INTERSPEECH 2017
August 20–24, 2017, Stockholm, Sweden

A Delay-Flexible Stereo Acoustic Echo Cancellation for DFT-Based In-Car Communication (ICC) Systems

Jan Franzen, Tim Fingscheidt

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

Abstract

In-car communication (ICC) systems supporting speech communication in noise by reproducing amplified speech from the car cabin in the car cabin ask for low-delay acoustic echo cancellation (AEC). In this paper we propose a delay-flexible DFT-based stereo AEC capable of cancelling also the echoes stemming from the audio player or FM radio. For the price of a somewhat higher complexity we are able to reduce the 32 ms delay of the baseline down to 4 ms, loosing only 1 dB in ERLE while even preserving system distance properties.

Index Terms: in-car communication, acoustic echo cancellation, frequency domain adaptive filtering, low delay

1. Introduction

Speech communication between rear and front seat passengers in cars is often difficult due to ambient noise, especially at higher speeds. In-car communication (ICC) systems typically use the car's existing microphones and loudspeakers to provide an easier communication in the car cabin. The speech signals are acquired, additionally amplified, and sent to the loudspeakers at the listening passenger positions. The core of an ICC system is—as it is for hands-free telephony systems—an acoustic echo cancellation (AEC). The AEC estimates the loudspeaker-enclosure-microphone (LEM) system including the acoustic echo path. Subsequently, the echo signal is estimated and subtracted from the microphone signal to obtain a widely echo-free 'uplink' signal.

This field of research offers a variety of AEC approaches that were developed in the past decades. As shown in [1], the challenging computational complexity has led from delayless time domain algorithms, as for example seen in [2], to block processing algorithms, e.g. [3,4], and sub-band processing algorithms using, e.g., the discrete Fourier transform (DFT) or filterbanks [5–10]. Though the latter approaches introduce delay to the signal path and may decrease the time resolution, high-order adaptive filtering becomes computationally feasible without the restriction to solely update subsets of the filter coefficients as proposed in [11]. Furthermore, frequency domain adaptive filtering (FDAF) allows for a state-space formulation according to the Kalman filter theory [12–14]. Partitioned block-based systems increasing convergence speed, reducing delay, or allowing the estimation of much longer impulse responses have been proposed in [15–18].

Originating from monophonic AEC algorithms, similar approaches were developed for stereophonic applications [19–22]. Stereo AECs (SAECs) are capable of estimating two separate acoustic echo paths, for two differing yet highly cross-correlated excitation signals. In an ICC system we find this case if stereo music is played while the passengers are having a conversation.

Furthermore, ICC systems must fulfill strict delay constraints to ensure that the amplified speech is not perceived as distinguishable echo: According to [23], so-called 'class A' systems need to keep mouth-to-ear delays below 12 ms. In contrast, the algorithmic delay is the biggest disadvantage of block-based frequency domain algorithms [24]. The delay constraints of ICC systems usually ask for filterbanks or delay-reducing methods like the use of asymmetric windows in an overlap-add structure [25]. However, the good performance of DFT-based approaches such as the state-space FDAF SAEC [26, 27] in hands-free telephony suggests a good performance for ICC systems as well, under the condition that the delay problem can be solved while considering the larger computational load [28].

In this paper, we present a delay-flexible frequency domain SAEC for ICC systems. Strict delay constraints are fulfilled by massively decreasing the frame shift. Unlike partitioned block-based approaches, we compensate the increasing complexity by decoupling the adaptation from the actual echo cancellation. A less frequent estimation of the echo path is then advantageously accompanied by an additional look-back in the microphone input signal buffer. The proposed system allows for a consistent configuration both for low-delay ICC systems and hands-free telephony, while maintaining a comparable performance.

The paper is structured as follows: In Section 2 the SAEC approach is explained in detail and the proposed ICC modifications with delay flexibility are introduced. The experimental setup is presented in Section 3, followed by the evaluation and discussion of results. Section 4 provides conclusions.

2. Algorithmic Approach

The general algorithmic approach of acoustic echo cancellation (AEC) for automotive hands-free telephony and in-car communication (ICC) can be chosen quite similar. While an ICC system sends the signal output to the loudspeakers superimposed with, e.g., stereo music, it is transmitted to the far-end speaker in a hands-free system. For simplicity, we regard a unidirectional ICC system amplifying the 'near-end' speaker's voice (driver) and reproducing it at the 'far-end', i.e., in the rear of the car.

Starting point for our stereo AEC for ICC systems is the state-space frequency domain adaptive filtering (FDAF) SAEC as presented in [27], being a faster converging derivate of [26] and [29]. An illustration of this SAEC in the ICC context is depicted in Figure 1. On the top left of the figure, four FM radio channels are connected to the left and right front and rear loudspeakers (FL, FR, RL, RR). Both left and both right channels are combined to provide the SAEC reference channels and *virtual* loudspeaker signals $x_j(n)$ with discrete sample index n and channel index $j \in \{1, 2\}$. The SAEC's LEM system

http://dx.doi.org/10.21437/Interspeech.2017-1084

model consists of the two impulse responses[1] $h_j(n)$ providing the echo signals $d_j(n)$. The echo signals are superimposed with the driver's ('near-end') speech $s(n)$ and noise $n(n)$ to form the microphone signal $y(n)$. The SAEC itself is a block-based approach to compute the error signal $e(n)$. It is then processed by further ICC functions such as noise reduction, equalization, and amplification, and then mixed with the FM radio signals to be output at the rear loudspeakers. However, this further ICC processing and consequent mixing is—without loss of generality—not part of this work.

The proposed SAEC/ICC approach is now explained in Section 2.1. The proposed delay-flexible scheduling for applicability in an ICC system is then introduced in Section 2.2.

2.1. Proposed SAEC/ICC System

The acoustic echo cancellation can be split into two parts: The frame-wise *echo cancellation* on the one side and the *coefficient adaptation* for the estimated impulse responses on the other side. In the following, both parts are described in detail. The so-called baseline approach can as well be seen in Figure 1, with the constraint that microphone input frame length L (see figure description) is set equal to the frame shift R. The usefulness and necessity of introducing a separate microphone input frame length L will be explained in Section 2.2 for the newly proposed scheduling scheme.

2.1.1. Echo Cancellation

Buffers A and B contain the loudspeaker signals of the current frame with index $\ell \in \{1, 2, \ldots\}$:

$$\begin{aligned}\mathbf{x}_j(\ell) = \Big[&x_j\big((\ell-1)\cdot R + R - K\big),\ \ldots,\\ &x_j\big((\ell-1)\cdot R\big),\ldots,\\ &x_j\big((\ell-1)\cdot R + R - 1\big)\Big]^T,\end{aligned} \tag{1}$$

with $K-R$ previous samples followed by R new samples ($x_j(n) = 0, n < 0$). Here, R denotes the frame shift and $[\,]^T$ the transpose. Applying a K-point DFT on each of the buffers leads to $X_j(\ell,k)$, $j \in \{1, 2\}$, or $\mathbf{X}_j(\ell)$ in vectorial notation.

To obtain the estimated echo signals we first introduce the $K \times K$ overlap-save constraint matrix

$$\mathbf{G} = \mathbf{F}_{K\times K}\mathbf{Q}\mathbf{Q}^T\mathbf{F}^{-1}_{K\times K} \tag{2}$$

with the K-point DFT matrix $\mathbf{F}_{K\times K}$ and the $K \times L$ projection matrix $\mathbf{Q} = (\mathbf{0}_{L\times K-L}\ \mathbf{I}_{L\times L})^T$. Here, $\mathbf{0}$ stands for the zero matrix and $\mathbf{I}$ for the unity matrix. The estimated echo signals are then obtained via estimated LEM system functions $\hat{H}_j(\ell,k)$ as

$$\hat{\mathbf{D}}_j(\ell) = \mathbf{G}\cdot\big(\mathbf{X}_j(\ell)\circ\hat{\mathbf{H}}_j(\ell)\big) \tag{3}$$

with $(\circ)$ indicating an *element-wise* multiplication of the two vectors (Hadamard product).

Matching the given overlap-save constraints, buffer C contains $K-L$ zeros, followed by $L-R$ previous and R new samples of the microphone signal:

$$\mathbf{y}(\ell) = \big[y\big((\ell-1)\cdot R-(L-R)\big),\ \ldots,\ y\big((\ell-1)\cdot R+R-1\big)\big]^T.$$

After applying $\mathbf{Y}(\ell) = \mathbf{F}_{K\times K}\cdot\mathbf{y}(\ell)$, the enhanced error signal

$$\mathbf{E}(\ell) = \mathbf{Y}(\ell) - \hat{\mathbf{D}}_1(\ell) - \hat{\mathbf{D}}_2(\ell) \tag{4}$$

[1]Note that for better readability, we omit the time variance in the notation of the impulse responses $h_j(n)$.

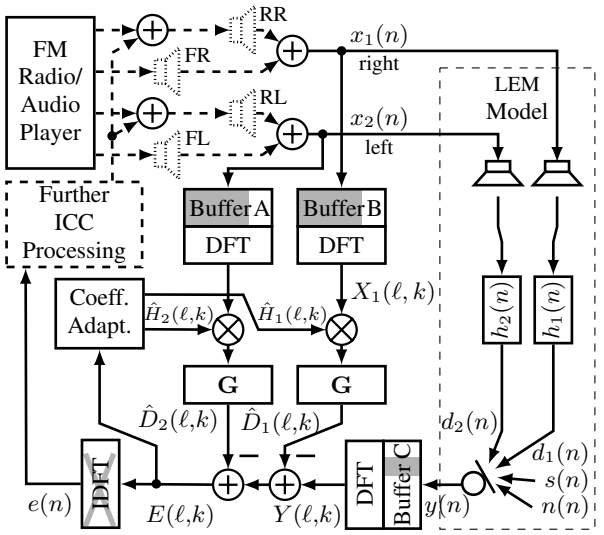

Figure 1: *Proposed unidirectional SAEC/ICC system model allowing low-delay* and *low-complexity scheduling. Box* ■ *symbolizes $K-R$ previous samples in buffers A and B, which are followed by R new samples. Buffer C consists of $K-L$ zeros, followed by* ■*-marked $L-R$ previous samples, and R new samples.* ✗ *in the IDFT marks the removal of $K-R$ samples to form the overlap-save valid output of R samples.*

is computed. As final step, $\mathbf{E}(\ell)$ is subject to an inverse DFT (IDFT). According to the overlap-save constraint, the first $K-L$ samples of its output are invalid and simply discarded, indicated by the gray cross in Figure 1. However, corresponding to the frame shift, only the last R samples are taken to form the valid $R\times 1$ time-domain output $\mathbf{e}(\ell)$ of the current frame.

2.1.2. Coefficient Adaptation

The second part of the algorithm is the coefficient adaptation. In case of the baseline approach it is carried out *once* per frame shift as follows: With the underlying Markov model, variables for the current frame are predicted based on the values of previous frames and prediction coefficient a. Predicted variables are marked as $(\cdot)^+$. This includes the estimated system functions

$$\hat{\mathbf{H}}_j^+(\ell) = a\hat{\mathbf{H}}_j(\ell-1) \qquad j\in\{1,2\}, \tag{5}$$

while $\hat{\mathbf{H}}_j(\ell=0)$ is initialized as zero vector. Furthermore, the two intra-channel process noise covariance matrices are

$$\mathbf{\Psi}^{\Delta}_{j,j}(\ell-1) = (1-a^2)\big[\hat{\mathbf{H}}_j(\ell-1)\hat{\mathbf{H}}_j^H(\ell-1)+\mathbf{P}_{j,j}(\ell-1)\big], \tag{6}$$

with $(\cdot)^H$ denoting the Hermitian transpose, and $\mathbf{\Psi}^{\Delta}_{j,i\neq j}(\ell-1) = \mathbf{0}_{K\times K}$. Finally, the intra-channel $(j,i) \in \{(1,1),(2,2)\}$ and cross-channel $(j,i) \in \{(1,2),(2,1)\}$ state error covariance matrices are predicted

$$\mathbf{P}^+_{j,i}(\ell) = a^2\mathbf{P}_{j,i}(\ell-1) + \lambda\mathbf{\Psi}^{\Delta}_{j,i}(\ell-1), \tag{7}$$

with an overestimation factor λ, and initializations $\mathbf{P}_{j,i}(\ell=0) = \mathbf{I}_{K\times K}$ and $\mathbf{\Psi}^{\Delta}_{j,i}(\ell=0) = \mathbf{0}_{K\times K}$.

Subsequent to the predictions, the current frame is taken into account. A preliminary error signal

$$\begin{aligned}\tilde{\mathbf{E}}(\ell) = \mathbf{Y}(\ell) - \big[&\mathbf{G}\cdot\big(\mathbf{X}_1(\ell)\circ\hat{\mathbf{H}}_1^+(\ell)\big)+\\ &\mathbf{G}\cdot\big(\mathbf{X}_2(\ell)\circ\hat{\mathbf{H}}_2^+(\ell)\big)\big]\end{aligned} \tag{8}$$

is used to compute the measurement noise covariance matrix

$$\begin{aligned}\boldsymbol{\Psi}^S(\ell) = (1-\beta)\cdot\Big(&\tilde{\mathbf{E}}(\ell)\tilde{\mathbf{E}}^H(\ell)+\\ &\tfrac{L}{K}\big[\mathbf{X}_1(\ell)\mathbf{P}^+_{1,1}(\ell)\mathbf{X}_1^H(\ell)+\\ &\mathbf{X}_1(\ell)\mathbf{P}^+_{1,2}(\ell)\mathbf{X}_2^H(\ell)+\\ &\mathbf{X}_2(\ell)\mathbf{P}^+_{2,1}(\ell)\mathbf{X}_1^H(\ell)+\\ &\mathbf{X}_2(\ell)\mathbf{P}^+_{2,2}(\ell)\mathbf{X}_2^H(\ell)\big]\Big)+\beta\cdot\boldsymbol{\Psi}^S(\ell-1)\end{aligned} \tag{9}$$

which is smoothed over time with factor $\beta = 0.5$. The matrix $\boldsymbol{\Psi}^S(\ell)$ is then taken to calculate the diagonal matrix

$$\begin{aligned}\mathbf{D}(\ell) = \tfrac{L}{K}\big[&\mathbf{X}_1(\ell)\mathbf{P}^+_{1,1}(\ell)\mathbf{X}_1^H(\ell)+\\ &\mathbf{X}_1(\ell)\mathbf{P}^+_{1,2}(\ell)\mathbf{X}_2^H(\ell)+\\ &\mathbf{X}_2(\ell)\mathbf{P}^+_{2,1}(\ell)\mathbf{X}_1^H(\ell)+\\ &\mathbf{X}_2(\ell)\mathbf{P}^+_{2,2}(\ell)\mathbf{X}_2^H(\ell)\big]+\boldsymbol{\Psi}^S(\ell)\end{aligned} \tag{10}$$

which in turn is used to compute the four diagonal near-optimal step-size matrices

$$\boldsymbol{\mu}_{j,i}(\ell) = \tfrac{L}{K}\mathbf{P}^+_{j,i}(\ell)\mathbf{D}^{-1}(\ell) \tag{11}$$

for $(j,i) \in \{(1,1),(1,2),(2,1),(2,2)\}$. Now the Kalman gains can be computed as

$$\begin{aligned}\mathbf{K}_1(\ell) &= \boldsymbol{\mu}_{1,1}(\ell)\mathbf{X}_1^H(\ell)+\boldsymbol{\mu}_{1,2}(\ell)\mathbf{X}_2^H(\ell)\\ \mathbf{K}_2(\ell) &= \boldsymbol{\mu}_{2,1}(\ell)\mathbf{X}_1^H(\ell)+\boldsymbol{\mu}_{2,2}(\ell)\mathbf{X}_2^H(\ell),\end{aligned} \tag{12}$$

the state error covariance matrices are updated to

$$\mathbf{P}_{j,i}(\ell) = \mathbf{P}^+_{j,i}(\ell) - \tfrac{R}{K}\mathbf{K}_j(\ell)\big[\mathbf{X}_1(\ell)\mathbf{P}^+_{1,i}(\ell)+\mathbf{X}_2(\ell)\mathbf{P}^+_{2,i}(\ell)\big],$$

and finally both estimated impulse responses are updated to

$$\hat{\mathbf{H}}_j(\ell) = \hat{\mathbf{H}}^+_j(\ell) + \mathbf{K}_j(\ell)\circ\tilde{\mathbf{E}}(\ell), \tag{13}$$

which are then available for the echo estimation in (3).

2.2. Proposed Delay-Flexible Scheduling

To motivate the following, it is important to first take a look at the SAEC subsystem and its delay. For a sampling frequency $f_s = 16\,\text{kHz}$, typical values for DFT length and frame shift are $K = 1024$ and $R = 256$, respectively [27]. The algorithmic delay is directly tied to the choice of the frame shift. Since a number of R samples has to be collected at the input buffers before a new frame is completely available for processing, the algorithmic delay as well equals R samples. For real-time systems the total delay τ_d is even higher, since the computation needs some time until the signal output is available. Typically, a buffering structure is applied where the output is delayed for another up to R samples, before it is read from an output buffer. In total this makes up a delay of $\tau_\text{d} = 2R$ samples, which we will consider in the rest of the paper. With $R = 256$ and a sample rate of 16 kHz this equals a total delay of 32 ms, which would be by far too high for an ICC system.

The (total) delay can now easily be decreased to 8 or 4 ms by reducing the frame shift to $R = 64$ or $R = 32$. Though, realizing that all processing steps are carried out with each frame shift, the computational complexity *in*creases by the same factor that the frame shift is *de*creased. Therefore, we aim for a way to again reduce the computational complexity while maintaining the low delay.

Inspired by similar approaches from other fields (e.g., [30] with a separate estimation of the feedback path), the idea is to decouple the actual echo cancellation from the coefficient adaptation. By doing so, a small frame shift R can be chosen while the coefficient adaptation could be performed more rarely at a lower update rate. However, simply performing the coefficient adaptation not *once* per frame shift but for example once per 8 frame shifts leads to a massive performance drop-down.

Instead, we propose to add a *look-back* into the microphone input buffer that corresponds to the lowered update rate of the coefficient adaptation. This is achieved with the help of three parameters (all defined in samples):

- the already known (and delay-causing) frame shift R,
- our introduced microphone input frame length L,
- and coefficient adaptation update rate U.

An illustrating example: The coefficient adaptation is called only once per 256 samples ($U = 256$). The frame shift, instead, shall be $R = 32$. We therefore set $L = 256$ as well. Thus, microphone input buffer C in Figure 1 contains at any time the R newest samples and additionally the $L - R$ previous samples as look-back. By doing so, the coefficient adaptation still processes all samples of the input without any discontinuing gaps in between[2]. Since the coefficient adaptation is now based on previous samples as well, our new scheduling scheme cannot react immediately to changes in the echo path. Therefore, we may observe a slight degradation in the echo cancellation performance, but much lower delay and complexity.

At this point it is important to note that the three parameters shown above offer some further degrees of freedom. For instance, if the complexity of the frequent coefficient adaptation at smaller frame shifts is *not* a problem, the additional look-back can as well be used with the perspective to achieve even higher echo suppression at lower delay. We will as well investigate the effect of this in the experimental evaluation.

3. Experimental Evaluation

The presented approach is evaluated in terms of echo return loss enhancement (ERLE), system distance, total delay, and relative complexity. ITU-T Recommendation P.501 [31] test signals are used for the simulations.

3.1. Experimental Setup

The DFT size is set to $K = 1024$. The simulations are carried out at three different frame shifts: On the one hand the baseline [27] with $R = L = U = 256$. On the other hand, different combinations of coefficient update rates U and microphone input frame lengths L at $R = 64$ and $R = 32$. Avoiding circular convolution artifacts, $K - L$ coefficients remain to cover the LEM system impulse responses. At a sampling rate of $f_s = 16\,\text{kHz}$ this equals a length of 48 ms ($L = 256$), 60 ms ($L = 64$), and 62 ms ($L = 32$), respectively. The first-order Markov model prediction coefficient a is derived from the value $a_{256} = 0.998$ at $U = R = 256$ [27] by ensuring an equal memory span even at different coefficient update rates according to $a = \exp\left(\frac{U}{256}\cdot\ln(0.998)\right)$. However, in the special cases of an introduced look-back ($L = 256$ while $U = 64$ or $U = 32$), the prediction coefficient a has to be set slightly higher to compensate for the look-back and thus maintain a comparable behavior over time.

ITU-T Recommendation P.501 [31, Secs. 7.3.5, 7.3.7] test signals are used for simulating a challenging speech scenario: An FM radio speaker's speech signal is made up of the 10 s

[2] In contrast: If no look-back were used in this example, the coefficient adaptation would only use 32 out of every 256 samples!

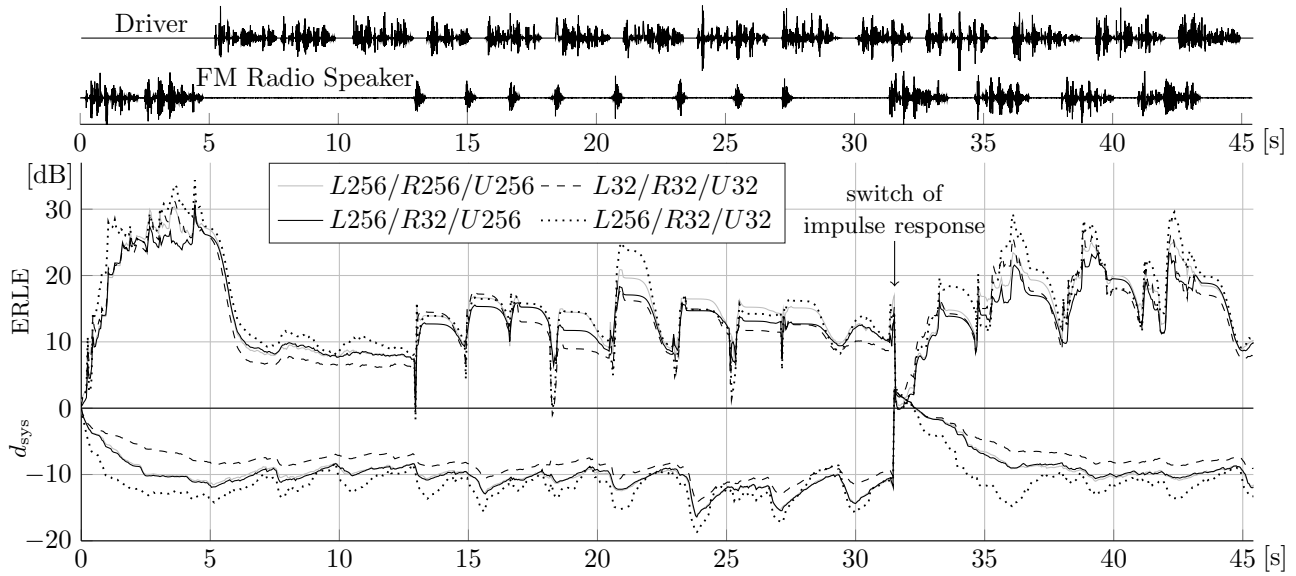

Figure 2: *Speech waveforms (top) and results for the baseline scheduling ($L = R = U = 256$) and the experiments with $R = 32$. ERLE (center curves) and system distance (bottom curves) over time: single-talk (0-10 s), single-talk with barge-ins (10-30 s), double-talk (30-45 s).*

short conditioning sequence I followed by the ca. 35 s *double-talk sequence*, both at a level of -26 dBov. It is convolved with randomly generated impulse responses[3] with exponential energy decay and a reverberation time of $T_{60} = 50$ ms. Uncorrelated white Gaussian noise is added with a level of -66 dBov as sensor noise, yielding the loudspeaker signals $x_i(n)$. The LEM path impulse responses[3] $h_i(n)$ are created in the same way and are convolved with the loudspeaker signals. By adding *short conditioning sequence II* followed by the *single-talk sequence* as $s(n)$ at a level of -26 dBov and *in-car noise* as $n(n)$ with an SNR of 15 dB, the microphone signal $y(n)$ is obtained. After 31.5 s, the near-end impulse responses are switched during double-talk. As mentioned before, we do not intend to evaluate a particular ICC system processing, but an SAEC in an ICC context. Therefore, we omit the mixing of $e(n)$ and the FM radio signal, thereby following ITU-T Recommendation P.501.

3.2. Experiments and Discussion

Figure 2 shows the speech waveforms on the top and the experimental results of four experiments below. The center curves show the ERLE in [dB] over time. The lower curves depict the course of the system distance

$$d_{\text{sys}}(n) = 10 \log \left(\frac{\sum_{j=1}^{2} ||\boldsymbol{h}_j(n) - \hat{\boldsymbol{h}}_j(n)||^2}{\sum_{j=1}^{2} ||\boldsymbol{h}_j(n)||^2} \right). \qquad (14)$$

Additionally, Table 1 provides the mean values as an overview of all experiments. These also include the results for $R=64$ and furthermore the total delay τ_{d} in ms. The last column in the table shows the relative complexity c_{rel} (averaged over five runs each) as ratio of processing time against the baseline.

The baseline ($L=R=U=256$) reaches about 30 dB ERLE in single-talk, ca. 15 dB at 'far-end' barge-ins, and around 20 dB during harsh double-talk. The mean system distance is -9.56 dB. Going from the baseline to the ca. 8 times more complex low-delay case with $L=R=U=32$ leads to a decrease in ERLE in nearly all speech sections and especially during barge-ins with a drop of up to ca. 6 dB. The system distance is constantly worse and on average 1.78 dB lower than the baseline. The newly proposed low-delay approach with $R=32$ and $L=U=256$ instead, compensates most of these drawbacks: While the mean ERLE increases again a bit by 0.24 dB but remains (as expected) at all speech sections below the baseline, the system distance is about the same as for the baseline approach. A further major advantage is the relative complexity which is about 45 percent less than $L=R=U=32$. Finally, we investigate the look-back while maintaining the high complexity, i.e., $L=256$ and $R=U=32$. The results show a major increase in both $\overline{\text{ERLE}}$ and system distance, also being clearly better than the baseline. The same set of experiments, but with $R=64$, show comparable results and are displayed in Table 1.

[3]Note that we shorten impulse responses to a car-typical value of $T_{60} = 50$ ms.

Table 1: *Mean ERLE [dB], system distance [dB], total delay [ms], and relative complexity for different microphone frame length L, frame shift R, and coefficient update every U samples.*

L	R	U	a	$\overline{\text{ERLE}}$	$\bar{d}_{\text{sys}}$	τ_{d}	c_{rel}
256	256	256	0.998	14.95	−9.56	32	1.00
64	64	64	0.9995	14.03	−8.34	8	4.03
256	64	256	0.998	14.05	−9.62	8	2.48
256	64	64	0.9996	16.30	−11.13	8	3.99
32	32	32	0.99975	13.68	−7.78	4	8.13
256	32	256	0.998	13.92	−9.60	4	4.49
256	32	32	0.99985	15.85	−11.24	4	8.15

4. Conclusions

In this paper, we proposed a delay-flexible stereo acoustic echo cancellation for in-car communication (ICC) systems employing frequency domain adaptive filtering. The approach achieves a low delay of 8 ms or even 4 ms by small frame shifts. The complexity is kept feasible by a less frequent coefficient adaptation. By including an additional look-back, only a slight degradation compared to the high-delay baseline is observed, while the system distances are about equal.

5. References

[1] E. Hänsler and G. Schmidt, *Acoustic Echo and Noise Control: A Practical Approach*. Hoboken, NJ, USA: Wiley-Interscience, 2004.

[2] 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.

[3] J. Lee and C. Un, "Block Realization of Multirate Adaptive Digital Filters," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 1, pp. 105–117, Feb. 1986.

[4] S. Malik and G. Enzner, "State-Space Frequency-Domain Adaptive Filtering for Nonlinear Acoustic Echo Cancellation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 7, pp. 2065–2079, Sep. 2012.

[5] A. Gilloire and M. Vetterli, "Adaptive Filtering in Subbands with Critical Sampling: Analysis, Experiments, and Application to Acoustic Echo Cancellation," *IEEE Transactions on Signal Processing*, vol. 40, no. 8, pp. 1862–1875, Aug. 1992.

[6] J. Chen, H. Bes, J. Vandewalle, and P. Janssens, "A New Structure for Sub-Band Acoustic Echo Canceler," in *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing*, New York, NY, USA, Apr. 1988, pp. 2574–2577.

[7] G. Schmidt, "Acoustic Echo Control in Subbands - An Application of Multirate Systems," in *Proc. of 9th European Signal Processing Conference*, Rhodes, Greece, Sep. 1998, pp. 1961–1964.

[8] W. Kellermann, "Analysis and Design of Multirate Systems for Cancellation of Acoustical Echoes," in *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing*, New York, NY, USA, Apr. 1988, pp. 2570–2573.

[9] K. Steinert, M. Schönle, C. Beaugeant, and T. Fingscheidt, "Subband Speech Enhancement System for Echo and Noise Reduction with Reduced Signal Delay," in *Proc. of ITG Conference on Speech Communication*, Aachen, Germany, Oct. 2008.

[10] ——, "Hands-free System with Low-Delay Subband Acoustic Echo Control and Noise Reduction," in *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Las Vegas, NV, USA, Apr. 2008, pp. 1521–1524.

[11] T. Schertler, "Selective Block Update of NLMS Type Algorithms," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*, Seattle, WA, USA, May 1998, pp. 1717–1720.

[12] 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, Jun. 2006.

[13] 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, Sep. 2011.

[14] ——, "A Wideband Automotive Hands-Free System for Mobile HD Voice Services," in *Smart Mobile In-Vehicle Systems – Next Generation Advancements*, G. Schmidt *et al.*, Eds. Springer, 2014, pp. 81–96.

[15] E. Moulines, O. A. Amrane, and Y. Grenier, "The Generalized Multidelay Adaptive Filter: Structure and Convergence Analysis," *IEEE Transactions on Signal Processing*, vol. 43, no. 1, pp. 14–28, Jan. 1995.

[16] B. H. Nitsch, "A Frequency-Selective Stepfactor Control for an Adaptive Filter Algorithm Working in the Frequency Domain," *Signal Processing (Elsevier)*, vol. 80, no. 9, pp. 1733–1745, Sep. 2000.

[17] J. S. Soo and K. K. Pang, "Multidelay Block Frequency Domain Adaptive Filter," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 2, pp. 373–376, Feb. 1990.

[18] F. Kuech, E. Mabande, and G. Enzner, "State-Space Architecture of the Partitioned-Block-Based Acoustic Echo Controller," in *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Florence, Italy, May 2014, pp. 1295–1299.

[19] J. Benesty, D. R. Morgan, and M. M. Sondhi, "A Hybrid Mono/Stereo Acoustic Echo Canceler," *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 5, pp. 468–475, Sep. 1998.

[20] H. Buchner and W. Kellermann, "Acoustic Echo Cancellation for Two and More Reproduction Channels," in *Proc. of International Workshop on Acoustic Echo and Noise Control*, Darmstadt, Germany, Sep. 2001, pp. 99–102.

[21] S. Makino, K. Strauss, S. Shimauchi, Y. Haneda, and A. Nakagawa, "Subband Stereo Echo Canceller Using the Projection Algorithm with Fast Convergence to the True Echo Path," in *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Munich, Germany, Apr. 1997, pp. 299–302.

[22] C. Stanciu, C. Paleologu, J. Benesty, S. Ciochina, and F. Albu, "Variable-Forgetting Factor RLS for Stereophonic Acoustic Echo Cancellation with Widely Linear Model," in *Proc. of 20th European Signal Processing Conference*, Bucharest, Romania, Aug. 2012, pp. 1960–1964.

[23] A. Theiß, G. Schmidt, J. Withopf, and C. Lüke, "Instrumental Evaluation of In-Car Communication Systems," in *Proc. of ITG Conference on Speech Communication*, Erlangen, Germany, Sep. 2014.

[24] P. Eneroth, "Joint Filterbanks for Echo Cancellation and Audio Coding," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 4, pp. 342–354, Jul. 2003.

[25] C. Lüke, H. Özer, G. Schmidt, A. Theiß, and J. Withopf, "Signal processing for in-car communication systems," in *5th Biennial Workshop on DSP for In-Vehicle Systems*, Kiel, Germany, Sep. 2011.

[26] 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.

[27] M. A. Jung, S. Elshamy, and T. Fingscheidt, "An Automotive Wideband Stereo Acoustic Echo Canceler using Frequency-Domain Adaptive Filtering," in *Proc. of 22nd European Signal Processing Conference*, Lisbon, Portugal, Sep. 2014, pp. 1452–1456.

[28] G. Enzner, H. Buchner, A. Favrot, and F. Kuech, "Acoustic Echo Control," in *Academic Press Library in Signal Processing*, R. Chellappa and S. Theodoridis, Eds. Elsevier/Academic Press, 2013, vol. 4, pp. 807–877.

[29] S. Malik and J. Benesty, "Variationally Diagonalized Multichannel State-Space Frequency-Domain Adaptive Filtering for Acoustic Echo Cancellation," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, Canada, May 2013, pp. 595–599.

[30] A. Pandey and V. J. Mathews, "Low-Delay Signal Processing for Digital Hearing Aids," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 4, pp. 699–710, May 2011.

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