Change Prediction for Low Complexity Combined Beamforming and Acoustic Echo Cancellation

1st Matthias Schrammen
Inst. of Communication Systems
RWTH Aachen University
Aachen, Germany
schrammen@iks.rwth-aachen.de

2nd Alexander Bohlender
IDLab, Dept. of Electronics and Information Systems
Ghent University - imec
Ghent, Belgium
alexander.bohlender@ugent.be

3rd Stefan Kühl, 4th Peter Jax
Inst. of Communication Systems
RWTH Aachen University
Aachen, Germany
{kuehl, jax}@iks.rwth-aachen.de

***Abstract*—Time-variant beamforming (BF) and acoustic echo cancellation (AEC) are two techniques that are frequently employed for improving the quality of hands-free speech communication. However, the combined application of both is quite challenging as it either introduces high computational complexity or insufficient tracking. We propose a new method to improve the performance of the low-complexity beamformer first (BF-first) structure, which we call *change prediction* (ChaP). ChaP gathers information on several BF changes to predict the effective impulse response seen by the AEC after the next BF change. To account for uncertain data and convergence states in the predictions, reliability measures are introduced to improve ChaP in realistic scenarios.**

***Index Terms*—combined beamforming and acoustic echo cancellation, low complexity, beamformer first, pseudo inverse**

I. Introduction

For the realization of a hands-free speech communication system, BF and AEC are usually needed to guarantee a sufficient quality of the conversation. However, the combined operation of AEC and BF at the same time causes problems. One basic combination scheme is echo canceller first (AEC-first), where AEC is performed before BF [1]. In AEC-first one adaptive filter including its own adaptation control is needed for each channel. Another basic combination scheme is BF-first, where BF is performed before the AEC [2]. Only one echo canceller is needed and hence the complexity can be reduced by a factor N_{mic} compared to AEC-first, where N_{mic} is the number of microphones. However, in BF-first the adaptive filter of the AEC has to track the time-variant BF weights in addition to the room impulse response (RIR). This usually results in an unsatisfactory echo reduction, because after every change of the BF weights, the AEC suffers from a large misalignment and needs a long time to reconverge again.

Other schemes exist, besides the simple concatenation used in AEC-first and BF-first. They incorporate the AEC after the fixed BF or in the interference canceller of a Generalized-Sidelobe-Canceller [1], or they jointly optimize the BF and AEC filter coefficients [3]. Recent approaches try to reduce the complexity of the AEC-first structure by exploiting similarities across channels for the adaptation control and design of the multi-microphone AEC [4], [5]. There also exist approaches that focus on enhancing the AEC adaptation by BF rather than reducing the complexity [6].

In this paper we focus on the low-complexity BF-first structure and propose a method to alleviate its shortcomings in tracking fast BF changes. Besides its lower complexity compared to AEC-first, BF-first can be beneficial for the adaptation of the AEC, because the echo-to-noise ratio can enhanced by the BF [1], [5]. One straightforward approach to ensure a quick reconvergence after the BF changes is to use the knowledge about the time instants when the BF weights are changed. At these times the AEC can be reinitialized by setting the adaptation speed to maximum. This approach will be termed rapid recovery (RR) in the following. However, one drawback of this approach is that the misalignment of the AEC is very high at each change of the BF weights. This results in sudden drops of the echo reduction performance leading to an annoying listening experience. Therefore, we propose a new algorithm termed ChaP that maintains a high performance of echo reduction. The outline of the paper is as follows. We first introduce the signal model in Section II and describe the core of the ChaP algorithm in Section III. In Section IV-B this core algorithm is improved by using reliability information on the result from the ChaP estimation in the AEC. Further improvement is achieved by using reliability information from the AEC in the ChaP estimation in Section IV-C. Finally, in Section IV-D, time-variant RIRs are taken into account by introducing a Markov model into ChaP. In Section V the ChaP algorithm and its improved versions are evaluated and the ChaP method is applied to a practical example.

II. Signal Model

Fig. 1 shows the block diagram of the BF-first structure. The signal $y_n(k)$ at the n-th microphone is defined by

$$y_n(k) = d_n(k) + s_n(k) + v_n(k), \tag{1}$$

where $d_n(k)$ is the echo signal, $s_n(k)$ is the near-end speech and $v_n(k)$ is the near-end noise. k and n are the discrete time and channel indices, respectively. In the BF-first structure the AEC identifies the composed system consisting of the acoustic transmission from the loudspeaker to the microphones, modeled by the RIRs $\mathbf{h}_n(k)$ of length $N_{\mathbf{h}}$, and the BF weights $\mathbf{w}_n(k)$ of length $N_{\mathbf{w}}$. For time-variant filters $\mathbf{h}_n(k)$ and $\mathbf{w}_n(k)$, it is not mathematically accurate to describe the combination with an impulse response. Nevertheless, the

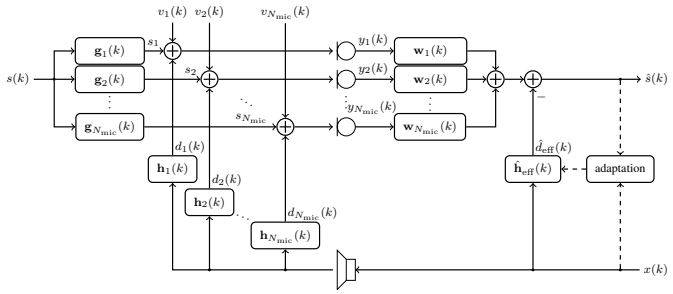

Fig. 1. Block diagram of the BF-first structure, where a BF with weights $\mathbf{w}_n$ is followed by an AEC with impulse response $\hat{\mathbf{h}}_{\text{eff}}$.

effective impulse response

$$\mathbf{h}_{\text{eff}}(k_0) = \sum_{n=1}^{N_{\text{mic}}} \mathbf{w}_n(k_0) * \mathbf{h}_n(k_0) \tag{2}$$

can serve as a simple yet sufficiently precise model for the true system. The convolution operator $*$ used in (2) performs a convolution over the entries of two vectors for a fixed time k_0. The result is thus again a vector. In this case $\mathbf{h}_{\text{eff}}(k)$ has a length of $N_{\mathbf{h}_{\text{eff}}} = N_{\mathbf{h}} + N_{\mathbf{w}} - 1$. An increased filter length is therefore required, to achieve a similar echo cancellation performance. This in turn reduces the speed of convergence of the AEC accordingly. Furthermore, the AEC has to track the time-variance not only of $\mathbf{h}_n(k)$ but also of $\mathbf{w}_n(k)$, where the latter can change very abruptly when the steering direction of the BF is altered. The BF, the AEC and the proposed ChaP algorithm work in the short-term Fourier domain. Therefore, the discrete Fourier transform (DFT) domain equivalent of (2) is now considered. Assuming a DFT length of $M > N_{\mathbf{h}_{\text{eff}}}$ and impulse responses $\mathbf{h}_1(k), \ldots, \mathbf{h}_{N_{\text{mic}}}(k)$ that do not change at least for the duration of one frame, the DFT of (2) is

$$H_{\text{eff}}(\mu, \lambda) = \sum_{n=1}^{N_{\text{mic}}} W_n(\mu, \lambda) \cdot H_n(\mu, \lambda), \tag{3}$$

where μ and λ are the frequency bin index and the frame index, respectively. With $\mathbf{W} = [W_1, W_2, \ldots, W_{N_{\text{mic}}}]^{\text{T}}$ and $\mathbf{H} = [H_1, H_2, \ldots, H_{N_{\text{mic}}}]^{\text{T}}$ (3) can be compactly written as

$$H_{\text{eff}}(\mu, \lambda) = \mathbf{W}^{\text{T}}(\mu, \lambda) \cdot \mathbf{H}(\mu, \lambda). \tag{4}$$

For the identification of H_{eff} we use the diagonalized Kalman filter adaptation in the frequency domain with an overlap-save framework [7], [8]. The fundamental concept of the Kalman based AEC is to model $H_{\text{eff}}(\mu, \lambda)$ as a first order Markov model with forgetting factor A. For brevity only the time update equations are recalled here (full derivation in [7], [8])

$$\hat{H}_{\text{eff}}(\mu, \lambda + 1) = A \cdot \hat{H}^{+}_{\text{eff}}(\mu, \lambda) \tag{5}$$

$$P_{\mu\mu}(\lambda + 1) = A^2 \cdot P^{+}_{\mu\mu}(\lambda) + M \cdot \Phi_{\Delta\Delta,\mu}, \tag{6}$$

where $\hat{H}_{\text{eff}}$ is an estimate of H_{eff}. $P_{\mu\mu}(\lambda)$ and $\Phi_{\Delta\Delta,\mu}$ are the estimation error covariance and the process noise covariance, respectively. The superscript $^{+}$ indicates an a-posteriori estimate resulting from the Kalman equations. In our experiments we use either a delay-and-sum (DS) BF or a minimum variance distortionless response (MVDR) BF with diffuse noise optimization to adjust the BF weights in the frequency domain [1], [9]. However, ChaP is in general independent of the specific BF principle.

III. Change Prediction (ChaP)

As already pointed out, a simple reset of the AEC at each BF change does not provide a sufficient echo reduction. Therefore, starting with (4) and keeping in mind the block diagram in Fig. 1, we investigate the effects of BF changes on H_{eff} and derive a solution to predict H_{eff} for a better initialization of the AEC after a change of the BF. Let λ_1 denote the index of the frame right before the BF weights $\mathbf{W}(\mu, \lambda)$ are changed from set 1 to set 2. As (4) holds for both $\lambda = \lambda_1$ and $\lambda = \lambda_1 + 1$, the system of two equations

$$\begin{cases} H_{\text{eff}}(\mu, \lambda_1) = \mathbf{W}^{\text{T}}(\mu, \lambda_1) \cdot \mathbf{H}(\mu, \lambda_1) \\ H_{\text{eff}}(\mu, \lambda_1 + 1) = \mathbf{W}^{\text{T}}(\mu, \lambda_1 + 1) \cdot \mathbf{H}(\mu, \lambda_1 + 1) \\ \qquad\qquad \approx \mathbf{W}^{\text{T}}(\mu, \lambda_1 + 1) \cdot \mathbf{H}(\mu, \lambda_1) \end{cases} \tag{7}$$

is obtained. The approximation results from the assumption that the change of the RIRs in one frame can be neglected compared to the change of the BF weights.

Now we would like to predict $H_{\text{eff}}(\mu, \lambda_1 + 1)$, but the entries $\mathbf{H}(\mu, \lambda_1)$ are not known. Therefore, a total of $N_{\text{mic}} + 1$ unknowns are present in these two equations and the problem is strongly ill-posed. Therefore, additional simplifying assumptions must be made in order to obtain an estimate of the new effective impulse response.

In the following, it is studied how a solution for the problem at hand can be obtained when the assumption is made that the RIRs $\mathbf{h}_1(k), \ldots, \mathbf{h}_{N_{\text{mic}}}(k)$ do not change for the duration of N_Δ changes of the BF filters. In Section IV-D we discuss how this restriction can be alleviated. For now, $\mathbf{H}(\mu, \lambda) = \mathbf{H}(\mu)$. From each change of the BF filters, which occurs after the frame with index λ_i ($i = 1, \ldots, N_\Delta$), one observation is obtained according to (4). Combining all yields the system of linear equations

$$\begin{bmatrix} H_{\text{eff}}(\mu, \lambda_1) \\ H_{\text{eff}}(\mu, \lambda_2) \\ \vdots \\ H_{\text{eff}}(\mu, \lambda_{N_\Delta}) \end{bmatrix} = \begin{bmatrix} \mathbf{W}^{\text{T}}(\mu, \lambda_1) \\ \mathbf{W}^{\text{T}}(\mu, \lambda_2) \\ \vdots \\ \mathbf{W}^{\text{T}}(\mu, \lambda_{N_\Delta}) \end{bmatrix} \cdot \mathbf{H}(\mu). \tag{8}$$

This is written in a compact form as

$$\mathbf{H}_{\text{eff}}(\mu, \boldsymbol{\lambda}_{N_\Delta}) = \boldsymbol{\mathcal{W}}(\mu, \boldsymbol{\lambda}_{N_\Delta}) \cdot \mathbf{H}(\mu), \tag{9}$$

where $\mathbf{H}_{\text{eff}}(\mu, \boldsymbol{\lambda}_{N_\Delta})$ is a $(N_\Delta \times 1)$-vector, $\boldsymbol{\mathcal{W}}(\mu, \boldsymbol{\lambda}_{N_\Delta})$ is a matrix of size $(N_\Delta \times N_{\text{mic}})$ and the N_Δ entries of the vector $\boldsymbol{\lambda}_{N_\Delta}$ identify the frames right after which a change of the BF occurs. While the exact entries of the vector $\mathbf{H}_{\text{eff}}(\mu, \boldsymbol{\lambda}_{N_\Delta})$ are not known, estimates thereof are provided by the AEC. These estimates can be considered sufficiently accurate as long as the time between two consecutive changes of the BF weights is not too short for the AEC to converge. How inaccurate estimates of $\mathbf{H}_{\text{eff}}$ can be handled will be shown in Section IV-C. Replacing $\mathbf{H}_{\text{eff}}(\mu, \boldsymbol{\lambda}_{N_\Delta})$ in (9) by its estimate yields

$$\hat{\mathbf{H}}_{\text{eff}}(\mu, \boldsymbol{\lambda}_{N_\Delta}) = \boldsymbol{\mathcal{W}}(\mu, \boldsymbol{\lambda}_{N_\Delta}) \cdot \hat{\mathbf{H}}(\mu). \tag{10}$$

In the special case of $N_\Delta = N_{\text{mic}}$, (10) can be solved by

$$\hat{\mathbf{H}}(\mu) = \boldsymbol{\mathcal{W}}^{-1}(\mu, \boldsymbol{\lambda}_{N_\Delta}) \cdot \hat{\mathbf{H}}_{\text{eff}}(\mu, \boldsymbol{\lambda}_{N_\Delta}), \tag{11}$$

if the matrix $\boldsymbol{\mathcal{W}}(\mu, \boldsymbol{\lambda}_{N_\Delta})$ has full rank. In the general case, where N_Δ is arbitrary and $\boldsymbol{\mathcal{W}}(\mu, \boldsymbol{\lambda}_{N_\Delta})$ might not have full

rank, a least-squares (LS) estimate

$$\hat{\mathbf{H}}(\mu) = \boldsymbol{\mathcal{W}}^{\dagger}(\mu, \boldsymbol{\lambda}_{N_\Delta}) \cdot \hat{\mathbf{H}}_{\text{eff}}(\mu, \boldsymbol{\lambda}_{N_\Delta}) \tag{12}$$

can be obtained from (10) when the Moore-Penrose pseudoinverse (MPP) (†) is used instead of the inverse.This expression can now be inserted back into (4).The result

$$\begin{aligned}
&\hat{H}^{-}_{\text{eff}}(\mu,\, \lambda_{N_\Delta}+1) = \\
&= \mathbf{W}^{\mathrm{T}}(\mu, \lambda_{N_\Delta}+1) \cdot \hat{\mathbf{H}}(\mu) \\
&= \mathbf{W}^{\mathrm{T}}(\mu, \lambda_{N_\Delta}+1) \cdot \boldsymbol{\mathcal{W}}^{\dagger}(\mu, \boldsymbol{\lambda}_{N_\Delta}) \cdot \hat{\mathbf{H}}_{\text{eff}}(\mu, \boldsymbol{\lambda}_{N_\Delta})
\end{aligned} \tag{13}$$

is the predicted μ-th DFT bin of the effective impulse response *after* the N_Δ-th change of the BF filters. If this is done for each frequency bin, the desired full estimate of the effective impulse response is obtained which serves the AEC as an initial estimate after the change of the BF. As indicated by the superscript of $\hat{H}^{-}_{\text{eff}}(\mu, \lambda_{N_\Delta}+1)$, this is merely the a-priori estimate for the following frame prior to the regular adaptation iteration.

This newly proposed algorithm is called *change prediction* (ChaP) in the following. ChaP is independent of RR introduced in Section I, but it can certainly be used in conjunction with it. This is done by reinitializing the weight vector with the ChaP estimate and the error covariance matrix $\mathbf{P} = \operatorname{diag}\{P_{\mu\mu}\}$ of the Kalman algorithm with the identity matrix. The latter initializes the AEC with maximum adaptation speed. This configuration will be referred to as change prediction with rapid recovery (ChaP/RR).

IV. Improvements of ChaP

To improve the ChaP algorithm and make it applicable to practical situations, several extensions are proposed in the following. We first improve the robustness of the calculation of the MPP used in (12) with the use of the effective rank in Section IV-A. After that we incorporate reliability information from the ChaP estimation to guide the adaptation of the AEC in Section IV-B. To handle erroneous estimates of a non-converged AEC, we use reliability information from the AEC to improve the robustness of the ChaP algorithm in Section IV-C. Finally, we propose a way to handle a moderate time-variance of the RIR in Section IV-D.

A. Regularization

If the BF weights of different observations are very similar, the matrix $\boldsymbol{\mathcal{W}}(\mu, \boldsymbol{\lambda}_{N_\Delta})$ usually has singular values close to zero. Then, the calculation of the MPP will amplify the noise at these frequencies. To circumvent this, the effective rank measure is applied to calculate the number of singular values that are used for the inversion [10], [11]. The effective rank of a matrix $\mathbf{A}$ is defined as

$$\mathcal{R}_{\text{eff}}(\mathbf{A}) = \exp\left(-\sum_{i=1}^{q} p_i \cdot \ln(p_i)\right) \tag{14}$$

where $p_i = \frac{\sigma_i}{\sum_{i=1}^{q}\sigma_i}, i = 1, \ldots, q$ are the normalized singular values of the matrix $\mathbf{A}$ with singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_q > 0$. The MPP is then calculated on the basis of a truncated singular value decomposition (SVD) where the largest $i_{\max} = \operatorname{Round}\left[\mathcal{R}_{\text{eff}}(\boldsymbol{\mathcal{W}}(\mu, \boldsymbol{\lambda}_{N_\Delta}))\right]$ singular values are used.

B. Directed Recovery

If the new vector of BF weights $\mathbf{W}(\mu, \lambda_{N_\Delta}+1)$ has been seen before, ChaP will generally provide a good estimate for the effective impulse response after the change. However, if the BF weights are very different from what has been observed before, the estimation of ChaP might not be accurate. We take this phenomenon into account by deriving an empiric reliability measure of the ChaP estimate and using this reliability measure to control the adaptation behavior of the AEC. To measure how much new information is introduced by the new BF weight vector, we calculate the difference of the effective rank of the BF weight matrix with and without the new weight vector as

$$\begin{aligned}
&\Delta\mathcal{R}_{\text{eff}}(\mu, \boldsymbol{\lambda}_{N_\Delta}) = \\
&= \mathcal{R}_{\text{eff}}\left(\begin{bmatrix} \boldsymbol{\mathcal{W}}(\mu, \boldsymbol{\lambda}_{N_\Delta}) \\ \mathbf{W}^{\mathrm{T}}(\mu, \lambda_{N_\Delta}+1) \end{bmatrix}\right) - \mathcal{R}_{\text{eff}}\left(\boldsymbol{\mathcal{W}}(\mu, \boldsymbol{\lambda}_{N_\Delta})\right)
\end{aligned} \tag{15}$$

$$= \mathcal{R}_{\text{eff}}\left(\boldsymbol{\mathcal{W}}(\mu, \boldsymbol{\lambda}_{N_\Delta+1})\right) - \mathcal{R}_{\text{eff}}\left(\boldsymbol{\mathcal{W}}(\mu, \boldsymbol{\lambda}_{N_\Delta})\right). \tag{16}$$

If $\Delta\mathcal{R}_{\text{eff}}$ is zero, no new information was added by the new weight vector. If the new weight vector provides new information, independent of the previous observations, $\Delta\mathcal{R}_{\text{eff}}$ will be one. Because ChaP estimates can be assumed more reliable when the new weight vector has been seen before, we map $\Delta\mathcal{R}_{\text{eff}}$ to the effective rank difference (ERD) reliability measure with $\text{ERD} = \max\{1 - \Delta\mathcal{R}_{\text{eff}}, 0\}$. The ERD measure is now used to adjust the initialization of the estimation error covariance matrix $\mathbf{P}$ of the Kalman algorithm according to

$$\begin{aligned}
P^{-}_{\mu\mu}(\lambda_{N_\Delta}+1) = &\min_{i=1,\ldots,N_\Delta}\left(P_{\mu\mu}(\lambda_i)\right) \cdot \text{ERD}(\mu, \boldsymbol{\lambda}_{N_\Delta}) \\
&+1 \cdot \left(1 - \text{ERD}(\mu, \boldsymbol{\lambda}_{N_\Delta})\right),
\end{aligned} \tag{17}$$

where the minimum is calculated over all observed error covariances. If ChaP is very sure about its estimate, the effective rank difference (ERD) measure is 1 and the Kalman algorithm uses a small $P_{\mu\mu}$ to adapt the AEC to a small system distance. If the ChaP estimate is potentionally erroneous, the effective rank difference (ERD) measure is 0 and the error covariance is set to 1 ensuring a fast convergence of the Kalman algorithm. This extension of the ChaP algorithm will be called directed recovery (DR) and the resulting combination with the core algorithm is termed change prediction with directed recovery (ChaP/DR).

C. Directed ChaP

Up to now the ChaP algorithm assumes that the estimates of $\mathbf{H}_{\text{eff}}$ from the AEC are correct. However, if the AEC is not yet converged, due to frequent changes of the BF weights or a high time-variance of the RIR, the estimates of $\mathbf{H}_{\text{eff}}$ might be more erroneous. To account for this varying reliability of the observations, we modify (13) to a weighted least-squares (WLS) solution with weighting matrix $\boldsymbol{\Psi}$ according to

$$\hat{H}^{-}_{\text{eff}} = \mathbf{W}^{\mathrm{T}}\left(\boldsymbol{\Psi}\boldsymbol{\mathcal{W}}\right)^{\dagger}\boldsymbol{\Psi}\hat{\mathbf{H}}_{\text{eff}}. \tag{18}$$

The same weighting is also applied to $\boldsymbol{\mathcal{W}}$ in (15). The weighting matrix $\boldsymbol{\Psi}$ is chosen as a diagonal matrix according to $\boldsymbol{\Psi} = \operatorname{diag}\{\psi_1, \ldots, \psi_{N_\Delta}\}$ with entries

$$\psi_i(\mu) = 1 - \sqrt{P_{\mu\mu}(\lambda_i)}, i = 1, \ldots, N_\Delta. \tag{19}$$

The ChaP estimate $\hat{H}_{\text{eff}}^{-}$ in (18) now results from a WLS solution, where reliable observations have a larger impact on $\hat{H}_{\text{eff}}^{-}$ than unreliable ones. This extension of the ChaP algorithm will be called directed change prediction (dChaP) and the resulting combinations with the core ChaP algorithm is termed directed change prediction with directed recovery (dChaP/DR) and directed change prediction with rapid recovery (dChaP/RR).

D. Markov model for ChaP

In the current version, the influence of old observations collected by the ChaP algorithm remains the same for all times. However, as the RIRs are time-variant, old observations should have a smaller impact on the final ChaP estimate. To account for this time-variance the same Markov model used in the Kalman algorithm is also used for the weighting matrix $\boldsymbol{\Psi}$ applied to $\boldsymbol{\mathcal{W}}$ in (15) and (18). Each entry ψ_i is updated according to

$$\psi_i(\mu,\lambda) = 1 - \sqrt{P_{i,\mu\mu}^{\text{ChaP}}(\lambda)}, i = 1,\ldots,N_\Delta \tag{20}$$

$$P_{i,\mu\mu}^{\text{ChaP}}(\lambda) = A^2 \cdot P_{i,\mu\mu}^{\text{ChaP}}(\lambda-1) + M \cdot \Phi_{\Delta\Delta,\mu}. \tag{21}$$

An estimate of the process noise covariance $\Phi_{\Delta\Delta,\mu}$ is provided by the AEC [7] and the initial value $P_{i,\mu\mu}^{\text{ChaP}}(\lambda) = P_{\mu\mu}(\lambda_i)$ is used.

V. Evaluation

For evaluating the ChaP algorithm and its extension a simulation is carried out. A circular array of 13 microphones is placed on the top of a rigid cylindrical mockup. The microphone signals are generated with a RIR simulator using the image source method and measured device-specific shadowing. The room dimensions are $5.5\,\text{m} \times 4.5\,\text{m} \times 2.7\,\text{m}$ and the microphone array is located in the center of the room. The echo (far-end) source is located $0.6\,\text{m}$ away from the array at an angle of $\varphi_{\text{echo}} = -20°$. The near-end source is located relative to the array at a constant distance of $0.6\,\text{m}$ but at a varying azimuth angle that is identical to the steering direction φ_{BF} of the BF. The frame size and DFT size are both set to $M = 1024$ and the frame shift is $R = 256$. The BF filter length is $N_{\mathbf{w}} = 64$. The AEC filter length is restricted to $N_{\mathbf{h}} = 704$. This guarantees a resulting length of the effective impulse response of $N_{\mathbf{h}_{\text{eff}}} = N_{\mathbf{h}} + N_{\mathbf{w}} - 1 \leq M - R + 1$, which prevents artifacts caused by cyclic convolution. To exclude effects of undermodelling in the evaluation, the simulated RIRs are truncated to $N_{\mathbf{h}}$ taps. For both, the desired near-end signal and the far-end signal, white Gaussian noise is used. For the near-end disturbance white Gaussian noise is used, too. All signals are sampled at $f_s = 16\,\text{kHz}$. In the first scenario we investigate the performance of ChaP and its extensions for a time-invariant RIR. The reverberation time is set to $T_{60} = 0.6\,\text{s}$. The forgetting factor is set to $A = 0.998$ for both, the Kalman algorithm and the ChaP extension in (21). The echo-to-signal-plus-noise ratio (ESNR) was set to 30 dB and the signal-to-noise ratio (SNR) was set to 0 dB at the input of the BF. As a quality metric the system distance (SysDis) calculated via $\text{SysDis}(k)[\text{dB}] = 10\log_{10}(||\mathbf{h}_{\text{eff}}(k) - \hat{\mathbf{h}}_{\text{eff}}(k)||^2 / ||\mathbf{h}_{\text{eff}}(k)||^2)$ is used. Fig. 2 shows the results for ChaP and its extension

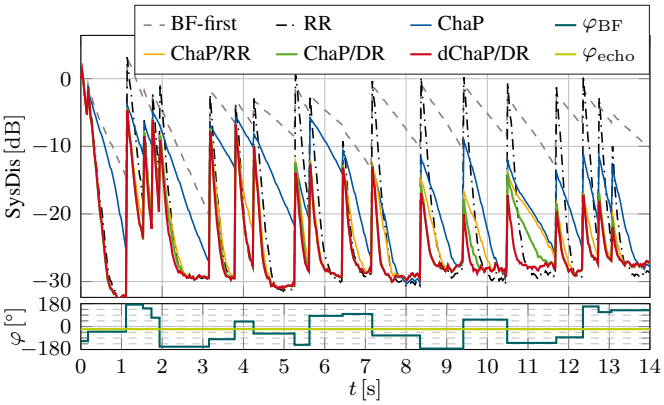

Fig. 2. SysDis for BF-first, RR and ChaP with its extensions in combination with a DS BF (top). At the bottom, the BF steering directions and the direction of the echo signal are shown. The RIRs are time-invariant.

for a DS BF. The bottom plot shows φ_{BF} and φ_{echo} over time. With the conventional BF-first approach (- -) the AEC barely converges between two changes of the BF steering direction and shows a high misalignment (SysDis). For RR (-·) the AEC converges to the ESNR of 30 dB. However, at each change of the BF, the misalignment of the AEC becomes very large resulting in an insufficient echo reduction. If the core ChaP algorithm is used (—), the SysDis is lower at points of BF changes compared to RR. However, the reconvergence is slow and prevents a complete convergence between most BF changes. This drawback is alleviated when ChaP/RR (—) is used. Further improvements can be achieved with ChaP/DR (—), because it takes into the accuracy of the ChaP estimates and adjusts the stepsize of the AEC accordingly. However, if ChaP/DR has received unreliable estimates from the AEC before, a small stepsize set by ChaP/DR can deteriorate the performance as can be seen at $t = 11\,\text{s}$. This case is correctly handled by dChaP/DR (—). By taking into account reliability information from the AEC, the reconvergence is fast and the misalignment of the AEC at BF changes can be further reduced. All in all, dChaP/DR significantly improves the BF-first structure by maintaining a system distance below $-17\,\text{dB}$ for $t > 8\,\text{s}$, irrespective of changes of the BF.

In the second scenario we investigate the performance of dChaP/DR under more challenging conditions with a time-variant RIR. For the near-end and far-end source a speech signal is used. Diffuse white noise is used for the near-end noise. The microphone array and the loudspeaker are moving together at a constant speed of $1\,\text{cm/s}$ through the room. The near-end source, and hence the steering directions of the BF, switches between three directions, $\varphi_{\text{BF}} = \{30°, 60°, 90°\}$, around the microphone array. At the bottom of Fig. 3 the azimuth φ_{BF} and elevation angle θ_{BF} of the near-end source (the BF) as well as the azimuth angle of the echo source φ_{echo} are shown over time. For this scenario the forgetting factor is set to $A = 0.997$. The reverberation time is set to $T_{60} = 0.3\,\text{s}$. The ESNR is set to 30 dB at the input of an MVDR BF and the SNR is set to 5 dB. In addition to the SysDis, we also measure the performance in terms of the echo return loss enhancement (ERLE). It is more related to

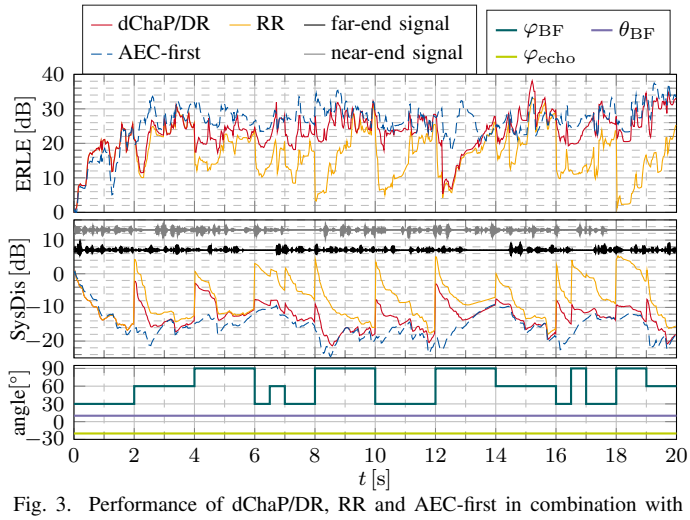

Fig. 3. Performance of dChaP/DR, RR and AEC-first in combination with an MVDR BF for a time-variant scenario and speech as far-end signal.

the subjective quality of the echo reduction and is calculated by $\mathrm{ERLE}(k)[\mathrm{dB}] = 10\log_{10}(E\{d_1^2(k)\}/E\{\tilde{d}^2(k)\})$, where $\tilde{d}(k)$ is the residual echo signal after BF and AEC. The expectation $E\{\cdot\}$ is estimated by recursive averaging [12]. The evaluation of the ERLE and SysDis in Fig. 3 shows that dChaP/DR consistently outperforms RR. Especially when the BF is steered to directions that have been seen before (e.g. $t = \{6, 8, 10\}$ s), dChaP/DR improves both, the ERLE and SysDis, by approx. 10 dB. Furthermore, dChaP/DR approaches the performance of AEC-first, which can be regarded as an upper bound.

We motivated the use of the BF-first structure with its lower complexity compared to the AEC-first structure, because only one AEC is needed instead of N_{mic} ones. However, the proposed dChaP/DR algorithm adds complexity to the BF-first structure. This is mainly caused by the calculation of the MPP of the $N_\Delta(\lambda) \times N_{\mathrm{mic}}$ matrix in (12), where $N_\Delta(\lambda)$ is the number of stored changes up to frame λ. This calculation can be done in $\mathcal{O}(mn^2)$, where $m = \max(N_\Delta(\lambda), N_{\mathrm{mic}})$ and $n = \min(N_\Delta(\lambda), N_{\mathrm{mic}})$ [13]. The number of stored changes $N_\Delta(\lambda)$ is implicitly limited by the Markov model in (21), because the corresponding row of $\boldsymbol{\Psi}\boldsymbol{\mathcal{W}}$ will become very small after some time determined by the forgetting factor A. In addition, we replace an old observation by a new one, if the distance between the corresponding BF weights is sufficiently small. It has to be noted that the MPP only has to be calculated in frames with changes of the BF. Therefore, the ratio between the number of frames with changes $N_\Delta(\lambda)$ and the total number of frames N_{frames} determines the complexity introduced by dChaP/DR. In addition one single channel diagonalized Kalman algorithm is used in the BF-first structure. Its complexity per frame can be approximated with $\mathcal{C}_{\mathrm{Kalman}} = \mathcal{O}(M) + \mathcal{O}(M\log_2 M)$ [14]. With these considerations the complexity per frame of dChaP/DR $\mathcal{C}_{\mathrm{ChaP}}$ and of AEC-first $\mathcal{C}_{\mathrm{AEC-first}}$ can be approximated by

$$\mathcal{C}_{\mathrm{ChaP}} = (M/2+1)\cdot\mathcal{O}(mn^2)\cdot N_\Delta(\lambda)/N_{\mathrm{frames}} + \mathcal{C}_{\mathrm{Kalman}}$$
$$\mathcal{C}_{\mathrm{AEC-first}} = N_{\mathrm{mic}}\cdot\mathcal{C}_{\mathrm{Kalman}}.$$

For the scenario presented in Fig. 3 we used $N_{\mathrm{mic}} = 13$, $N_{\mathrm{frames}} = 1250$ and $M = 1024$. Because only three different BF directions are occurring, $N_\Delta(\lambda) \leq 3$ is used. This results in a relative complexity of $\mathcal{C}_{\mathrm{ChaP}}/\mathcal{C}_{\mathrm{AEC-first}} \approx 10\,\%$.

VI. Conclusion

In this contribution we proposed a new system for combined BF and AEC. Starting from the low-complexity BF-first structure we introduced ChaP to prevent a large misalignment of the AEC at each change of the BF. With this the BF-first structure becomes applicable to realistic scenarios, where the BF typically is changed very abruptly. By extending ChaP to dChaP/DR we introduced reliability information for and from the AEC to better guide its adaptation. Furthermore, we used a first-order Markov model to account for time-variant RIRs in dChaP/DR. Finally we showed that the BF-first method, including the computational overhead for dChaP/DR, significantly reduces the complexity compared to AEC-first.

References

[1] Wolfgang Herbordt, *Sound Capture for Human/Machine Interfaces*, vol. 315 of *Lecture Notes in Control and Information Sciences*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.

[2] R. Martin, S. Gustafsson, and M. Moser, "Acoustic Echo Cancellation for Microphone Arrays using Switched Coefficient Vectors," in *International Workshop on Acoustic Echo and Noise Control*, London, Sept. 1997, pp. 160–163.

[3] M. H. Maruo, J. C. M. Bermudez, and L. S. Resende, "Statistical Analysis of a Jointly Optimized Beamformer-Assisted Acoustic Echo Canceler," *IEEE Transactions on Signal Processing*, vol. 62, no. 1, pp. 252–265, Jan. 2014.

[4] A. Cohen, A. Barnov, S. Markovich-Golan, and P. Kroon, "Joint Beamforming and Echo Cancellation Combining QRD Based Multichannel AEC and MVDR for Reducing Noise and Non-Linear Echo," in *2018 26th European Signal Processing Conference (EUSIPCO)*, Sept. 2018, pp. 6–10.

[5] M. Luis Valero and E. A. P. Habets, "Low-Complexity Multi-Microphone Acoustic Echo Control in the Short-Time Fourier Transform Domain," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 3, pp. 595–609, Mar. 2019.

[6] Gal Reuven, Sharon Gannot, and Israel Cohen, "Joint noise reduction and acoustic echo cancellation using the transfer-function generalized sidelobe canceller," *Speech Communication*, vol. 49, no. 7–8, pp. 623–635, July 2007.

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

[8] Gerald Enzner, *A Model-Based Optimum Filtering Approach to Acoustic Echo Control: Theory and Practice*, Dissertation, IND, RWTH Aachen, Aachen, Germany, Apr. 2006.

[9] Michael Brandstein and Darren Ward, Eds., *Microphone Arrays: Signal Processing Techniques and Applications*, Digital Signal Processing. Springer Berlin Heidelberg, Berlin, Heidelberg, 2001.

[10] Vladimir Tourbabin and Boaz Rafaely, "Theoretical Framework for the Optimization of Microphone Array Configuration for Humanoid Robot Audition," *IEEE/ACM Trans. Audio, Speech and Lang. Proc.*, vol. 22, no. 12, pp. 1803–1814, Dec. 2014.

[11] Olivier Roy and Martin Vetterli, "The effective rank: A measure of effective dimensionality," in *Signal Processing Conference, 2007 15th European*. 2007, pp. 606–610, IEEE.

[12] Peter Vary and Rainer Martin, *Digital speech transmission: Enhancement, coding and error concealment*, John Wiley & Sons, 2006.

[13] Lloyd N. Trefethen and David Bau, *Numerical linear algebra*, Society for Industrial and Applied Mathematics, Philadelphia, 1997.

[14] Sarmad Malik and Gerald 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, Sept. 2012.