# Maximum Likelihood Measurement Noise Estimation for Block-Time Domain Kalman Filters

*Till Hardenbicker, Jan Schneider, and Peter Jax*
Institute of Communication Systems, RWTH Aachen University, Germany
{hardenbicker, jax}@iks.rwth-aachen.de

***Abstract*—The performance of Kalman filters in acoustic system identification depends on an accurate estimation of the (co)variance of the measurement noise. When no prior knowledge on the noise is available, it needs to be estimated online, which is well researched for the Frequency Domain Kalman Filter (FDKF). Recent publications however, rely on the fast convergence and tracking performance of the Block-Time Domain Kalman Filter (BTKF). Hence, this paper addresses the online estimation of the measurement noise covariance for the BTKF. Although online maximum likelihood estimation relies on the error vector's outer product, a Toeplitz structure is imposed onto the noise covariance matrix which makes its inversion more stable and is motivated by short-time stationarity of the measurement noise. An additional onset detection mechanism enhances robustness against non-stationary near-end noise levels. Our exemplary evaluation on using real-world test signals from the ICASSP Acoustic Echo Cancellation challenge data set demonstrates that the proposed approach shows the same steady-state performance like the FDKF but allows to use the BTKFs faster convergence. Compared to the straightforward maximum likelihood approach, the performance is significantly improved.**

***Index Terms*—Kalman filter, Maximum likelihood, Measurement noise**

## I. Introduction

Tracking an acoustic system is a common task in digital signal processing. It refers to the identification of the parameters of an unknown system when only the input to the system and a noisy observation of its output are available. As a distinction to the more general case of acoustic system identification, tracking refers to the case that the system is time-variant. In most applications, the system is modeled as a Finite Impulse Response (FIR) filter, such that the parameters to be identified are the filters' coefficients. In Acoustic Echo Cancellation (AEC) applications, this filter represents the acoustic path between the loudspeaker and a microphone of a hands-free communication device. Tracking this filter allows to remove echo components in the microphone signal and thus at the far-end side [1]–[4]. In the application of acoustic feedback cancellation, the signal recorded by the microphone is played back in the same enclosure [5], [6]. In the field of Active Noise Cancellation (ANC) in hearables, acoustic tracking can be used to identify the so-called secondary path between the hearable's loudspeaker and its error microphone [7]–[9]. The filter coefficients of this path are then used for the design of cancellation filters. A similar principle is applied in Crosstalk Cancellation (CTC). Here, the loudspeakers are anywhere in the same room, and the error microphones are again at the ears of a listener. Like with ANC, the path between speaker and ear is tracked to design filters with which a target signal at the ears can be obtained [10], [11]. Lastly, a similar setup can be used for fast measurements of Head-Related Transfer Functions (HRTFs). When the acoustic paths between a loudspeaker and ear microphones are tracked, the listeners HRTFs are obtained with a high spatial resolution [12], [13].

One solution to the tracking problem is given by adaptive filters. Regarding the various applications, there exist many realizations of adaptive filters, all of which show different performances and complexities [14]. The (Normalized) Least Mean Squares (NLMS) algorithm at one end requires only few operations for its filter update [15]–[17], making it suitable for the application in battery-powered devices like hearables [7]–[9]. Its performance is, however, limited due to its scalar step size. At the other end, the Kalman Filter (KF) combines statistics of the measurement noise, statistics of the system changes and possibly prior knowledge about the filter coefficients' distribution, to replace the scalar step size by an optimal step-size matrix [18]. A popular implementation of the KF is given by the Frequency Domain Kalman Filter (FDKF), which usually assumes all covariance matrices to be diagonal in the frequency domain [3]. Although this significantly increases the algorithm's efficiency and is meaningful for the measurement noise covariance, some applications benefit when this assumption is alleviated for the state error covariance. In the example of fast HRTF measurements, prior knowledge on typical HRTF's properties can be incorporated into the non-diagonal state error covariance to improve the adaptation. To leverage the computational burden of the KF in the time domain, it can be formulated in block-time domain, such that multiple microphone signal samples are processed at once [19]. A recently proposed class of adaptive filters performs the filter update in a latent space with a dimension lower than the number of filter coefficients, yielding a non diagonal state covariance in the original space [20], [21]. These filters also exploit the Block-Time Domain Kalman Filter (BTKF).

Although the KF solves the problem of finding an adaptive step size (matrix) by incorporating the second-order statistics of a state-space system, it merely shifts the problem from estimating a step size to estimating these statistics. This includes the covariance of measurement and process noise. While both have been well understood in the context of the FDKF [4], [6], [22], the BTKF has been studied less. Measurement noise estimators [17], [23], [24] and a process noise estimator [25]

 In *International Conference on Acoustics, Speech and Signal Processing*, May. 2026, pp 16017–16021,
DOI: 10.1109/ICASSP55912.2026.11464581

exist for adaptive filters without block processing. In this case, the estimation problem reduces to estimating a scalar variance. While the process noise estimator can be applied without modification to the BTKF, to the authors' knowledge, there is yet no established estimator for the measurement noise covariance matrix.

The aim of this paper is to develop an online estimation procedure for the measurement noise covariance matrix in a BTKF. Section II recaps the signal model and the resulting formulation of the BTKF. In Section III, the principle of online maximum likelihood learning is applied to the problem at hand. While this approach yields a unique solution, it is possible to impose different structures on the resulting covariance matrix. The effect of these choices is evaluated in Section IV.

## II. Kalman Filter in Block-Time Domain

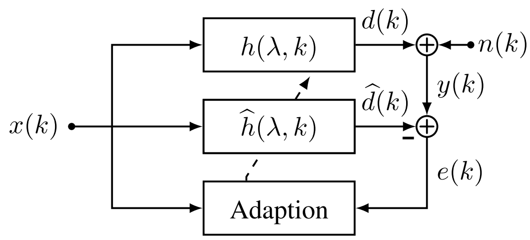


**Fig. 1:** Adaptive system identification

Throughout this paper, bold letters denote column vectors and bold capital letters denote matrices. The element of a matrix $\mathbf{Q}$ in row $i$ and column $j$ is denoted by $\left[\mathbf{Q}\right]_{i,j}$, and similarly $\left[\mathbf{q}\right]_i$ is the $i^{\text{th}}$ element in a vector $\mathbf{q}$. A Gaussian distribution is denoted by $\mathcal{N}$ and $\mathbb{E}$ denotes the expectation operator. An identity matrix is denoted by $\mathbf{I}$. Fig. 1 depicts the signal model in discrete time. The excitation signal $x(k)$ is convolved with a time-variant impulse response $h(\lambda, k)$ and superimposed with noise $n(k)$ to yield the microphone signal

$$y(k) = x(k) * h(\lambda, k) + n(k) \tag{1}$$
$$= \sum_{\lambda=0}^{l-1} x(k-\lambda) \cdot h(\lambda, k) \quad + n(k) \ . \tag{2}$$

Here, $\lambda$ is the index of the filter tap, while $k$ is the global time index. The aim of the filter adaptation is to identify $h(\lambda, k)$ only based on $x(k)$ and $y(k)$. To allow for block-wise processing instead of processing each sample of $y(k)$ separately, we formulate

$$\mathbf{y}_m = \left[y(mr{-}r{+}1)\,,\, y(mr{-}r{+}2)\,,\, \dots\, y(mr)\right]^{\mathrm{T}} \tag{3}$$
$$\mathbf{h}_m \approx \left[h(0, mr)\,,\, h(1, mr)\,,\, \dots\, h(l{-}1, mr)\right]^{\mathrm{T}} \tag{4}$$
$$\mathbf{x}_k = \left[x(k)\,,\, x(k{-}1)\,,\, \dots\, x(k{-}l{+}1)\right]^{\mathrm{T}} \tag{5}$$
$$\mathbf{X}_m = \left[\mathbf{x}_{mr-r+1},\, \mathbf{x}_{mr-r+2},\, \dots \mathbf{x}_{mr}\right]^{\mathrm{T}} \tag{6}$$

where $r$ is the frame shift, $m$ is the block index, $l$ is the length of the modeled impulse response and the matrix $\mathbf{X}_m$ is the convolution matrix constructed from the excitation signal. These definitions allow to formulate the state space system

$$\mathbf{y}_m = \mathbf{X}_m \mathbf{h}_m + \mathbf{n}_m \tag{7}$$
$$\mathbf{h}_{m+1} = \gamma \mathbf{h}_m + \boldsymbol{\delta}_m \ . \tag{8}$$

The desired impulse response vector $\mathbf{h}_m$ is the unknown state of the system. The observation equation (7) is a re-formulation of (1) where $\mathbf{n}_m$ is defined similar to $\mathbf{y}_m$. The state transition equation (8) is commonly modeled as a Markov model with a fading factor $\gamma$ and process noise $\boldsymbol{\delta}$ [3].

The Kalman filter aims to track the state of this state space system. The time update describes the state transition equation:

$$\widehat{\mathbf{h}}_m^- = \gamma \widehat{\mathbf{h}}_{m-1}^+ \tag{9a}$$
$$\mathbf{P}_m^- = \gamma^2 \mathbf{P}_{m-1}^+ + \mathbf{Q}_{\boldsymbol{\delta},m} \ . \tag{9b}$$

Here, $\mathbf{P}$ is the state error covariance and $\mathbf{Q}_{\boldsymbol{\delta},m}$ is the process noise covariance. The superscripts $(\cdot)^-$ and $(\cdot)^+$ denote prior and posterior estimates, respectively. The measurement update reflects the observation equation and reads

$$\mathbf{K}_m = \mathbf{P}_m^- \mathbf{X}_m^{\mathrm{T}} \left(\mathbf{X}_m \mathbf{P}_m^- \mathbf{X}_m^{\mathrm{T}} + \mathbf{Q}_{\mathbf{n},m}\right)^{-1} \tag{10a}$$
$$\mathbf{e}_m = \mathbf{y}_m - \mathbf{X}_m \widehat{\mathbf{h}}_m^- \tag{10b}$$
$$\widehat{\mathbf{h}}_m^+ = \widehat{\mathbf{h}}_m^- + \mathbf{K}_m \mathbf{e}_m \tag{10c}$$
$$\mathbf{P}_m^+ = \left(\mathbf{I}_l - \mathbf{K}_m \mathbf{X}_m\right) \mathbf{P}_m^- \ . \tag{10d}$$

Here, $\mathbf{Q}_{\mathbf{n},m}$ denotes the measurement noise covariance. The Kalman gain matrix $\mathbf{K}$ acts as an optimal step size matrix when the second order moments $\mathbf{P}_m$ and $\mathbf{Q}_{\mathbf{n},m}$ are known [18]. While $\mathbf{P}_m$ is tracked by the KF itself, $\mathbf{Q}_{\mathbf{n},m}$ needs to be set explicitly. In the scenario of an HRTF measurement, where $n(k)$ is mostly microphone noise, $\mathbf{Q}_{\mathbf{n},m}$ can be estimated from a calibration measurement. In the AEC scenario, however, the measurement noise originates from the near-end speaker such that the covariance has to be estimated in real time. Otherwise, when the noise covariance is underestimated, the elements in the step size will be too high, such that the filter erroneously considers the measurement noise to originate from the unknown system. In turn, when the noise covariance is overestimated, the step size is reduced and so is the filter's convergence speed. In both cases, the adaptive filter fails to estimate the impulse response correctly and residual echoes will remain.

## III. Online Maximum Likelihood Estimation of the Measurement Noise

The objective of maximum likelihood estimation is to find the parameters of a probability distribution which maximize the likelihood of a set of observations. A detailed derivation for the FDKF is given in [22] while [26] gives a domain-independent explanation which does not consider time-varying covariances.

The goal of this publication is to estimate the covariance matrix $\mathbf{Q}_{\mathbf{n},m}$ of the measurement noise such that the likelihood $p(\mathbf{y}_m | \mathbf{h}_m, \mathbf{X}_m)$ of the observations $\mathbf{y}_m$ is maximized [22],

 In *International Conference on Acoustics, Speech and Signal Processing*, May. 2026, pp 16017–16021,
DOI: 10.1109/ICASSP55912.2026.11464581

[26]. Usually, the log-likelihood is considered, which for the measurement noise reads

$$\mathcal{L}(\widehat{\mathbf{Q}}_{\mathbf{n},m}) \stackrel{c}{=} -\frac{1}{2}\log\left|\widehat{\mathbf{Q}}_{\mathbf{n},m}\right| - \frac{1}{2}\mathbb{E}_{\mathbf{h}_m}\left\{\mathbf{e}_m\widehat{\mathbf{Q}}_{\mathbf{n},m}^{-1}\mathbf{e}_m^{\mathrm{T}}\right\} \quad (11)$$

where $p(\mathbf{y}_m|\mathbf{h}_m, \mathbf{X}_m) = \mathcal{N}(\mathbf{y}_m|\mathbf{X}_m\mathbf{h}_m, \mathbf{Q}_{\mathbf{n},m})$ is used. The expectation in (11) arises because $\mathbf{h}_m$ is not known precisely, only that it is distributed as $\mathbf{h}_m \sim \mathcal{N}(\mathbf{h}_m|\widehat{\mathbf{h}}_m^-, \mathbf{P}_m^-)$. This distribution can analogously be formulated using the posterior estimates $\widehat{\mathbf{h}}_m^+$ and $\widehat{\mathbf{P}}_m^+$ [22], which is avoided here for two reasons: First, $\widehat{\mathbf{Q}}_{\mathbf{n},m}$ is used to compute the Kalman gain matrix, where the posterior estimates are not yet available. Secondly, preliminary experiments have shown that using the posterior estimates does not alter the metrics significantly. Finally, maximizing (11) with respect to $\widehat{\mathbf{Q}}_{\mathbf{n},m}$ yields

$$\widehat{\mathbf{Q}}_{\mathbf{n},m} = \mathbf{e}_m\mathbf{e}_m^{\mathrm{T}} + \mathbf{X}_m\mathbf{P}^-\mathbf{X}_m^{\mathrm{T}} \quad (12)$$

$$= \mathbf{Q}_{\mathbf{e},m} + \mathbf{X}_m\mathbf{P}^-\mathbf{X}_m^{\mathrm{T}} \ . \quad (13)$$

Here, $\mathbf{e}_m\mathbf{e}_m^{\mathrm{T}} =: \mathbf{Q}_{\mathbf{e},m}$ is a (biased) covariance estimate of the error vector $\mathbf{e}_m$. The second addend in (12) originates from the expectation with respect to the distribution of $\mathbf{h}_m$ and is sometimes discarded [6], but not in this paper.

Although (12) is the closed-form solution to the maximum likelihood approach, it performs suboptimal in practical applications, since the outer product $\mathbf{e}_m\mathbf{e}_m^{\mathrm{T}}$ aims to estimate an $r \times r$ covariance matrix, based on only $r$ new observations, making it an under-determined problem. As a result $\widehat{\mathbf{Q}}_{\mathbf{n},m}$ will have low eigenvalues, making its inversion unstable. To circumvent this problem, one can take a look at the FDKF, which renders all covariance matrices diagonal in frequency domain. Regarding $\widehat{\mathbf{Q}}_{\mathbf{n},m}$, this assumption holds when the signal $n(k)$ is a wide-sense stationary random process [27]. In practical situations, where only short-time stationarity can be assumed, $\widehat{\mathbf{Q}}_{\mathbf{n},m}$ is approximately diagonal in frequency domain when a short time interval of $n(k)$ is considered. In the time domain short-time stationarity means that $\mathbb{E}\{n(i)\, n(j)\} = \mathbb{E}\{n(i \pm \lambda)\, n(j \pm \lambda)\}$ when $\lambda$ is less than the time span in which stationarity holds. For the estimate of the vector covariance this means that $\left[\mathbf{Q}_{\mathbf{e},m}\right]_{i,j} = \left[\mathbf{Q}_{\mathbf{e},m}\right]_{i\pm\lambda,j\pm\lambda}$ which is the definition of a symmetric Toeplitz matrix. Such a matrix is fully occupied but has only $r$ independent variables. This structure is applied to $\mathbf{Q}_{\mathbf{e},m}$, instead of using the outer product $\mathbf{e}_m\mathbf{e}_m^{\mathrm{T}}$. To do so, we define the auxiliary signals

$$e^*_{\Delta k}(k) = e(k)\, e(k - \Delta k) \quad (14)$$

for $0 \leq \Delta k < r$ and their recursively smoothed versions

$$\tilde{e}_{\Delta k}(k) = \alpha_{\text{fast}}\tilde{e}_{\Delta k}(k-1) + (1 - \alpha_{\text{fast}})e^*_{\Delta k}(k) \ . \quad (15)$$

Assuming $e(k)$ to be ergodic, we can set

$$\left[\mathbf{Q}_{\mathbf{e},m}\right]_{i,j} = \mathbb{E}\{n(i)\, n(j)\} = \tilde{e}_{|i-j|}(mr) \ . \quad (16)$$

It is common practice to average (12) along the time axis, which is usually done by computing

$$\widehat{\widehat{\mathbf{Q}}}_{\mathbf{n},m} = \alpha_{\text{slow}}\widehat{\widehat{\mathbf{Q}}}_{\mathbf{n},m-1} + (1 - \alpha_{\text{slow}})\widehat{\mathbf{Q}}_{\mathbf{n},m} \quad (17)$$

and using $\widehat{\widehat{\mathbf{Q}}}_{\mathbf{n},m}$ instead of $\widehat{\mathbf{Q}}_{\mathbf{n},m}$ in (10a) [4], [6]. This technique is adopted here, even though temporal smoothing has already been performed in (15), since the smoothing is only applied to $\mathbf{Q}_{\mathbf{e},m}$ and not to the entire $\widehat{\mathbf{Q}}_{\mathbf{n},m}$. In (15) a lower (faster) time constant $\alpha_{\text{fast}}$ is used than $\alpha_{\text{slow}}$ in (17).

Although the temporal smoothing may improve the accuracy of the covariance estimation, it can be detrimental when there are sudden increases in the near-end noise level, such as impulsive noise or speech onsets. In this case, $\widehat{\mathbf{Q}}_{\mathbf{n},m}$ is first underestimated, such that the entries of the Kalman gain matrix are overestimated and the noisy observation is weighted too high in the adaptive filter update. Next, the estimated impulse response $\widehat{\mathbf{h}}_m^+$ is erroneous and the situation changes to the opposite: The level of the error signal rises as well as the trace of $\widehat{\mathbf{Q}}_{\mathbf{n},m}$ such that the Kalman gain is now underestimated and the KF needs time to re-converge from the faulty estimate. In [24], a conservative estimator was proposed for $r = 1$ (scalar measurement noise variance). This mechanism can be generalized for $r > 1$ as follows:

$$\big[\widehat{\widehat{\mathbf{Q}}}_{\mathbf{n},m}\big]_{i,i} \Leftarrow \begin{cases} \big[\widehat{\mathbf{Q}}_{\mathbf{n},m}\big]_{i,i} & \big[\widehat{\mathbf{Q}}_{\mathbf{n},m}\big]_{i,i} > \theta\big[\widehat{\widehat{\mathbf{Q}}}_{\mathbf{n},m}\big]_{i,i} \\ \big[\widehat{\widehat{\mathbf{Q}}}_{\mathbf{n},m}\big]_{i,i} & \text{else.} \end{cases} \quad (18)$$

With this mechanism, the smoother reacts instantly and overwrites the diagonal of $\widehat{\widehat{\mathbf{Q}}}_{\mathbf{n},m}$, when the novel estimate exceeds the smoothed estimate by the threshold $\theta \geq 1$.

## IV. Evaluation

This section evaluates the proposed method in the context of an AEC application. As test signals, the `test real` subset from the ICASSP AEC data set [28] is used. All signals are sampled with $f_{\mathrm{s}} = 16\,\mathrm{kHz}$. To be able to compute the Echo Return Loss Enhancement (ERLE), the recordings with far-end single talk are used such that a recording of the echo signal $d(k)$ is available with only few disturbances. Some of the recordings contain a low level of environmental noise, disturbances when the device is moved, and loudspeaker nonlinearities. To avoid the problem of delay estimation, all microphone signals are pre-aligned to the excitation signals with a margin of 80 samples. Each recording is mixed with speech and environmental noise. The speech is taken from the synthetic subset of the same data set. In these recordings, the near-end speech is not present all the time, such that there are periods of single talk and double talk. The noise signals are taken from the ETSI database [29], including both instationary noise types like cafeteria and kindergarten, and stationary noise types like car or aircraft. Binaural recordings are averaged to obtain mono signals. We added the noise signals directly to the echo recording and did not aim to obtain a specific echo-to-noise ratio. After presenting the resulting signals to the KF under test, the echo return loss enhancement

$$\mathrm{ERLE} = 10\log_{10}\frac{\mathbb{E}_k\{d^2(k)\}}{\mathbb{E}_k\{(d(k) - \widehat{d}(k))^2\}\ .} \quad (19)$$

is evaluated. The expectations are estimated using recursive smoothing.


DOI: 10.1109/ICASSP55912.2026.11464581

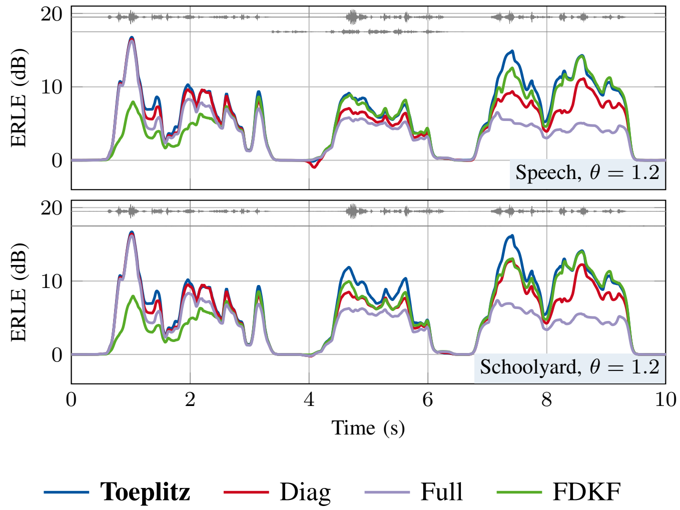


**Fig. 2:** Course of the ERLE for speech and noise. The upper gray audio envelope visualizes the excitation $x(k)$, and the lower envelope visualizes the near-end speech or noise $n(k)$, respectively.

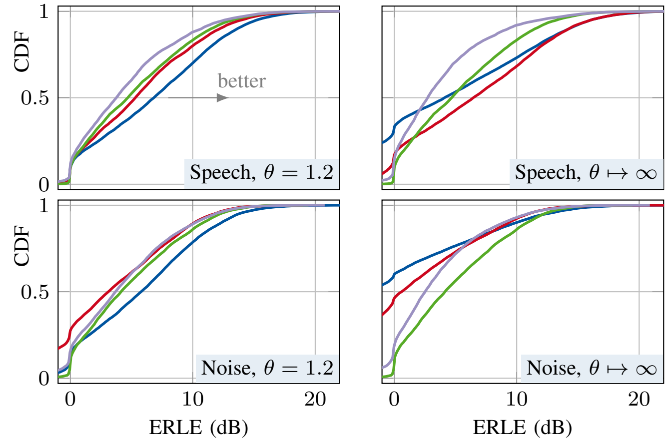


**Fig. 3:** CDF of the aggregated ERLE

To validate the performance of the proposed estimator (16), the following baselines are considered:

$$\texttt{Full} \quad \mathbf{Q}_{\mathbf{e},m} = \mathbf{e}_m^{\mathrm{T}}\mathbf{e}_m$$
$$\texttt{Diag} \quad \left[\mathbf{Q}_{\mathbf{e},m}\right]_{i,i} = \left[\mathbf{e}_m\right]_i^2$$

These options do not assume stationarity of the signal explicitly. Instead, each of the $r$ elements in $\mathbf{e}_m$ is considered a separate sensor signal. Using the option `Full`, any covariance between these signals is considered. As an established reference, we consider the FDKF using the measurement noise estimator of [22] and without post filter. Other than in the original publication we use temporal smoothing and the prior error vector in line with the proposed estimator. These modifications were also used in [4]. For all variants, the onset detection mechanism with $\theta = 1.2$ is used, except for the FDKF, which performed best without this mechanism. For sample-wise smoothing in (15) $\alpha_{\text{fast}} = \mathrm{e}^{-1/(0.02\,\mathrm{s}\cdot f_s)} \approx 0.997$ is used while $\alpha_{\text{slow}} = \mathrm{e}^{-r/(0.1\,\mathrm{s}\cdot f_s)} \approx 0.961$ is used for frame-wise smoothing in (17). For all algorithms, we set $r = 64$ ($4\,\mathrm{ms}$), $l = 1000$ ($62.5\,\mathrm{ms}$) and $\gamma = \mathrm{e}^{-r/(8\,\mathrm{s}\cdot f_s)} \approx 0.9995$. The (smoothed) measurement noise covariance is initialized with $\mathbf{I}$ in the time domain and $r \cdot \mathrm{diag}(\mathbf{I})$ in the frequency domain. To estimate the process noise [25] is used in the time domain and [22] in the frequency domain, both smoothed with $\alpha_{\text{slow}}$. For the filters in the time domain, $\mathbf{P}_0$ is initialized heuristically as $\mathrm{diag}[10^{-1.2}, \ldots 10^{-3.7}]$ which accounts for a reverberation time of $0.2\,\mathrm{s}$. For the FDKF, chirp recordings in the training set are used to extract impulse responses and compute their diagonal covariance in the frequency domain. Although optimal, this procedure is not applicable to the time domain, since all test signals are pre-aligned to have the same delay.

Fig. 2 shows the course of the ERLE for near-end speech and near-end noise in one exemplary test scenario. The results show that the proposed estimator (——, Toeplitz) consistently shows a good performance, especially after the first pause of the far-end speaker, when there was time to properly observe the noise signal, e.g. after $4\,\mathrm{s}$ in the lower part of Fig. 2. The `Full` estimator (——) starts well since it is initialized with a diagonal noise covariance. However, this structure fades away as soon as there is significant near-end noise, and the performance deteriorates thereafter. The FDKF (——) together with its measurement noise estimator performs as well, or better than the proposed algorithm, but the BTKF converges much faster, initially, due to the fully-occupied state error covariance.

Since Fig. 2 shows only one echo signal and two noise types, the Cumulative Density Function (CDF) of the ERLE of all test signals is considered in Fig. 3. Frames without voice activity in the echo signal are discarded. The medians and the quantiles of the distributions can be determined by the curves' intersections with the corresponding y-values. When the ERLE is higher on average, the curves are shifted to the right. The results reveal that the proposed measurement noise estimator outperforms the reference approaches by around $2\,\mathrm{dB}$ in terms of ERLE, when the threshold $\theta$ is set to $1.2$. This holds even for near-end speech, which is non-Gaussian. The right column reveals that the proposed method and the `Diag` method strongly rely on the onset threshold. When that mechanism is not applied, more than 50% of the ERLE values are below $0\,\mathrm{dB}$. When tuning $\theta$ for this experiment, the estimators performed well for $1 \leq \theta \leq 1.5$.

## V. Summary

In this paper, we have proposed a method to estimate the measurement noise in a block time domain Kalman filter. We derived an online estimator using the method of maximum likelihood. The result relies on the outer product of the current error vector. By assuming short-time stationarity of the measurement noise, we replaced this expression with a Toeplitz matrix. To be robust against non-stationary noise levels, we adapted an existing technique for the case $r > 1$. The proposed estimator allows to use the appealing convergence speed of the Kalman filter in block-time domain in situations with varying near-end noise levels.

© IEEE. In *International Conference on Acoustics, Speech and Signal Processing*, May. 2026, pp 16017–16021,
DOI: 10.1109/ICASSP55912.2026.11464581

## REFERENCES


[1] M. M. Sondhi, D. R. Morgan, and J. L. Hall, "Stereophonic acoustic echo cancellation - An overview of the fundamental problem," *IEEE Signal Processing Letters*, vol. 2, no. 8, pp. 148–151, 1995.

[2] G. Schmidt, "Applications of acoustic echo control - an overview," in *European Signal Processing Conf.*, 2004, pp. 9–16.

[3] G. Enzner and P. Vary, "Frequency-domain adaptive Kalman filter for acoustic echo control in hands-free telephones," *Signal Processing*, vol. 86, no. 6, pp. 1140–1156, Jun. 2006.

[4] M.-A. Jung, S. Elshamy, and T. Fingscheidt, "An automotive wideband stereo acoustic echo canceler using frequency-domain adaptive filtering," in *European Signal Processing Conf.*, 2014, pp. 1452–1456.

[5] T. Van Waterschoot and M. Moonen, "Fifty years of acoustic feedback control: State of the art and future challenges," *Proceedings of the IEEE*, vol. 99, no. 2, pp. 288–327, 2010.

[6] J. Franzen and T. Fingscheidt, "Improved measurement noise covariance estimation for n-channel feedback cancellation based on the frequency domain adaptive Kalman filter," in *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, 2019, pp. 965–969.

[7] S. Kuo and D. Morgan, "Active noise control: a tutorial review," *Proceedings of the IEEE*, vol. 87, no. 6, pp. 943–973, 1999.

[8] M. Zhang, H. Lan, and W. Ser, "Cross-updated active noise control system with online secondary path modeling," *IEEE Transactions on speech and audio processing*, vol. 9, no. 5, pp. 598–602, 2001.

[9] J. Fabry, F. König, S. Liebich, and P. Jax, "Online secondary path estimation with masked auxiliary noise for active noise control," in *Int. Workshop on Acoustic Echo and Noise Control*, Sep. 2018, pp. 331–335.

[10] B. Masiero and M. Vorländer, "A framework for the calculation of dynamic crosstalk cancellation filters," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 9, pp. 1345–1354, 2014.

[11] T. Kabzinski and P. Jax, "An adaptive crosstalk cancellation system using microphones at the ears," in *Audio Engineering Society Convention*, Oct 2019.

[12] G. Enzner, "Analysis and optimal control of LMS-type adaptive filtering for continuous-azimuth acquisition of head related impulse responses," in *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Mar. 2008, pp. 393–396.

[13] T. Kabzinski and P. Jax, "Towards faster continuous multi-channel hrtf measurements based on learning system models," in *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, May 2022, pp. 436–440.

[14] S. Haykin, *Adaptive Filter Theory*, 5th ed. Pearson Education Ltd., 2014.

[15] S. Makino, Y. Kaneda, and N. Koizumi, "Exponentially weighted stepsize NLMS adaptive filter based on the statistics of a room impulse response," in *Transactions on Speech and Audio Processing*, vol. 1, no. 1, 1993, pp. 101–108.

[16] C. Antweiler, J. Grunwald, and H. Quack, "Approximation of optimal step size control for acoustic echo cancellation," in *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, vol. 1, 1997, pp. 295–298.

[17] T. Strutz, "Estimation of measurement-noise variance for variable-step-size NLMS filters," in *European Signal Processing Conf.*, 2019, pp. 1–5.

[18] D. P. Mandic, S. Kanna, and A. G. Constantinides, "On the intrinsic relationship between the least mean square and Kalman filters [lecture notes]," *IEEE Signal Processing Mag.*, vol. 32, no. 6, pp. 117–122, Nov. 2015.

[19] T. Kabzinski and P. Jax, "A unified perspective on time-domain and frequency-domain Kalman filters for acoustic system identification," in *European Signal Processing Conf.*, Aug. 2022, pp. 90–94.

[20] T. Hardenbicker and P. Jax, "Online system identification on learned acoustic manifolds using an extended Kalman filter," in *Int. Workshop on Acoustic Echo and Noise Control*, 2024, pp. 339–343.

[21] T. Hardenbicker, J. Hahn, and P. Jax, "Network architectures for manifold learning in MIMO acoustic system identification," in *European Signal Processing Conf.*, Sep. 2025.

[22] S. Malik and G. Enzner, "Online maximum-likelihood learning of time-varying dynamical models in block-frequency-domain," in *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, 2010, pp. 3822–3825.

[23] S. Liebich, J. Fabry, P. Jax, and P. Vary, "Time-domain Kalman filter for active noise cancellation headphones," in *European Signal Processing Conf.*, 2017, pp. 593–597.

[24] J. Fabry, "Least-squares methods for individualization of hearables with active noise cancellation," Ph.D. dissertation, IKS, RWTH Aachen University, 2024.

[25] C. Paleologu, J. Benesty, S. Ciochina, and S. L. Grant, "A Kalman filter with individual control factors for echo cancellation," in *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, May 2014, pp. 5974–5978.

[26] C. M. Bishop, *Pattern recognition and machine learning*. Springer, 2006.

[27] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*, 5th ed. John Wiley & Sons, 2015.

[28] K. Sridhar, R. Cutler, A. Saabas, T. Parnamaa, M. Loide, H. Gamper, S. Braun, R. Aichner, and S. Srinivasan, "ICASSP 2021 acoustic echo cancellation challenge: Datasets, testing framework, and results," in *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Jun. 2021, pp. 151–155.

[29] ETSI EG 202 396-1, "Speech quality performance in the presence of background noise," Oct. 2010.



DOI: 10.1109/ICASSP55912.2026.11464581