Adaptive Sampling Rate Correction for Acoustic Echo Control in Voice-Over-IP

Matthias Pawig, Gerald Enzner, *Member, IEEE*, and Peter Vary, *Fellow, IEEE*

***Abstract*—Hands-free terminals for speech communication employ adaptive filters to reduce echoes resulting from the acoustic coupling between loudspeaker and microphone. When using a personal computer with commercial audio hardware for teleconferencing, a sampling frequency offset between the loudspeaker output D/A converter and the microphone input A/D converter often occurs. In this case, state-of-the-art echo cancellation algorithms fail to track the correct room impulse response. In this paper, we present a novel least mean square (LMS-type) adaptive algorithm to estimate the frequency offset and resynchronize the signals using arbitrary sampling rate conversion. In conjunction with a normalized LMS-type adaptive filter for room impulse response tracking, the proposed system widely removes the deteriorating effects of a frequency offset up to several Hz and restores the functionality of echo cancellation.**

***Index Terms*—Acoustic signal processing, adaptive filters, echo suppression, interpolation, least mean squares methods, resampling, teleconferencing.**

I. Introduction

In recent years, Voice-over-IP (VoIP) systems based on a personal computer (PC) have become very popular. Users of hands-free teleconferencing devices expect reliable acoustic echo attenuation with full duplex capability, which is a particularly challenging signal processing task.

Usually, a linear echo path is assumed and an adaptive echo canceler is applied with identical sampling frequencies for all signals [1]–[6]. Sophisticated control mechanisms are available to handle double-talk and to achieve fast and robust adaptation of the echo canceller coefficients in time-varying and noisy environments [7]–[10].

Unfortunately, the excellent performance of these algorithms is degraded dramatically if the sampling frequencies of the D/A and the A/D converters are not exactly the same. This is often not guaranteed even on the same audio hardware. The frequency offset Δf of cheap sound devices may be in the range of up to 10 Hz. This offset causes nonlinear time-varying disturbances of the effective echo path including D/A converter, loudspeaker room microphone impulse response, and A/D converter. The different sampling frequencies in the microphone and loudspeaker path cause a drift of the effective echo path, as well as buffer over- or underflows and therefore jumps of the effective impulse response which deteriorate the performance of the adaptive filter.

It can be observed that the offset of the sampling frequencies in a particular system remains constant, at least for the duration of a telephone conversation. A change of the frequency difference may however occur when changing the PC system configuration. Since changes in the configuration are generally unknown to the VoIP software, a safe and reliable correction of the frequency offset requires an automatic frequency offset inference at each start of the software. The straightforward solution is to estimate the frequency offset by using a pilot-based technique, e.g., by generating a sinusoid of a predefined frequency at the D/A converter and measuring the frequency of the samples after the A/D converter. Since this estimation would only be performed once in advance, extreme accuracy would be required, as even a very small residual offset would result in buffer over- or underflow from time to time.

In order to avoid the user to be bothered by repeated, pilot-based calibration of the VoIP system, this paper presents an LMS-type adaptive algorithm to estimate the frequency offset during the course of conversation. We will show that such frequency offset estimation at runtime requires knowledge of the echo path impulse response. Reliable echo path estimation, in turn, requires the correction of the frequency offset in the system. In the paper, we will therefore derive an iterative approach for jointly estimating both quantities. We will demonstrate the convergence of the algorithm in terms of successful resynchronization of the input and output signals of the acoustic echo path.

Generally, the proposed solution can be adopted to other applications which suffer from sampling frequency offsets, such as distributed audio processing [11], or it might serve as a refinement stage for larger sampling rate conversions [12].

Section II of this contribution explains the fundamental problem. Section III introduces the new algorithm in detail, where Section III-C contains a gradient-based derivation of the adaptive algorithm for frequency offset estimation. In Section IV, the performance of the algorithm is analyzed in conjunction with an NLMS adaptive filter for echo path impulse response tracking.

II. The Fundamental Problem Caused by a Frequency Offset

The principle structure of an acoustic echo canceler (AEC) on a PC platform is shown in Fig. 1. The echo $d(k)$ caused by the acoustic coupling between loudspeaker and microphone is canceled by an adaptive echo canceler with impulse response

Manuscript received January 06, 2009; accepted June 25, 2009. First published July 21, 2009; current version published December 16, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Danilo P. Mandic. This work was supported by VSN Systemen BV, Venray, the Netherlands.

M. Pawig and P. Vary are with the Institute of Communication Systems and Data Processing, RWTH Aachen University, 52065 Aachen, Germany (e-mail: pawig@ind.rwth-aachen.de; vary@ind.rwth-aachen.de).

G. Enzner is with the Institute of Communication Acoustics, Ruhr-University Bochum, 44780 Bochum, Germany (e-mail: gerald.enzner@rub.de).

Digital Object Identifier 10.1109/TSP.2009.2028187

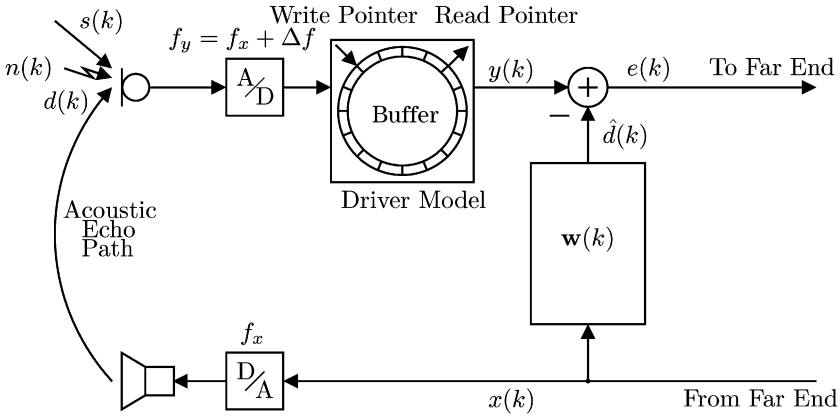

Fig. 1. System model with input sample buffer.

$\mathbf{w}(k)$ which estimates a copy $\hat{d}(k)$ of the echo from the loudspeaker signal $x(k)$. This estimated echo is subtracted from the microphone signal $y(k)$, which also contains the near speaker speech signal $s(k)$ and near end noise $n(k)$. In our analysis, the normalized least mean square (NLMS) algorithm was employed for the adaptation of the echo canceler [5]. The possibly differing sampling frequencies f_x of the D/A converter and f_y of the A/D converter of cheap PC audio hardware cause a time varying nonlinear behavior, such as an increasing or decreasing delay of the effective echo path which deteriorates its estimation. The driver model in Fig. 1 includes the finite sound driver buffer. In the model a finite circular buffer is written at a frequency $f_y = f_x + \Delta f$ and read at a frequency f_x. Once the write pointer and read pointer of this buffer overlap, a full buffer length of samples will be lost or repeated. If the circular buffer is written with higher frequency than it is read, once the delay between writing and reading exceeds the buffer length, new samples are overwritten before they are read out. On the other hand, if the buffer is read with higher frequency, after the delay exceeds the buffer length, samples will be repeated before they are written with new data.

A. Optimum Step-Size NLMS Algorithm

A well known algorithm for adaptive filtering is the NLMS algorithm. The update of the filter coefficients $\mathbf{w}(k) = [w_0(k), w_1(k), \ldots, w_{N-1}(k)]^T$ is given by

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \alpha(k)\frac{\mathbf{x}(k)}{\|\mathbf{x}(k)\|^2}e(k) \tag{1}$$

with an adaptive step-size $0 \leq \alpha(k) \leq 1$ and

$$e(k) = y(k) - \hat{d}(k) = y(k) - \mathbf{w}^T(k)\mathbf{x}(k) \tag{2}$$

where $\mathbf{x}(k) = [x(k), x(k-1), \ldots, x(k-(N-1))]^T$ is the far end speaker signal and N is the adaptive filter length. Instead of an additive regularization parameter in this contribution the norm $\|\mathbf{x}(k)\|^2$ is lower bounded to $\max(\|\mathbf{x}(k)\|^2, \epsilon)$ to avoid numerical problems for very small input levels of $x(k)$. The constant $\epsilon \ll \mathrm{E}\{x^2(k)\}$ is chosen small in comparison to the expected signal power. In practice, this limitation does mostly not apply for realistic scenarios with natural far end signals $x(k)$ which contain a noise floor.

An adaptive step-size is required to slow down adaptation in case of high local noise or—more importantly—the double-talk situation with a near speaker signal $s(k)$. In this contribution, the optimum step-size factor $\alpha_{\mathrm{opt}}(k)$ with respect to the system mismatch is taken according to [8]:

$$\alpha_{\mathrm{opt}}(k) = \frac{\mathrm{E}\left\{b^2(k)\right\}}{\mathrm{E}\left\{e^2(k)\right\}} \tag{3}$$

where $b(k) = d(k) - \hat{d}(k)$ is the echo estimation error signal. The echo signal can be expressed as $d(k) = \mathbf{h}^T(k)\mathbf{x}_M(k)$ with the effective room impulse response $\mathbf{h} = [h_0, h_1, \ldots, h_{M-1}]$ and the input vector $\mathbf{x}_M(k) = [x(k), x(k-1), \ldots, x(k-(M-1))]^T$ of length M. In order to approximate the expectations $\mathrm{E}\{b^2(k)\}$ and $\mathrm{E}\{e^2(k)\}$ by accessible quantities, the equation is transformed using the filter estimation error $\mathbf{g}(k) = \mathbf{h}(k) - \mathbf{w}_{\mathbf{M}}(k)$ where $\mathbf{w}_M$ denotes the vector $\mathbf{w}$ zero-padded to length M. Assuming $x(k)$, $s(k)$, and $n(k)$ to be independent and uncorrelated (white) Gaussian signals, it follows:

$$\begin{aligned}\alpha_{\mathrm{opt}}(k) &= \frac{\mathrm{E}\left\{\left(d(k) - \hat{d}(k)\right)^2\right\}}{\mathrm{E}\left\{\left(d(k) - \hat{d}(k) + s(k) + n(k)\right)^2\right\}} \\ &= \frac{\mathrm{E}\left\{\left(\left(\mathbf{h}^T(k) - \mathbf{w}_M^T(k)\right)\mathbf{x}_M(k)\right)^2\right\}}{\mathrm{E}\left\{\left(\left(\mathbf{h}^T(k) - \mathbf{w}_M^T(k)\right)\mathbf{x}_M(k) + s(k) + n(k)\right)^2\right\}} \\ &= \frac{\mathrm{E}\left\{\|\mathbf{g}(k)\|^2\right\}\mathrm{E}\left\{x^2(k)\right\}}{\mathrm{E}\left\{\|\mathbf{g}(k)\|^2\right\}\mathrm{E}\left\{x^2(k)\right\} + \mathrm{E}\left\{\left(s(k)+n(k)\right)^2\right\}}.\end{aligned} \tag{4}$$

As the system distance

$$\|\mathbf{g}(k)\|^2 = \|\mathbf{h}(k) - \mathbf{w}_M(k)\|^2 = \mathbf{g}^T(k)\mathbf{g}(k) \tag{5}$$

is inaccessible, it is approximated by a steady state value g^2 after convergence and set to a constant value. This parameter controls the adaptation speed as well as the sensitivity for double-talk. A reasonable compromise between speed of convergence and steady state performance was found empirically around $g^2 \approx 0.01$ for unit-norm echo paths. The expectations of the signals are replaced by the short term powers $\sigma_x^2(k)$ and $\sigma_{s+n}^2(k)$ of the

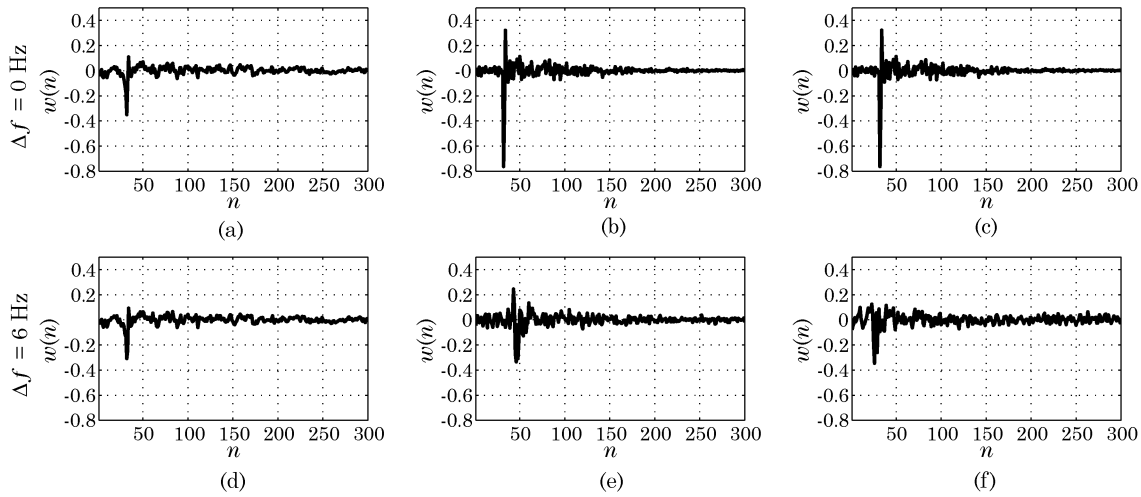

Fig. 2. Adaptive filter coefficient development at $t_0 = 0.1$ s (a), (d), $t_1 = 3$ s (b), (e) and $t_2 = 5$ s (c), (f); $N = 300$, $f_S = 8$ kHz.

signals $x(k)$ and $s(k) + n(k)$, respectively, which are defined by

$$\sigma_x^2(k) = \frac{1}{P} \sum_{\kappa=k-P+1}^{k} x^2(\kappa). \tag{6}$$

If the adaptive filter is in a steady state, $\mathbf{w}(k) \approx \mathbf{h}(k)$ is a valid assumption. In this case $e(k) = y(k) - \hat{d}(k)$ is approximately $s(k) + n(k)$ and $\sigma_e^2(k) \approx \sigma_{s+n}^2(k)$ can be used to approximate the adaptive step-size factor defined in (4) as follows:

$$\alpha_{\text{opt}}(k) \approx \alpha(k) = \frac{g^2 \sigma_x^2(k)}{g^2 \sigma_x^2(k) + \sigma_e^2(k)}, \quad g^2 \approx 0.01. \tag{7}$$

The short term powers $\sigma_x^2(k)$ and $\sigma_e^2(k)$ are calculated by recursive averaging as

$$\sigma_{e,x}^2(k) = (1 - 0.99)(e,x)^2(k-1) + 0.99\sigma_{e,x}^2(k) \tag{8}$$

approximating the short term power for $P \approx 200$ samples. The simplifications in the derivation are supported by the application examples in Section IV.

B. The Impact of the Frequency Offset Δf

The effect of a frequency offset Δf between the signals $x(k)$ and $y(k)$ can be represented by a time stretching/compression factor

$$a = \frac{f_x}{f_y} = \frac{f_x}{f_x + \Delta f}. \tag{9}$$

A sampling frequency offset causes the NLMS algorithm to fail in identifying the effective acoustic echo path. Echo cancelation by subtracting the estimated echo is no longer successful. The effect of a sampling frequency offset between the two signals can be understood as a time-variable delay increasing or decreasing for each sample. This nonlinear effect cannot be compensated by the NLMS algorithm.

Fig. 2 shows for $\Delta f = 0$ Hz and $\Delta f = 6$ Hz snapshots of the estimated echo path impulse response in case of no near end disturbance, i.e., $(n(k) = s(k) = 0)$ and a steady echo path of length $M = 300$ with a fixed step-size $\alpha(k) = 0.5$ for the NLMS algorithm and an adaptive filter length of $N = 300$. White noise input $x(k)$ was used for this figure. The adaptive filter coefficients are shown at $t_0 = 0.1$ s, $t_1 = 3$ s, and $t_2 = 5$ s.

In the case of $\Delta f = 6$ Hz, the maximum of the estimated impulse response moves with time to compensate for the time-varying delay of the effective echo path. This can be seen in Fig. 2(d) and (e), where the main peak has moved to the right. No stable and accurate echo path estimate will be found because of this movement. Without a frequency offset, as shown in Fig. 2(a), (b), and (c), the adaptive filter is converging and suitable to cancel the acoustic echo.

The second degrading effect due to $\Delta f \neq 0$ is caused by time jumps due to the finite buffer length L_B. The buffer will be written with a different sampling period than the sampling period for reading. If the difference between the pointers for writing and reading exceeds the buffer length, there will be a repetition of a buffer or one buffer of the input signal will be ignored, as explained in the beginning of this section. In the previous example, the effect can be observed in Fig. 2(f). Instead of moving further to the right, the filter impulse response jumps back to the left, representing the repetition of a buffer ($L_B = 32$ in this example). These jumps severely deteriorate the performance and cause major readaptation of the filter.

The described effects of a frequency offset have an increasing impact on the performance of the AEC with increasing values for Δf. However, even at very small offsets, the effect is non-negligible because of the systematic misadaptation of the filter. It should be noted that the same deteriorating effects occur in frequency domain adaptive filtering (FDAF) [13]. For block adaptation the effect of a time-variable delay due to a sampling frequency offset is even more severe.

III. Frequency Offset Detection and Correction

A. The Combined Estimation Concept

The proposed algorithm for frequency offset estimation is an adaptive waveform-based approach. The principle is illustrated in Fig. 3. The idea is to estimate the offset by comparing the

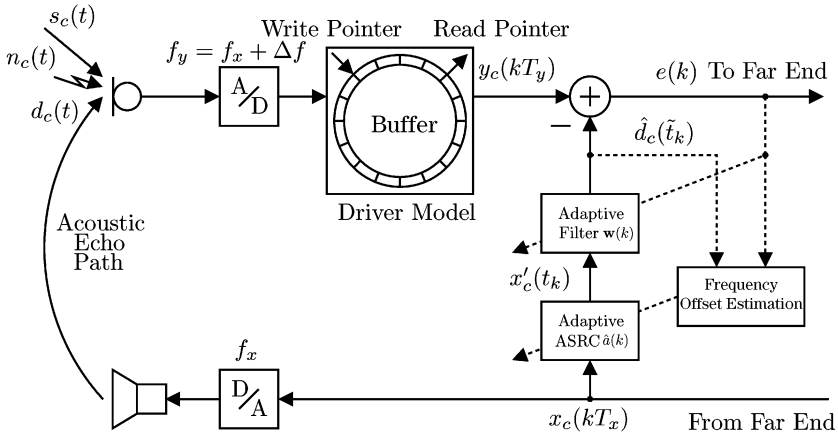

Fig. 3. Frequency estimation and correction principle.

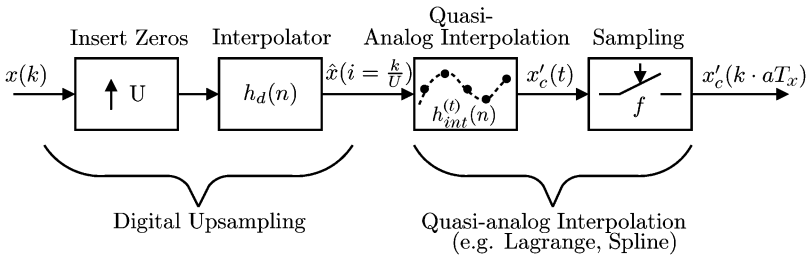

Fig. 4. ASRC principle.

waveforms of the near end signal $y(k) = y_c(kT_y)$ and the far end signal $x(k) = x_c(kT_x)$, or its filtered version $\hat{d}(k)$ to each other and minimizing the error signal $e(k)$. Note that subscript c denotes the continuous-time versions of the signals and T_x and T_y denote the sampling periods according to the frequencies f_x and f_y.

As the echo signal $d(k)$ contained in the signal $y(k)$ is produced from $x(k)$ by the unknown echo path, it is necessary to use an adaptive filter to match the signals prior to frequency offset estimation (FOE). Therefore, the estimated echo signal $\hat{d}(k)$ is used for the frequency offset estimation algorithm instead of $x(k)$. For the echo cancelation filter, the NLMS algorithm as explained in Section II is used.

B. Arbitrary Sampling Rate Conversion (ASRC)

In order to correct a frequency offset Δf, adaptive resampling is employed. Usually, resampling is based on upsampling by an integer factor U and subsequent downsampling by an integer factor D using constant interpolation filters, obtaining a total rational resampling factor of U/D.

In the context of our algorithm, rational resampling is not suitable, because the required U/D ratios are very close to one, e.g., at a sampling frequency of $f_x = 8$ kHz and a frequency offset $\Delta f = 1$ Hz, the resampling factor would be $U/D = 1/a = 8001\ \text{Hz}/8000\ \text{Hz} = 1.000125$. When using ratios like these, or adaptive resampling with time-variable ratios, extremely high up- and downsampling factors U and D would be required, causing extraordinary complexity and delay.

Instead, arbitrary sampling rate conversion (ASRC) (e.g., [14], [15]) is used. As shown in Fig. 4, it consists of two stages. First, the signal $x(k)$ is upsampled by a digital upsampling stage with moderate upsampling factors, in the course of this contribution $U = 4$, and a digital interpolation filter $h_d(n)$. This upsampling significantly increases the accuracy of the next interpolation stage. Then, to reach the desired rate conversion $1/a$, the upsampled signal $\hat{x}(i)$ is interpolated by a quasi-analog interpolator at the necessary sampling time instants $k' = ak$. It should be noted that the signal $x'_c(t)$ in Fig. 4 is not evaluated on a continuous-time scale. Instead, in such a system the interpolation and sampling are carried out in one unit by a continuous-time interpolator $h_{int}^{(t)}(n)$ which uses a discrete number of samples of $\hat{x}(i)$ to approximate the samples of a continuous-time signal $x'_c(t)$. As shown in Fig. 5 for $U = 4$, the sampling instants are calculated by compressing the time scale according to the factor a (here, $a < 1$). In the figure, $T_x = 1/f_x$ represents the sampling period of the input signal. The samples $x'_c(k \cdot aT_x)$ are created by evaluating the continuous-time interpolation at exactly these instants. In our investigation, Lagrange interpolation was employed.

Lagrange interpolation with $L+1$ samples of $\hat{x}(i)$ uses polynomials to approximate the signal, e.g., [16]. For $L = 3$, four samples of the upsampled signal $\hat{x}(i)$ are used. The coefficients $l_n(\Delta_k)$ of the interpolation filter are

$$\begin{aligned} l_{-1}(\Delta_k) &= \frac{1}{6}\Delta_k(\Delta_k - 1)(\Delta_k - 2) \\ l_0(\Delta_k) &= \frac{1}{2}(\Delta_k + 1)(\Delta_k - 1)(\Delta_k - 2) \\ l_1(\Delta_k) &= -\frac{1}{2}(\Delta_k + 1)\Delta_k(\Delta_k - 2) \\ l_2(\Delta_k) &= -\frac{1}{6}(\Delta_k + 1)\Delta_k(\Delta_k - 1) \end{aligned} \tag{10}$$

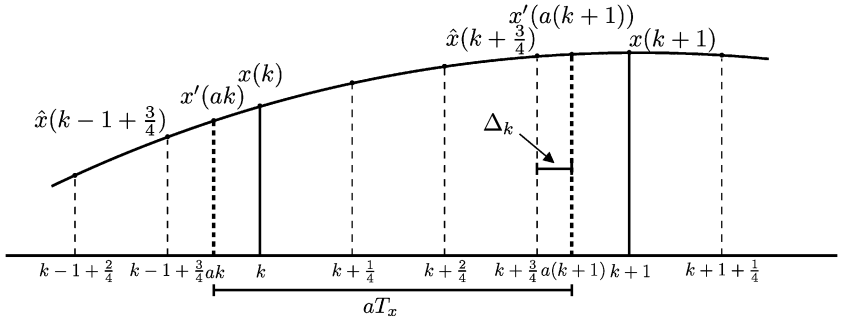

Fig. 5. Principle of continuous-time interpolation.

where Δ_k is defined in Fig. 5. The interpolation is

$$x'(ak = i_k + \Delta_k) = \sum_{n=-1}^{2} l_n(\Delta_k)\hat{x}(i_k + n) \tag{11}$$

where ak is the desired new sampling time instant, i_k is the sampling instant in the digitally upsampled signal $\hat{x}(i)$ which is closest to ak, and Δ_k is the relative distance between the two, as demonstrated in Fig. 5. This can be also be understood as digital filtering with the time varying filter

$$h_{int}^{(t)}(n) = l_n; \quad n = -1, 0, 1, 2. \tag{12}$$

The complexity of the interpolation increases when using more samples L, but the dominating factor in complexity considerations is the digital upsampling filter $h_d(n)$.

C. Novel LMS FOE Algorithm

The new algorithm is based on the structure in Fig. 3. Instead of a direct estimation of Δf we sequentially adapt the time stretching/compression factor a. Due to the LMS-type approach to be derived in the following, the estimated factor $\hat{a}(k)$ is time variant. Thus the sampling instances of the resampled signal $x_c(t_k)$ can be described as an accumulation of nonuniform sampling periods up to the kth sampling event:

$$t_k = \sum_{l=0}^{k} \hat{a}(l)T_x. \tag{13}$$

In the steady state of the sequential estimation, we expect $\hat{a}(l) = a$.

The signal $\hat{d}_c(\tilde{t}_k) = \mathbf{w}^T(k)\mathbf{x}'(t_k)$ is a filtered version of the resampled loudspeaker signal $x_c(t_k)$. The adaptive filter $\mathbf{w}(k)$ matches the waveform of $\hat{d}_c(\tilde{t}_k)$ to $y_c(kT_y)$.

At time $(k-1)T_y$, the relationship

$$t_{k-1} = \sum_{l=0}^{k-1} \hat{a}(l)T_x = (k-1)T_y + \Delta T_{k-1} \tag{14}$$

can be established in which ΔT_{k-1} is considered as a slowly varying sampling time mismatch at time t_{k-1} between samples $y_c((k-1)T_y)$ and $\hat{d}_c(t_{k-1})$. Since this sampling time mismatch is compensated by the phase of the adaptive filter $\mathbf{w}(k)$ up to time $(k-1)T_y$, the effective time instant $\tilde{t}_k$ can be written as a function of the current time stretching factor $\hat{a}(k)$

$$\tilde{t}_k = (k-1)T_y + \hat{a}(k)T_x. \tag{15}$$

The optimization criterion for estimating the factor $\hat{a}(k)$ is the minimization of the *mean-square* error, i.e.

$$\begin{aligned} \mathrm{E}\left\{e^2(k)\right\} &= \mathrm{E}\left\{\left(y_c(kT_y) - \hat{d}_c(\tilde{t}_k)\right)^2\right\} \\ &= \mathrm{E}\left\{\left(y_c(kT_y) - \hat{d}_c\left((k-1)T_y + \hat{a}(k)T_x\right)^2\right\} \\ &\rightarrow \min \end{aligned} \tag{16}$$

in analogy to the NLMS algorithm of the echo cancellation filter. For the purpose of the derivation, the noise-free single-talk case is assumed, meaning $y(k) = d(k)$.

The gradient used for minimizing $\mathrm{E}\{e^2(k)\}$ is, thus, given by

$$\begin{aligned} \nabla_f &= \frac{\partial \mathrm{E}\left\{e^2(k)\right\}}{\partial \hat{a}(k)} \\ &= -2\mathrm{E}\left\{e(k)\frac{\partial \hat{d}_c\left((k-1)T_y + \hat{a}(k)T_x\right)}{\partial\left((k-1)T_y + \hat{a}(k)T_x\right)} \right. \\ &\qquad \left. \cdot \frac{\partial\left((k-1)T_y + \hat{a}(k)T_x\right)}{\partial \hat{a}(k)}\right\} \\ &= -2\mathrm{E}\left\{e(k)\,\frac{\partial}{\partial t}\hat{d}_c(t)\Big|_{t=(k-1)T_y+\hat{a}(k)T_x=\tilde{t}_k} T_x\right\}. \end{aligned} \tag{17}$$

By using the abbreviation

$$d'(k) = T_x \frac{\partial}{\partial t}\hat{d}_c(t)|_{t=\tilde{t}_k} \tag{18}$$

the gradient is expressed as

$$\nabla_f = -2\mathrm{E}\left\{e(k)d'(k)\right\}. \tag{19}$$

The gradient ∇_f is then approximated by the instantaneous gradient

$$\hat{\nabla}_f(k) = -2e(k)d'(k). \tag{20}$$

With the abbreviation $\hat{d}(k) = \hat{d}_c(\tilde{t}_k)$, the straigthforward approximation of the gradient is $d'(k) = \hat{d}(k) - \hat{d}(k-1)$. However, when substituting $e(k) = y(k) - \hat{d}(k)$ in (20), this approximation would lead to

$$\begin{aligned} \hat{\nabla}_f(k) = -2\Big(&y(k)\hat{d}(k) - y(k)\hat{d}(k-1) \\ &-\hat{d}^2(k) + \hat{d}(k)\hat{d}(k-1)\Big). \end{aligned} \tag{21}$$

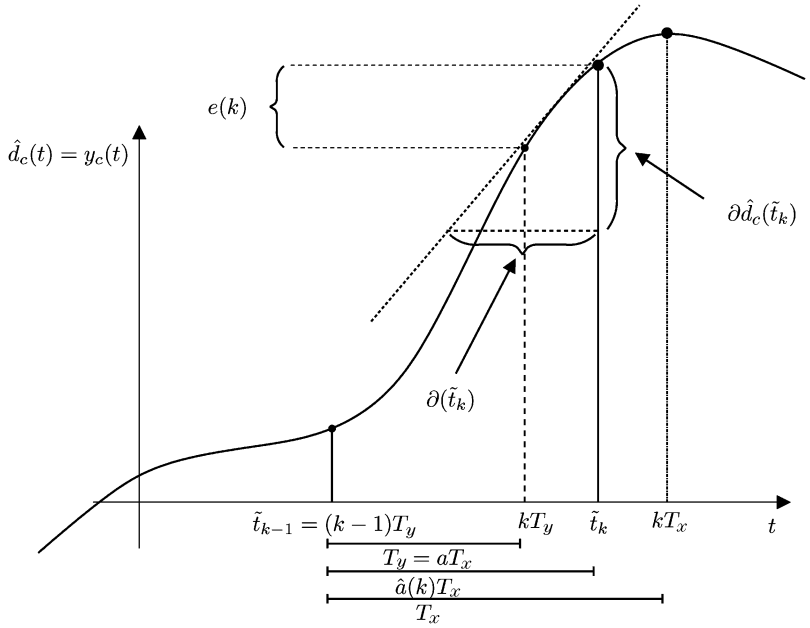

Fig. 6. Illustration of the components for gradient descent, demonstrating the sampling error in case of identical continuous-time waveforms $\hat{d}_c = y_c$.

It has been observed that in the case where the adaptive filter has not yet synchronized the signals $y(k)$ and $\hat{d}(k)$ good enough, the correlation is very low, so that the quadratic term $\hat{d}^2(k)$ is the only term in the gradient which is not close to zero and thus causes a constant bias.

To avoid the occurence of the term $\hat{d}^2(k)$, the gradient will instead be approximated by the difference between the two adjacent samples previous and subsequent to $\tilde{t}_k$ as

$$d'(k) \approx T_x \frac{\hat{d}(k+1) - \hat{d}(k-1)}{2T_x} = \frac{1}{2}\left(\hat{d}(k+1) - \hat{d}(k-1)\right) \tag{22}$$

causing a delay of one sample.

Given the above approximation of the derivative, the adaptation is performed in the direction of the negative gradient by

$$\hat{a}(k+1) = \hat{a}(k) - \mu_f(k)\hat{\nabla}_f(k) \tag{23}$$

where $\mu_f(k)$ is the time-variable step-size which will be analyzed in detail in Section III-D.

A graphical illustration of the adaptation by gradient descent can be seen in Fig. 6. The situation depicted here, idealized for the purpose of understanding the algorithm, is as follows: There is no near end signal $s(t)$ or $n(t)$. The two waveforms $y(t) = d(t)$ and $\hat{d}(t)$ are perfectly synchronized by the adaptive filter at the time instant $(k-1)T_y$, such that their waveforms are exactly the same. The remaining error $e(k)$ is then caused by the different sampling time instants kT_y of the signal $y(t)$ and $\tilde{t}_k$ of $\hat{d}(t)$, respectively.

In the example given in Fig. 6, it can be observed that the currently estimated factor $\hat{a}(k)$ has to be corrected in the negative direction to meet the correct value a with respect to the actual frequency offset Δf. The sampling time $\tilde{t}_k$ is greater than the ideal sampling time kT_y. The error $e(k)$ is less than zero and the derivative $d'(k)$ is positive. Hence, according to (20) the gradient $\nabla_f(k)$ is also positive, meaning $a(k+1)$ is correctly modified in the negative direction. This reduces the effective frequency offset between the signals $\hat{d}(k)$ and $y(k)$.

D. Adaptive Step-Size for Double-Talk Robustness

Like every other adaptive algorithm, the LMS frequency offset estimation critically depends on a suitable step-size. The proposed time-variable step-size $\mu_f(k)$ introduced in (23) consists of two factors:

$$\mu_f(k) = \mu_{\text{fix}} \cdot \mu_{\text{opt}}(k). \tag{24}$$

The fixed step-size factor μ_{fix} causes averaging over time. This is useful since the reliability of single instants of the gradient $\nabla_f(k)$ is not very high, because the algorithm relies on the ability of the adaptive filter to synchronize the signals under every circumstance. The synchronization of the signals $y(k)$ and $\hat{d}(k)$ is not exact in practice. For this reason, and because of the high sensitivity of the system to changes of $\hat{a}(k)$, the averaging by μ_{fix} has to consider a large number of values for $\hat{a}(k)$, which motivates a fixed step-size factor $\mu_{\text{fix}} \in [10^{-7}; 10^{-6}]$.

Because the presence of significant near end disturbance or double-talk leads to incorrect adaptation, the step-size has to be decreased in these circumstances. Therefore, a step-size factor μ_{opt} similar to the optimum step-size factor (7) of the NLMS algorithm for room impulse response tracking is introduced. Comparing the update equation for the NLMS filter (1) with the update equations of the proposed algorithm (20) and (23) suggests that the role of $\mathbf{x}(k)$ in (1) is similar to the role of $d'(k)$ in (20), (23), thus in analogy to (7) an auxiliary factor is calculated as

$$\tilde{\mu}_{\text{opt}}(k) = \frac{g_f^2 \sigma_{d'}^2(k)}{g_f^2 \sigma_{d'}^2(k) + \sigma_e^2(k)}. \tag{25}$$

The factor g_f^2 which corresponds to the system distance (5) controls the adaptation speed and the sensitivity for double-talk. It

represents the remaining degree of freedom to change the behavior of the algorithm. The short term power $\sigma^2_{d'}(k)$ is defined by

$$\sigma^2_{d'}(k) = \frac{1}{P} \sum_{\kappa=k-P+1}^{k} d'^2(\kappa). \tag{26}$$

This auxiliary factor $\tilde{\mu}_{\text{opt}}(k)$ slows down adaptation significantly in case of double-talk or other considerable near end disturbances. In case of low power of the near end signals $s(k)$ and $n(k)$, the power $\sigma^2_e(k)$ of the remaining error is close to zero and the step-size factor $\tilde{\mu}_{\text{opt}}$ is close to one. In cases of high near end signal powers, the factor $\tilde{\mu}_{\text{opt}}(k)$ is close to zero because of the high power $\sigma^2_e(k)$ relative to $\sigma^2_{d'}(k)$.

Because of the nonstationarity of speech signals, the changes $\Delta\hat{a}(k)$ to the time stretching factor $\hat{a}(k)$ have to be normalized. Otherwise, the influence of high amplitude parts of the input signal would be overrated. Normalization is performed by using the instantaneous power of the derivative $d'^2(k)$, i.e., $\mu_{\text{opt}} = \tilde{\mu}_{\text{opt}}/d'^2(k)$. To avoid numerical instability, the term $d'^2(k)$ is lower bounded to $\epsilon_f \ll \mathrm{E}\{d'^2\}$ by using $\max(d'^2(k), \epsilon_f)$. To simplify the required equations, the short-term power $\sigma^2_{d'}(k)$ is approximated by the instantaneous power $d'^2(k)$ and the step-size factor becomes

$$\mu_{\text{opt}}(k) = \frac{g_f^2 \sigma^2_{d'}(k)}{g_f^2 \sigma^2_{d'}(k) + \sigma^2_e(k)} \cdot \frac{1}{d'^2(k)} \tag{27}$$

$$\approx \frac{g_f^2}{g_f^2 d'^2(k) + \sigma^2_e(k)}. \tag{28}$$

The selection of the control parameter g_f^2 will be discussed in the next section.

E. Increasing the Estimation Accuracy While Preserving Fast Convergence

The performance analysis of the novel algorithm indicates that small values for the system distance g_f^2 are necessary to reach good steady state performance in case of near end disturbances. However, in the start up period of the algorithm, slow convergence is observed for these small values of g_f^2. During this period, the residual error signal $e(k)$ has significant power. Since the power $\sigma^2_e(k)$ is used to detect double-talk, this causes the step-sizes of both the adaptive filter as well as the LMS frequency offset detection to be very small and thus the adaptation is very slow. To increase the convergence speed, the influence of the remaining echo power $\sigma^2_e(k)$ has to be reduced, which directly increases the impact of double-talk in the adapted case. In order to solve this dilemma, the specific systems properties have to be considered.

In our investigation of acoustic echo control on PC systems, measurements showed that different systems have different frequency offsets, but the offset specific to a system remains stable. The only change of the offset behavior observed in these systems was due to different audio driver configurations.

The assumption of an unknown but constant or very slowly changing frequency offset encourages the use of a time-variable parameter $\hat{g}_f^2(k)$. The general adaptation of the offset works

TABLE I
LMS-FOE ALGORITHM

for each k**:**

resample the far end input signal with current ratio $\hat{a}(k)$

$$x'(k) = \text{ASRC}(x(k), \hat{a}(k))$$

create the input vector

$$\mathbf{x}'(k) = [x'(k-N), x'(k-N+1), \ldots, x'(k)$$

create the estimate of the echo

$$\hat{d}(k) = \mathbf{w}^T(k)\mathbf{x}'(k)$$

calculate the output signal

$$e(k) = y(k) - \hat{d}(k)$$

calculate the power estimates ($\sigma^2_{e,x'}(0) = 0$)

$$\sigma^2_e(k) = (1-0.99)e^2(k) + 0.99\sigma^2_e(k-1)$$

$$\sigma^2_{x'}(k) = (1-0.99)x'^2(k) + 0.99\sigma^2_{x'}(k-1)$$

calculate the adaptive step-size

$$\alpha(k) = \frac{g^2\sigma^2_{x'}(k)}{g^2\sigma^2_{x'} + \sigma^2_e(k)}$$

perform filter update

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \alpha(k)e(k)\frac{\mathbf{x}(k)}{\max(||\mathbf{x}(k)||^2, \epsilon)}$$

calculate derivative (delay by one sample)

$$d'(k) = \frac{1}{2\hat{a}(k)}(d(k+1) - d(k-1))$$

calculate LMS-FOE step-size

$$\mu_{opt}(k) = \frac{\hat{g}_f(k)^2}{\hat{g}_f^2(k)\max(d'(k), \epsilon_f) + \sigma_e(k)}$$

$$\mu_f(k) = \mu_{fix}\mu_{opt}$$

update $\hat{g}_f^2(k+1)$

$$\hat{g}_f^2(k+1) = (1 - \mu_f d'^2(k)\gamma_g)\hat{g}_f^2(k) + \mu_k(k)\gamma_g g^2_{end}$$

update time stretching factor

$$\hat{a}(k+1) = \hat{a}(k) + \mu_f \cdot e(k)d'(k)$$

even for large values, e.g., $g_f^2 \approx 0.2$. The main problem caused by these large values is the instability of the estimated time stretching factor $\hat{a}(k)$ in case of near end disturbance. With the assumption of improving convergence of the FOE algorithm, the parameter $\hat{g}_f^2$ is replaced by a time-variable parameter

$$\hat{g}_f^2(k) = \left(1 - \mu'_f(k)\gamma_g\right) \cdot \hat{g}_f^2(k-1) + \mu'_f(k)\gamma_g \cdot g^2_{\text{end}}. \tag{29}$$

Here, the parameter $\hat{g}_f^2(k)$ is initiated to $\hat{g}_f^2(0) = 0.2$ to employ the fast convergence at the start of the adaptation. Using a fixed time-constant $\gamma_g \in [10^{-4}; 10^{-3}]$, the parameter $\hat{g}_f^2(k)$ is reduced over time, in the best case reaching the end value $g^2_{\text{end}} = 0.001$ as soon as the adaptation of the frequency offset is converged.

In addition to the fixed time-constant γ_g, the adaptive step-size for frequency offset estimation is employed in the denormalized form

$$\mu'_f(k) = \mu_f(k) \cdot d'^2(k) \in [0; 1] \tag{30}$$

to stop the descent of the parameter $\hat{g}_f^2(k)$ in case of near end disturbance. This ensures that $\hat{g}_f^2(k)$ is not decreased during strong double-talk, where no convergence of the FOE algorithm or the room impulse response tracking filter can be assumed. Once the parameter $\hat{g}_f^2(k)$ is close to the small final value g^2_{end}, the offset estimation step-size $\mu_{\text{opt}}(k)$ is very small in case of double-talk or other near end disturbance, and the estimation remains stable. Finally, the proposed algorithm is summarized in Table I.

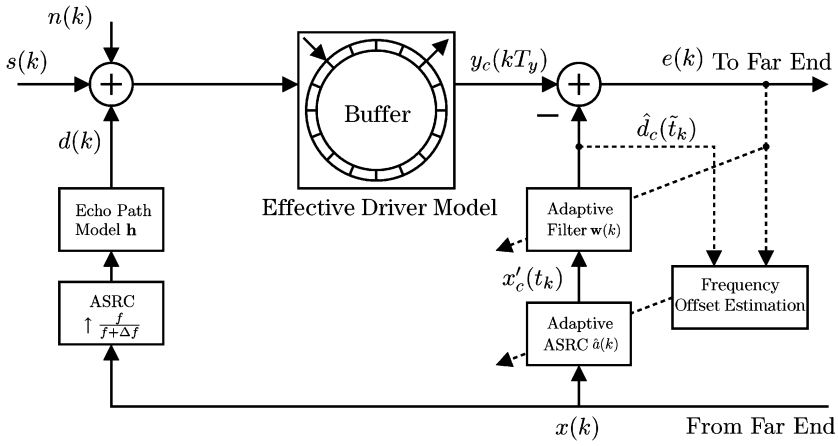

Fig. 7. Simulation model.

F. Computational Complexity

The additional algorithmic complexity of the algorithm is mostly due to the digital interpolation filter in the ASRC. In this paper, the interpolation employs a polyphase filter implementation with filter lengths of 50 taps. The remainder of the LMS-FOE algorithm consists of 27 multiplications, 13 summations, and one maximum operation per iteration and, thus, does not add much complexity in comparison to an adaptive filter of length $N \geq 300$ in acoustic echo control.

IV. Simulation Results

The performance of the proposed algorithm was analyzed in simulations according to Fig. 7. The employed far end input signals $x(k)$ were recorded speech by male and female speakers, taken from a German audio book. Different room impulse responses $\mathbf{h}$ were generated randomly by

$$h(i) = \begin{cases} 0.01 r(i) & i < \Delta_h \\ r(i)(e^{-i} + 0.1)e^{-\frac{i}{0.15M}} & i \geq \Delta_h \end{cases} \tag{31}$$

where $r(i)$, $0 \leq i \leq M$ is a white noise random variable of Gaussian distribution and variance $\sigma_r^2 = 1$, Δ_h is a direct sound delay and M is the desired filter length. These random room impulse responses resemble measured room impulse responses. The effect of the frequency offset Δf was modeled by ASRC. The possible time-scale jump effects of the audio driver were considered by the driver model in the microphone path. This model was implemented by observing the current sampling frequency difference and the according relative position of the read and write pointers on a buffer. Whenever they meet, a buffer length is repeated or removed from the signal, simulating the additional time jump effect. The buffer length did not have significant impact on the performance of the algorithm, as long as it was smaller than the adaptive filter length of the echo canceler. A near end white noise signal was added to the echo signal to adjust the echo-to-noise ratio

$$\frac{\mathrm{ENR}}{\mathrm{dB}} = 10 \log_{10} \frac{\mathrm{E}\left\{d^2(k)\right\}}{\mathrm{E}\left\{n^2(k)\right\}} \tag{32}$$

in the cases denoted as *single-talk*. The *double-talk* case was considered by adding a near end speech signal with a ratio of 0 dB between near end and far end signal powers, in addition to a noise signal of 25 dB ENR. Performance of the echo canceler was evaluated in terms of echo return loss enhancement

$$\frac{\mathrm{ERLE}}{\mathrm{dB}} = 10 \log_{10} \frac{\mathrm{E}\left\{d^2(k)\right\}}{\mathrm{E}\left\{\left(d(k) - \hat{d}(k)\right)^2\right\}}. \tag{33}$$

The gain of all room impulse responses used in the simulations was normalized to 0 dB.

A. Basic Algorithm Results

Simulation results of the LMS-FOE algorithm in single-talk are shown in Fig. 8 for a relatively short room impulse response of length $M = 500$ and an NLMS filter length of $N = 300$. It can be observed that the frequency estimation converges to the correct value of $\Delta f = 2$ Hz after about 30 s in case of 60 dB ENR and 45 s in case of 25 dB ENR. The convergence time varies with different speech input signals as well as different near end disturbances. The ERLE reached after compensating for the frequency offset is close to the expected limits if using the NLMS filter in a system with $\Delta f = 0$ as stated in e.g., [17], Chapter 13. The echo canceler performance is limited by near end disturbance and by the length of the echo canceler N, which is mostly shorter than the actual room impulse response length M.

The synergy between adaptive filtering by NLMS and adaptive frequency offset correction is of special interest. The FOE starts working as soon as the adaptive filter has found a very rough estimate of the actual echo path impulse response. The most important feature of the filter to be found is the direct sound peak, representing the essential delay of the echo path. Once this is estimated roughly, the LMS algorithm for frequency offset estimation is able to detect the correct value for Δf. While the remaining frequency error is reduced, the adaptive filter is able to improve the estimate of the room impulse response, which in turn leads to a better estimate of the frequency offset. This way, both adaptive components improve the performance of each other until a stable estimate of both echo path impulse response as well as the frequency offset is found.

As illustrated in Fig. 8, the offset estimation slows down in case of more near end noise and thus a lower ENR. This is due

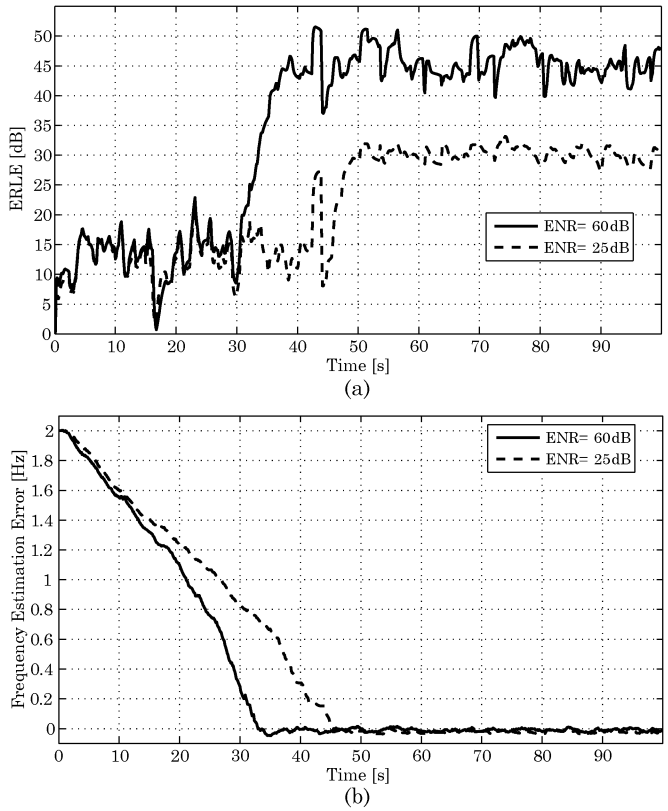

Fig. 8. Simulation results for LMS-FOE. (a) ERLE results. (b) Remaining frequency estimation error; speech input without double-talk, $N = 300$, $M = 500$, $\Delta f = 2$ Hz, $g_f^2 = 0.01$.

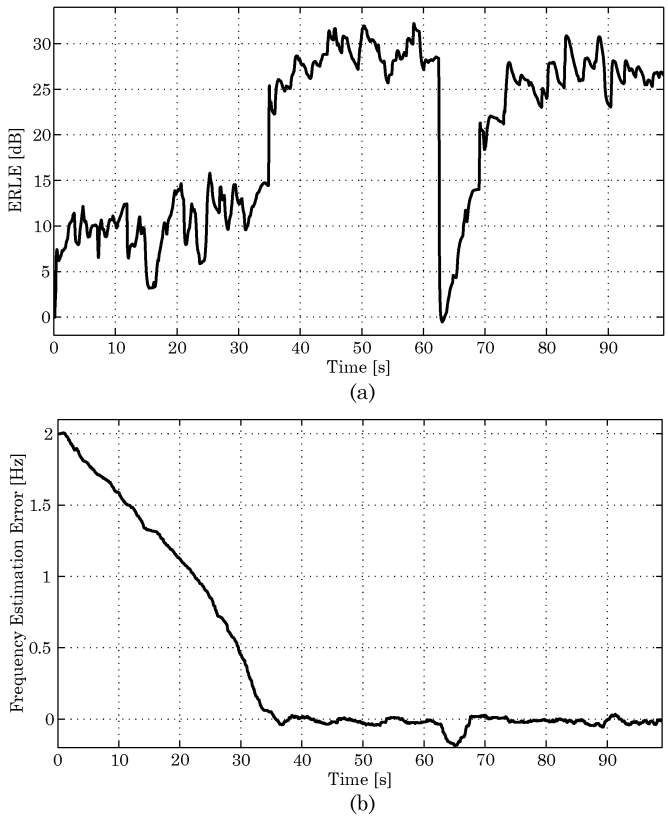

Fig. 9. Simulation results for LMS-FOE and abruptly changing room impulse response after 62 s; speech input without double-talk, (a) ERLE results, (b) remaining frequency estimation error, $N = 300$, $M = 500$, $\Delta f = 2$ Hz, $g_f^2 = 0.01$.

to the adaptive step-size $\mu_{\text{opt}}(k)$, which considers the lower significance of single estimated values for $a(k)$. Another obvious reason for the slower convergence is the slower convergence of the adaptive filter under the influence of near end noise, caused by the adaptive step-size of the NLMS algorithm. However, despite slower detection of a frequency offset in case of near end disturbance, the frequency detection remains stable.

The stability of the algorithm needs to be ensured in case of sudden changes of the echo path impulse response. The result of an abrupt change of the impulse response can be observed in Fig. 9. The room impulse response in this simulation was replaced by a completely different one at the time of 62 s. In case of such a sudden change, the illustration shows that the resynchronization of the NLMS algorithm used for the echo cancellation and room impulse tracking filter is fast enough not to cause serious deterioration of the frequency estimation. This is a very important feature. In practice the room impulse response might not jump, but in any case will be time-variable. If the algorithm can deal with abrupt changes like this, it can also deal with slower changes of the impulse response.

B. Time-Variable Parameter g_f^2 for Fast Convergence

The behavior of the frequency estimation error and the associated parameter $g_f^2(k)$ are illustrated in Fig. 10. The 60 and 25 dB ENR signals represent the single-talk case with only near end noise while the 0 dB double-talk scenario consists of a near end speaker at the same mean energy as the far end speaker with added white noise at 25 dB. It can be observed that the

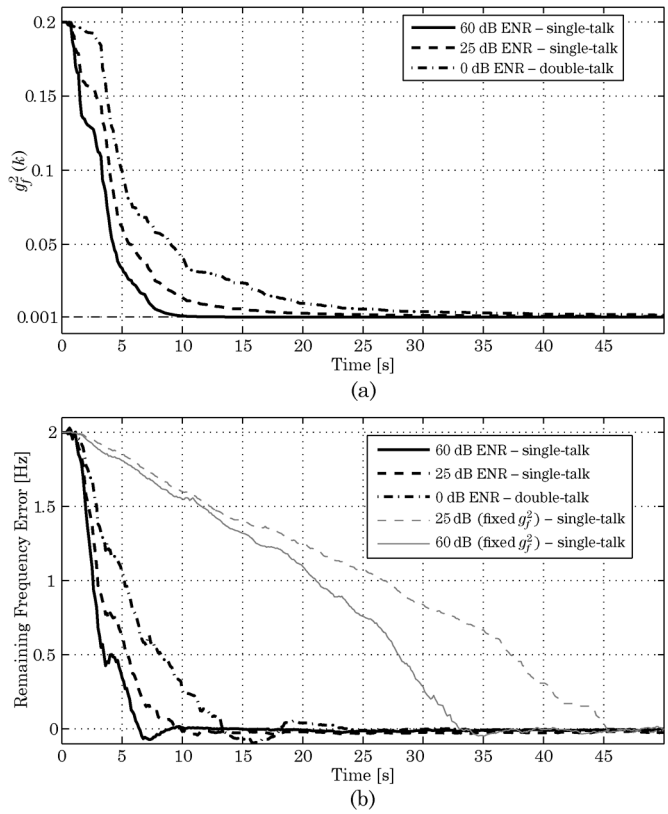

Fig. 10. Time-variable g_f^2 (a) and remaining frequency estimation error (b), $N = 300$, $M = 500$, $\Delta f = 2$ Hz.

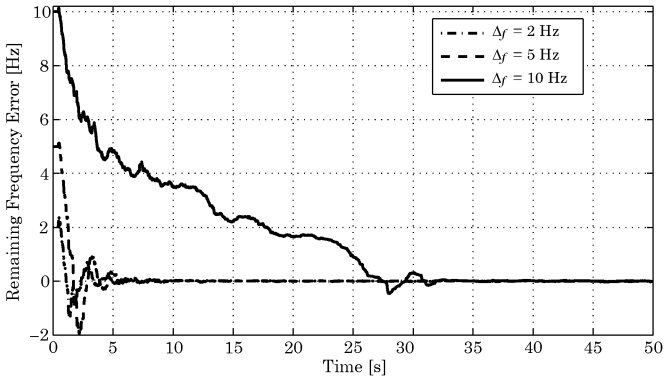

Fig. 11. Frequency estimation for different Δf, $N = 300$, $M = 500$.

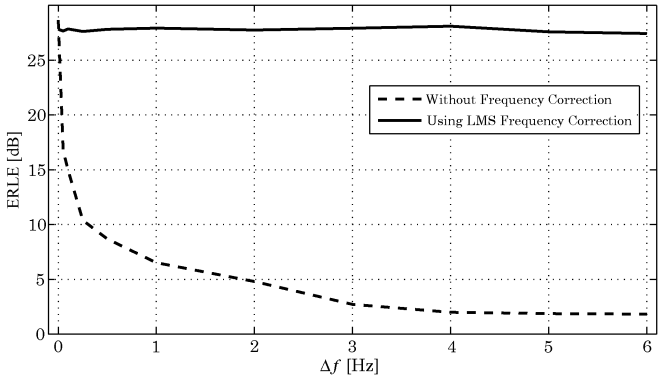

Fig. 12. ERLE development versus frequency offset Δf; $M = 1500$, $N = 1000$, ENR $= 25$ dB.

convergence is considerably faster than using a constant parameter $g_f^2 = 0.01$ as illustrated in the gray plots. For fixed g_f^2 and double-talk, convergence is not reached within the scope of this plot. Ideally, the descent of $g_f^2(k)$ is adjusted to reach the final value $g_{\text{end}}^2 = 0.001$ a short time after the point of actual convergence to maximize both steady-state stability and double-talk robustness, as well as convergence speed. The illustration also shows the slower descent in cases where adaptation is disturbed by double-talk influence. With this time-variable approach, the convergence time can be reduced to similar values as when using large parameters g_f^2, while preserving the high accuracy and high ERLE of slow convergence speeds.

The course of the frequency offset estimation for larger frequency offsets can be observed in Fig. 11 for a speech signal. It should be noted that the fixed step-size has been slightly increased compared to Fig. 10, which leads to some overshooting at the start of the estimation for smaller offsets, but decreases the convergence time for larger offsets.

C. Practical Value of Adaptive Resampling by LMS-FOE

When using the adaptive sampling rate correction as described in the paper, the performance of AEC in the presence of a frequency offset Δf can increase dramatically compared to the pure NLMS algorithm. In order to provide realistic simulation results according to the probable setup for this algorithm in practice, the simulated echo path impulse response uses $M = 1500$ coefficients to represent a regular office room. To reach acceptable ERLE values, the adaptive filter length is set to $N = 1000$. With a PC-fan in the acoustic environment of the system, a constant near end disturbance $n(k)$ with an ENR of 25 dB is considered realistic.

Fig. 12 shows the average ERLE in the steady state for different frequency offsets Δf. Without any offset compensation, the ERLE deteriorates severely with increasing offset Δf. The LMS frequency estimation algorithm combined with an NLMS adaptive filter introduced in the paper achieves that the performance of the AEC remains nearly constant for different frequency offsets. In this simulation, the loss of performance in case of $\Delta f = 0$ Hz due to the unavoidable frequency estimation error is less than 1 dB.

Quantitative simulation results of the convergence time are presented in Table II for an offset of $\Delta f = 2$ Hz. Here, ERLE means the average ERLE after convergence is reached. The convergence time is measured as the time until the ERLE reaches 3 dB less than the steady-state. The reason for the slower convergence in case of noise input is the stationarity of the white noise signal. In this case, the step-size remains constantly low at the beginning of the adaptation. In case of speech input, however, high amplitude sections of the input signal yield a big step-size, causing the algorithm to converge fast.

TABLE II
SIMULATION RESULTS FOR TIME-VARIABLE $g_f^2(k)$; $M = 1500$, $N = 1000$, $\Delta f = 2$ Hz

input signal	ENR [dB]	ERLE [dB]	convergence time [s]
white noise	60, single-talk	37.0883	25.64
	25, single-talk	27.5620	25.97
	0, double-talk	23.0163	32.73
speech	60, single-talk	39.3500	10.86
	25, single-talk	28.0201	10.55
	0, double-talk	17.8953	18.78

In conclusion, these simulations prove that the introduced LMS algorithm is suitable for frequency offset detection and correction in the simulated environment. The proposed algorithm works as long as the adaptive filter is able to find a rough estimate of the actual echo path impulse response. With the NLMS filter as described in Section II, convergence can be reached for offsets of at least $\Delta f \in [-10\text{ Hz}; 10\text{ Hz}]$. However, the convergence time grows with the frequency offset Δf.

V. CONCLUSION

In this paper, we proposed a novel LMS-type algorithm to estimate and correct a sampling frequency offset between the microphone and loudspeaker signal in an acoustic echo control environment. The resulting echo return loss enhancement of a system with a small frequency offset is dramatically improved.

First, the fundamental problems caused by the differing sampling periods and thus, time-variable delay between the input and output signals of the system were described in detail. It was shown that the deteriorating effects can be compensated by adaptive resampling. The proposed algorithm uses the principle of gradient descent to adapt the time-stretching/-compression parameter needed for the compensation.

Simulations proved that the proposed LMS-type adaptive resampling algorithm is suitable for frequency offset detection and correction in the simulated environment. The algorithm works

as long as the adaptive filter is able to find a rough estimate of the actual echo path impulse response. Naturally, the convergence time grows with the frequency offset Δf.

REFERENCES

[1] G. Schmidt, "Applications of acoustic echo control—An overview," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, 2004, pp. 9–16.

[2] E. Hänsler, "The hands-free telephone problem—An annotated bibliography update," *Ann. Telecommun.*, vol. 49, no. 7–8, pp. 360–367, 1994.

[3] 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, 1999.

[4] *Acoustic Signal Processing for Telecommunications*, S. L. Gay and J. Benesty, Eds.. Dordrecht, Germany: Kluwer Accademic, 2000.

[5] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1996.

[6] J. Benesty, T. Gänsler, D. Morgan, M. Sondhi, and S. Gay, *Advances in Network and Acoustic Echo Cancellation*. Berlin, Germany: Springer, 2001.

[7] E. Hänsler and G. Schmidt, *Acoustic Echo and Noise Control: A Practical Approach*. New York: Wiley, 2004.

[8] A. Mader, H. Puder, and G. Schmidt, "Step-size control for acoustic echo cancellation filters—An overview," *Signal Process.*, vol. 80, no. 9, pp. 1697–1719, Sep. 2000.

[9] J. Benesty, D. R. Morgan, and J. H. Cho, "A new class of doubletalk detectors based on cross-correlation," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 2, pp. 168–172, 2000.

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

[11] R. Lienhart, I. Kozintsev, S. Wehr, and M. Yeung, "On the importance of exact synchronization for distributed audio signal processing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '03)*, 2003, vol. 4, pp. IV-840–IV-843.

[12] J. W. Stokes and H. S. Malvar, "Acoustic echo cancellation with arbitrary playback sampling rate," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '04)*, 2004, vol. 4, pp. iv-153–iv-156.

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

[14] J. G. Proakis, C. M. Rader, F. Ling, and C. L. Nikias, *Advanced Digital Signal Processing*, Maxwell Macmillan Int. Editions ed. New York: Macmillan, 1992.

[15] G. Evangelista, "Design of digital systems for arbitrary sampling rate conversion," *Signal Process. Elsevier*, vol. 83, no. 2, pp. 377–387, Oct. 2003.

[16] U. Zölzer, *Digitale Audiosignalverarbeitung*. Stuttgart, Germany: B. G. Teubner, 1996.

[17] P. Vary and R. Martin, *Digital Speech Transmission*. New York: Wiley, 2006.

Matthias Pawig received the Dipl.-Ing. degree in information and communication technology from RWTH Aachen University, Aachen, Germany, in 2006.

He is currently with the Institute of Communication Systems and Data Processing, RWTH Aachen University, where he is also pursuing the Dr.-Ing. degree. His research interests cover the areas of hands-free communication, acoustic echo control, and adaptive filter algorithms.

Gerald Enzner (S'00–M'06) received the Dipl.-Ing. degree from University of Erlangen-Nuremberg, Germany, in 2000, and the Dr.-Ing. degree from RWTH Aachen University, Germany, in 2006, both in electrical engineering.

Since 2006, he is a Principal Scientist at the Institute of Communication Acoustics, Ruhr-University Bochum, Germany. His research interests include dynamical modeling, adaptive filtering, detection and estimation, speech and audio enhancement, and signal processing for spatial sound control.

Peter Vary (M'85–SM'04–F'09) received the Dipl.-Ing. degree in electrical engineering from the University of Darmstadt, Darmstadt, Germany, in 1972 and the Dr.-Ing. degree from the University of Erlangen-Nuremberg, Germany, in 1978.

In 1980, he joined Philips Communication Industries (PKI), Nuremberg, where he became head of the Digital Signal Processing Group. Since 1988, he has been a Professor with RWTH Aachen University, Aachen, Germany, and Head of the Institute of Communication Systems and Data Processing. His main research interests are speech coding, joint source-channel coding, error concealment, and speech enhancement including noise suppression, acoustic echo cancellation, and artificial wideband extension.