Improved prediction error filters for adaptive feedback cancellation in hearing aids

Ngo, Kim; van Waterschoot, Toon; Christensen, Mads Græsbøll; Moonen, Marc; Jensen, Søren Holdt

Published in:
Signal Processing

DOI (link to publication from Publisher):
10.1016/j.sigpro.2013.03.042

Publication date:
2013

Document Version
Publisher's PDF, also known as Version of record

Signal Processing 93 (2013) 3062–3075

Contents lists available at SciVerse ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Improved prediction error filters for adaptive feedback cancellation in hearing aids

Kim Ngo [a,1], Toon van Waterschoot [a,*], Mads Græsbøll Christensen [b], Marc Moonen [a], Søren Holdt Jensen [c]

[a] KU Leuven, Department of Electrical Engineering (ESAT-SCD)/IBBT Future Health Department, Kasteelpark Arenberg 10, 3001 Leuven, Belgium
[b] Aalborg University, Department of Architecture, Design & Media Technology (AD:MT), Audio Analysis Lab, Sofiendalsvej 11, 9200 Aalborg, Denmark
[c] Aalborg University, Department of Electronic Systems, Fredrik Bajers Vej 7, 9220 Aalborg, Denmark

ARTICLE INFO

Article history:
Received 27 September 2012
Received in revised form
29 March 2013
Accepted 30 March 2013
Available online 17 April 2013

Keywords:
Acoustic feedback
Adaptive feedback cancellation
Hearing aids
Prediction error filter
Linear prediction
Sinusoidal modeling

ABSTRACT

Acoustic feedback is a well-known problem in hearing aids, caused by the undesired acoustic coupling between the hearing aid loudspeaker and microphone. Acoustic feedback produces annoying howling sounds and limits the maximum achievable hearing aid amplification. This paper is focused on adaptive feedback cancellation (AFC) where the goal is to adaptively model the acoustic feedback path and estimate the feedback signal, which is then subtracted from the microphone signal. The main problem in identifying the acoustic feedback path model is the correlation between the near-end signal and the loudspeaker signal caused by the closed signal loop, in particular when the near-end signal is spectrally colored as is the case for a speech signal. This paper adopts a prediction-error method (PEM)-based approach to AFC, which is based on the use of decorrelating prediction error filters (PEFs). We propose a number of improved PEF designs that are inspired by harmonic sinusoidal modeling and pitch prediction of speech signals. The resulting PEM-based AFC algorithms are evaluated in terms of the maximum stable gain (MSG), filter misadjustment, and computational complexity. Simulation results for a hearing aid scenario indicate an improvement up to 5–7 dB in MSG and up to 6–8 dB in terms of filter misadjustment.

1. Introduction

Acoustic feedback is a well-known problem in hearing aids. It is caused by the undesired acoustic coupling between the hearing aid loudspeaker and microphone. It has become an even more important problem due to two recent trends in hearing aids design, both of which further increase the loudspeaker–microphone coupling: (1) the use of open fittings, in which the ear canal is intentionally left open to avoid the occlusion effect and hence improve the user comfort, (2) the use of smaller form factors which implicitly reduce the hearing aid dimensions, including the loudspeaker–microphone distance. The acoustic loudspeaker–microphone coupling results in a closed signal loop which may become unstable, resulting in acoustic oscillations known as howling. Therefore, as a consequence of acoustic feedback, the speech intelligibility and listening comfort for hearing aid users is compromised in two ways: acoustic feedback may result in

* Corresponding author. Tel.: +32 16 321788; fax: +32 16 321970.
E-mail addresses: kim.ngo@citrix.com (K. Ngo), toon.vanwaterschoot@esat.kuleuven.be (T. van Waterschoot), mgc@create.aau.dk (M. Græsbøll Christensen), marc.moonen@esat.kuleuven.be (M. Moonen), shj@es.aau.dk (S. Holdt Jensen).
[1] Present address: Citrix Online GmbH, Theaterstrasse 6, 01067 Dresden, Germany.

0165-1684/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.sigpro.2013.03.042

howling artifacts that interfere with desired speech components, and it may severely constrain the maximum hearing aid amplification that can be used if howling, due to instability, is to be avoided. In many cases this maximum amplification is too small to compensate for the hearing loss, i.e., the auditory loss in the user, and therefore feedback cancellation is considered a crucial component in present-day hearing aids [1–3].

The goal of adaptive feedback cancellation (AFC) is to adaptively model the acoustic feedback path and estimate the feedback signal, which is then subtracted from the microphone signal. The main problem in identifying the acoustic feedback path model is the correlation between the near-end signal and the loudspeaker signal caused by the closed signal loop, in particular when the near-end signal is spectrally colored, as it is the case for a speech signal. This correlation problem causes standard adaptive filtering algorithms to converge to a biased solution. A major challenge is therefore to reduce the correlation between the near-end signal and the loudspeaker signal. Typically, there exist two approaches to achieve this decorrelation [4], i.e., decorrelation in the closed signal loop and decorrelation in the adaptive filtering circuit. Recently proposed methods for decorrelation in the closed signal loop consist of inserting all-pass filters in the forward path of the hearing aid [5], applying a clipping operation to the feedback signal arriving at the microphone [6], or inserting a probe noise signal into the closed signal loop [7]. However, decorrelation in the closed signal loop implicitly affects the desired (near-end) speech component, hence a trade-off between signal decorrelation and perceptual degradation is unavoidable [4].

Alternatively, an unbiased identification of the feedback path model can be achieved by applying decorrelation in the adaptive filtering circuit, i.e., by first prefiltering the loudspeaker and microphone signals with the inverse near-end signal model before feeding these signals to the adaptive filtering algorithm [8–10]. In this way, the desired (near-end) speech component remains unaffected and so the signal decorrelation does not induce any perceptual degradation. The near-end signal model and the feedback path model can then be jointly estimated using the prediction error method (PEM) [11]. For PEM-based AFC with near-end speech signals, a linear prediction (LP) model is commonly used [8,12]. Other near-end speech signal models have been based on a pole-zero LP (PZLP), a warped LP, or a pitch prediction model, cascaded with an LP model [9].

Recently, the use of a harmonic sinusoidal near-end signal model for PEM-based AFC has been proposed by the authors [13,14] and was shown to improve the AFC performance compared to using a PZLP near-end signal model. The main difference with the PZLP model of [9] is that the near-end signal model in [13,14] and the corresponding pitch estimation [15,16] rely on harmonicity, i.e., the sinusoidal frequencies are assumed to be integer multiples of a fundamental frequency, which in the near-end speech case follows naturally from voiced speech being quasi-periodic, whereas the sinusoidal frequencies in a PZLP model are estimated independently [17]. In [13] it has been shown how different pitch estimation techniques based on subspace shift-invariance, subspace orthogonality, and optimal filtering [15] can be employed to improve the resulting AFC performance. In [14] it has further been shown how the harmonic sinusoidal near-end signal model and the corresponding design of the prediction error filter (PEF) can be improved by including a variable model order (corresponding to the number of near-end signal harmonics) and a variable amplitude, next to a variable pitch.

In this paper, different designs for the PEF are analyzed and it is shown that a more accurate modeling of the near-end signal generally results in a significant performance improvement in PEM-based AFC in terms of the achieved maximum stable gain (MSG) and filter misadjustment. As compared to our previous work in [13,14], two improved PEF designs are presented, which are inspired by harmonic sinusoidal speech models, as used in speech applications other than AFC, for the extraction, separation, and enhancement of periodic signals [15,16]. The first improvement is based on a refinement of the harmonic sinusoidal model such as to incorporate a number of typical speech features into the PEF design. Since speech is highly non-stationary, the PEF should be able to adapt quickly both in terms of tracking the pitch, number of harmonics, and amplitude changes as well as in terms of characterizing voiced versus unvoiced frames. To this end, the PEM-AFC algorithm proposed in [14] is extended by including a non-intrusive voiced–unvoiced detection algorithm in the PEF design, and the different impact of voiced and unvoiced speech frames on the resulting PEF design is investigated. The second improvement is based on the use of so-called amplitude and phase estimation (APES) filters in the PEF design, which are specifically suited for periodic signals and are optimal and signal-adaptive given the observed signals [18]. Both improvements are then evaluated in terms of the resulting PEM-based AFC performance. Simulation results for a hearing aid scenario indicate an improvement up to 5–7 dB in MSG and up to 6–8 dB in terms of filter misadjustment. Finally, a computational complexity analysis of the competing PEF designs is conducted, which in particular supports the computational benefit of including a voiced–unvoiced detector.

The paper is organized as follows. Section 2 defines the notation and reviews the PEM-based AFC concept, while Section 3 introduces the harmonic sinusoidal near-end signal model. Section 4 elaborates on the different existing and proposed approaches to PEF design. Section 5 shows the effect of voiced–unvoiced detection on the PEF performance, and Section 6 contains a computational complexity analysis. In Section 7 experimental results are presented. The work is summarized in Section 8.

2. Adaptive feedback cancellation

The typical AFC set-up is shown in Fig. 1. The microphone signal is given by

$$y(t) = v(t) + x(t) = v(t) + F(q,t)u(t) \tag{1}$$

where q denotes the time shift operator, e.g., $q^{-k}u(t) = u(t-k)$, and t is the discrete time variable. $F(q,t) = f_0(t) + f_1(t)q^{-1} + \cdots + f_{n_F}(t)q^{-n_F}$ represents a linear finite impulse

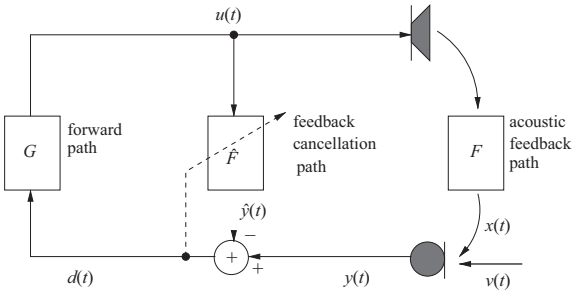

Fig. 1. Concept of an adaptive feedback cancellation (AFC) algorithm.

response (FIR) model of the feedback path between the loudspeaker and the microphone, where n_F is the feedback path model order, $v(t)$ is the near-end signal, $x(t)$ is the feedback signal, and $u(t)$ is the loudspeaker signal. The forward path $G(q,t)$ maps the microphone signal, possibly after AFC, to the loudspeaker signal. It typically consists of an amplifier with a time-varying gain $K(t)$ cascaded with a linear equalization filter $J(q,t)$, such that

$$G(q,t) = K(t)J(q,t) \tag{2}$$

The aim of the AFC is to place an FIR adaptive filter $\hat{F}(q,t)$ in parallel with the feedback path, having the loudspeaker signal as its input and the microphone signal as its desired output. The feedback canceller $\hat{F}(q,t)$ provides an estimate[2] of $F(q,t)$ and also produces an estimate of the feedback signal which is then subtracted from the microphone signal. The feedback-compensated signal is given by

$$d(t) = v(t) + [F(q,t) - \hat{F}(q,t)]u(t) \tag{3}$$

2.1. Prediction error method

The main problem in identifying the feedback path model is the correlation between the near-end signal and the loudspeaker signal, due to the forward path $G(q,t)$, which causes standard adaptive filtering algorithms to converge to a biased solution [4,10]. This means that the adaptive filter does not only predict and cancel the feedback component in the microphone signal, but also part of the near-end signal. This generally results in a distorted feedback-compensated signal. The PEM-based AFC is shown in Fig. 2. It relies on a linear model for the near-end signal, which is specified as

$$v(t) = H(q,t)e(t) \tag{4}$$

where $e(t)$ is a white noise signal. An unbiased identification of the feedback path model can be achieved by applying decorrelation in the adaptive filtering circuit, i.e., by first prefiltering the loudspeaker signal and the microphone signal with the inverse near-end signal model $H^{-1}(q,t)$ (see Fig. 2 with $\hat{H}^{-1}(q,t) = H^{-1}(q,t)$) before feeding these signals to the adaptive filtering algorithm. As $H^{-1}(q,t)$ is obviously unknown, the near-end signal model and the feedback path model have to be jointly identified. The PEM [4,8,9,11]

[2] In the sequel ^ is used to indicate estimates, i.e., $\hat{a}$ denotes an estimate of a.

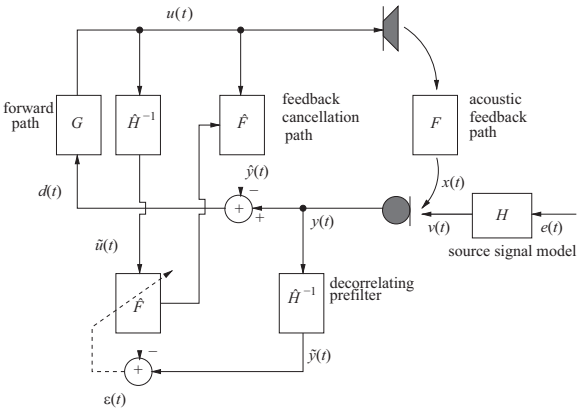

Fig. 2. AFC with decorrelating prefilters in the adaptive filtering circuit.

delivers an estimate of the models $F(q,t)$ and $H(q,t)$, by minimization of the prediction error criterion

$$\min_{\hat{\boldsymbol{\theta}}(t)} \sum_{k=1}^{t} \varepsilon^2(k, \hat{\boldsymbol{\theta}}(t)) \tag{5}$$

where the prediction error is defined as

$$\varepsilon(k, \hat{\boldsymbol{\theta}}(t)) = \hat{H}^{-1}(q,t)[y(k) - \hat{F}(q,t)u(k)] \tag{6}$$

and the parameter vector estimate $\hat{\boldsymbol{\theta}}(t) = [\hat{\mathbf{f}}^T(t), \hat{\mathbf{h}}^T(t)]^T$ contains the parameters of the feedback path model estimate

$$\hat{F}(q,t) = \hat{f}_0(t) + \hat{f}_1(t)q^{-1} + \cdots + \hat{f}_{n_F}(t)q^{-n_F} \tag{7}$$

and the near-end signal model estimate $\hat{H}(q,t)$, the parametrization of which will be discussed later. Note that, throughout the paper, we assume a sufficient-order condition for the feedback path model (i.e., $n_{\hat{F}} = n_F$). We will furthermore assume sufficiently slowly time-varying systems so that $\varepsilon(k, \hat{\boldsymbol{\theta}}(t))$ can also be computed as

$$\varepsilon(k, \hat{\boldsymbol{\theta}}(t)) = [\hat{H}^{-1}(q,t)y(k)] - \hat{F}(q,t)[\hat{H}^{-1}(q,t)u(k)] \tag{8}$$

A final assumption is that the near-end signal $v(t)$ is piecewise stationary, which implies that the near-end signal model $H(q,t)$ does not need to be re-estimated at each time instant t. Instead, $H(q,t)$ will be estimated in a frame-based manner, using a frame length M that approximates the stationarity interval [9,19], and will be denoted by $H(q,i)$, with $i = \lceil t/M \rceil$ the frame index. Unless otherwise stated, we will use a frame length of $M = 320$ samples, which corresponds to 20 ms at a 16 kHz sampling rate. The model structure of $H(q,i)$ and the estimation of its parameters are discussed in detail in the next sections.

2.2. Near-end signal model

A common approach in PEM-based AFC for hearing aids is to model the near-end signal with an LP model [8,12], i.e.,

$$v(t) = H_{\mathrm{LP}}(q,i)e(t), \quad t = (i-1)M + 1, \ldots, iM$$
$$= \frac{1}{1 + \sum_{k=1}^{n_C} c_k(i)q^{-k}} e(t) \tag{9}$$

where $e(t)$ is a white noise signal and n_C is the LP model order. The prediction error for the PEM-based AFC using an

LP model is then

$$\varepsilon(k,\hat{\boldsymbol{\theta}}(t)) = \hat{H}_{\mathrm{LP}}^{-1}(q,i)[y(k)-\hat{F}(q,t)u(k)] \tag{10}$$

where the parameter vector estimate $\hat{\boldsymbol{\theta}}(t)$ is now defined as $\hat{\boldsymbol{\theta}}(t) = [\hat{\mathbf{f}}^T(t), \hat{\mathbf{c}}^T(i)]^T$ with $\hat{\mathbf{c}}^T(i) = [\hat{c}_1(i), \ldots, \hat{c}_{n_c}(i)]$. The drawback with an LP model for the near-end signal is that the white noise assumption is not valid for (quasi-)periodic signals such as voiced speech, where the LP model excitation signal $e(t)$ is an impulse train rather than a white noise signal [8].

2.3. Cascaded near-end signal model

The idea behind using a cascaded near-end signal model is that tonal near-end signal components can be represented by one model and broadband components by another model. In [9] it has been shown that a PZLP model of order $2P$ [20] can be used to represent P tonal components. Still, by constraining the poles and the zeros to lie on a pairwise common radial line and at a specified distance from the origin in the z-plane, the number of unknown parameters in the pole-zero model can be limited to P and the LP parameters can be uniquely related to the unknown frequencies [17]. The constrained PZLP (CPZLP) model can be written as

$$\begin{aligned} v(t) &= H_{\mathrm{CPZLP}}(q,i)e(t), \quad t = (i-1)M+1, \ldots, iM \\ &= \left(\prod_{n=1}^{P} \frac{1-2\rho_n \cos\omega_n(i) q^{-1} + \rho_n^2 q^{-2}}{1-2\nu_n \cos\omega_n(i) q^{-1} + \nu_n^2 q^{-2}}\right) e(t) \end{aligned} \tag{11}$$

where $\omega_n(i)$ denote the frequencies, ν_n the zero radii, and ρ_n the pole radii[3] for $n = 1, \ldots, P$. When $v(t)$ is a known signal, a minimization criterion for estimating the CPZLP model parameters is given by

$$\min_{\hat{\boldsymbol{\omega}}(i)} \frac{1}{M} \sum_{t=(i-1)M+1}^{iM} e^2(t, \hat{\boldsymbol{\omega}}(i)) \tag{12}$$

The residual signal is defined as the output from the PEF,

$$\begin{aligned} e(t,\hat{\boldsymbol{\omega}}(i)) &= \hat{H}_{\mathrm{CPZLP}}^{-1}(q,i)v(t) \\ &= \left(\prod_{n=1}^{P} \frac{1-2\nu_n \cos\hat{\omega}_n(i) q^{-1} + \nu_n^2 q^{-2}}{1-2\rho_n \cos\hat{\omega}_n(i) q^{-1} + \rho_n^2 q^{-2}}\right) v(t) \end{aligned} \tag{13}$$

and $\hat{\boldsymbol{\omega}}(i) = [\hat{\omega}_1(i) \cdots \hat{\omega}_P(i)]^T$. The nonlinear minimization in (12)–(13) can be solved in a decoupled fashion, using an iterative line search optimization algorithm [17]. Using the CPZLP model for the tonal near-end signal components and an LP model for the broadband near-end signal components, the prediction error using a cascaded near-end signal model can be written as

$$\varepsilon(k,\hat{\boldsymbol{\theta}}(t)) = \hat{H}_{\mathrm{LP}}^{-1}(q,i)\hat{H}_{\mathrm{CPZLP}}^{-1}(q,i)[y(k)-\hat{F}(q,t)u(k)] \tag{14}$$

where the PEF for the broadband near-end components $\hat{H}_{\mathrm{LP}}^{-1}(q,i)$ is obtained by LP of the output signal of the tonal components PEF $\hat{H}_{\mathrm{CPZLP}}^{-1}(q,i)$. The parameter vector estimate $\hat{\boldsymbol{\theta}}(t)$ is in this case defined as $\hat{\boldsymbol{\theta}}(t) = [\hat{\mathbf{f}}^T(t), \hat{\mathbf{c}}^T(i), \hat{\boldsymbol{\omega}}^T(i)]^T$.

[3] Note that ρ_n is referred to as the pole radius as it appears in the denominator of the corresponding PEF, see also (13). A similar comment applies to the zero radius ν_n.

3. Harmonic sinusoidal near-end signal model

In [13,14] a harmonic sinusoidal near-end signal model has been proposed where the tonal near-end signal components are represented as a sum of real harmonically related sinusoids. This means that the sinusoids are assumed to have frequencies $\omega_n(i)$ that are integer multiples of a fundamental frequency $\omega_0(i)$, i.e., $\omega_n(i) = n\omega_0(i)$. This follows naturally from, e.g., voiced speech being quasi-periodic. The near-end signal $v(t)$ consisting of a sum of real harmonically related sinusoids and additive noise can be written as

$$\begin{aligned} v(t) &= \sum_{n=1}^{P} \alpha_n(i)\cos(n\omega_0(i)t + \phi_n(i)) + r(t), \\ & t = (i-1)M+1, \ldots, iM \end{aligned} \tag{15}$$

where $\omega_0(i) \in [0, \frac{\pi}{P}]$ is the fundamental frequency, also referred to as the pitch, $\alpha_n(i)$ is the amplitude, and $\phi_n(i) \in [0, 2\pi)$ the phase of the nth sinusoid, and $r(t)$ is the noise which is assumed to be autoregressive, i.e., $r(t) = H_{\mathrm{LP}}^{-1}(q,i)e(t)$. Compared to the CPZLP model in (11), the harmonic sinusoidal near-end signal model offers less modeling flexibility in terms of the tonal component frequencies (which are restricted to be harmonically related), but is more flexible in modeling the tonal components' amplitudes and phases (which are frame-dependent, as opposed to the time-invariant pole and zero radii in the CPZLP model).

The estimation of the harmonic sinusoidal near-end signal model parameters is based on a so-called optimal filtering [15] of the feedback-compensated signal $d(t)$, which is an approximation of the near-end signal $v(t)$. The optimal filtering approach relies on an LP model for the tonal near-end signal components,

$$v(t) = \frac{1}{1-A(q,i)} r(t) \tag{16}$$

such that the filter to be optimized in frame i is an FIR "enhancement filter" $A(q,i)$ producing an output signal $A(q,i)d(t) = \mathbf{a}^T(i)\mathbf{d}(t)$ in which the tonal components have been enhanced w.r.t. the broadband components, with

$$\mathbf{d}(t) = [d(t), \ldots, d(t-N+1)]^T \tag{17}$$

$$\mathbf{a}(i) = [a_0(i), \ldots, a_{N-1}(i)]^T \tag{18}$$

For each pitch candidate $\hat{\omega}_0(i)$ an FIR filter $\hat{A}(q,i)$ can be designed that passes power undistorted at the harmonic frequencies $n\hat{\omega}_0(i)$, while minimizing the output power at all other frequencies. This filter design problem can then be formulated as follows [15]:

$$\min_{\hat{\mathbf{a}}(i)} \quad \hat{\mathbf{a}}^T(i)\mathbf{R}(i)\hat{\mathbf{a}}(i) \tag{19}$$

$$\text{s.t.} \quad \mathbf{Z}^H(\hat{\omega}_0(i))\hat{\mathbf{a}}(i) = \mathbf{1}_{2P\times 1} \tag{20}$$

where

$$\mathbf{Z}(\hat{\omega}_0(i)) = [\mathbf{z}(\hat{\omega}_0(i)) \;\; \mathbf{z}^*(\hat{\omega}_0(i)) \;\; \cdots \;\; \mathbf{z}(P\hat{\omega}_0(i)) \;\; \mathbf{z}^*(P\hat{\omega}_0(i))] \tag{21}$$

with $\mathbf{z}(\omega) = [e^{-j\omega} \cdots e^{-j\omega(N-1)}]^T$, is a Vandermonde matrix containing $2P$ harmonically related complex sinusoids, which come in complex-conjugate pairs due to the involved signals being real, $\mathbf{1}_{2P\times 1}$ is a $2P \times 1$ vector of all

ones, $(\cdot)^*$ denotes complex conjugation, $(\cdot)^H$ denotes Hermitian transposition, and $\mathbf{R}(i)$ is the $N \times N$ sample autocorrelation matrix of the feedback-compensated signal, defined as

$$\mathbf{R}(i) = \frac{1}{M} \sum_{t=(i-1)M+1}^{iM} \mathbf{d}(t)\,\mathbf{d}^T(t) \tag{22}$$

The filter obtained from the linearly constrained quadratic optimization problem in (19) and (20) is optimal, signal-adaptive, and robust to the presence of multiple pitches [15].

Using the Lagrange multiplier method, the optimization problem in (19) and (20) can be shown to have a closed-form solution, which is given by

$$\hat{\mathbf{a}}(i) = \mathbf{R}^{-1}(i)\mathbf{Z}(\hat{\omega}_0(i))(\mathbf{Z}^H(\hat{\omega}_0(i))\mathbf{R}^{-1}(i)\mathbf{Z}(\hat{\omega}_0(i)))^{-1}\mathbf{1}_{2P\times 1} \tag{23}$$

This filter is signal-adaptive through $\mathbf{R}(i)$ and also depends on the fundamental frequency $\hat{\omega}_0(i)$ via $\mathbf{Z}(\hat{\omega}_0(i))$. A pitch estimate can now be obtained by filtering the signal using the optimal filters for various values of $\hat{\omega}_0(i)$ and then picking the one for which the output power is maximized, i.e.,

$$\begin{aligned}\hat{\omega}_0(i) &= \arg\max_{\hat{\omega}_0(i)} \frac{1}{M} \sum_{t=(i-1)M+1}^{iM} \left|\hat{\mathbf{a}}^T(i)\,\mathbf{d}(t)\right|^2 \\ &= \arg\max_{\hat{\omega}_0(i)} \mathbf{1}_{2P\times 1}^T(\mathbf{Z}^H(\hat{\omega}_0(i))\mathbf{R}^{-1}(i)\mathbf{Z}(\hat{\omega}_0(i)))^{-1}\mathbf{1}_{2P\times 1}\end{aligned} \tag{24}$$

This method has demonstrated to have a number of desirable features, namely excellent statistical performance and robustness against periodic interference [15]. Once $\hat{\omega}_0(i)$ is available, the complex amplitudes (i.e., the phases and the amplitudes) of the sinusoids can be estimated using a least squares approach, i.e.,

$$\hat{\boldsymbol{\alpha}}(i) = (\overline{\mathbf{Z}}^H(\hat{\omega}_0(i))\overline{\mathbf{Z}}(\hat{\omega}_0(i)))^{-1}\overline{\mathbf{Z}}^H(\hat{\omega}_0(i))\overline{\mathbf{d}}(i) \tag{25}$$

with

$$\hat{\boldsymbol{\alpha}}(i) = [\hat{\alpha}_1(i)e^{j\hat{\phi}_1(i)}\ \ \hat{\alpha}_1(i)e^{-j\hat{\phi}_1(i)}\ \ \ldots\ \hat{\alpha}_P(i)e^{j\hat{\phi}_P(i)}\ \ \hat{\alpha}_P(i)e^{-j\hat{\phi}_P(i)}]^T \tag{26}$$

$$\overline{\mathbf{Z}}(\hat{\omega}_0(i)) = [\overline{\mathbf{z}}(\hat{\omega}_0(i))\ \ \overline{\mathbf{z}}^*(\hat{\omega}_0(i))\ \ \ldots\ \ \overline{\mathbf{z}}(P\hat{\omega}_0(i))\ \ \overline{\mathbf{z}}^*(P\hat{\omega}_0(i))] \tag{27}$$

$$\overline{\mathbf{z}}(\omega) = [e^{-j\omega}\ \ \ldots\ \ e^{-j\omega(M-1)}]^T \tag{28}$$

$$\overline{\mathbf{d}}(i) = [d((i-1)M+1)\ \ \ldots\ \ d(iM)]^T \tag{29}$$

or, alternatively, using optimal filtering techniques (see, e.g., [18]). Finally, the number of harmonics P can be estimated by using a maximum a posteriori (MAP) criterion [15,21],

$$\hat{P} = \arg\min_{\hat{P}}(M\log\hat{\sigma}_{\hat{P}}^2 + \hat{P}\log M) \tag{30}$$

where the first term is a log-likelihood term which comprises a noise variance estimate $\hat{\sigma}_{\hat{P}}^2$ depending on the candidate model order $\hat{P}$, and the second term is the order-dependent penalty associated with the $\hat{P}$ amplitudes and phases.

4. Prediction error filter design

4.1. Including pitch and amplitude in PEF design

The harmonic sinusoidal near-end signal model in (15) cannot directly be inverted with the aim of designing a PEF. However, by exploiting the observation that a PZLP model can asymptotically (as the poles and zeros tend to lie on the unit circle) provide an exact representation of a sum of sinusoids, we can use the CPZLP model in (13) to construct an approximate PEF for the harmonic sinusoidal near-end signal model in (15). By enforcing harmonicity in the CPZLP model and inserting the estimated pitch $\hat{\omega}_0(i)$, the PEF $\hat{H}_{\text{Pitch}}^{-1}(q,i)$ related to the tonal near-end signal components can be written as a cascade of second-order sections,

$$\hat{H}_{\text{Pitch}}^{-1}(q,i) = \prod_{n=1}^{P} \frac{1-2\nu_n\cos n\hat{\omega}_0(i)q^{-1}+\nu_n^2q^{-2}}{1-2\rho_n\cos n\hat{\omega}_0(i)q^{-1}+\rho_n^2q^{-2}} \tag{31}$$

where, as before, the poles and zeros are on the same radial lines, with the zeros positioned between the poles and the unit circle, i.e., $0 \ll \rho_n < \nu_n \leq 1$. The prediction error using a harmonic sinusoidal model for the tonal near-end components cascaded with an LP model for the broadband near-end components can then be written as

$$\varepsilon(k,\hat{\boldsymbol{\theta}}(t)) = \hat{H}_{\text{LP}}^{-1}(q,i)\hat{H}_{\text{Pitch}}^{-1}(q,i)[y(k)-\hat{F}(q,t)u(k)] \tag{32}$$

with $\hat{\boldsymbol{\theta}}(t) = [\hat{\mathbf{f}}^T(t), \hat{\mathbf{c}}^T(i), \hat{\omega}_0(i)]^T$.

In [13], the application of the pitch PEF (31) in a PEM-based AFC algorithm was evaluated, using pitch estimation techniques based on optimal filtering and subspace methods. An improved AFC performance was found when computing the prediction error using the pitch PEF as in (32) as compared to using the CPZLP tonal near-end components model as in (14). For this evaluation, the PEF in (31) was designed with the zero radii fixed to $\nu_n = 1$ and the pole radii fixed to $\rho_n = 0.95$, and with a fixed order P. This results in a PEF that applies equal (infinity) suppression for the estimated tonal component frequencies by placing all the zeros on the unit circle. However, speech is a non-stationary signal with time-varying pitch, amplitude and number of harmonics. Therefore, in [14] the pitch PEF was further improved by including the estimated amplitudes $\hat{\alpha}_n(i)$ and the estimated order $P = \hat{P}$, which was again found to result in a better PEM-based AFC performance. The pitch PEF design using variable amplitudes and order will now be explained in more detail.

For an example speech frame, with a spectrum as shown in Fig. 3, the corresponding PEF response (displayed as a superposition of second-order section filter responses) is shown in Fig. 4 for the CPZLP PEF (13) and in Fig. 5 (top) for the pitch PEF (31). A first observation is that, in both cases, each of the PEF second-order sections behaves as a notch filter, with an infinite notch depth due to placing the PEF zeros on the unit circle. Moreover, a comparison of Figs. 4 (bottom) and 5 (top) shows that the PEF has a more dense notch structure in the low frequency region when harmonicity is assumed.

Since the PEF response should ideally correspond to the inverse near-end signal spectrum, it was conjectured in

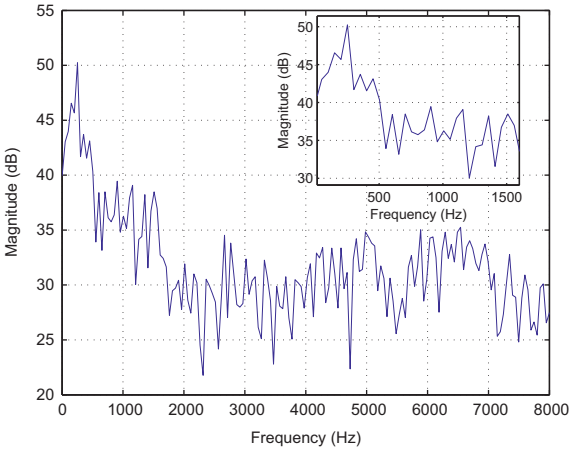

Fig. 3. Example speech spectrum used to estimate the PEF.

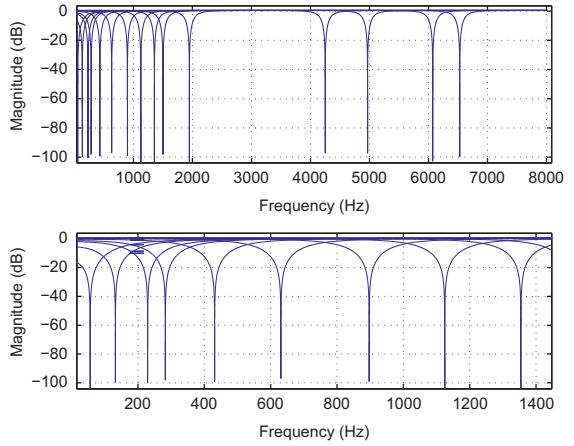

Fig. 4. CPZLP PEF designed for example speech frame: notch filters up to 8000 Hz (top) and notch filters up to 1400 Hz (bottom).

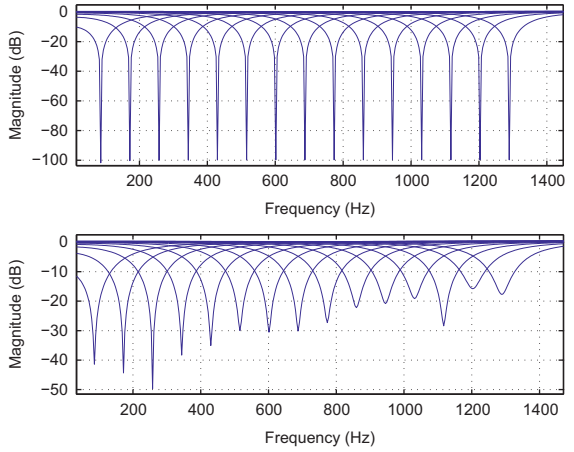

Fig. 5. Pitch PEF designed for example speech frame: without incorporating amplitude information (top) and with incorporating amplitude information (bottom).

[14] that the pitch PEF design can be improved by adjusting the notch depth to be the reciprocal of the estimated amplitudes. From the literature on parametric equalizer

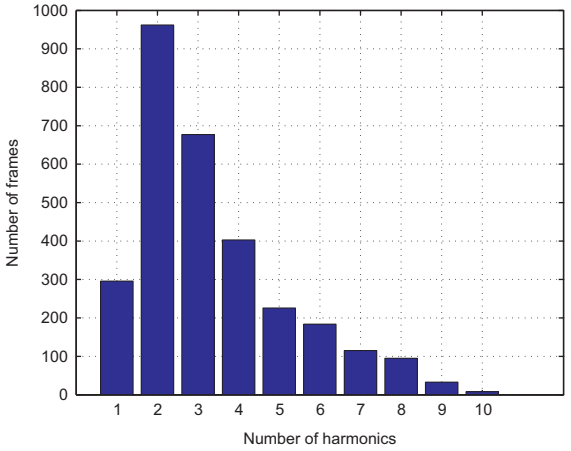

Fig. 6. Histogram of estimated number of harmonics $\hat{P}(i)$ in 3000 frames of example speech signal.

design [22] it is known that the notch depth is primarily determined by the zero radius. Consequently, a simple design rule for computing the zero radius given the (fixed) pole radius ρ_n and the estimated amplitude $\hat{\alpha}_n(i)$ was proposed in [14], i.e.,

$$\hat{\iota}_n(i) = \max\left(\rho_n, 1-\frac{1-\rho_n}{\hat{\alpha}_n(i)}\right). \tag{33}$$

When substituting the zero radii computed using (33) in the pitch PEF in (31), the resulting PEF response for the example speech frame defined earlier is shown in Fig. 5 (bottom). Here, the estimated amplitudes resulting from the optimal filtering approach in (25) have been used in the zero radius design rule (33). From a comparison of Fig. 5 (top) and (bottom), the improved PEF is clearly seen to provide a better approximation to the inverse of the example speech spectrum shown in Fig. 3.

Two further improvements to the pitch PEF design based on the harmonic sinusoidal near-end signal model can be made, by exploiting the observation that a speech frame can generally be classified as either voiced or unvoiced. A first improvement related to this observation was proposed in [14], and consists of replacing the fixed model order P by an estimated and frame-dependent model order $\hat{P}(i)$. A histogram of the estimated number of harmonics (resulting from the optimal filtering approach in (30)) for a speech signal consisting of 3000 speech frames (as used in the evaluation in Section 7) is shown in Fig. 6. This example indeed suggests that the harmonic sinusoidal near-end signal model order varies across different speech frames and that the fixed model order of $P=15$ used in previous work [13] is generally too high. A second improvement, which will be treated in more detail in Section 5, consists in incorporating a voiced–unvoiced detector in the PEM-based AFC algorithm, and pursuing a different PEF design strategy depending on the detection outcome.

4.2. APES-Based PEF design

While the pitch estimation in Section 3 was based on an optimal filtering approach, we did not actually use the

optimal filter $\hat{\mathbf{a}}(i)$ in (23) in the subsequent PEF design. Indeed, the pitch PEF in (31) only depends on the pitch estimate $\hat{\omega}_0(i)$, which is obtained by solving the optimization problem in (24), without explicitly computing $\hat{\mathbf{a}}(i)$. In this Section, we present an alternative PEF design procedure, which is based on the direct use of the optimal filter rather than plugging the pitch estimate into a second-order sections pole-zero model as in (31). However, even if the optimal filter proposed in (23) has desirable properties w.r.t. the estimation of the harmonic sinusoidal near-end signal model parameters, it has been demonstrated to suffer from a number of problems when applied as a signal extraction filter [18]. Therefore, we adopt a slightly modified approach to the optimal filter design problem, the result of which is referred to as an amplitude and phase estimation (APES) filter [18]. The APES filter inherits the desirable properties of the optimal filter in (23), being optimal, signal-adaptive, and specifically aimed at periodic signals, however, it has also demonstrated to have an excellent performance in extracting the desired signal while rejecting noise and periodic interferences. Consequently, a PEF based on the APES filter can be expected to be a valuable alternative to the pitch PEF in (31) for being used in PEM-based AFC, as will be experimentally validated in Section 7.

For a given pitch candidate $\hat{\omega}_0(i)$, the APES filter design can be formulated as [18]

$$\min_{\hat{\mathbf{a}}(i)} \quad \hat{\mathbf{a}}^T(i)\mathbf{Q}(i)\hat{\mathbf{a}}(i) \tag{34}$$

$$\text{s.t.} \quad \mathbf{Z}^H(\hat{\omega}_0(i))\hat{\mathbf{a}}(i) = \mathbf{1}_{2P\times 1} \tag{35}$$

where, compared to the optimal filter design in (20), the sample autocorrelation matrix $\mathbf{R}(i)$ has been replaced by a modified correlation matrix

$$\mathbf{Q}(i) = \mathbf{R}(i) - \mathbf{G}^H(i)\mathbf{W}^{-1}(i)\mathbf{G}(i) \tag{36}$$

that is formed by subtracting the contribution of the estimated harmonics from the sample autocorrelation matrix $\mathbf{R}(i)$, with $\mathbf{G}(i)$ and $\mathbf{W}(i)$ defined as

$$\mathbf{G}(i) = \frac{1}{M}\sum_{t=(i-1)M+1}^{iM} \mathbf{w}(t)\,\mathbf{d}^T(t) \tag{37}$$

where $\mathbf{w}(t) = [e^{j\omega_0 t}\ e^{-j\omega_0 t}\ \dots\ e^{jP\omega_0 t}\ e^{-jP\omega_0 t}]^T$ and

$$\mathbf{W}(i) = \frac{1}{M}\sum_{t=(i-1)M+1}^{iM} \mathbf{w}(t)\mathbf{w}^H(t) \tag{38}$$

Using the Lagrange multiplier method, the solution of (34) and (35) is found to be

$$\hat{\mathbf{a}}(i) = \mathbf{Q}^{-1}(i)\mathbf{Z}(\hat{\omega}_0(i))(\mathbf{Z}^H(\hat{\omega}_0(i))\mathbf{Q}^{-1}(i)\mathbf{Z}(\hat{\omega}_0(i)))^{-1}\mathbf{1}_{2P\times 1} \tag{39}$$

The difference between the optimal filter in (23) and the APES filter in (39) is in the definition of the sample autocorrelation matrix $\mathbf{R}(i)$ and the modified correlation matrix $\mathbf{Q}(i)$. Furthermore, the output of the APES filter in (39) is periodic, i.e., the output of the filter resembles a sum of harmonically related sinusoids. The pitch can then again be estimated by maximizing the filter output power, i.e.,

$$\hat{\omega}_0(i) = \arg\max_{\hat{\omega}_0(i)} \hat{\mathbf{a}}^T(i)\mathbf{R}(i)\hat{\mathbf{a}}(i) \tag{40}$$

$$\hat{\omega}_0(i) = \mathbf{1}_{2P\times 1}^T(\mathbf{Z}^H(\hat{\omega}_0(i))\mathbf{Q}^{-1}(i)\mathbf{Z}(\hat{\omega}_0(i)))^{-1}\mathbf{Z}^H(\hat{\omega}_0(i))\mathbf{Q}^{-1}(i)\mathbf{R}(i) \times\mathbf{Q}^{-1}(i)\mathbf{Z}(\hat{\omega}_0(i))(\mathbf{Z}^H(\hat{\omega}_0(i))\mathbf{Q}^{-1}(i)\mathbf{Z}(\hat{\omega}_0(i)))^{-1}\mathbf{1}_{2P\times 1} \tag{41}$$

A PEF designed using the APES filter can now be defined as

$$\hat{H}_{\text{APES}}^{-1}(q,i) = 1 - \hat{A}(q,i) \tag{42}$$

where

$$\hat{A}(q,i) = [1\ \ q^{-1}\ \ \dots\ \ q^{-N+1}]\hat{\mathbf{a}}(i) \tag{43}$$

Apart from using a modified correlation matrix in the pitch estimation, there is a second issue that discriminates the PEF design based on the APES filter from the PEF design based on the optimal filter. Due to the subtractive nature of the APES PEF in (42), the APES filter design does not only depend on the signal of which the pitch is to be estimated (i.e., the feedback-compensated signal $d(t)$), but also on the signal that is to be filtered using the APES PEF (i.e., the loudspeaker and microphone signals $u(t)$ and $y(t)$). The APES PEF design in the ith frame can thus be summarized as follows. In a first step, the pitch estimate $\hat{\omega}_0(i)$ is computed from the feedback-compensated $d(t)$ signal using (41). In a second step, two PEFs $H_{u,\text{APES}}^{-1}(q,i)$ and $H_{y,\text{APES}}^{-1}(q,i)$ are designed according to (42) and (43), one for filtering the loudspeaker signal $u(t)$ and one for filtering the microphone signal $y(t)$, by using the APES filters $\hat{\mathbf{a}}_u(i)$ and $\hat{\mathbf{a}}_y(i)$ computed in (39) with two different modified correlation matrices:

$$\mathbf{Q}_u(i) = \left(\frac{1}{M}\sum_{t=(i-1)M+1}^{iM} \mathbf{u}(t)\mathbf{u}^T(t)\right) - \mathbf{G}^H(i)\mathbf{W}^{-1}(i)\mathbf{G}(i) \tag{44}$$

$$\mathbf{Q}_y(i) = \left(\frac{1}{M}\sum_{t=(i-1)M+1}^{iM} \mathbf{y}(t)\mathbf{y}^T(t)\right) - \mathbf{G}^H(i)\mathbf{W}^{-1}(i)\mathbf{G}(i) \tag{45}$$

where

$$\mathbf{u}(t) = [u(t)\ \ \dots\ \ u(t-N+1)]^T \tag{46}$$

$$\mathbf{y}(t) = [y(t)\ \ \dots\ \ y(t-N+1)]^T. \tag{47}$$

The prediction error when using the APES PEFs cascaded with an LP PEF for modeling the broadband near-end components can then be written as

$$\varepsilon(k,\hat{\boldsymbol{\theta}}(t)) = \hat{H}_{\text{LP}}^{-1}(q,i)[\hat{H}_{y,\text{APES}}^{-1}(q,i)y(k) - \hat{F}(q,t)\hat{H}_{u,\text{APES}}^{-1}(q,i)u(k)] \tag{48}$$

with $\hat{\boldsymbol{\theta}}(t) = [\hat{\mathbf{f}}^T(t), \hat{\mathbf{c}}^T(i), \hat{\mathbf{a}}_u^T(i), \hat{\mathbf{a}}_y^T(i)]^T$.

5. Voiced–unvoiced detection

In Section 2.3, we have introduced a cascaded near-end signal model that allows to independently represent the tonal and broadband components in the near-end signal. The focus of Sections 3 and 4 has mainly been on the estimation of the tonal near-end components and the related PEF design. While the cascaded near-end signal model may be highly appropriate for representing voiced speech frames, presumably it does not offer an improved modeling capability in the case of unvoiced speech, which generally does not contain any tonal components. In this Section, we will illustrate that the impact of using either an

LP near-end signal model or a cascaded near-end signal model, is significant for voiced speech frames but negligible for unvoiced speech frames. Consequently, it makes sense to discrimimate between voiced and unvoiced speech frames in the near-end signal, and to choose the appropriate PEF structure accordingly. To this end, a non-intrusive voiced–unvoiced detection algorithm which operates on the feedback-compensated signal (and hence does not require the near-end signal to be available) will be proposed.

5.1. Spectral flatness of the prediction error

The purpose of the PEF in PEM-based AFC is to maximally whiten the near-end signal component in the loudspeaker and microphone signals. In an attempt to evaluate the whitening capability of the PEF outside of the PEM-based AFC framework, we will calculate the spectral flatness measure (SFM) [23] of the PEF output when using the near-end signal as the PEF input, i.e.,

$$\mathrm{SFM}(i)=\frac{\exp\left[\frac{1}{M}\sum_{k=0}^{M-1}\ln|E(e^{j2\pi k/M},i)|\right]}{\frac{1}{M}\sum_{k=0}^{M-1}|E(e^{j2\pi k/M},i)|} \tag{49}$$

Here, $E(e^{(j2\pi k/M)},i)$, $k=0,\ldots,M-1$ is the M-point Discrete Fourier Transform (DFT) of the ith frame $(t=(i-1)M+1,\ldots,iM)$ of the PEF output $e(t)$, which is defined and labeled as follows for the four PEFs under consideration:

$$e(t)=\hat{H}_{\mathrm{LP}}^{-1}(q,i)v(t) \quad \text{(LP)} \tag{50}$$

$$e(t)=\hat{H}_{\mathrm{LP}}^{-1}(q,i)\hat{H}_{\mathrm{CPZLP}}^{-1}(q,i)v(t) \quad \text{(CPZLP)} \tag{51}$$

$$e(t)=\hat{H}_{\mathrm{LP}}^{-1}(q,i)\hat{H}_{\mathrm{Pitch}}^{-1}(q,i)v(t) \quad \text{(optfilt)} \tag{52}$$

$$e(t)=\hat{H}_{\mathrm{LP}}^{-1}(q,i)\hat{H}_{\mathrm{APES}}^{-1}(q,i)v(t) \quad \text{(APES)} \tag{53}$$

The SFM is expressed on a dB-scale where 0 dB corresponds to a flat spectrum, and a more negative value indicates a more colored spectrum.

Fig. 7 shows the SFM for a selection of 30 voiced speech frames, and for the corresponding PEF output frames. A first observation is that the SFM for the voiced frames at the PEF input tends to have a highly negative value, which is due to the contribution of the tonal components in the near-end spectrum. Second, it can be seen that the SFM improvement (i.e., the reduction in the SFM absolute value) is consistently higher for the cascaded near-end signal models (CPZLP, optfilt, and APES) compared to the LP near-end signal model. On average, the APES PEF shows slightly better whitening properties than the other PEFs based on a cascaded near-end signal model.

On the other hand, the SFM plotted in Fig. 8 for a selection of 30 unvoiced speech frames, and for the corresponding PEF output frames, shows that the use of a cascaded near-end signal model in this case does not improve the PEF whitening capability, compared to the use of a single LP model. Moreover, it has been observed that applying a cascaded near-end signal model to an unvoiced speech frame is often undesirable, since the notch filter behavior of the CPZLP or pitch PEF may actually result in a PEF output that is spectrally less flat than the corresponding input, the effect of which will then be compensated by the LP model intended to represent the broadband near-end component.

5.2. ZCR- and energy-based voiced-unvoiced detection

The voiced and unvoiced frames of a speech signal can be detected using features such as the zero crossing rate (ZCR) and the energy in a given frame [24–28]. The ZCR indicates the number of times the amplitude of a speech

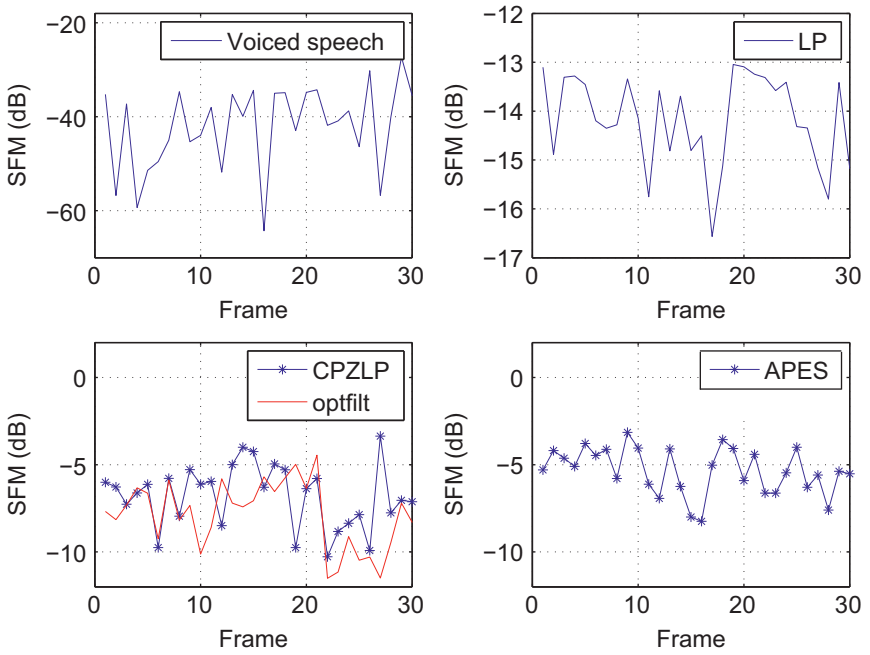

Fig. 7. SFM for selected voiced speech frames (top left) and for the corresponding PEF output frames.

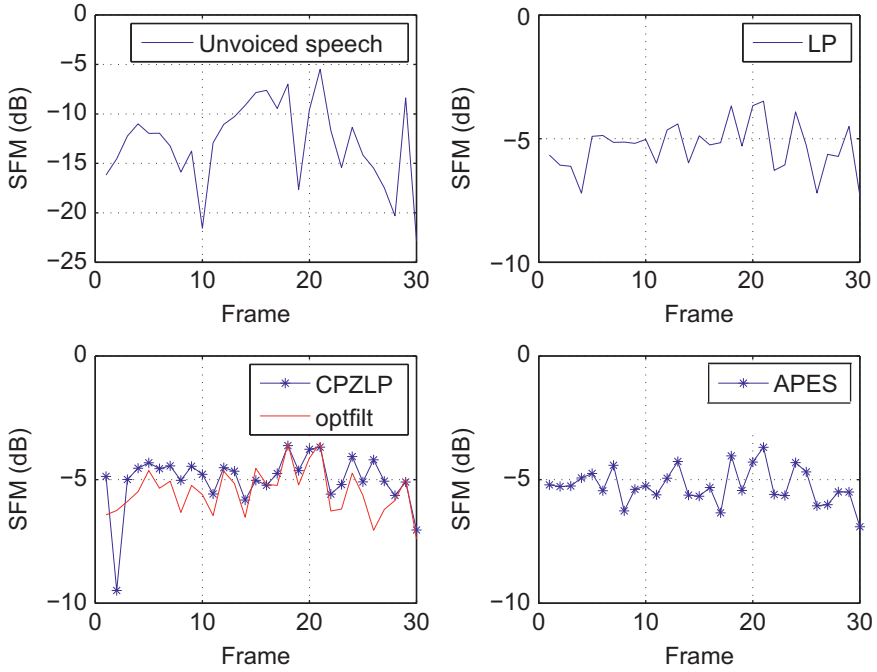

Fig. 8. SFM for selected unvoiced speech frames (top left) and for the corresponding PEF output frames.

signal for a given frame undergoes a sign change, i.e.,

$$\mathrm{ZCR}(i) = \sum_{t=(i-1)M+1}^{iM} \frac{1-\mathrm{sgn}[d(t)]\mathrm{sgn}[d(t+1)]}{2} \tag{54}$$

where sgn[·] is the *signum* function. The short-time energy of a speech frame is given by

$$\mathrm{Energy}(i) = \frac{1}{M} \sum_{t=(i-1)M+1}^{iM} d^2(t) \tag{55}$$

The ZCR is typically low for voiced speech and high for unvoiced speech. On the other hand, the short-time energy is high for voiced speech (due to its periodicity) and low for unvoiced speech.

The ZCR and Energy features in (54) and (55) allow for a non-intrusive voiced–unvoiced detection of the near-end signal $v(t)$ in an AFC context, since the availability of the near-end signal is not required. The motivation for using the feedback-compensated signal $d(t)$ in the feature calculation in (54) and (55) is that this signal closely approximates the near-end signal $v(t)$ as the PEM-based AFC feedback path estimate approaches the true acoustic feedback path.

Examples of the ZCR for selected voiced and unvoiced frames are shown in Fig. 9 (top), and confirm that the ZCR is indeed much lower for voiced speech compared to unvoiced speech. Fig. 9 (bottom) shows the time-domain waveform for a selected voiced and unvoiced frame, from which it can be deduced that the short-time energy of a voiced speech frame is indeed significantly higher than the energy in an unvoiced frame.

For the speech signal used in this simulation, 751 out of 3000 frames (≈25%) are classified as unvoiced. This implies that by using a simple LP model whenever an unvoiced frame has been detected, the overall computational complexity can be significantly reduced compared to our previous work in [13,14] where the voiced–unvoiced detection was not included.

6. Computational complexity

In this Section, we will briefly discuss the computational complexity related to the application of the different PEFs in a PEM-based AFC framework. Table 1 shows the order of the required number of multiplications in one signal frame, both for the PEF design and for one filtering operation with the resulting PEF. The first row in Table 1 corresponds to the complexity of the LP model that is used to represent the broadband near-end signal components. The following rows show the additional complexity when including different models for the tonal near-end signal components, cascaded with the LP model. Before interpreting the different complexity expressions in Table 1, the four parameters determining the PEF design will be discussed. The frame size M ideally corresponds to the average stationarity interval of speech, which is here chosen to be $M=320$ (corresponding to 20 ms at a 16 kHz sampling rate). From Fig. 6, the number of harmonics P included in the tonal near-end signal components model can be determined to usually lie in the range $P=1$–5. The optimal/APES filter length N is typically chosen as $N=M/4=80$ [15,18]. The number L of candidate pitch values $\hat{\omega}_0$ that are used in the optimal/APES filter design requires a trade-off between the PEF accuracy and the resulting computational complexity. We will typically use about 50–100 uniformly spaced frequencies in the frequency range 40–300 Hz. Roughly, the different parameters can be assumed to follow the ordering $P \ll N \approx L \ll M$.

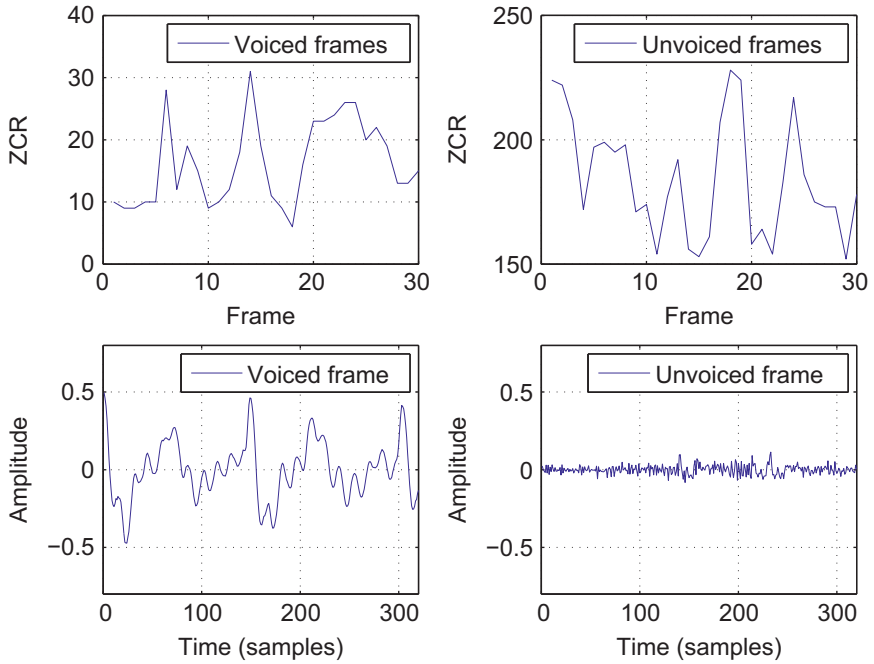

Fig. 9. ZCR for selected voiced and unvoiced speech frames (top) and time domain waveform of a typical voiced and unvoiced speech frame (bottom).

Table 1
Computational complexity of PEF design and filtering: order of number of multiplications per frame.

PEF	PEF design	Filtering
LP	$O(MP)+O(P^2)$	$2MP$
CPZLP	$+O(\overline{\kappa}\overline{\beta}MP)$	$+4MP$
optfilt	$+O(N(M+LN\log N))+O(PM\log M)+O(LP^2)$	$+4MP$
APES	$+O(LM(N+\log M))+O(LN^2(\log N+log\ P))$ $+O(LPN\log N)+O(LP^2)$	$+MN$

Due to this ordering, the first $O(\cdot)$ term in each of the rows in Table 1 dominates the subsequent $O(\cdot)$ terms, hence we can restrict ourselves to the first term in each row for comparing the complexity of the different models. A first observation, which confirms the conclusions in our previous work [9,13,14], is that the PEF design related to the tonal near-end signal components requires more computations than the LP PEF design. On the other hand, as illustrated in [9,13,14] as well as in Section 7, this increased complexity often does pay off in terms of the AFC performance improvement. However, if we include the non-intrusive voiced–unvoiced detection algorithm proposed in Section 5, we can reduce the additional complexity related to the CPZLP/optfilt/APES PEF design with approximately 25%, which was found to be the relative amount of unvoiced frames in a typical speech signal. A second observation, which results from substituting the above values for L,M,N,P in the dominant $O(\cdot)$ term for the different models in Table 1 (and using $\overline{\kappa}\overline{\beta}=O(10^3)$ for the CPZLP model [17]), is that the complexity of the PEF design for the CPZLP, optfilt, and APES models is of the same order.

7. Experimental results

In this Section, simulation results are presented in which the impact of the different PEF designs in a PEM-based AFC algorithm using a cascaded near-end signal model is evaluated for a hearing aid scenario.

7.1. Experimental set-up

The near-end noise model order is fixed to $n_C=30$. All near-end signal models are estimated using 50% overlapping signal frames of length $M=320$ samples. The optimal/APES filter length is set to $N=M/4=80$, and the related pitch estimation is performed on a uniform $L=66$-point frequency grid ranging from 40 to 300 Hz. The PEM-based estimation of the acoustic feedback path, which corresponds to a measured hearing aid feedback path, is based on the normalized least mean squares (NLMS) algorithm with an adaptive filter length equal to the acoustic feedback path length, i.e., $n_F=200$. The near-end signal is a 30 s male speech signal sampled at $f_s=16$ kHz. The forward path gain $K(t)$ is set to a value of 3 dB below the maximum stable gain (MSG) without feedback cancellation.

7.2. Performance measures

To assess the AFC performance, the following measures are used. The achievable amplification before instability

occurs is measured by the MSG, which is defined as

$$\mathrm{MSG}(t) = -20\log_{10}[\max_{\omega\in\mathcal{P}}|J(\omega,t)[F(\omega,t)-\hat{F}(\omega,t)]|] \tag{56}$$

where $J(q,t)=G(q,t)/K(t)$ denotes the forward path transfer function without the amplification gain $K(t)$, and $\mathcal{P}$ denotes the set of frequencies at with the feedback signal $x(t)$ is in phase with the near-end signal $v(t)$. The misadjustment (MA) between the estimated feedback path impulse response $\hat{\mathbf{f}}(t)$ and the true feedback path impulse response $\mathbf{f}$ represents the accuracy of the feedback path estimation and is defined as

$$\mathrm{MA}_F(t) = 20\log_{10}\frac{\|\hat{\mathbf{f}}(t)-\mathbf{f}\|_2}{\|\mathbf{f}\|_2} \tag{57}$$

7.3. Simulation results

The first simulation illustrates the performance of a state-of-the-art PEM-based AFC algorithm that uses a single LP near-end signal model (AFC-LP) [8]. The instantaneous value of the MSG(t) and the corresponding misadjustment is shown in Fig. 10. The MSG(t) curves have been smoothed with a one-pole low-pass filter to improve the clarity of the figures. The instantaneous value of the forward path gain $20\log_{10}K(t)$ and the MSG without acoustic feedback cancellation (MSG $F(q)$) are also shown. The performance of the AFC-LP algorithm is clearly affected by the lack of a proper model for the tonal near-end signal components, since only 751 out of 3000 speech frames are classified as unvoiced. Even though the feedback path estimate converges to a misadjustment of around −9 dB, the estimation appears to be not sufficiently accurate at one or more of the closed-loop system's critical frequencies and hence the hearing aid is bound to operate very close to instability, as can be seen from the MSG curve.

The second simulation is performed with a cascaded near-end signal model in which the PEF design related to the tonal components is based on a second-order sections pole-zero model. Two such PEF designs are compared, the first one based on the non-harmonic CPZLP model in (13) in which the frequencies are estimated using the CPZLP method [17] (AFC-CPZLP), and the second one based on the harmonic CPZLP model in (31) with the pitch estimated using the optimal filtering approach (AFC-optfilt). We evaluate the AFC-CPZLP and AFC-optfilt performance for two different orders $P=5$ and $P=10$, as well as with and without the use of the non-intrusive voiced–unvoiced detection algorithm, see Figs. 11 and 12. In general the

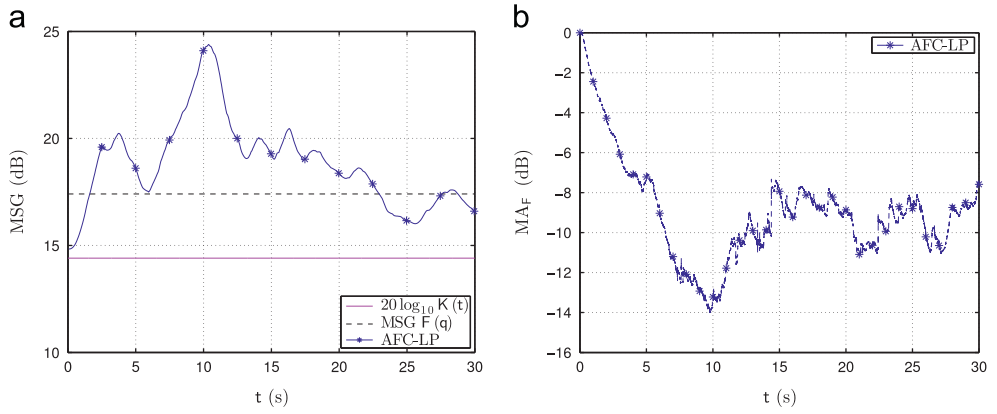

Fig. 10. Performance of state-of-the-art AFC-LP algorithm: (a) instantaneous MSG vs. time and (b) feedback path misadjustment vs. time.

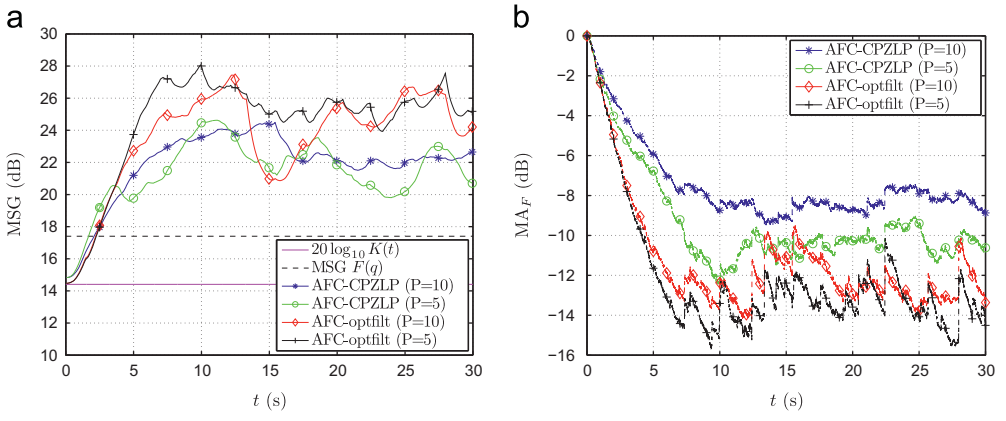

Fig. 11. Performance of AFC-CPZLP and AFC-optfilt algorithms for different orders P, *without* voiced–unvoiced detection: (a) instantaneous MSG vs. time and (b) feedback path misadjustment vs. time.

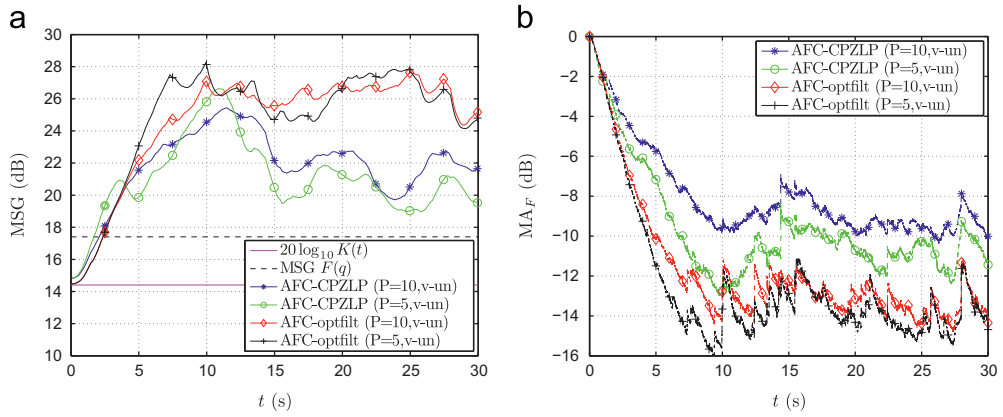

Fig. 12. Performance of AFC-CPZLP and AFC-optfilt algorithms for different orders P, *with* voiced–unvoiced detection: (a) instantaneous MSG vs. time and (b) feedback path misadjustment vs. time.

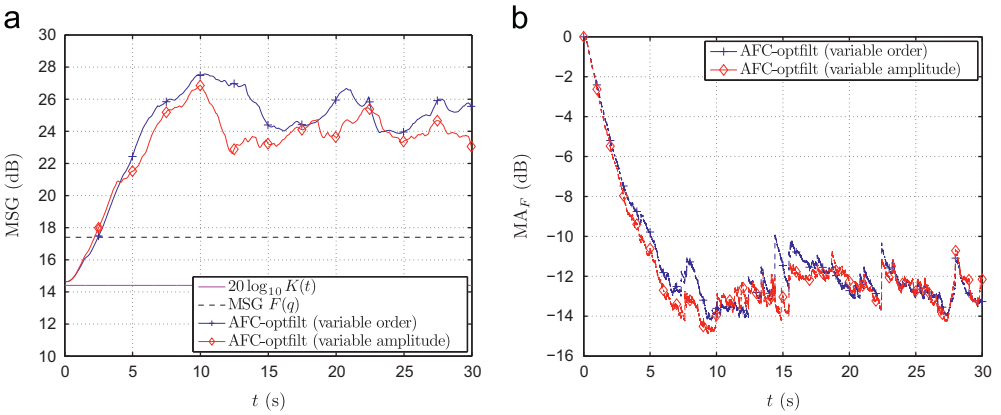

Fig. 13. Performance of AFC-optfilt algorithm with variable order/amplitude: (a) instantaneous MSG vs. time and (b) feedback path misadjustment vs. time.

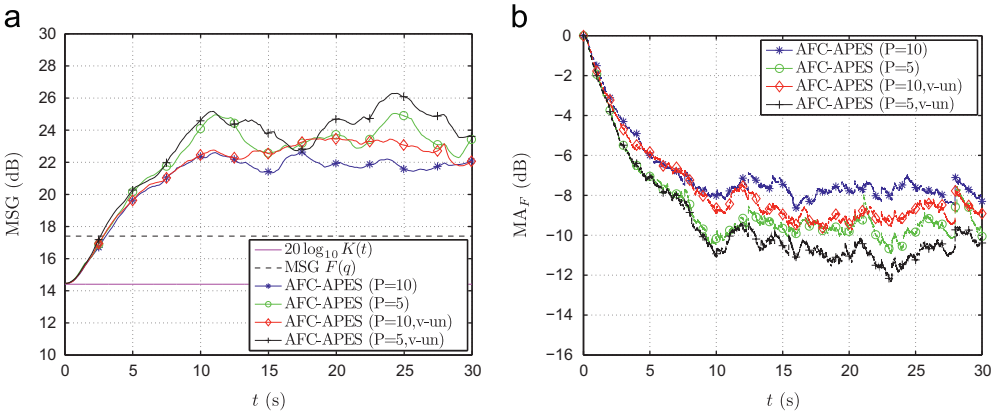

Fig. 14. Performance of AFC-APES algorithm for different orders P, with and without voiced–unvoiced detection: (a) instantaneous MSG vs. time and (b) feedback path misadjustment vs. time.

MSG is higher for AFC-optfilt compared to AFC-CPZLP and the corresponding misadjustment is lower for AFC-optfilt. For both AFC-optfilt and AFC-CPZLP a lower order seems to perform better both in terms of MSG and misadjustment. AFC-optfilt with $P=10$ performs poorly around $t=15$ s but this problem is eliminated when the voiced–unvoiced detector is used. Overall, the use of the voiced–unvoiced detector improves the AFC-optfilt MSG performance with up to 2–4 dB. For the AFC-CPZLP algorithm, the use of a voiced–unvoiced detector does not appear to be beneficial, and in fact results in an MSG decrease up to 1–2 dB. Finally, it is also important to note that the AFC-optfilt

algorithm provides a much higher and more sustainable MSG increase compared to the state-of-the-art AFC-LP algorithm.

In the third simulation, a variable amplitude and order, estimated using the optimal filtering approach, are incorporated in the PEF using the design rule in (33). Fig. 13 shows the results when using a variable order with fixed amplitude, and using a variable amplitude with fixed order. The performance when both variable order and variable amplitude are used is not shown since the performance is similar to the case when only a variable amplitude is used (as some of the estimated amplitudes are close to zero in this case). From comparing Fig. 13 with Figs. 11 and 12, it can be concluded that the use of a variable amplitude/order slightly improves the AFC-optfilt performance, however, the use of a voiced–unvoiced detector (with a fixed amplitude/order) improves the performance even more. This is an important result, since the use of a voiced–unvoiced detector is computationally much more interesting than the use of a variable amplitude/order (see Section 6).

Finally, in the fourth simulation, the APES filter is used to represent the tonal near-end components in the cascaded near-end signal model (AFC-APES) for different orders P, with and without a voiced–unvoiced detector, see Fig. 14. The best performance is achieved with $P=5$ and using a voiced–unvoiced detector, which is also the computationally most interesting configuration out of the four AFC-APES configurations evaluated here. Even though the AFC-APES performance is slightly worse compared to the AFC-optfilt performance shown in Fig. 12, the APES PEF may often be preferable from an implementation point of view due to its FIR structure. In particular, one weak point of the AFC-CPZLP and AFC-optfilt algorithms is that a lack of numerical precision in the pole-zero placement may have a large impact on the resulting PEF response.

8. Conclusion

In this paper we have investigated the impact of the PEF design in PEM-based AFC, and proposed two novel components in the PEF design that are beneficial in terms of AFC performance, computational complexity, and ease of implementation. From an experimental analysis of the state-of-the-art PEM-based AFC algorithms, we concluded that a single near-end signal model based on LP fails to provide sufficient decorrelation for voiced speech frames, whereas in the case of a cascaded near-end signal model based on an LP and CPZLP model, the accuracy of the tonal near-end signal components model seems to play a crucial role since these components have a more significant share in the AFC correlation problem. This has motivated us to look for improvements in the modeling and estimation of the tonal near-end signal components and in the related PEF design.

To this end, we have adopted the use of a harmonic sinusoidal near-end signal model, and derived two approaches for designing an appropriate PEF. The first approach is based on a pole-zero second-order sections PEF structure, where information such as pitch, amplitude and the number of harmonics is included. The second approach is based on APES filters which are specifically aimed at periodic signals and are optimal and signal-adaptive given the observed signals. Furthermore, a non-intrusive voiced–unvoiced detection algorithm is included in the PEF design to switch between a single near-end signal model and a cascaded near-end signal model.

Experimental results show that the PEM-based AFC performance can indeed be improved significantly with an accurate modeling of the near-end signal. A cascaded near-end signal model outperforms a single near-end signal model with up to 5–7 dB in MSG and up to 5–8 dB in filter misadjustment. The use of a voiced–unvoiced detector has shown to be the most important step in arriving at a more accurate near-end signal model for AFC, and moreover it reduces the overall computational complexity with about 25%.

Acknowledgments

This research work was carried out at the ESAT Laboratory of KU Leuven, in the frame of the EST-SIGNAL Marie-Curie Fellowship program under contract no. MEST-CT-2005-021175, the KU Leuven Research Council CoE EF/05/006 “Optimization in Engineering” (OPTEC) and PFV/10/002 (OPTEC), the Concerted Research Action GOA-MaNet, the Belgian Programme on Interuniversity Attraction Poles initiated by the Belgian Federal Science Policy Office IUAP P6/04 “Dynamical systems, control and optimization” (DYSCO) 2007–2011, the IWT Project “Signal processing and automatic fitting for next generation cochlear implants”, and the Research Project FWO nr. G.0600.08 “Signal processing and network design for wireless acoustic sensor networks”, and was supported by a Postdoctoral Fellowship of the Research Foundation Flanders (FWO–Vlaanderen, T. van Waterschoot). The scientific responsibility is assumed by its authors.

References

[1] H. Dillon, Hearing Aids, Thieme, Stuttgart, 2001.
[2] M. Kahrs, K. Brandenburg, Applications of Digital Signal Processing to Audio and Acoustics, Kluwer, Dordrecht, 1998.
[3] J.M. Kates, Digital Hearing Aids, Plural Publishing, San Diego, 2008.
[4] T. van Waterschoot, M. Moonen, Fifty years of acoustic feedback control: state of the art and future challenges, Proceedings of the IEEE 99 (2) (2011) 288–327.
[5] C. Boukis, D.P. Mandic, A.G. Constantinides, Towards bias minimization in acoustic feedback cancellation systems, Journal of the Acoustical Society of America 121 (3) (2007) 1529–1537.
[6] D.J. Freed, Adaptive feedback cancellation in hearing aids with clipping in the feedback path, Journal of the Acoustical Society of America 123 (3) (2008) 1618–1626.
[7] M. Guo, S.H. Jensen, J. Jensen, Novel acoustic feedback cancellation approaches in hearing aid applications using probe noise and probe noise enhancement, IEEE Transactions on Audio, Speech and Language Processing 20 (9) (2012) 2549–2563.
[8] A. Spriet, I. Proudler, M. Moonen, J. Wouters, Adaptive feedback cancellation in hearing aids with linear prediction of the desired signal, IEEE Transactions on Signal Processing 53 (10) (2005) 3749–3763.
[9] T. van Waterschoot, M. Moonen, Adaptive feedback cancellation for audio applications, Signal Processing 89 (11) (2009) 2185–2201.
[10] J. Hellgren, U. Forssell, Bias of feedback cancellation algorithms in hearing aids based on direct closed loop identification, IEEE Transactions on Speech and Audio Processing 9 (7) (2001) 906–913.
[11] L. Ljung, System Identification: Theory for the User, Prentice-Hall, Englewood Cliffs, New Jersey, 1987.

[12] A. Spriet, G. Rombouts, M. Moonen, J. Wouters, Adaptive feedback cancellation in hearing aids, Journal of The Franklin Institute 343 (2006) 545–573.
[13] K. Ngo, T. van Waterschoot, M.G. Christensen, M. Moonen, S.H. Jensen, J. Wouters, Adaptive feedback cancellation in hearing aids using a sinusoidal near-end signal model, in: Proceedings of the 2010 IEEE International Conference on Acoustics, Speech, Signal Process. (ICASSP '10), Dallas, TX, USA, 2010, pp. 181–184.
[14] K. Ngo, T. van Waterschoot, M.G. Christensen, M. Moonen, S.H. Jensen, J. Wouters, Prediction-error-method-based adaptive feedback cancellation in hearing aids using pitch estimation, in: Proceedings of the 18th European Signal Processing Conference (EUSIPCO '10), Aalborg, Denmark, 2010, pp. 40–44.
[15] M.G. Christensen, A. Jakobsson, Multi-Pitch Estimation, Morgan & Claypool Publishers, 2009.
[16] M.G. Christensen, J.H. Jensen, A. Jakobsson, S.H. Jensen, On optimal filter designs for fundamental frequency estimation, IEEE Signal Processing Letters 15 (2008) 745–748.
[17] T. van Waterschoot, M. Moonen, Constrained pole-zero linear prediction: an efficient and near-optimal method for multi-tone frequency estimation, in: Proceedings of 16th European Signal Processing Conference (EUSIPCO '08), Lausanne, Switzerland, 2008.
[18] M.G. Christensen, A. Jakobsson, Optimal filter designs for separating and enhancing periodic signals, IEEE Transactions on Signal Processing 58 (12) (2010) 5969–5983.
[19] G. Rombouts, T. van Waterschoot, K. Struyve, M. Moonen, Acoustic feedback suppression for long acoustic paths using a nonstationary source model, IEEE Transactions on Signal Processing 54 (9) (2006) 3426–3434.
[20] Y.T. Chan, J.M.M. Lavoie, J.B. Plant, A parameter estimation approach to estimation of frequencies of sinusoids, IEEE Transactions on Acoustics, Speech, and Signal Processing ASSP-29 (2) (1981) 214–219.
[21] P.M. Djuric, Asymptotic map criteria for model selection, IEEE Transactions on Signal Processing 46 (7) (1998) 2726–2735.
[22] T. van Waterschoot, M. Moonen, A pole-zero placement technique for designing second-order IIR parametric equalizer filters, IEEE Transactions on Audio, Speech, and Language Processing 15 (8) (2007) 2561–2565.
[23] J.D. Markel, A.H. Gray Jr., Linear Prediction of Speech, Springer-Verlag, New York, 1976.
[24] R.W. Schafer, L.R. Rabiner, Digital representations of speech signals, Proceedings of the IEEE 63 (4) (1975) 662–677.
[25] B. Atal, L. Rabiner, A pattern recognition approach to voiced-unvoiced-silence classification with applications to speech recognition, Transactions on Acoustics, Speech, and Signal Processing ASSP-24 (3) (1976) 201–212.
[26] L. Rabiner, M. Sambur, Voiced-unvoiced-silence detection using the Itakura LPC distance measure, in: Proceedings of the 1977 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '77), Hartford, CT, USA, 1977, pp. 323–326.
[27] P.C. Loizou, Speech Enhancement: Theory and Practice, CRC Press, Boca Raton, 2007.
[28] R.G. Bachu, S. Kopparthi, B. Adapa, B.D. Barkana, Voiced/unvoiced decision for speech signals based on zero-crossing rate and energy, in: K. Elleithy (Ed.), Advanced Techniques in Computing Sciences and Software Engineering, 2010, pp. 279–282.