Tracking Theory of Adaptive Filters with Input-Output Sampling Rate Offset

Philipp Thüne and Gerald Enzner
Ruhr-Universität Bochum, Department of Electrical Engineering and Information Technology,
Institute of Communication Acoustics (IKA), D-44780 Bochum, Germany
Email: {philipp.thuene, gerald.enzner}@rub.de

***Abstract*—Adaptive filter theory for supervised identification of linear time-invariant (LTI) systems is an established and fruitful discipline in digital signal processing. In certain applications, however, the input and output signals of an LTI system may be asynchronously sampled at slightly different sampling frequencies resulting in a small input-output sampling rate offset (IO-SRO). In this contribution, we argue that an LTI system with IO-SRO is seen as a linear time-variant system by the adaptive filter. By conducting a convergence-in-the-mean analysis, we propose a model to capture the influence of IO-SRO on the tracking properties of the adaptive filter. Eventually, we validate our model by reconstruction of the IO-SRO based on the proposed model and the observable adaptive filter behavior. The model-based IO-SRO reconstruction turns out to be highly precise and robust against noise and excitation bandwidth limitations when compared to a state-of-the-art method.**

***Index Terms*—Adaptive filter theory, supervised system identification, sampling rate offset, asynchronous sampling.**

I. Introduction

Adaptive filter theory has been a topic of active research in digital signal processing for several decades [1]–[3] and the use of adaptive filters for supervised system identification has found numerous successful applications in acoustic echo cancellation (AEC) for hands-free communication [4]–[8], feedback cancellation, e.g., in public address systems [5], [9], and network echo cancellation [4]. In the vast majority of cases, the system under consideration is modeled in the digital domain entirely. For the identification of continuous-time physical models under clock-skew, however, the effects of analog-to-digital (A/D) and digital-to-analog (D/A) conversion have to be taken into account. Even if quantization can be neglected due to sufficient resolution, the discrete-time sampling itself may present a challenge if D/A and A/D conversion is asynchronous, i.e., not controlled by the same oscillator. The ensuing input-output sampling rate offset (IO-SRO) may often be very small, i.e., in the order of parts per million (ppm), yet the effect can be detrimental in several applications [10], [11].

While the issue of sampling rate offset in general has recently become the focus of research in applications of distributed acoustic sensing [12]–[15], adaptive filter performance with IO-SRO has mostly been tackled in the area of acoustic interference cancellation [16] and AEC. While the authors in [17] mostly consider the problem of sampling rate conversion for deliberate multi-rate AEC, a time-domain algorithm for error-signal-based estimation and compensation of the small, yet unknown IO-SRO is proposed in [11]. Another approach to asynchronous frequency-domain AEC (async-FDAEC) is presented in [18]. A fundamental analysis of the behavior of the adaptive filter under the influence of IO-SRO is, however, not available in the respective literature.

In this contribution, we argue that an analog linear time-invariant (LTI) system with two independent and asynchronous D/A and A/D converters can be described in the time-discrete domain as a linear time-variant (LTV) system comprised of a digital-to-digital (D/D) converter and the time-discrete representation of the original LTI system. An adaptive filter designed for system identification based on the asynchronously sampled input and output signals will thus attempt to track the composite LTV system. We derive a model for this tracking behavior of the adaptive filter in a convergence-in-the-mean analysis of the filter coefficients. The model eventually suggests a time-invariant *lag filter* that captures the tracking bias in magnitude and phase. As the IO-SRO is among the main constituent factors of the lag filter's phase, we can validate our theoretical model by reconstruction of the underlying IO-SRO, for instance, by fitting the model phase to that of the cross-power spectral density (CPSD) between the system output and the adaptive filter output. While our focus is on the convergence analysis and the ensuing model for adaptive filtering with IO-SRO, the extracted IO-SRO information would need to be compensated, e.g., via resampling [19]–[21] in order to obtain a fully functional AEC or interference cancellation system. The investigation of IO-SRO compensation is, however, beyond the scope of this paper.

Our contribution is structured as follows. In Sec. II, we derive the equivalent LTV representation of an LTI system with IO-SRO before analyzing the tracking of the adaptive filter culminating in the model of the respective time-invariant lag filter. Sec. III then describes a procedure for model verification by reconstructing the IO-SRO based on the observable signals. Computer experiments to validate our model are presented in Sec. IV before we conclude our contribution in Sec. V. Throughout this paper, capital letters denote frequency-domain quantities while bold and bold underlined characters indicate vectors and matrices, respectively.

This work has been supported by *Deutsche Forschungsgemeinschaft* (DFG) under grant EN 869/3-1 within the Research Unit FOR2457 "Acoustic Sensor Networks".

978-9-0827-9703-9/19/$31.00 ©2019 IEEE

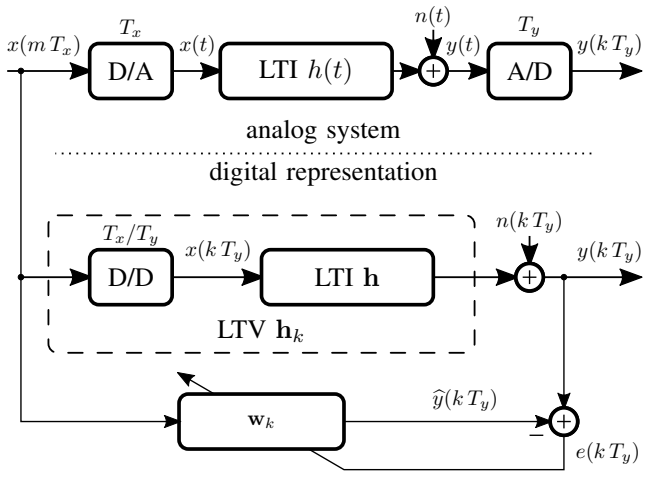

Fig. 1. Analog and digital representation of a linear system with IO-SRO and supervised system identification by an adaptive filter.

II. Adaptive Filters with Input-Output SRO

A. Equivalent System Model with Input-Output SRO

We first consider the digital-to-digital signal model in Fig. 1. A signal $x(m\,T_x)$ at discrete time instants m with sampling interval T_x is converted into an analog signal $x(t)$ to be processed by an analog LTI system $h(t)$. The system output is corrupted by additive noise $n(t)$ before the resulting signal $y(t)$ is converted into a time-discrete signal $y(k\,T_y)$ at time instant k and sampling interval T_y. The mismatching sampling intervals introduce a small IO-SRO $\varepsilon = T_y/T_x - 1$. Note that, due to the different sampling intervals, all signals here are represented with a continuous time scale.

As depicted in the middle of Fig. 1, we can represent the system in an all-digital form by shifting the A/D converter across the LTI system and merging the mismatched D/A and A/D converters. In this digital representation, the resulting D/D converter [19] adjusts the sampling interval T_x to T_y. Accordingly, the new input signal $x(k\,T_y)$ of the discrete-time LTI system $\mathbf{h}$ is modeled as the output of an ideal resampler via sinc interpolation [19], i.e.,

$$x(k\,T_y) = \sum_m x(m\,T_x)\,\mathrm{sinc}(k\,(1+\varepsilon) - m)\,, \tag{1}$$

where we have assumed $T_x > T_y$ without loss of generality. With input $x(k\,T_y)$ and output $y(k\,T_y)$ at the same sampling frequency, we then describe the output of the LTI system via a time-discrete convolution

$$\begin{aligned} y(k\,T_y) &= \sum_l x\left((k-l)\,T_y\right)\, h(l\,T_y) \\ &= \sum_l \sum_m x(m\,T_x)\,\mathrm{sinc}\left((k-l)(1+\varepsilon) - m\right)\, h(l\,T_y) \\ &= \sum_m x(m\,T_x) \sum_l h(l\,T_y)\,\mathrm{sinc}\left((k-l)(1+\varepsilon) - m\right), \end{aligned} \tag{2}$$

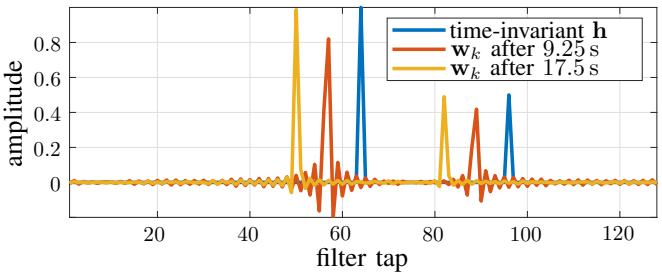

Fig. 2. Example of an adaptive filter $\mathbf{w}_k$ tracking the impulse response $\mathbf{h}$ of an LTI system at 16 kHz sampling rate with IO-SRO $\varepsilon = 50$ ppm.

where (1) is employed to relate $y(k\,T_y)$ back to the original input $x(m\,T_x)$. For the inner sum $\mathcal{S}(m,k)$ in (2), we recall $T_y = (1+\varepsilon)\,T_x$ and apply a resampling similar to (1) to find

$$\begin{aligned} \mathcal{S}(m,k) &= \sum_l h(l\,T_y)\,\mathrm{sinc}\left((k-l)(1+\varepsilon) - m\right) \\ &= \sum_l h(l\,T_y)\,\mathrm{sinc}\left(k - m + \varepsilon\,k - (1+\varepsilon)\,l\right) \\ &= h\left((k - m + \varepsilon\,k)\,T_x\right) \end{aligned} \tag{3}$$

and hence, inserting (3) back into (2) yields

$$y(k\,T_y) = \sum_m x(m\,T_x)\,h\left((k-m)T_x + \varepsilon\,k\,T_x\right)\,. \tag{4}$$

We may interpret the result in (4) in the following way. If the time-discrete samples of $x(m\,T_x)$ and $y(k\,T_y)$ are considered as input and output of a linear system, then this system is the time-discrete representation of the original LTI system, but with a uniformly increasing delay $\varepsilon\,k\,T_x$, resulting in an overall "moving" LTV system $\mathbf{h}_k$.

B. Model of Adaptive Filter Tracking With IO-SRO

Our conclusion regarding (4) can be verified by running an adaptive filter $\mathbf{w}_k = [w_{0,k} \cdots w_{N-1,k}]^T$ of length N in parallel to the system as depicted in Fig. 1 in order to perform system identification. In this setup, the adaptive filter treats input and output signal samples as if they originated from the same time basis. Consider Fig. 2 for a simple example of the impulse response $\mathbf{h} = [h_0 \cdots h_{L-1}]^T$ of length $L = N$ of the LTI system. The adaptive filter $\mathbf{w}_k$ shifts with respect to the true LTI system $\mathbf{h}$ over time just as predicted by (4) for the "moving" LTV system $\mathbf{h}_k$. Note that the sinc-like structure is more pronounced, the farther the real-valued delay $\varepsilon\,k\,T_x$ is from an integer number.

To quantitatively model the tracking of the adaptive filter, we consider the update equation of an LMS-type algorithm [1]–[3] and conduct a convergence-in-the-mean analysis. For this derivation we assume $N = L$. In the update, the input signal vector $\mathbf{x}_{k-1} = [x((k-1)\,T_x)\,x((k-2)\,T_x) \cdots x((k-N)\,T_x)]^T$ is defined on the original time basis via T_x while output signal $y_{k-1} = y((k-1)\,T_y)$ and error signal $e_{k-1} =$

$e((k-1)\,T_y)$ are based on T_y as in Fig. 1. The update from discrete time instant $k-1$ to k with step size μ then becomes

$$\begin{aligned}\mathbf{w}_k &= \mathbf{w}_{k-1} + \mu\, e_{k-1}\, \mathbf{x}_{k-1} \\ &= \mathbf{w}_{k-1} + \mu \left(y_{k-1} - \mathbf{w}_{k-1}^T \mathbf{x}_{k-1}\right)\, \mathbf{x}_{k-1} \\ &= \left(\underline{\mathbf{I}} - \mu\, \mathbf{x}_{k-1}\mathbf{x}_{k-1}^T\right) \mathbf{w}_{k-1} + \mu\, \mathbf{x}_{k-1}\, y_{k-1}\,, \end{aligned} \tag{5}$$

where $\underline{\mathbf{I}}$ is an $N \times N$ identity matrix. For a closed-form analysis of $\mathbf{w}_k$, we restate (5) in a non-recursive form [3], starting with an arbitrarily chosen time with signals y_0 and $\mathbf{x}_0$, i.e.,

$$\begin{aligned}\mathbf{w}_k &= \sum_{i=0}^{k-1} \left(\underline{\mathbf{I}} - \mu\, \mathbf{x}_i \mathbf{x}_i^T\right)^{k-1-i} \mu\, \mathbf{x}_i\, y_i \\ &= \sum_{i=0}^{k-1} \left(\underline{\mathbf{I}} - \mu\, \mathbf{x}_i \mathbf{x}_i^T\right)^{k-1-i} \mu\, \mathbf{x}_i \left(\mathbf{x}_i^T \mathbf{h}_i + n_i\right)\,. \end{aligned} \tag{6}$$

In the last line of (6), we have made use of (4) to express the system output y_i as a discrete-time convolution between the input signal and the time-varying impulse response $\mathbf{h}_i = [h_{0,i} \cdots h_{L-1,i}]^T$, where $h_{l,i} = h(l\,T_x + \varepsilon\, i\, T_x)$ as per (4).

We now perform a convergence-in-the-mean analysis of the adaptive filter $\mathbf{w}_k$ in (6) by considering the statistical expectation of the filter coefficients in the sense of the *direct-averaging method* [3], [22],

$$E\left\{\mathbf{w}_k\right\} \approx \sum_{i=0}^{k-1} \left(\underline{\mathbf{I}} - \mu\, \underline{\mathbf{R}}_{xx}\right)^{k-1-i} \mu\, \underline{\mathbf{R}}_{xx} \mathbf{h}_i\,, \tag{7}$$

where $\underline{\mathbf{R}}_{xx} = E\left\{\mathbf{x}_i \mathbf{x}_i^T\right\}$ is the correlation matrix of a stationary input signal and we have assumed $E\left\{\mathbf{x}_i\, n_i\right\} = \mathbf{0}$.

We further assume a Fourier matrix $\underline{\mathbf{F}}$ with $\underline{\mathbf{F}}\,\underline{\mathbf{F}}^H = \underline{\mathbf{I}}$ to diagonalize the input correlation $\underline{\mathbf{R}}_{xx} = \underline{\mathbf{F}}^H\, \underline{\mathbf{D}}_{xx}\, \underline{\mathbf{F}}$ with $\underline{\mathbf{D}}_{xx} = \mathrm{diag}\left(\mathbf{\Phi}_{xx}\right)$ [3], where $\mathbf{\Phi}_{xx}$ is a vector containing the power spectral density (PSD) of the input signal as a function of the discrete frequency. With $\underline{\mathbf{F}}$, we can express (7) as

$$E\left\{\mathbf{w}_k\right\} = \sum_{i=0}^{k-1} \underline{\mathbf{F}}^H \left(\underline{\mathbf{I}} - \mu\, \underline{\mathbf{D}}_{xx}\right)^{k-1-i} \underline{\mathbf{F}}\,\underline{\mathbf{F}}^H \mu\, \underline{\mathbf{D}}_{xx}\, \underline{\mathbf{F}}\, \mathbf{h}_i\,. \tag{8}$$

Left-multiplying (8) by $\underline{\mathbf{F}}$ and defining frequency-domain quantities $\mathbf{H}_i = \underline{\mathbf{F}}\, \mathbf{h}_i$ and $\mathbf{W}_k = \underline{\mathbf{F}}\, \mathbf{w}_k$, we obtain

$$E\left\{\mathbf{W}_k\right\} = \sum_{i=0}^{k-1} \left(\underline{\mathbf{I}} - \mu\, \underline{\mathbf{D}}_{xx}\right)^{k-1-i} \mu\, \underline{\mathbf{D}}_{xx}\, \mathbf{H}_i\,. \tag{9}$$

As (9) describes the evolution of all frequencies in $\mathbf{W}_k$ independently, we focus on a single discrete frequency Ω, i.e.,

$$E\left\{W_k(\Omega)\right\} = \sum_{i=0}^{k-1} \left(1 - \nu(\Omega)\right)^{k-1-i} \nu(\Omega) H_i(\Omega) \tag{10}$$

with $\nu(\Omega) = \mu\, \Phi_{xx}(\Omega)$. To express $H_i(\Omega)$ w.r.t. the state of the LTV system $H_k(\Omega)$ at the current time instant k, the time-varying delay implied by (4) can equivalently be expressed in frequency domain as $H_i(\Omega) = \exp\left(-j\,\Omega\,(k-i)\,\varepsilon\right) H_k(\Omega)$. Accordingly, (10) can be cast into an equation relating $W_k(\Omega)$ directly to the LTV system $H_k(\Omega)$, namely,

$$\begin{aligned} E\left\{W_k(\Omega)\right\} &= \sum_{i=0}^{k-1} \left(1 - \nu(\Omega)\right)^{k-1-i} \nu(\Omega)\, e^{-j\,\Omega\,(k-i)\,\varepsilon}\, H_k \\ &= H_k(\Omega)\, \nu(\Omega)\, e^{-j\,\Omega\,\varepsilon} \\ &\quad \times \sum_{i=0}^{k-1} \left(\left(1 - \nu(\Omega)\right) e^{-j\,\Omega\,\varepsilon}\right)^{k-1-i}\,. \end{aligned} \tag{11}$$

Eventually, we assume the current time instant k to be already arbitrarily far away from the initial time $i = 0$, hence approximating the sum in (11) by an infinite sum, i.e.,

$$\begin{aligned} E\left\{W_k(\Omega)\right\} &\approx H_k(\Omega)\, \nu(\Omega)\, e^{-j\,\Omega\,\varepsilon} \sum_{n=0}^{\infty} \left(\left(1 - \nu(\Omega)\right) e^{-j\,\Omega\,\varepsilon}\right)^n \\ &= H_k(\Omega) \cdot \frac{\nu(\Omega)}{e^{j\,\Omega\,\varepsilon} - 1 + \nu(\Omega)}\,, \end{aligned} \tag{12}$$

where we have exploited the geometric sum with $\left|\left(1 - \nu(\Omega)\right) \exp(-j\,\Omega\,\varepsilon)\right| < 1$ for a small μ or for an NLMS step size [3]. Thus, the bias between $W_k(\Omega)$ and $H_k(\Omega)$ is represented by a time-invariant *lag filter*

$$L(\Omega) = \frac{\nu(\Omega)}{e^{j\,\Omega\,\varepsilon} - 1 + \nu(\Omega)}\,. \tag{13}$$

For small IO-SRO $\varepsilon \ll 1$, we may further approximate the complex exponential $\exp(j\,\Omega\,\varepsilon) \approx 1 + j\,\Omega\,\varepsilon$ which yields

$$L(\Omega) \approx \frac{\nu(\Omega)}{\nu(\Omega) + j\,\Omega\,\varepsilon}\,. \tag{14}$$

The lag filter (14) is time-invariant, i.e., independent of k and is determined entirely by the degree of time-variance of the system expressed via IO-SRO ε and by the adaptation process itself that is controlled by step size μ and input PSD $\Phi_{xx}(\Omega)$.

III. Verification via IO-SRO Reconstruction

While the proposed lag filter (14) models the tracking behavior of the adaptive filter, it will also affect the CPSD

$$\begin{aligned} \Phi_{Y\widehat{Y}}(\Omega) &= E\left\{Y_k(\Omega)\, \widehat{Y}_k^*(\Omega)\right\} \\ &= E\left\{\left(H_k(\Omega)\, X_k(\Omega) + N_k(\Omega)\right)\, W_k^*(\Omega)\, X_k^*(\Omega)\right\} \\ &= H_k(\Omega) E\left\{W_k^*(\Omega)\right\} \Phi_{xx}(\Omega)\,, \end{aligned} \tag{15}$$

between the observed signal $Y(\Omega)$ and the adaptive filter output $\widehat{Y}(\Omega)$ with $E\left\{X(\Omega)\, X^*(\Omega)\right\} = \Phi_{xx}(\Omega)$. In the last step of (15), we have assumed statistical independence between $W_k(\Omega)$ and $X_k(\Omega)$. Using (12), we can readily express the CPSD (15) as

$$\Phi_{Y\widehat{Y}}(\Omega) = \left|H(\Omega)\right|^2\, \Phi_{xx}(\Omega) \cdot L^*(\Omega)\,, \tag{16}$$

with $|H(\Omega)|^2 = |H_k(\Omega)|^2$, as the time-variance is only a shift.

From (16) we observe that the phase of the CPSD is the same as the negative phase of the lag filter, i.e., $\angle\Phi_{Y\widehat{Y}}(\Omega) = -\angle L(\Omega) = \mathrm{atan}(\Omega\,\varepsilon/\nu(\Omega))$ and hence

$$\tan\angle\Phi_{Y\widehat{Y}}(\Omega) = \frac{\Omega}{\nu(\Omega)} \cdot \varepsilon\,, \tag{17}$$

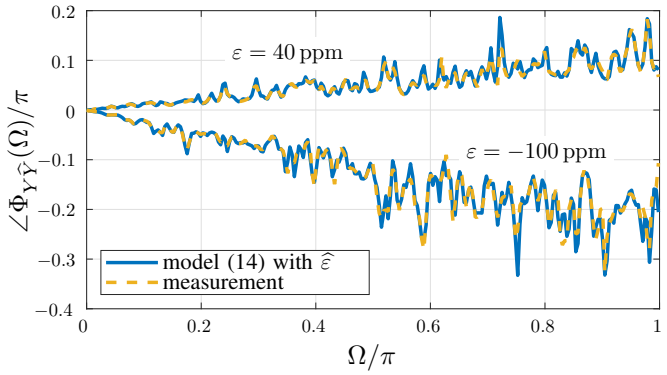

Fig. 3. Measured CPSD phase and phase according to the lag-filter model (14) with reconstructed $\widehat{\varepsilon}$ via weighted least-squares solution from (17).

where we have used the approximation in (14).

With relation (17) we can directly fit the phase of the model lag filter to the phase of the CPSD between system output and adaptive filter output. We can therefore verify the lag filter model by calculating a reconstructed $\widehat{\varepsilon}$ of the IO-SRO from the CPSD. This reconstruction can then be used either to compare directly to a ground-truth ε or to predict the CPSD phase based on the model (14).

Since we are interested in a single broadband reconstruction $\widehat{\varepsilon}$ based on the available $\Phi_{Y\widehat{Y}}(\Omega)$ and $\nu(\Omega)$ at individual Ω, we propose to solve (17) for ε via a weighted least-squares procedure. We specifically use $\Phi_{\widehat{Y}\widehat{Y}}(\Omega) = E\left\{\widehat{Y}_k\widehat{Y}_k^*\right\}$ as a weighting function since it will put zero weights on frequencies Ω with $W_k(\Omega) = 0$ and $X_k(\Omega) = 0$ and accordingly undefined phase. For a simple experiment without additive noise, we find $\widehat{\varepsilon}$ based on (17) and reconstruct the CPSD phase via the model (14). The result is depicted in Fig. 3 and indicates that the approximation by the lag filter based on $\widehat{\varepsilon}$ is valid for the full frequency range. Further analysis regarding the quality of the reconstructed $\widehat{\varepsilon}$ is presented in the computer experiments in Sec. IV.

IV. Experimental Evaluation

We aim to validate the proposed lag filter model (14) in computer experiments by assessing its ability to extract the IO-SRO from the CPSD between system output and adaptive filter output via (17) as proposed in Sec. III.

As discrete-time LTI system $\mathbf{h}$ in Fig. 1, we use a random impulse response of length $L = 128$ with the first 64 taps set to zero. Input sequence $x(k\,T_y)$ is generated as pseudo-random multi-tone [23] with random phase and amplitudes for the individual tones at $T_y^{-1} = 16\,\mathrm{kHz}$ sampling with a total duration of 20 s. We calculate the output of $\mathbf{h}$ as a discrete-time convolution and add white noise $n(k\,T_y)$ at different signal-to-noise ratios (SNR) to obtain the output $y(k\,T_y)$. The input signal for the adaptive filter $x(m\,T_x)$ is generated with the same phases and amplitudes as $x(k\,T_y)$, but with a slightly different sampling to introduce a desired IO-SRO ε.

The update of $\mathbf{w}_k$ is performed by an NLMS algorithm [2], [3] with a time-varying step size $\mu_k = \mu_0/\|\mathbf{x}_k\|^2$ and normalized step size $\mu_0 = 0.01$ and a filter length of $N = 256$. Notice that in this setup, the adaptive filter $\mathbf{w}_k$ has 64 taps headroom to drift to the *left* (positive ε) as seen in Fig. 2 and 128 taps to drift to the *right* (negative ε). In practical application scenarios, such as AEC or interference cancellation, where the extracted IO-SRO could be compensated by resampling [19], this would not cause a problem. But in our experiments, we do not compensate for the reconstructed IO-SRO to assess the model independently of the compensation method. This introduces a slightly asymmetrical reconstruction performance because of the finite headroom for the uncompensated drift.

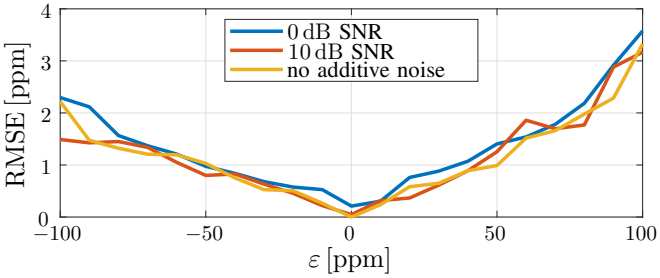

Fig. 4. RMSE between ε and $\widehat{\varepsilon}$ obtained using NLMS adaptation.

Using 256 ms blocks of $y(k\,T_y)$ and $\widehat{y}(k\,T_y)$, we calculate the CPSD in (17) and find a broadband reconstruction $\widehat{\varepsilon}$ via the weighted least-squares procedure described in Sec. III. The final value $\widehat{\varepsilon}$ is obtained by averaging over the 20 s of input data. All experiments are conducted for $\varepsilon = -100 \ldots 100$ ppm at 0 dB SNR, 10 dB SNR, and without any additive noise. For each setup, we repeat the experiments $R = 20$ times and evaluate the root-mean-square error (RMSE) for each ε as

$$\mathrm{RMSE}(\varepsilon) = \sqrt{\frac{1}{R}\sum_{r=1}^{R}\left(\varepsilon - \widehat{\varepsilon}^{(r)}\right)^2}. \tag{18}$$

The results in terms of RMSE are depicted in Fig. 4. Generally, we observe an asymmetrical tilt due to the aforementioned imbalance of filter lengths for positive and negative ε. In either way, the RMSE values are in the range of single-digit ppm and decline towards $\varepsilon = 0$, which is an important characteristic if the resulting IO-SRO were to be successively compensated via resampling. Overall, the model allows the extraction of $\widehat{\varepsilon}$ over a wide range of ε and also in adverse noise conditions without significant loss of accuracy.

So far, we have only considered the influence of ε, but the model in (14) is also determined by the excitation PSD $\Phi_{xx}(\Omega)$. In a second experiment, we therefore introduce different upper cut-off frequencies Ω_c for the excitation signal $x(k\,T_y)$ that introduce a low-pass effect on $\Phi_{xx}(\Omega)$. Since the RMSE in (18) turned out to be dependent on ε as seen in Fig. 4, we consider instead the relative root-mean-square error $\mathrm{relRMSE}(\varepsilon) = \mathrm{RMSE}(\varepsilon)/|\varepsilon|$ in the second experiment and average across all ε. We further compare the performance of the proposed model in terms of IO-SRO reconstruction with that of a state-of-the-art method proposed for async-FDAEC in [18]. For this comparison, we employ a normalized step-

TABLE I
RESULTS IN TERMS OF relRMSE FOR $\widehat{\varepsilon}$.

	SNR [dB]	proposed model	async-FDAEC [18]
$\Omega_c = \pi$	0	0.035 ± 0.010	0.059 ± 0.047
	10	0.026 ± 0.013	0.050 ± 0.035
	no noise	0.027 ± 0.013	0.051 ± 0.040
$\Omega_c = \pi/2$	0	0.019 ± 0.005	0.148 ± 0.132
	10	0.016 ± 0.003	0.139 ± 0.134
	no noise	0.015 ± 0.003	0.145 ± 0.148
$\Omega_c = \pi/4$	0	0.032 ± 0.014	0.385 ± 0.408
	10	0.027 ± 0.006	0.403 ± 0.443
	no noise	0.028 ± 0.006	0.350 ± 0.370

size $\mu_0 = 0.01$ for both methods and the step-size for the SRO update in [18] is set to 10^{-12} for stable convergence. All other parameters, such as frame size and shift, are chosen as suggested by the authors in [18].

The results including the standard deviation across all investigated IO-SROs ($R = 20$ trials each) are presented in Tab. I. In the majority of setups, the relative RMSE of the proposed method is around or below $3\,\%$, indicating that the model predicts accurately for the considered range of ε, SNR, and excitation bandwidth. The state-of-the-art async-FDAEC method performs only slightly worse with fullband-excitation, but seems to suffer from insufficient excitation bandwidth. While the results in Tab. I suggest superior performance of the proposed method, it should be noted that the method proposed in [18] also compensates the estimated IO-SRO and the ensuing drift via phase rotation in the STFT domain and is hence suitable for asynchronous AEC. Our proposed model and reconstruction, however, would require an additional drift compensation that is beyond the scope of this paper.

V. CONCLUSION

In this contribution, we have presented a novel model for adaptive filter behavior under IO-SRO conditions. By arguing that an adaptive filter sees a system with mismatched input and output sampling rates as a time-varying system, we have developed an expression for a time-invariant lag-filter that captures the tracking bias resulting from the IO-SRO. This lag filter is a function of the IO-SRO itself as well as the adaptive filter configuration via the update step size and the PSD of the excitation. It can be related to the phase of the CPSD between system output and adaptive filter output. Through this link, we have been able to extract the IO-SRO from the phase of the lag filter and have assessed the quality of this IO-SRO reconstruction as an indicator of model accuracy in several computer experiments. A comparison with a state-of-the-art technique shows good and robust performance of the proposed model in various configurations.

REFERENCES

[1] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Eaglewood Cliffs, NJ, USA: Prentice Hall, 1985.

[2] A. Sayed, *Adaptive Filters*. Hoboken, NJ, USA: John Wiley & Sons, 2008.

[3] S. Haykin, *Adaptive Filter Theory*, 5th ed. Upper Saddle River, NJ, USA: Pearson, 2014.

[4] J. Benesty, T. Gänsler, D. R. Morgan, M. M. Sondhi, and S. L. Gay, *Advances in Network and Acoustic Echo Cancellation*. Berlin, Heidelberg, New York: Springer, 2001.

[5] E. Hänsler and G. Schmidt, *Acoustic Echo and Noise Control: A Practical Approach*. Hoboken, NJ, USA: John Wiley & Sons, 2004.

[6] P. Vary and R. Martin, *Digital Speech Transmission: Enhancement, Coding and Error Concealment*. Hoboken, NJ, USA: John Wiley & Sons, 2006.

[7] E. Hänsler and G. Schmidt, Eds., *Topics in Acoustic Echo and Noise Control*. Berlin, Heidelberg, New York: Springer, 2006.

[8] G. Enzner, H. Buchner, A. Favrot, and F. Kuech, "Acoustic echo control," in *Academic Press Library in Signal Processing, Volume 4: Image, Video Processing and Analysis, Hardware, Audio, Acoustic and Speech Processing*, S. Theodoridis and R. Chellappa, Eds. Cambridge, MA, USA: Academic Press, 2014.

[9] T. van Waterschoot and M. Moonen, "Fifty years of acoustic feedback control: State of the art and future challenges," *Proc. IEEE*, vol. 99, no. 2, pp. 288 – 327, Feb. 2011.

[10] E. Robledo-Arnuncio, T. S. Wada, and B.-H. Juang, "On dealing with sampling rate mismatches in blind source separation and acoustic echo cancellation," in *IEEE Workshop Appl. Signal Process. Audio Acoustics (WASPAA)*, New Paltz, NY, USA, Oct. 2007.

[11] M. Pawig, G. Enzner, and P. Vary, "Adaptive sampling rate correction for acoustic echo control in voice-over-IP," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 189 – 199, Jan. 2010.

[12] R. Lienhart, I. Kozintsev, S. Wehr, and M. Yeung, "On the importance of exact synchronization for distributed audio signal processing," in *IEEE Int. Conf. Acoustics, Speech, Signal Process. (ICASSP)*, Hong Kong, China, Apr. 2003.

[13] S. Miyabe, N. Ono, and S. Makino, "Blind compensation of interchannel sampling frequency mismatch for ad hoc microphone array based on maximum likelihood estimation," *Elsevier Signal Process.*, vol. 107, no. 2, pp. 185 – 196, Feb. 2015.

[14] L. Wang and S. Doclo, "Correlation maximization-based sampling rate offset estimation for distributed microphone arrays," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 24, no. 3, pp. 571 – 582, Mar. 2016.

[15] M. H. Bahari, A. Bertrand, and M. Moonen, "Blind sampling rate offset estimation for wireless acoustic sensor networks through weighted least-squares coherence drift estimation," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 3, pp. 1 – 13, Mar. 2017.

[16] H. Ding and D. I. Havelock, "Drift-compensated adaptive filtering for improving speech intelligibility in cases with asynchronous inputs," *EURASIP Journal Advances Signal Process.*, vol. 2010, pp. 1 – 12, Aug. 2010.

[17] J. Stokes and H. Malvar, "Acoustic echo cancellation with arbitrary playback sampling rate," in *IEEE Int. Conf. Acoustics, Speech, Signal Process. (ICASSP)*, Montreal, Canada, May 2004.

[18] M. Abe and M. Nishiguchi, "Frequency domain acoustic echo canceller that handles asynchronous A/D and D/A clocks," in *IEEE Int. Conf. Acoustics, Speech, Signal Process. (ICASSP)*, Florence, Italy, May 2014.

[19] R. E. Crochiere and L. R. Rabiner, *Multirate Digital Signal Processing*. Eaglewood Cliffs, NJ, USA: Prentice Hall, 1983.

[20] J. Schmalenstroeer and R. Haeb-Umbach, "Efficient sampling rate offset compensation - an overlap-save based approach," in *Europ. Sig. Process. Conf. (EUSIPCO)*. Rome, Italy, Sep. 2018.

[21] A. Chinaev, P. Thüne, and G. Enzner, "Low-rate Farrow structure with discrete-lowpass and polynomial support for audio resampling," in *Europ. Sig. Process. Conf. (EUSIPCO)*, Rome, Italy, Sep. 2018.

[22] H. Kushner, *Approximation and Weak Convergence Methods for Random Processes with Applications to Stochastic Systems Theory*. Cambridge, MA, USA: MIT Press, 1984.

[23] J. Schoukens, R. Pintelon, E. van der Ouderaa, and J. Renneboog, "Survey of excitation signals for FFT based signal analyzers," *IEEE Trans. Instrumentation Measurement*, vol. 37, no. 3, pp. 342 – 352, Sep. 1988.