EFFICIENT MULTICHANNEL NONLINEAR ACOUSTIC ECHO CANCELLATION BASED ON A COOPERATIVE STRATEGY

Mhd Modar Halimeh and Walter Kellermann

Multimedia Communications and Signal Processing. Friedrich-Alexander-Universität Erlangen-Nürnberg, Cauerstr. 7, D-91058 Erlangen, Germany. {mhd.m.halimeh, walter.kellermann}@fau.de

ABSTRACT

While a common approach to address nonlinear distortions, emitted by multiple loudspeakers and observed by multiple microphones, is to use post-filtering techniques, this paper proposes a cooperative strategy to rather model and then cancel such distortions. In this approach, the overall problem of modeling distortions emitted by a number of loudspeakers is divided into multiple simpler and easier tasks of estimating distortions emitted by subsets of loudspeakers. This approach allows also the exploitation of the physical configuration of the loudspeakers and microphones to select certain microphone signals for estimating the nonlinearity of loudspeakers that contribute the predominant part of the acoustic echo to this microphone signal. The proposed strategy is realized using the elitist resampling particle filter and the Gaussian particle filter. Both variants are evaluated and compared to a linear approach using synthesized and real recordings.

Index Terms— Multichannel NAEC, Mutlichannel nonlinear system identification, Particle filtering

1. INTRODUCTION

Acoustic Echo Cancellation (AEC) is a topic that has been studied extensively in the last few decades [1, 2]. This interest has been motivated by the need for full-duplex communication channels in telephone systems. As a result of the wide-spread use of mobile devices, smart speakers, and teleconferencing systems, a typical AEC scenario nowadays includes multiple loudspeakers driven at distortions-inducing levels and observed by multiple microphones.

Many linear multichannel AEC approaches are proposed in the literature. [3] introduced a multichannel extension of the frequency domain adaptive filtering algorithm [4], while [5, 6] proposed a frequency domain-based approximation of the recursive least squares algorithm. In [7], an efficient multichannel AEC approach is introduced by utilizing the knowledge of the rendering system driving the loudspeakers. [8] uses a beamformer coupled with a linear acoustic echo canceler to suppress both the signal echo and distortions.

The authors would like to thank the Deutsche Forschungsgemeinschaft (DFG) for supporting this work (contract number KE 890/9-1).

Nonlinear AEC (NAEC) has also been, and continues to be, studied extensively. A typical approach for single-channel NAEC consists of modeling the loudspeaker nonlinearities by a parametric nonlinear preprocessor that uses a set of basis functions to approximate the underlying nonlinear function, e.g., [9–14]. To optimize the parameters of the nonlinear preprocessor, many approaches have been proposed [15–17]. On the other hand, for multichannel sound rendering, no methods are known that model the nonlinearities of the individual loudspeakers. Therefore, an obvious approach to address nonlinear distortions in multichannel scenarios would be to apply postfilters to each microphone signal as in, e.g., [18, 19].

In this paper, we propose a novel and efficient online method to indeed model the nonlinear behavior of individual loudspeakers in multichannel sound rendering systems. The computational efficiency is achieved by exchanging information between different estimators. This approach is verified for two particle filtering methods for identifying nonlinear systems, the Elitist Resampling Particle Filter (ERPF) [20] and the Gaussian Particle Filter (GPF) [21]. Both variants are compared to a linear approach for synthesized and real recordings.

2. PROBLEM STATEMENT

The multichannel NAEC problem is depicted in Fig. 1. At time instant n, each of the far-end signals $\{s_i(n)\}_{i=1}^{N}$ is first nonlinearly distorted and emitted by the corresponding loudspeaker. The nonlinearly distorted signals $\{d_i(n)\}_{i=1}^{N}$ are then propagated through the acoustic enclosure, which is characterized by the block matrix $\mathbf{H}$ that represents the linear transmission characteristics as expressed by the impulse responses between each loudspeaker and microphone. Afterwards, the distorted signals are picked up by the M microphones as signals $\{y_j(n)\}_{j=1}^{M}$.

In order to approximate the nonlinearly distorted signal $d_i(n)$ produced by the i-th loudspeaker, a parametric Memoryless Nonlinear Preprocessor $(\text{MNLP})_i$ is used such that

$$\hat{d}_i(n) = \sum_{l=1}^{L_a} \hat{a}_{i,l} \cdot f_{i,l}(s_i(n)) = \hat{\boldsymbol{a}}_i^{\mathrm{T}} \boldsymbol{f}_i(s_i(n)), \qquad (1)$$

where the L_a nonlinearly distorted signals captured by the

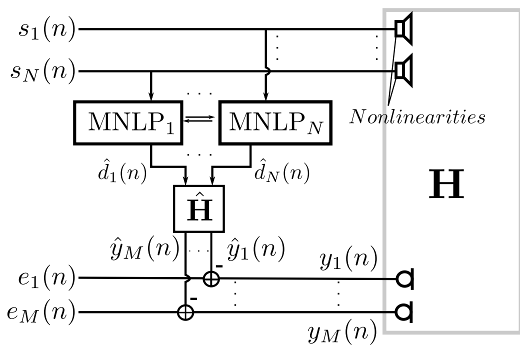

Fig. 1. The multichannel AEC problem.

vector $\boldsymbol{f}_i(s_i(n)) = [f_{i,1}(s_i(n)), ..., f_{i,L_a}(s_i(n))]^{\mathrm{T}}$ are scaled by the parameter vector $\hat{\boldsymbol{a}}_i = [\hat{a}_{i,1}, ..., \hat{a}_{i,L_a}]^{\mathrm{T}}$. The distorted signals $\{\hat{d}_i(n)\}_{i=1}^{N}$ are then passed to the $N \times M$ block matrix $\hat{\mathbf{H}}(n)$, which has the elements $\hat{\boldsymbol{h}}_{i,j} = [\hat{h}_{i,j}(n), ..., \hat{h}_{i,j}(n-L+1)]^{\mathrm{T}}$ denoting a L-tap Finite Impulse Response (FIR) filter modeling the echo path between the i-th loudspeaker and j-th microphone. Finally, the estimated microphone signals $\{\hat{y}_j(n)\}_{j=1}^{M}$ are given by

$$\hat{y}_j(n) = \sum_{i=1}^{N} \hat{\boldsymbol{h}}_{i,j}^{\mathrm{T}} \hat{\boldsymbol{d}}_i(n), \tag{2}$$

with $\hat{\boldsymbol{d}}_i(n) = \left[\hat{d}_i(n), ..., \hat{d}_i(n-L+1)\right]^{\mathrm{T}}$ denoting the estimated nonlinearly distorted signal vector.

3. MULTICHANNEL NONLINEAR ACOUSTIC ECHO CANCELLATION

3.1. Estimating The Linear Subsystem

To estimate the linear part of the system, i.e., the matrix $\hat{\mathbf{H}}$, the unconstrained version of the Generalized Frequency Domain Adaptive Filtering (GFDAF) algorithm is used [5, 6]. However, this choice is not unique to the proposed approach and the GFDAF can be replaced by any other linear system identification method such as [3, 4, 7]. In this subsection, the unconstrained GFDAF is reviewed and described briefly.

Let $\hat{\mathbf{D}}_i(k) = [\hat{\boldsymbol{d}}_i(nk), ..., \hat{\boldsymbol{d}}_i(nk-L+1)]^{\mathrm{T}}$ denote the i-th channel estimated distorted signal Toeplitz matrix, with k denoting the block index. The block error signal for the j-th microphone is then defined as

$$\boldsymbol{e}_j(k) = \boldsymbol{y}_j(k) - \sum_{i=1}^{N} \mathbf{D}_i^{\mathrm{T}}(k) \hat{\boldsymbol{h}}_{i,j}. \tag{3}$$

Let $\mathbf{F}$ denote the discrete Fourier transform matrix of size $2L \times 2L$. Then we define the following quantities

$$\underline{\boldsymbol{e}}_j(k) = \mathbf{F} \begin{bmatrix} \mathbf{0}_{L\times 1} \\ \boldsymbol{e}_j(k) \end{bmatrix}, \qquad \underline{\boldsymbol{y}}_j(k) = \mathbf{F} \begin{bmatrix} \mathbf{0}_{L\times 1} \\ \boldsymbol{y}_j(k) \end{bmatrix},$$
$$\mathbf{W} = \begin{bmatrix} \mathbf{0}_{L\times L} & \mathbf{0}_{L\times L} \\ \mathbf{0}_{L\times L} & \mathbf{I}_{L\times L} \end{bmatrix}, \qquad \hat{\underline{\boldsymbol{h}}}_{i,j} = \mathbf{F} \begin{bmatrix} \hat{\boldsymbol{h}}_{i,j} \\ \mathbf{0}_{L\times 1} \end{bmatrix},$$
$$\mathbf{G} = \mathbf{F}\mathbf{W}\mathbf{F}^{-1}.$$

Let $\widehat{\mathbf{D}}_i$ denote a diagonal matrix whose diagonal elements are the discrete Fourier transform of the last two frames of the signal $\hat{\boldsymbol{d}}_i$. Then by using the above definitions, (3) is written in the frequency domain as

$$\underline{\boldsymbol{e}}_j(k) = \underline{\boldsymbol{y}}_j(k) - \mathbf{G} \sum_{i=1}^{N} \widehat{\mathbf{D}}_i(k)^{\mathrm{T}} \hat{\underline{\boldsymbol{h}}}_{i,j} = \underline{\boldsymbol{y}}_j(k) - \mathbf{G}\widehat{\mathbf{D}}(k)\hat{\underline{\boldsymbol{h}}}_j,$$

where $\widehat{\mathbf{D}}(k) = [\widehat{\mathbf{D}}_1(k), ..., \widehat{\mathbf{D}}_N(k)]$ is the matrix containing all the distorted signals diagonal matrices while $\hat{\underline{\boldsymbol{h}}}_j = [\hat{\underline{\boldsymbol{h}}}_{1,j}^{\mathrm{T}}, ..., \hat{\underline{\boldsymbol{h}}}_{N,j}^{\mathrm{T}}]^{\mathrm{T}}$ denotes the transfer functions associated with the j-th microphone which are recursively estimated via

$$\hat{\underline{\boldsymbol{h}}}_j(k) = \hat{\underline{\boldsymbol{h}}}_j(k-1) + \mu \mathbf{S}^{-1}(k) \widehat{\mathbf{D}}^{\mathrm{H}}(k) \underline{\boldsymbol{e}}_j(k), \tag{4}$$

with μ denoting the adaptation step size. $\mathbf{S}(k)$ denotes a matrix of diagonal block matrices each containing estimates for power and cross-power spectra of the far-end signals [5, 6].

3.2. The Cooperative Multichannel NAEC Strategy

Unlike estimating the impulse responses, when estimating the parameters of the MNLPs, the different microphones observe the same nonlinear system to be identified. As a consequence, a naive approach of employing $N \times M$ MNLPs yields an inefficient class of algorithms due to the redundancy in the estimation task. In other words, a single universal set of N MNLPs would model the system sufficiently. This estimation problem is summarized as estimating the parameters matrix

$$\hat{\mathbf{A}} = [\hat{\boldsymbol{a}}_1, ..., \hat{\boldsymbol{a}}_N] = \begin{bmatrix} \hat{a}_{1,1} & \cdots & \hat{a}_{N,1} \\ \vdots & \ddots & \vdots \\ \hat{a}_{1,L_a} & \cdots & \hat{a}_{N,L_a} \end{bmatrix}. \tag{5}$$

Estimating the parameter vectors jointly, i.e., the matrix $\hat{\mathbf{A}}$, is a computationally expensive task. Alternatively, one can partition the matrix $\hat{\mathbf{A}}$ into U submatrices $\hat{\mathbf{A}}_1, ..., \hat{\mathbf{A}}_U$. The submatrix $\hat{\mathbf{A}}_u$ contains the parameter vectors characterizing the distortions emitted by the u-th loudspeaker set $\Theta_u = \{i_u, ..., I_u\}$ where i_u and I_u denote the lowest and the largest index of the used loudspeakers in the u-th subset, respectively. Since the different loudspeaker nonlinearities are mutually independent, the overall model accuracy is not altered by this partitioning of the estimation task. The partitioning is done such that $\cup_{u=1}^{U} \Theta_u = \{1, ..., N\}$ and $\cap_{u=1}^{U} \Theta_u = \emptyset$. Each submatrix $\hat{\mathbf{A}}_u = [\hat{\boldsymbol{a}}_{i_u}, ..., \hat{\boldsymbol{a}}_{I_u}]$ is adapted by a separate estimator that uses the j_u-th microphone signal. Consequently, compared to the naive approach, the proposed method is $\frac{N \times M}{U}$ times more efficient. This partitioning of parameter vectors can be further motivated when considering the positions of the microphones and loudspeakers, which can lead to a certain microphone signal capturing mainly the nonlinear distortions produced by a specific loudspeaker or a subset of loudspeakers. Adaptive microphone assignment is also possible by considering, e.g., coherence measures [22]. Each of the U estimators is characterized by the latent parameter submatrix $\hat{\mathbf{A}}_u = [\hat{\boldsymbol{a}}_{i_u}, ..., \hat{\boldsymbol{a}}_{I_u}]$ specifying the set of loudspeakers tracked by this estimator, and the error signal $\boldsymbol{e}_{j_u}$ used for adapting the nonlinearities of subset u. Using separate estimators for different MNLPs requires the estimators to share their estimates. This is due to the fact that in order to evaluate the error signal $\boldsymbol{e}_{j_u}$, one needs the entire matrix $\hat{\mathbf{A}}$. Therefore, estimates produced by one estimator are passed as prior knowledge to the other estimators and the quality of one estimate is directly related to the quality of the other estimates. Hence, the proposed strategy is denoted by the Cooperative Multichannel AEC (CM-AEC). Due to their capability to track time-varying nonlinear systems, the CM-AEC is realized using the ERPF [20], denoted (CM-ERPF), and the Gaussian Particle Filter (GPF) [21], denoted (CM-GPF), with the emphasis being on the ERPF due to its superior performance.

3.3. The Elitist Resampling Particle Filter

In this subsection, the realization of one ERPF [16, 20] unit, out of U in total, is detailed.

Let $\hat{\boldsymbol{z}}_u$ denote the latent variable constructed by a vector concatenation of the submatrix $\hat{\mathbf{A}}_u$, i.e., $\hat{\boldsymbol{z}}_u = [\hat{\boldsymbol{a}}_{i_u}^{\mathrm{T}}, ..., \hat{\boldsymbol{a}}_{I_u}^{\mathrm{T}}]^{\mathrm{T}}$. For each block k, the ERPF approximates the posterior density by a set of Q_u particles and their associated weights $\{\hat{\boldsymbol{z}}_u^{(q)}(k), w_u^{(q)}(k)\}_{q=1}^{Q_u}$. The ERPF considers two subsets of particles, an elitist subset Φ_{E} and non-elitist subset Φ_{NE}, based on their weights

$$\hat{\boldsymbol{z}}_u^{(q)}(k) \in \begin{cases} \Phi_{\mathrm{E}} & \text{if } w_u^{(q)}(k) \geq w_{\mathrm{th}}, \\ \Phi_{\mathrm{NE}} & \text{if } w_u^{(q)}(k) < w_{\mathrm{th}}, \end{cases} \tag{6}$$

with w_{th} denoting a weight threshold separating the two subsets. At each block k, the particles' weights are updated

$$w_u^{(q)}(k) = \begin{cases} p(\boldsymbol{y}_{j_u}(k)|\hat{\boldsymbol{z}}_u^{(q)}(k)) & \text{if } \hat{\boldsymbol{z}}_u^{(q)}(k) \in \Phi_{\mathrm{NE}}, \\ w_u^{(q)}(k-1)p(\boldsymbol{y}_{j_u}(k)|\hat{\boldsymbol{z}}_u^{(q)}(k)) & \text{if } \hat{\boldsymbol{z}}_u^{(q)}(k) \in \Phi_{\mathrm{E}} \end{cases} \tag{7}$$

The likelihood $p(\boldsymbol{y}_{j_u}(k)|\hat{\boldsymbol{z}}_u^{(q)}(k))$ is evaluated by $\mathcal{N}(\overline{e}_{j_u}^{(q)}(k)|\sigma_u^2)$, where $\overline{e}_{j_u}^{(q)}(k)$ is the mean of the block error signal, while $\sigma_{j_u}^2$ is the noise variance.

To evaluate $\boldsymbol{e}_{j_u}^{(q)}(k)$, estimates of the nonlinearly distorted signals (1) are calculated. As seen in (1), a full estimate of $\hat{\mathbf{A}}^{(q)}$, which $\hat{\boldsymbol{z}}_u^{(q)}(k)$ represents a submatrix of, is needed. Therefore, assuming that the overall system consists of U estimators, i.e., particle filters, each estimating a submatrix similar to $\hat{\boldsymbol{z}}_u^{(q)}(k)$, a complete coefficients matrix $\hat{\mathbf{A}}_u^{(q)}$ is obtained using $\hat{\boldsymbol{z}}_u^{(q)}(k)$ and $\{\hat{\boldsymbol{z}}_{u_o}(k-1)\}_{u_o \neq u}$, which are the estimates produced by the other $U-1$ particles filters, yielding an estimate of the nonlinearly distorted signal

$$\hat{d}_i^{(q)}(n) = \begin{cases} \hat{\boldsymbol{a}}_i^{(q)\mathrm{T}} \boldsymbol{f}_i(s_i(n)); \hat{\boldsymbol{a}}_i^{(q)} \xrightarrow{\mathrm{sv}} \hat{\boldsymbol{z}}_u^{(q)} \text{if } \ i \in \{i_u, ..., I_u\}, \\ \hat{\boldsymbol{a}}_i^{(q)\mathrm{T}} \boldsymbol{f}_i(s_i(n)); \hat{\boldsymbol{a}}_i^{(q)} \xrightarrow{\mathrm{sv}} \hat{\boldsymbol{z}}_{u_o} \text{if } \ i \notin \{i_u, ..., I_u\}, \end{cases} \tag{8}$$

where $\xrightarrow{\mathrm{sv}}$ denotes a vector being a subvector of another.

After estimating $\hat{d}_i^{(q)}(n)$, an estimate of the microphone signal $\hat{\boldsymbol{y}}_{j_u}^{(q)}(k)$ is obtained in (2) by using the estimated impulse responses provided by the GFDAF. Finally, the error signal block is obtained by (3). The particles' association is then updated by (6), and while the non-elitist particles are discarded, the elitist particles are preserved and used to construct the density $\mathcal{N}(\overline{\boldsymbol{z}}_u(k), \Sigma_u(k))$ characterized by the mean vector $\overline{\boldsymbol{z}}_u(k)$ and the covariance matrix $\Sigma_u(k)$. Finally, a set of substitution particles are drawn from the constructed density.

4. EXPERIMENTS

In this section, the two realizations of the proposed CM-AEC using the ERPF (CM-ERPF) and using the GPF (CM-GPF) are compared to the purely linear GFDAF. The different approaches are evaluated in two experiments, the first uses synthesized nonlinear distortions, while the other uses real recording done using a communication device. As a performance measure, the Echo Return Loss Enhancement (ERLE), across the multiple channels, is employed [23]

$$\mathrm{ERLE}(n) = 10\log_{10}\left(\frac{1}{M}\sum_{j=1}^{M}\frac{\mathrm{E}\{y_j^2(n)\}}{\mathrm{E}\{e_j^2(n)\}}\right), \tag{9}$$

where $\mathrm{E}(\cdot)$ refers to the expectation operator.

In both experiments, the FIR filters used to approximate the impulse responses are 1024 taps long.

4.1. Synthesized Distortions

In this experiment, a one-meter-long uniform linear array consisting of five loudspeakers is placed at a distance of 30cm from two microphones which are 10cm apart. The loudspeakers are driven by rendering a virtual source impinging from 210° on a control region centered on the microphone array using pressure matching [24]. The source signal is female speech. In order to simulate the nonlinear distortions in the loudspeakers, the loudspeaker signals are distorted by [25]

$$d_i(n) = \frac{2}{4i}\tanh(4\frac{i}{2}s_i(n)), \quad i = 1, ..., 5, \tag{10}$$

The nonlinearly distorted signals are convolved with simulated impulse responses [26], with a reverberation time $T_{60} = 450$ms, to generate the microphone signals

$$y_j(n) = \sum_{i=1}^{N} \boldsymbol{h}_{i,j}^{\mathrm{T}} \boldsymbol{d}_i(n); \quad j = 1, 2\ , \ i = 1, ..., 5. \tag{11}$$

The microphone signals are then corrupted by an additive white Gaussian noise resulting in a Signal-to-Noise Ratios (SNR) of $\{30, 20, 10\}$dB.

In all the different algorithms, a step size of $\mu = 0.02$ is used in (4). The CM-ERPF and CM-GPF use MNLPs that employ the first three Legendre polynomials of odd orders [13], i.e., $L_a = 3$ in (1).

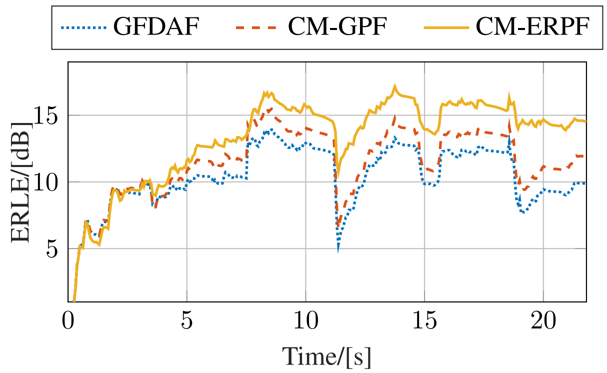

Fig. 2. ERLE [dB] for synthesized distortions at SNR = 30dB.

Experiment	SNR [dB]	GFDAF	MGPF	MERPF
Synthesized	30	10.4	11.5	13.4
	20	9.8	10.9	11.8
	10	7.1	7.4	8.1
Real	30	18.1	19.0	20.2

Table 1. Average ERLE [dB] for Sec. 4.1 and 4.2.

Regarding the concatenation of the estimation task, the CM-ERPF and CM-GPF use two estimators, i.e., $U = 2$ particle filters, where the first tracks the nonlinearities emitted by the loudspeakers $\{1, 2, 3\}$, i.e., $\hat{\mathbf{A}}_1 = [\hat{\boldsymbol{a}}_1, \hat{\boldsymbol{a}}_2, \hat{\boldsymbol{a}}_3]$, using the microphone signal $y_1(n)$. The second particle filter estimates the distortions of the loudspeakers $\{4, 5\}$, i.e., $\hat{\mathbf{A}}_2 = [\hat{\boldsymbol{a}}_4, \hat{\boldsymbol{a}}_5]$, using the microphone signal $y_2(n)$. Finally, each particle filter in the CM-ERPF and CM-GPF uses $Q_u = 60$ particles.

The ERLE values produced by the different algorithms at an SNR level of 30dB are depicted in Fig. 2. As seen by the results, the CM-AEC does outperform the purely linear GFDAF algorithm. Moreover, the CM-ERPF outperforms the CM-GPF considerably, especially after the convergence phase, i.e., $t > 5$s. Furthermore, By examining the average ERLE values across the different SNR levels in Table 1, we can see consistently better results for the different CM-AEC realizations compared to the purely linear GFDAF, indicating a high degree of robustness the CM-AEC has against unfavorable noisy conditions. The advantageous performance exhibited by the CM-ERPF in comparison to the CM-GPF is also consistent across the considered SNR range.

4.2. Real Recorded Distortions

The proposed approach is also evaluated for real recorded nonlinear distortions. The recordings are done using a circular smart speaker with 8 loudspeakers and 7 microphones. A 15 seconds-long music signal is emitted by the 8 loudspeakers simultaneously at an average sound pressure level of 94.1 dB. The microphone signals are recorded at a sampling frequency of 16kHz and an SNR level of around 35dB.

The different algorithms use a step size $\mu = 0.06$ in (4) to estimate the impulse responses. Moreover, the CM strategy uses MNLPs employing the first three Legendre polynomials of odd order [13], i.e., $L_a = 3$ in (1), to model the nonlinear

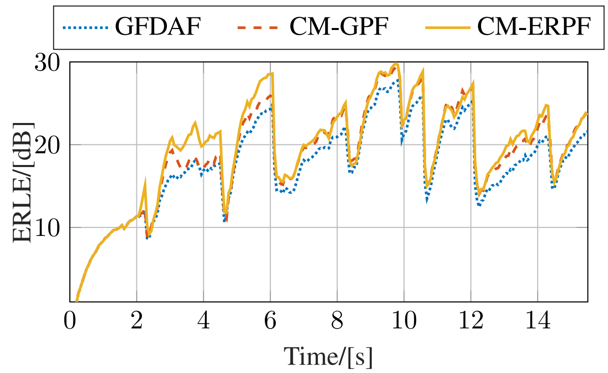

Fig. 3. ERLE [dB] for real recorded nonlinearities

distortions emitted by the different loudspeakers. The CM-GPF and the CM-ERPF use three particle filters to adapt the 8 MNLPs, each MNLP modeling a loudspeaker. The first particle filter adapts the MNLPs modeling the loudspeakers $\{1, 2, 3, 4\}$. The second adapts MNLPs $\{5, 6\}$, while the third tracks the distortions from loudspeakers $\{7, 8\}$.

The ERLE values resulting from the purely linear GFDAF and both variants, the CM-ERPF and CM-GPF, are depicted in Fig. 3. As seen in the figure, both CM-AEC realizations are consistent in outperforming the purely linear GFDAF, which indicates that the CM-AEC successfully cancels the nonlinear distortions emitted by the loudspeakers. The difference between the purely linear GFDAF and the CM-AEC realizations reaches a maximum of 5dB at $t \approx 4$s, indicating fast convergence w.r.t. estimating the nonlinear distortions. Large gains are also observed for distortions that occur after the linear part's convergence, e.g., at $t \approx 6$s, which is of significant importance, since such distortions are most noticeable.

The average ERLE values, averaged over 100 Monte Carlo runs, are provided in Table 1, where the CM-ERPF achieves ERLE levels that are on average 2dB higher than the linear GFDAF, and that are 1dB higher than those achieved by the CM-GPF. The consistency in which the CM-ERPF outperforms the CM-GPF is explained by the memory effect seen in the weights calculations in (7) (see, e.g., [20]).

5. CONCLUSIONS

In this paper, the cooperative multichannel nonlinear acoustic echo cancellation strategy was proposed. The CM-AEC exploits the fact that the several microphones are observing the same set of nonlinear systems to concatenate the estimation task into several simpler ones. This concatenation introduces the need of the several estimator to communicate with each other, i.e., cooperate. Beside being computationally more efficient, the concatenation also enables the exploitation of the physical distribution of the microphones and loudspeakers, by allowing certain microphones to track distortions emitted by specific loudspeakers only. The CM-AEC is realized using the ERPF (CM-ERPF) and GPF (CM-GPF), which are then verified and compared to a purely linear GFDAF using synthesized and recorded signals.

6. REFERENCES

[1] M. Sondhi, "An adaptive echo canceller," *Bell Syst. Tech. J.*, vol. 46, no. 3, pp. 497–511, 1967.

[2] C. Breining *et al.*, "Acoustic echo control - An application of very-high-order adaptive filters," *IEEE Signal Process. Mag.*, vol. 16, no. 4, pp. 42–69, July 1999.

[3] J. Benesty and D. R. Morgan, "Frequency-domain adaptive filtering revisited, generalization to the multichannel case, and application to acoustic echo cancellation," in *ICASSP*, Istanbul, Turkey, June 2000, vol. 2, pp. II789–II792 vol.2.

[4] J. J. Shynk, "Frequency-domain and multirate adaptive filtering," *IEEE Signal Processing Magazine*, vol. 9, no. 1, pp. 14–37, Jan. 1992.

[5] M . Schneider and W. Kellermann, "The generalized frequency-domain adaptive filtering algorithm as an approximation of the block recursive least-squares algorithm," *EURASIP Journal on Advances in Signal Processing*, vol. 1, pp. 6, Jan. 2016.

[6] H. Buchner *et al.*, "Generalized multichannel frequency-domain adaptive filtering: efficient realization and application to hands-free speech communication," *Signal Processing*, vol. 85, no. 3, pp. 549 – 570, 2005.

[7] C. Hofmann and W. Kellermann, "Source-specific system identification," in *ICASSP*, Shanghai, China, Mar. 2016, pp. 161–165.

[8] A. Cohen *et al.*, "Joint beamforming and echo cancellation combining QRD based multichannel aec and MVDR for reducing noise and non-linear echo," in *European Signal Processing Conference, EUSIPCO*, Rome, Italy, Sep. 2018, pp. 6–10.

[9] F. Kuech *et al.*, "Nonlinear acoustic echo cancellation using adaptive orthogonalized power filters," in *ICASSP*, Philadelphia, PA, USA, Mar. 2005, pp. 105–108.

[10] C. Huemmer *et al.*, "The elitist particle filter based on evolutionary strategies as novel approach for nonlinear acoustic echo cancellation," in *ICASSP*, Florence, Italy, May 2014, pp. 1315–1319.

[11] A. Stenger *et al.*, "Nonlinear acoustic echo cancellation with 2nd order adaptive Volterra filters," in *ICASSP*, Phoenix, AZ, USA, Mar. 1999, pp. 877–880.

[12] M. Zeller *et al.*, "Adaptive Volterra filters with evolutionary quadratic kernels using a combination scheme for memory control," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1449–1464, Apr. 2011.

[13] A. Carini *et al.*, "Introducing Legendre nonlinear filters," in *ICASSP*, Florence, Italy, May 2014, pp. 7939–7943.

[14] D. Comminiello *et al.*, "Functional link adaptive filters for nonlinear acoustic echo cancellation," *IEEE Trans. Audio, Speech, Language Process.*, vol. 21, no. 7, pp. 1502–1512, July 2013.

[15] E. Ferrara, "Fast implementations of LMS adaptive filters," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 28, no. 4, pp. 474–475, Aug. 1980.

[16] C. Huemmer *et al.*, "Estimating parameters of nonlinear systems using the elitist particle filter based on evolutionary," *IEEE/ACM Trans. on Audio, Speech and Language Process.*, vol. 26, pp. 595–608, 2018.

[17] C. Hofmann *et al.*, "Significance-aware Hammerstein group models for nonlinear acoustic echo cancellation," in *ICASSP*, Florence, Italy, May 2014, pp. 5934–5938.

[18] A. Schwarz *et al.*, "Spectral feature-based nonlinear residual echo suppression," in *2013 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, USA, Oct. 2013, pp. 1–4.

[19] A. Schwarz *et al.*, "Combined nonlinear echo cancellation and residual echo suppression," in *Speech Communication; 11. ITG Symposium*, Erlangen, Germany,, Sep. 2014, pp. 1–4.

[20] M. M. Halimeh *et al.*, "Nonlinear acoustic echo cancellation using elitist resampling particle filter," in *ICASSP*, April 2018, pp. 236–240.

[21] J. Kotecha and P. Djuric, "Gaussian particle filtering," *IEEE Trans. Signal Process.*, vol. 51, no. 10, pp. 2592–2601, Oct. 2003.

[22] A. Schwarz and W. Kellermann, "Unbiased coherent-to-diffuse ratio estimation for dereverberation," in *IWAENC*, Juan-les-Pins, France, Sep. 2014, pp. 6–10.

[23] E. Haensler and G. Schmidt, *Topics in acoustic echo and noise control*, Springer-Verlag, Berlin Heidelberg, 2006.

[24] M. Poletti, "An investigation of 2-D multizone surround sound systems," in *Proc. 125th Audio Eng. Soc. (AES) Conv.*, 2008.

[25] C. Huemmer *et al.*, "The significance-aware EPFES to estimate a memoryless preprocessor for nonlinear acoustic echo cancellation," in *GlobalSIP*, Atlanta, GA, USA, Dec. 2014, pp. 557–561.

[26] E. Habets, "Room impulse response generator," Tech. Rep., Technische Universiteit Eindhoven, 2010.