous coefficient adaptation is given by $G_{\mathrm{PF}}(\ell,k) = \big(1 - \sum_j X_j(\ell,k)\mu_j(\ell,k)X_j^*(\ell,k)\big)^{\lambda}$, where $\mu_j(\ell,k) = \Big(\Psi_{ss}(\ell,k) + \frac{R}{K}\big(\sum_i X_i(\ell,k)P^+_{i,i}(\ell,k)X_i^*(\ell,k)\big)\Big)^{-1}\cdot\frac{R}{K}P^+_{j,j}(\ell,k)$ is a separate stepsize definition used only for the postfilter and λ is a design parameter introduced in this work. In the single-channel case, this formulation is equivalent to [2, 4].

3. MEASUREMENT NOISE COVARIANCE ESTIMATION

The measurement noise covariance matrix $\mathbf{\Psi}^S$ has an important role within the FDAKF. It is basically a joint representation of the non-feedback signal components $\big(s(n)+n(n)\big)$ in the microphone signal and allows the algorithm to distinguish between feedback and non-feedback components. Compared to acoustic *echo* cancellation, in the case of acoustic *feedback* cancellation the estimation of $\mathbf{\Psi}^S$ becomes even more difficult and at the same time more relevant, since the desired signal component $s(n)$ and the undesired feedback component are highly correlated. In the following, the known approaches for the estimation of $\mathbf{\Psi}^S$ from literature are summarized and two new proposals are presented, the latter specifically matched towards acoustic *feedback* cancellation. All approaches will be evaluated experimentally in Section 4.

Oracle • First of all, an upper performance bound as influenced by $\mathbf{\Psi}^S$ can be determined with an oracle experiment. This can be obtained by simply setting

$$\mathbf{\Psi}^S(\ell) = (1-\beta)\cdot\Big(\mathbf{Y}(\ell) - \sum_{j\in\mathcal{N}}\mathbf{D}_j(\ell)\Big)\cdot \Big(\mathbf{Y}(\ell) - \sum_{j\in\mathcal{N}}\mathbf{D}_j(\ell)\Big)^H + \beta\cdot\mathbf{\Psi}^S(\ell-1), \tag{1}$$

using the known feedback components $\mathbf{D}_j$. The smoothing factor $\beta = 0.5$ avoids harsh changes over time which would degrade the performance.

ME'10 • In [18] (Malik and Enzner) and [19, 20], an online maximum-likelihood estimation rule has been derived that is based on the *a posteriori* error signal of the previous frame[1]. Again extending the given formula by a smoothing factor $\beta = 0.5$, the learning rule is then given as

$$\mathbf{\Psi}^S(\ell) = (1-\beta)\cdot\Big(\mathbf{E}(\ell-1)\mathbf{E}^H(\ell-1) + \frac{R}{K}\Big(\sum_{j\in\mathcal{N}}\sum_{i\in\mathcal{N}}\mathbf{X}_j(\ell-1)\mathbf{P}_{j,i}(\ell-1)\mathbf{X}_i^H(\ell-1)\Big)\Big) + \beta\cdot\mathbf{\Psi}^S(\ell-1). \tag{2}$$

JEF'14 • The general idea of **ME'10** has been taken up by Jung et al. in [5] and has been refined for the FDAKF algorithm in acoustic *echo* cancellation. Using the given formula, the authors propose to use the *preliminary* error signal of the current

[1]Note that the variable $\mathbf{P}_{j,i}(\ell)$ used in [18] is not yet available at the time when $\mathbf{\Psi}^S(\ell)$ is computed. Therefore we deploy $\mathbf{P}_{j,i}(\ell-1)$ here.

frame $\tilde{\mathbf{E}}(\ell)$ and the *predicted* state error covariance $\mathbf{P}^{+}_{j,i}(\ell)$. The estimation formula is then given as

$$\mathbf{\Psi}^{S}(\ell) = (1-\beta)\cdot\Big(\tilde{\mathbf{E}}(\ell)\tilde{\mathbf{E}}^{H}(\ell)+ \tfrac{R}{K}\big(\sum_{j\in\mathcal{N}}\sum_{i\in\mathcal{N}}\mathbf{X}_{j}(\ell)\mathbf{P}^{+}_{j,i}(\ell)\mathbf{X}^{H}_{i}(\ell)\big)\Big) + \beta\cdot\mathbf{\Psi}^{S}(\ell-1), \tag{3}$$

again with smoothing factor $\beta = 0.5$.

KME'14 • Interestingly, in the presentation of the partitioned FDAKF in [4], Kuech et al. avoid the explicit computation of $\mathbf{\Psi}^{S}$. Instead, they directly substitute the covariance $\mathbf{\Psi}_{EE}$ of the error signal into the stepsize definition as

$$\boldsymbol{\mu}_{j,i}(\ell) = \tfrac{R}{K}\mathbf{P}^{+}_{j,i}(\ell)\mathbf{\Psi}^{-1}_{EE}(\ell). \tag{4}$$

Since they have not explicitly stated how $\mathbf{\Psi}_{EE}$ is obtained, we assume it to be based on the preliminary error signal as $\mathbf{\Psi}_{EE}(\ell) = (1-\beta)\cdot\big(\tilde{\mathbf{E}}(\ell)\tilde{\mathbf{E}}^{H}(\ell)\big) + \beta\cdot\mathbf{\Psi}_{EE}(\ell-1)$. Here the smoothing factor is set to $\beta = 0.6$ to obtain best results within our experiments ($\beta = 0.5$ led to an instable system), and $\mathbf{\Psi}_{EE}(0) = \mathbf{0}_{K\times K}$.

YEY'17 • Although it is not directly used in a Kalman filter, a very interesting yet different approach is shown by Yang et al. in [21] for the FDAF. The authors obtain the power spectral density (PSD) of the measurement noise as

$$\Phi^{S}(\ell,k) = (1 - C_{\bar{x}e}(\ell,k))\cdot\Phi_{ee}(\ell,k). \tag{5}$$

Here, $C_{\bar{x}e}(\ell,k) = |\Phi_{\bar{x}e}(\ell,k)|^2 \,/\, \big(\Phi_{\bar{x}\bar{x}}(\ell,k)\Phi_{ee}(\ell,k)\big)$, whereas the used PSDs and cross-PSDs are being estimated recursively:
$\hat{\Phi}_{ee}(\ell,k) = \alpha\hat{\Phi}_{ee}(\ell-1,k) + (1-\alpha)|\tilde{E}(\ell,k)|^2$,
$\hat{\Phi}_{\bar{x}\bar{x}}(\ell,k) = \alpha\hat{\Phi}_{\bar{x}\bar{x}}(\ell-1,k) + (1-\alpha)|\overline{X}(\ell,k)|^2$, and
$\hat{\Phi}_{\bar{x}e}(\ell,k) = \alpha\hat{\Phi}_{\bar{x}e}(\ell-1,k) + (1-\alpha)\overline{X}^{*}(\ell,k)\tilde{E}(\ell,k)$.
In contrast to **ME'10** or **JEF'14**, the variable $\bar{\mathbf{x}}$ and its DFT $\overline{\mathbf{X}}(\ell)$ are only based on the current frame: $\bar{\mathbf{x}}(\ell) = \big[\mathbf{0}_{K-R}, x\big((\ell-1)\cdot R\big), \ldots, x\big((\ell-1)\cdot R + R-1\big)\big]^{T}$, 'in order to match the definition' [21] of the microphone signal component.

For this approach we found the smoothing factor $\alpha = 0.6$ to give the best results within our experiments. It should be noted that we will use the preliminary $\tilde{E}$ instead of E for the application in the FDAKF, since—as before—E is not yet available at that time instant. Furthermore, we obtain an extension to the N-channel case by
$C_{\bar{x}e}(\ell,k) = \sum_{j\in\mathcal{N}}|\Phi_{\bar{x}_{j}e}(\ell,k)|^2 \,/\, \big((\sum_{j\in\mathcal{N}}\Phi_{\bar{x}_{j}\bar{x}_{j}}(\ell,k))\Phi_{ee}(\ell,k)\big)$,
finally obtaining $\mathbf{\Psi}^{S}(\ell) = \mathrm{diag}\{[\Phi^{S}(\ell,1), \ldots, \Phi^{S}(\ell,K)]^{T}\}$.

Proposal 1 • Our first proposal is based on the idea of **YEY'17** *to only take the current frame of the reference signal into account*, however, with a very different motivation. While Yang et al. [21] did this to match the different PSDs, it has a significant and interesting influence in the case of feedback cancellation: It reduces the correlation between reference signals and the microphone signal. Since the reference signals contain a slightly delayed segment of $s(n)$, to some extent the signals already hurt the generally underlying assumption of statistical independence between these two components. Though noting that previous frames of the reference signals are obviously necessary in the *other* FDAKF computation steps to identify the feedback paths correctly, including these past frames within *this* computation step is more likely to result in a misinterpretation of the feedback components as desired signal. Conversely: Focusing the reference signal in this step to only contain the current frame should as well reduce the correlation and thus result in a better performance. Accordingly, we induce **YEY'17** into the Kalman-optimized formulation of **JEF'14** and just use the most recent frame of the reference signals, resulting in

$$\mathbf{\Psi}^{S}(\ell) = (1-\beta)\cdot\Big(\tilde{\mathbf{E}}(\ell)\tilde{\mathbf{E}}^{H}(\ell)+ \tfrac{R}{K}\big(\sum_{j}\sum_{i}\overline{\mathbf{X}}_{j}(\ell)\mathbf{P}^{+}_{j,i}(\ell)\overline{\mathbf{X}}^{H}_{i}(\ell)\big)\Big) + \beta\cdot\mathbf{\Psi}^{S}(\ell-1), \tag{6}$$

with smoothing factor $\beta = 0.5$.

Proposal 2 • Our second proposal is a more extreme continuation of our first proposal. We assume the *preliminary* error signal $\tilde{\mathbf{E}}(\ell)$ to be estimated perfectly already and therefore ignore the predicted state error covariances in this computation step. This allows us to drop the term of reference signals and predicted state error covariances completely, now significantly simplifying the originally derived solution of [18]. With

$$\mathbf{\Psi}^{S}(\ell) = (1-\beta)\cdot\tilde{\mathbf{E}}(\ell)\tilde{\mathbf{E}}^{H}(\ell) + \beta\cdot\mathbf{\Psi}^{S}(\ell-1) \tag{7}$$

and smoothing factor $\beta = 0.5$, we result in the computationally most efficient approach, since no further computation steps are included anymore.

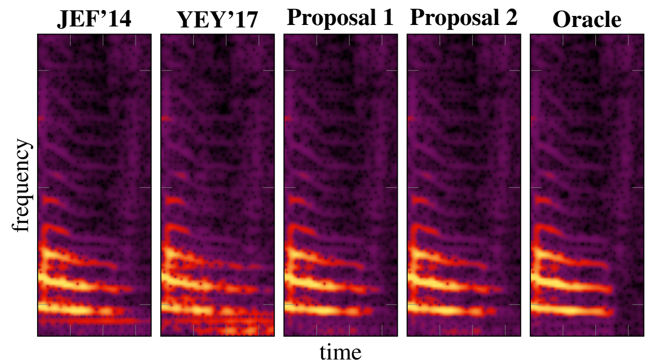

Fig. 2. Example spectrograms for the stable approaches: Section of the postfiltered enhanced signal $s_{\mathrm{PF}}(n)$. The excerpt is shown in the range from 0 to 2 kHz, with a duration of 0.5 s.

4. EXPERIMENTAL VALIDATION

The experimental validation is performed by deploying the FDAKF in different configurations (single- and multi-channel) in the practical context of an in-car communication system. To enhance communication between passengers inside a car cabin and to improve speech intelligibility for rear seat passengers, the driver's speech is amplified and reproduced in the rear of the car. The car cabin is modeled by one microphone and two loudspeakers (left/right) in the front, and two loudspeakers (left/right) in the rear. The four impulse responses (IRs) from loudspeakers to microphone are randomly generated with exponential energy decay and a car-typical reverberation time of $T_{60} = 50$ ms, and are cut off after 50 ms. Giving consideration to different distances and acoustic dampening between front and rear loudspeakers towards microphone, the two front IRs are multiplied with an empirically motivated factor 0.7 and the rear IRs with 0.3 [15].

Experiment	①		②		③	
	$\overline{\text{ERLE}}$	MOS	$\overline{\text{ERLE}}$	MOS	$\overline{\text{ERLE}}$	MOS
Oracle	13.30	3.62	9.81	1.42	16.01	2.69
ME'10	instable					
JEF'14	12.23	3.08	8.53	**1.35**	14.21	2.39
KME'14	12.13	3.02	instable			
YEY'17	10.88	2.73	**8.60**	1.33	14.12	2.31
Proposal 1	**12.69**	**3.17**	**8.59**	**1.35**	**14.76**	**2.46**
Proposal 2	**12.80**	**3.17**	8.54	**1.34**	**14.75**	**2.47**

Table 1. Mean ERLE and PESQ MOS on $s_{\text{PF}}(n)$. ① Mono FDAKF, only driver's speech. ② Mono FDAKF, driver's speech plus noise and additional music. ③ Stereo FDAKF, driver's speech plus noise and additional music. The two best approaches apart from the oracle are printed in **boldface** font.

The FDAKF system is set up to work with a sampling frequency of 16 kHz, DFT size $K=1024$, frame shift $R=64$, FDAKF forgetting factor $a=0.9995$, and the postfilter design parameter is set to $\lambda=2$. As shown in Figure 1 and done in [7], the postfilter is applied after the subtraction of the estimated feedback signals. Since low delay is usually essential for feedback cancellation, an asymmetric window structure based on [22] is used to obtain an overall delay of only 12 ms (typically underlying buffering structures included [15]): In each frame the most recent $4R$ samples of the FDAKF output $e(n)$ are multiplied with the analysis window and zero-padded to be subject to a K-point DFT. Now the K postfilter coefficients are applied, and the result is again subject to a K-point IDFT. Only the first $2R$ output samples are multiplied with the smaller synthesis window, and combined with the previous frame's output to yield the postfiltered enhanced signal $s_{\text{PF}}(n)$. The gain applied afterwards is set to fairly sufficient 12 dB.

The influence of all presented measurement noise covariance estimation rules will be investigated in three different experiments: ① Using a single-channel (*mono*) FDAKF while only driver's speech $s(n)$ is active and reamplified to the rear loudspeakers. ② Using a *mono* FDAKF and driver's speech as before, but with superimposed noise $n(n)$ at the microphone and stereo music as additional signal on all loudspeakers. ③ Considering the same scenario as ② , but deploying a *stereo*-channel FDAKF to validate the performance for more than one channel. For the stereo FDAKF both front and both rear channels are combined and multiplied with factor 0.5 to provide the FDAKF reference signals $x_j(n)$, for the mono case these are further combined to provide only one reference channel, respectively.

Intentionally, the first experiment ① is set up as simple as possible. By only considering driver's speech $s(n)$ (without noise at the microphone or additional signals on the loudspeakers), it is possible to focus solely on the speech quality within the feedback loop. In this setting ITU-T Recommendation P.501 [23, Secs. 7.3.5, 7.3.7] *short conditioning sequence II* followed by the *single-talk sequence* is used as driver's speech $s(n)$ at a level of -26 dBov. The results are given in the left part of Table 1. Note that convergence times are not explicitly listed, since they are nearly similar for all (stable) approaches. As first measure, the mean echo (and feedback) return loss enhancement (ERLE) is given in dB, and is computed over the entire sequence as done in [7, 24]. A higher value signifies more feedback suppression. The **oracle** estimation of $\mathbf{\Psi}^S$ is able to achieve a mean ERLE of 13.3 dB. The approach of **ME'10** is not capable to cancel the feedback sufficiently at this gain and results in an instable system. Therefore, no results can be given for the simple scenario and the approach will be omitted in the investigation of the more challenging scenarios. It can be seen that **JEF'14** and **KME'14** obtain similar ERLE values of about 12.2 dB, while **YEY'17** remains ca. 1.3 dB below that. In contrast, both **proposal** approaches obtain a higher ERLE value, ending up only ca. 0.5 and 0.6 dB below oracle performance, respectively.

As further measure, the recently updated wideband PESQ MOS LQO [25, 26] computed on the entire enhanced signal $s_{\text{PF}}(n)$ with $s(n)$ as reference is given (dubbed MOS). The PESQ results are in line with the mean ERLE values, revealing **YEY'17** with only 2.73 MOS points and **JEF'14** and **KME'14** more than 0.3 points above that. Both **proposal** approaches obtain an even about 0.1 points higher MOS score, thereby being closest to the **oracle** performance. In Figure 2, we provide an excerpt of the enhanced signal spectrograms to determine further differences. It can easily be seen that while the **oracle** speech component is rather clear and crisp, **JEF'14** and **YEY'17** are much more reverberant and distorted. In contrast, the **proposals** exhibit only minor differences, are much less reverberant and more similar to the **oracle** speech component. Note that the spectrogram of **KME'14** is not given here, since it is found to be instable in the next investigated scenario and therefore doesn't seem to be suitable for feedback applications.

In the second experiment ② , we now superimpose ITU-T P.501 signal *in-car noise* as $n(n)$ at an SNR of 15 dB with the driver's speech signal $s(n)$. Furthermore, we add stereo pop music at a challenging level of -26 dBov as left and right loudspeaker signals to interfere with the desired driver's speech and the re-amplified enhanced signal. The results are shown in the center of Table 1. As to be expected, the mean ERLE and PESQ values are generally lower than before. **KME'14** is not able to maintain a stable system, and will therefore be omitted in the last experiment. All other approaches obtain comparable instrumental results of around 8.55 dB ERLE and 1.34 MOS points. Interestingly, **YEY'17** now also obtains similar objective results as the proposed approaches. Though the spectrograms cannot be presented here for all experiments due to limited space, subjective impressions on the spectrograms reveal the same behaviours as for experiment ① , leaving **YEY'17** with a more distorted and both **proposals** with the most crisp speech component.

In the final experiment ③ we deploy the *stereo*-channel FDAKF (Section 2, $N=2$) in the same experimental setting as before. In general, this improves the performance massively. The subjective impressions on the spectrograms as well as the instrumental results given in the right of Table 1 confirm the findings of the previous experiments. While **JEF'14** now slightly outperforms **YEY'17** by 0.09 dB in mean ERLE and 0.08 points in PESQ MOS, both of our new **proposals** obtain a mean ERLE *and* PESQ MOS of more than 0.5 dB and ca. 0.17 MOS points, respectively, even above **JEF'14**.

5. CONCLUSIONS

In this paper, we investigated measurement noise covariance estimation methods for feedback cancellation based on the frequency domain adaptive Kalman filter. Considering the significant impact of the deployed estimation approach on the feedback cancellation performance, we summarized known methods from literature, applied them to the N-channel case, and provided two new proposals that are explicitly motivated for the use in acoustic *feedback* cancellation. The experimental validation was performed in the context of an in-car communication system. Instrumental measures as well as analysis of the spectrograms reveal a robust performance of the two proposals: They do not only obtain a much better speech quality compared to existing approaches, but also achieve a higher overall feedback suppression.

6. REFERENCES

[1] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, Upper Saddle River, NJ, USA, 4th edition, 2002.

[2] G. Enzner and P. Vary, "Frequency-Domain Adaptive Kalman Filter for Acoustic Echo Control in Hands-Free Telephones," *Signal Processing (Elsevier)*, vol. 86, no. 6, pp. 1140–1156, June 2006.

[3] S. Malik and G. Enzner, "Recursive Bayesian Control of Multichannel Acoustic Echo Cancellation," *IEEE Signal Processing Letters*, vol. 18, no. 11, pp. 619–622, Nov. 2011.

[4] F. Kuech, E. Mabande, and G. Enzner, "State-Space Architecture of the Partitioned-Block-Based Acoustic Echo Controller," in *Proc. of ICASSP*, Florence, Italy, May 2014, pp. 1295–1299.

[5] M. A. Jung, S. Elshamy, and T. Fingscheidt, "An Automotive Wideband Stereo Acoustic Echo Canceler Using Frequency-Domain Adaptive Filtering," in *Proc. of EUSIPCO*, Lisbon, Portugal, Sept. 2014, pp. 1452–1456.

[6] M. L. Valero, E. Mabande, and E. A. P. Habets, "A State-Space Partitioned-Block Adaptive Filter for Echo Cancellation Using Inter-Band Correlations in the Kalman Gain Computation," in *Proc. of ICASSP*, Brisbane, QLD, Australia, Apr. 2015, pp. 599–603.

[7] J. Franzen and T. Fingscheidt, "An Efficient Residual Echo Suppression for Multi-Channel Acoustic Echo Cancellation Based on the Frequency-Domain Adaptive Kalman Filter," in *Proc. of ICASSP*, Calgary, AB, Canada, Apr. 2018, pp. 226–230.

[8] H. Shin, A. H. Sayed, and W. Song, "Variable Step-Size NLMS and Affine Projection Algorithms," *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 132–135, Feb. 2004.

[9] F. Strasser and H. Puder, "Adaptive Feedback Cancellation for Realistic Hearing Aid Applications," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 12, pp. 2322–2333, Dec. 2015.

[10] S. Pradhan, V. Patel, D. Somani, and N. V. George, "An Improved Proportionate Delayless Multiband-Structured Subband Adaptive Feedback Canceller for Digital Hearing Aids," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 8, pp. 1633–1643, Aug. 2017.

[11] E. Lleida, E. Masgrau, and A. Ortega, "Acoustic Echo Control and Noise Reduction for Cabin Car Communication," in *Proc. of EUROSPEECH*, Aalborg, Denmark, Sept. 2001, vol. 3, pp. 1585–1588.

[12] P. Bulling, K. Linhard, A. Wolf, and G. Schmidt, "Stepsize Control for Acoustic Feedback Cancellation Based on the Detection of Reverberant Signal Periods and the Estimated System Distance," in *Proc. of INTERSPEECH*, Stockholm, Sweden, Aug. 2017, pp. 176–180.

[13] J. Franzen and T. Fingscheidt, "A Delay-Flexible Stereo Acoustic Echo Cancellation for DFT-Based In-Car Communication (ICC) Systems," in *Proc. of INTERSPEECH*, Stockholm, Sweden, Aug. 2017, pp. 181–185.

[14] F. Albu, L. T. T. Tran, and S. Nordholm, "The Hybrid Simplified Kalman Filter for Adaptive Feedback Cancellation," in *Proc. of International Conference on Communications*, Bucharest, Romania, June 2018, pp. 45–50.

[15] J. Franzen, I. Meyer zum Alten Borgloh, and T. Fingscheidt, "On the Benefit of a Stereo Acoustic Echo Cancellation in an In-Car Communication System," in *Proc. of 13. ITG Symposium Speech Communication*, Oldenburg, Germany, Oct. 2018, pp. 41–45.

[16] G. Bernardi, T. van Waterschoot, J. Wouters, and M. Moonen, "Adaptive Feedback Cancellation Using a Partitioned-Block Frequency-Domain Kalman Filter Approach With PEM-Based Signal Prewhitening," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 9, pp. 1784–1798, Sept. 2017.

[17] R. Marchthaler and S. Dingler, *Kalman-Filter: Einführung in die Zustandsschätzung und ihre Anwendung für eingebettete Systeme* (in German), Springer Vieweg, Wiesbaden, Germany, 2017.

[18] S. Malik and G. Enzner, "Online Maximum-Likelihood Learning of Time-Varying Dynamical Models in Block-Frequency-Domain," in *Proc. of ICASSP*, Dallas, TX, USA, Mar. 2010, pp. 3822–3825.

[19] S. Malik and J. Benesty, "Variationally Diagonalized Multichannel State-Space Frequency-Domain Adaptive Filtering for Acoustic Echo Cancellation," in *Proc. of ICASSP*, Vancouver, BC, Canada, May 2013, pp. 595–599.

[20] S. Malik, *Bayesian Learning of Linear and Nonlinear Acoustic System Models in Hands-free Communication*, Ph.D. thesis, Institute of Communication Acoustics, Ruhr-Universität Bochum, Bochum, Germany, Oct. 2012.

[21] F. Yang, G. Enzner, and J. Yang, "Statistical Convergence Analysis for Optimal Control of DFT-Domain Adaptive Echo Canceler," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 5, pp. 1095–1106, May 2017.

[22] C. Lüke, G. Schmidt, A. Theiß, and J. Withopf, "In-Car Communication," in *Smart Mobile In-Vehicle Systems – Next Generation Advancements*, G. Schmidt et al., Ed., pp. 97–118. Springer-Verlag, 2014.

[23] "ITU-T Recommendation P.501, Test signals for use in telephonometry," ITU, Jan. 2012.

[24] M.-A. Jung and T. Fingscheidt, "A Shadow Filter Approach to a Wideband FDAF-Based Automotive Handsfree System," in *5th Biennial Workshop on DSP for In-Vehicle Systems*, Kiel, Germany, Sept. 2011, pp. 60–67.

[25] "ITU-T Recommendation P.862.2, Wideband Extension to Recommendation P.862 for the Assessment of Wideband Telephone Networks and Speech Codecs," ITU, Nov. 2007.

[26] "ITU-T Recommendation P.862.2 Corrigendum 1, Wideband Extension to Recommendation P.862 for the Assessment of Wideband Telephone Networks and Speech Codecs," ITU, Oct. 2017.