Signal Processing 89 (2009) 710–723

Contents lists available at ScienceDirect

Signal Processing

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

An improved partial Haar dual adaptive filter for rapid identification of a sparse echo channel

P. Kechichian [*], B. Champagne

Department of Electrical and Computer Engineering, McGill University, 3480 University Street, Montréal (Quebec), Canada H3A 2A7

ARTICLE INFO

Article history:
Received 7 February 2008
Received in revised form
13 October 2008
Accepted 17 October 2008
Available online 7 November 2008

Keywords:
Line echo cancellation
Sparse system identification
Adaptive filters
Haar transform
Dezert–Smarandache theory
Fuzzy inference

ABSTRACT

Recently, a coupled echo canceller was proposed that uses two short adaptive filters for sparse echo cancellation. The first filter operates in the partial Haar domain and is used to locate the channel's dispersive region; the second filter is then centered around this location to cancel the echo in the time domain. In this paper, we propose feasible solutions to improve the performance of this *partial Haar dual adaptive filter* (PHDAF) in practical applications. These include: (1) alleviating the dependence of the PHDAFs performance on the echo-path impulse response's bulk delay; (2) improving the tracking performance of the PHDAF in response to abrupt changes in the echo path; and (3) extending the original PHDAF structure to support the cancellation of multiple echoes. The proposed algorithmic solutions exploit the Haar transform's polyphase representation and make use of a novel peak tendency estimator (PTE) based on Dezert–Smarandache theory (DSmT). The improved PHDAF is evaluated in terms of its mean-square error (MSE) curves and its mean time to properly locate a dispersive region for different SNRs. Results show that enhanced performance can be obtained using the proposed solutions at a minimal increase in computational cost.

1. Introduction

Line, or network echo is commonplace in today's expanding communications infrastructure. Unlike other types of echo (e.g. acoustic), line echo is sparse: the echo-path impulse response consists of an initial zero or *bulk delay* region, corresponding to the signal round-trip, followed by a non-zero or *dispersive region*, corresponding to the echo arrival. Line echo is usually caused by an impedance mismatch, as occurring in the hybrid circuits used for $\frac{4}{2}$-wire conversion [1]. In voice communications, when the bulk delay between callers exceeds 25 ms or so, the reflected signal is perceived as a distinct echo that can severely impede a conversation. The coding and signal processing functions of digital technologies may introduce delays in excess of 100 ms; while for long distance calls routed via satellites, the propagation delay may reach several 100 ms [2]. Recent advancements, such as Voice over Internet Protocol (VoIP) telephony and xDSL technologies for broadband data transmission, highlight the need to develop better echo cancellers for sparse line echo.

In one of the earliest works on sparse echo cancellation [3], the input and desired signals are bandpass-filtered, decimated, and used by a short adaptive filter to estimate the bulk delay. A second short filter operating at the original sampling rate is centered around the dispersive region to cancel the echo. This way, only two short adaptive filters are required (compared to one long filter), thus reducing overall system complexity. However, the use of bandpass filtering in this approach has important drawbacks. Firstly, in speech applications, it can remove

[*] Corresponding author. Present address: Philips Research, High Tech Campus 36, 5656 AE Eindhoven, The Netherlands.
E-mail addresses: patrick.kechichian@philips.com (P. Kechichian), benoit.champagne@mcgill.ca (B. Champagne).

0165-1684/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2008.10.033

important frequency components in the signals, preventing proper convergence of the subsampled adaptive filter. Secondly, bandpass filtering in effect smears the peak of the unknown echo-path impulse response, making it more difficult to correctly locate the dispersive region.

Recent literature is rich in adaptive filtering algorithms that exploit the sparse characteristics of line echo [4]. Most of these algorithms are based on finding ways to determine which filter coefficients are associated with the echo, and then adapting only these coefficients. An adaptive multiple echo (ME) canceller is proposed in [5], which uses a full-length primary adaptive filter in parallel with a group of short secondary adaptive filters. In [6], a two-stage adaptation process is proposed in which the first stage estimates the bulk delay while the second stage adapts filter coefficients using a constrained tap-selection approach [7]. A well-known class of sparse echo cancellers are based on the proportionate normalized least mean squares (PNLMS) algorithm [8] and its variants [9–11]. These algorithms allocate individual step-size gains in proportion to the magnitude of each filter coefficient.

Other recent attempts at improving sparse echo cancellation rely on applying orthogonal wavelet transforms to the input data. It is shown in [12] that the number of adaptive coefficients needed to cancel the echo can be reduced significantly by applying a Haar transform. In [13], the authors propose using a subset of Haar wavelets to detect the significant channel coefficients. By exploiting the hierarchical structure of the dyadic wavelet expansion, these significant coefficients are used to activate wavelets in the remaining Haar subsets that share the same non-zero time-support. The Haar transform is particularly attractive for sparse echo identification: it facilitates the location of the dispersive region through proper selection of its scale/translation parameters, and it is easily amenable to a digital implementation.

Recently, Bershad and Bist [14] have proposed a solution to the sparse echo cancellation problem that combines favourable attributes of [3,13] in a coupled configuration consisting of two short adaptive filters. In this approach, referred to here as the partial Haar dual adaptive filter (PHDAF), the first filter operates on a subset of input Haar coefficients, and is used by a peak delay estimator to locate the echo-path's dispersive region. The second filter is centered around this location to cancel the echo in the time domain. In cases where the bulk delay is large, the PHDAF provides a significant reduction in computational and memory requirements. In addition, by reducing the number of filter taps to an amount necessary to model the dispersive region, the convergence speed of the echo canceller is increased. An improved theoretical model of the LMS algorithm in [14] for low rank systems was recently developed in [15] to better predict the behaviour of first and second moment statistics of the partial Haar adaptive filter. In [16], a partial block wavelet transform is proposed to increase efficiency and improve peak detection of the system in [14]. It is shown that the block transform reduces to using Daubechies' biorthogonal 2.2 spline wavelet, which is claimed to have better properties for estimating a peak's location.

In this paper we identify, and propose feasible solutions to three inherent limitations of the PHDAF for sparse echo cancellation in [14]: (1) dependence of the PHDAFs performance on the echo-path impulse response's bulk delay; (2) degraded tracking performance of the PHDAF in response to abrupt changes in the echo-path impulse response; and (3) limitation of the original PHDAF to a single dispersive region. The proposed algorithmic solutions exploit the polyphase representation of the Haar transform and make use of a novel peak tendency estimator (PTE) based on Dezert–Smarandache theory (DSmT) and fuzzy inference [17,18]. The improved PHDAF is evaluated in terms of its mean-square error (MSE) curves as well as its mean time to properly locate a dispersive region under different SNRs. In experiments using normalized least mean squares (NLMS) for the filter coefficient adaptation, the proposed amendments to the original PHDAF are shown to yield significant performance gains at a minimal increase in computational cost.

This paper is organized as follows: The structure and main equations of the PHDAF in [14] are reviewed in Section 2, along with a discussion of its main limitations. The solutions that we propose to overcome the latter are developed in Section 3. A series of supporting computer experiments is presented in Section 4. Finally, Section 5 concludes the work.

2. Background and problem formulation

2.1. The partial Haar dual adaptive filter

2.1.1. Partial Haar transform

Let $N = 2^J$, where J is a positive integer. The N-dimensional discrete-time Haar wavelet transform can be represented by an $N \times N$ orthogonal matrix $\mathbf{H}$ with entries $h_m(n)$, where the row index m identifies a basis vector and the column index n represents discrete-time [19]. The elements of the first row are equal to $h_0(n) = 1/\sqrt{N}$, while the remaining rows are obtained by scaling and shifting a discrete-time wavelet filter $\psi(n)$ defined as

$$\psi(n) = \begin{cases} +1, & 0 \leqslant n < N/2, \\ -1, & N/2 \leqslant n < N, \\ 0 & \text{otherwise}. \end{cases} \tag{1}$$

Specifically, let $m = 2^j + k$, where $j \in \{0, \ldots, J-1\}$ is the scale index and $k \in \{0, \ldots, 2^j - 1\}$ is the translation index. We have

$$h_m(n) = \alpha_j \psi(2^j n - kN), \tag{2}$$

where the normalization factor $\alpha_j = \sqrt{2^j/N}$. The number of rows with scale index j is equal to 2^j and the corresponding basis vectors have non-overlapping time-support of length $N/2^j$. A *partial* Haar wavelet transform consists of using only a subset of size $q = 2^j$ of Haar basis vectors corresponding to scale index j, to transform a given input data vector. The corresponding transform matrix, denoted $\mathbf{H}_q$, is thus a $q \times N$ submatrix of $\mathbf{H}$, obtained by extracting its rows $h_m(n)$ for $m = q + k$ and $k = 0, \ldots, q-1$. As j increases, the time resolution of the transform improves at the cost of increasing the dimension of the output vector, which translates into a longer Haar-domain adaptive filter in the PHDAF (see below). Unlike the complete Haar transform matrix, the partial Haar matrix $\mathbf{H}_q$ $(q<N)$ is only row-wise orthonormal, i.e. $\mathbf{H}_q\mathbf{H}_q^{\mathrm{T}} = \mathbf{I}_q \neq \mathbf{H}_q^{\mathrm{T}}\mathbf{H}_q$.

2.1.2. PHDAF structure and operation

The structure of the PHDAF proposed in [14] is shown in Fig. 1, where discrete-time signals $u(n)$ and $d(n)$, respectively, denote the line input and the echo-contaminated reference used to drive the adaptive process.

The upper branch, consists of a partial Haar transform matrix $\mathbf{H}_q$ of size $q \times N$, followed by a length-q $(\leqslant N)$ partial Haar adaptive filter. The value of N is set to match the maximum length of the unknown echo-path impulse response. The selection of $q = 2^j$ involves a trade-off between the desired level of time resolution and complexity reduction. At time n, a new sample $u(n)$ is shifted into the input data vector $\mathbf{u}(n) = [u(n), u(n-1), \ldots, u(N-1)]^{\mathrm{T}}$ of length N and a new transformed input vector of length q is calculated, i.e.

$$\mathbf{z}(n) = \mathbf{H}_q\mathbf{u}(n). \tag{3}$$

Vector $\mathbf{z}(n)$ is used as input into a so-called partial Haar adaptive filter with coefficient vector at time n given by $\mathbf{v}(n)$. In this work, the partial Haar adaptive filter $\mathbf{v}(n)$ is updated by means of the NLMS algorithm, i.e.

$$\mathbf{v}(n+1) = \mathbf{v}(n) + \mu\|\mathbf{z}(n)\|^{-2}e_H(n)\mathbf{z}(n), \tag{4}$$

$$e_H(n) = d(n) - \mathbf{v}(n)^{\mathrm{T}}\mathbf{z}(n), \tag{5}$$

where μ denotes the step size and $e_H(n)$ is the error signal. Under appropriate conditions, $\mathbf{v}(n)$ converges to a scaled-down version of the true echo-path impulse response. The peak delay estimator in Fig. 1 tracks the location of the dispersive region by locating the peak coefficient magnitude of $\mathbf{v}(n)$.

In the lower branch of the PHDAF, the estimated peak location is used to offset a short time-domain adaptive filter $\mathbf{w}(n)$ of length L so as to center it around the dispersive region. The value of L is set to match the longest expected dispersive region in the echo path. Once the peak location of $\mathbf{v}(n)$ is properly estimated, the short time-domain adaptive filter can converge to the dispersive region and thus cancel the echo component of the reference signal $d(n)$. In this work, the NLMS algorithm is also used to update $\mathbf{w}(n)$. In practice, the residual signal after echo removal, i.e. $e(n) = d(n) - \mathbf{w}(n)^{\mathrm{T}}\mathbf{u}(n)$, is used for data transmission over the network.

2.1.3. Properties of the PHDAF

To simplify the discussion, let us assume that

$$d(n) = \mathbf{w}_0^{\mathrm{T}}\mathbf{u}(n) + v(n), \tag{6}$$

where $\mathbf{w}_0$ represents the true echo-path impulse response and $v(n)$ is an additive measurement noise. Also assume that the sequences $u(n)$ and $v(n)$ are stationary white, uncorrelated, with variances σ_u^2 and σ_v^2, respectively. Then, the optimum Wiener solution for the partial Haar filter can be obtained as

$$\mathbf{v}_0 = \mathbf{H}_q\mathbf{w}_0, \tag{7}$$

which is the partial Haar transform (i.e. a scaled-down version) of the true time-domain echo-path impulse response $\mathbf{w}_0$. Thus, when $\mathbf{w}_0$ features a single, localized dispersive region and the scale index $j = \log_2 q$ of the partial Haar transform provides adequate temporal resolution, the location of the dispersive region can be recovered from the peak coefficient in $\mathbf{v}_0$. It is also worth mentioning that $\mathbf{w}_0$ cannot be recovered from $\mathbf{v}_0$ unless it lies in the row-space of $\mathbf{H}_q$. Under proper operating conditions, the partial Haar adaptive filter will converge to the Wiener solution in (7). The steady-state MSE then takes the form $J_\infty = J_{\min} + \mathscr{M}$ where $\mathscr{M}$ is the misadjustment noise power of the adaptation algorithm and $J_{\min}$ is the minimum MSE corresponding to the Wiener solution

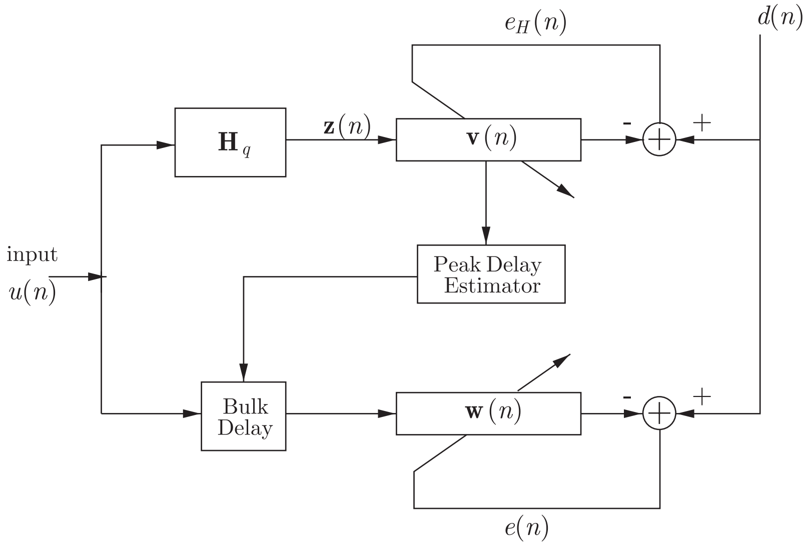

Fig. 1. PHDAF structure.

(7). The latter is given by

$$J_{\min} = \sigma_v^2 + \sigma_u^2(\|\mathbf{w}_0\|_2^2 - \|\mathbf{v}_0\|_2^2). \quad (8)$$

When $q = N$ (complete Haar transform), we have $\|\mathbf{w}_0\|_2 = \|\mathbf{v}_0\|_2$ and the second term on the right-hand side of (8) is equal to zero. When $q<N$, this term is positive, reflecting the rank-deficiency of the transform matrix $\mathbf{H}_q$. This increase in minimum MSE is acceptable if it does not impact the ability of the partial Haar adaptive filter to locate a dispersive region.

These considerations motivate the use of the converged vector $\mathbf{v}(n)$ as input to the peak delay estimator for the purpose of locating the dispersive region. It is shown in [14] that the greater the steady-state peak magnitude of $\mathbf{v}(n)$, the faster the short time-domain filter can be centered, and therefore, the overall PHDAF can adapt more quickly.

2.2. Limitations of the PHDAF

The simulation results in [14] show that the PHDAF may provide a drastic increase of the echo canceller's convergence speed in general, while keeping computational complexity low. However, our own experiments show that in certain cases, the PHDAF can require an extremely large number of input samples to converge, underlining inherent shortcomings of the original system. Below, we take a closer look at these drawbacks.

2.2.1. Sensitivity to bulk delay

The lack of shift-invariance of the Haar wavelet transform can greatly affect the time required by the peak delay estimator to properly locate a dispersive region. A wavelet transform whose basis vectors have a time-support of length $M = N/q$ is periodically shift-invariant with a period equal to M. As a result, depending on the bulk delay of the true echo-path impulse response, there exist M *distinct* partial Haar transformed Wiener solutions. Furthermore, the respective peak magnitudes of each of these solutions can differ, and consequently, the amount of time it takes the peak delay estimator to correctly locate a dispersive region.

To illustrate this point, Fig. 2 shows the transformed impulse responses of differently shifted versions of a common echo-path impulse response. In this example, the true echo-path vector of length $N = 1024$ is set to $\mathbf{w}_0 = [m_1(n-600-l)]^{\mathrm{T}}$, where $m_1(n)$ is one of the ITU-T G.168 hybrid impulse responses in [20] and shift index $l \in \{1, 5, 6, 7\}$ (see Section 4 for additional details). For each value of l, a transformed impulse response vector of size $q = 128$ is computed by applying the partial Haar matrix $\mathbf{H}_q$ as in (7). One can clearly observe that the second case ($l = 5$) displays a diminished peak.

Due to measurement noise and the rank-deficiency effect of the partial Haar transform on the minimum MSE in (8), the peak delay estimator takes a longer time to locate such small peaks. To support this claim, consider the family of experimental cumulative distribution function (CDF) curves in Fig. 3 corresponding to the time[1] to correctly locate the dispersive region for these four different bulk delays. It is apparent from this figure that it takes the peak delay estimator a larger number of input samples to correctly locate a smaller peak. The curve corresponding to a peak magnitude of 0.0166 shows that the peak delay estimator can barely locate the peak after 10^4 input samples.

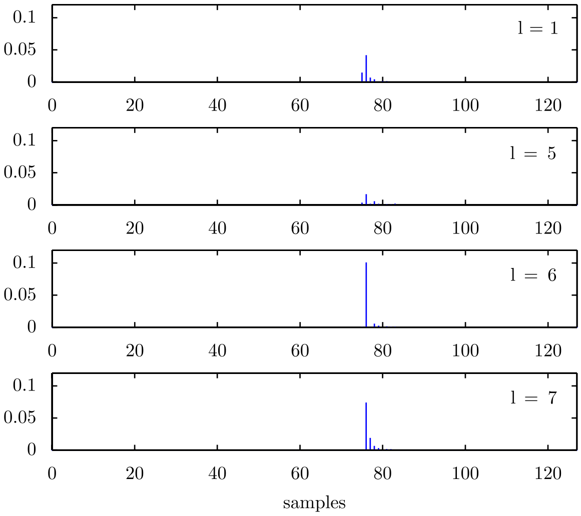

Fig. 2. Set of transformed impulse responses corresponding to $m_1(n-600-l)$ for shift index $l \in \{1, 5, 6, 7\}$.

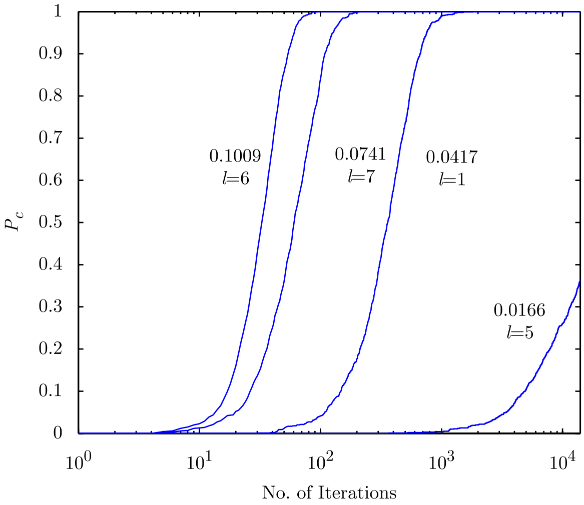

Fig. 3. Probability (P_c) of correctly estimating peak delays within a given number of time iterations (each curve is labelled by its peak magnitude in the Haar domain and the corresponding value of l).

[1] The peak delay estimator looks for the filter coefficient with the maximum instantaneous absolute value, and returns its location $\Delta(n)$. If $|\Delta(n) - \Delta_0| < \eta$ for n_c *consecutive* input samples starting with $n = n_d$, where Δ_0 is the true peak delay, then n_d is set as the time to correctly estimate the peak location. Integer η is chosen to avoid small estimate jitters around Δ_0. As pointed out by Ribas et al. [16], a similar solution to the jitter problem was used in [14]. In Fig. 3, we set $\eta = 10$, and $n_c = 950$.

2.2.2. Degraded tracking performance

In [14], the merits of using the coupled PHDAF configuration are demonstrated by calculating the echo canceller's mean time to correctly locate a dispersive region. It is shown on average that this time does not exceed more than 200 samples at 8 kHz when employing a partial Haar NLMS algorithm. These results, however, only represent the case when the partial Haar adaptive filter has been initialized to zero, i.e. $\mathbf{v}(0) = \mathbf{0}$, and the echo path is stationary. The case where the echo-path impulse response changes during the course of the echo canceller's steady-state operation is not considered. Our experiments show that the performance of the PHDAF can degrade significantly in the presence of sudden changes in the true echo path. Indeed, unlike the case where $\mathbf{v}(0) = \mathbf{0}$, and a peak only has to compete with neighbouring coefficient noise, after a sudden change in bulk delay, the new peak also has to compete with the residual of an older peak. In the case of the LMS adaptive algorithm for example, this older peak will undergo a slow exponential decay which may prevent the detection of the new peak for a large number of time iterations.

2.2.3. ME cancellation requirements

The peak delay estimator analyzed in [14] assumes the presence of a single dispersive region. This greatly simplifies the estimation process because the adaptive filter coefficient associated with a global maximum directly indicates the location of this region. In a multiple peak delay estimation scenario, the task does not only involve finding local maxima from the partial Haar adaptive filter, but also requires the classification of these maxima as peaks. A major difficulty arises from the fact that knowledge about the number of dispersive regions in a channel may not be available *a priori*. Clearly, this number must remain small if any advantages are to be gained from the use of a tailored ME canceller structure.

Another difficulty relates to the shift-variant property of the Haar wavelets. It was seen that, depending on the bulk delay, the partial Haar Wiener solutions can display peaks of varying magnitude, and this in turn determines how fast a dispersive region is located. The same is true for MEs, except now their corresponding peaks with largest magnitude need not correspond to the same shift within a period of M. Finally, an algorithm that cancels MEs will require a dynamic mechanism to allocate resources (e.g. short time-domain filters) to dispersive regions that have been correctly located.

3. Improving the coupled echo canceller

3.1. Redundant partial Haar transform

From the above discussion, it appears that some of the difficulties related to the rank-deficiency of the partial Haar transform, especially the sensitivity of the peak magnitude in the Wiener solution $\mathbf{v}_0$ (7) to the bulk delay in the true echo-path impulse response $\mathbf{w}_0$, could be overcome by driving the partial Haar adaptive filter towards a properly shifted version of $\mathbf{w}_0$. In effect, this can be achieved by modifying the upper Haar-domain adaptive filter in Fig. 1 in two different ways (without directly affecting the operation of the lower time-domain adaptive filter):

(1) Replacing the reference signal $d(n)$ in (5) by $d(n+l)$, where l is the desired integer shift.
(2) Replacing the transformed input vector $\mathbf{z}(n)$ by $\mathbf{z}(n-l)$ in (4)–(5).

Under the modelling assumptions presented at the beginning of Section 2.1.3, both approaches lead to the same *modified* optimum Wiener solution, i.e.

$$\mathbf{v}_0 = \mathbf{H}_q \mathbf{S}_l \mathbf{w}_0, \tag{9}$$

where $\mathbf{S}_l$ is an $N \times N$ shift matrix with (i,j)th entry equal to 1 for $j = i + l$ (i.e. lth diagonal) and 0 otherwise.

While both approaches have the equivalent effect of decreasing the effective bulk delay of $\mathbf{w}_0$ by l samples, in this work, we favour the use of the second approach. Indeed, as explained below, all $M = N/q$ shifted input vectors $\mathbf{z}(n-l)$, where $l \in \{0, 1, \ldots, M-1\}$, can be obtained naturally as the polyphase components of a so-called redundant partial Haar transform (RPHT), which trades arithmetic operations for memory. In addition, shifting $d(n)$ cannot help when dealing with MEs since a ME path channel can be written as a sum of single-echo-path channels, each with its own bulk delay, and whose contribution to $d(n)$ is unknown *a priori*.

A consequence of the wavelet transform's lack of shift-invariance is the requirement of calculating the partial Haar transform of the length-N input data block $\mathbf{u}(n)$ in (3) for every new input sample. The computational load for this operation is on the order of N arithmetic operations per iteration,[2] which is expensive. The length-N vector of RPHT coefficients at time n, denoted $\mathbf{s}(n)$, is defined by

$$\mathbf{s}(n) = [s(n), s(n-1), \ldots, s(n-N+1)]^{\mathrm{T}}, \tag{10}$$

where

$$s(n) = \mathbf{h}^{\mathrm{T}} \mathbf{u}'(n), \tag{11}$$

$\mathbf{h} = \alpha_j [\psi(0), \psi(q), \ldots \psi(q(M-1))]^{\mathrm{T}}$ is a vector of length M containing the non-zero portion of any row of $\mathbf{H}_q$, and $\mathbf{u}'(n) = [u(n), u(n-1), \ldots, u(n-M+1)]^{\mathrm{T}}$. Because this RPHT has a simple translational relationship with the input, it only requires the calculation of a single partial Haar coefficient $s(n)$ every iteration with a small cost on the order of M arithmetic operations.

To extract the q coefficients corresponding to the standard partial Haar transform $\mathbf{z}(n) = \mathbf{H}_q \mathbf{u}(n)$ from $\mathbf{s}(n)$, we simply observe that

$$\mathbf{z}(n) = [s(n), s(n-M), \ldots, s(n-(q-1)M)]^{\mathrm{T}}. \tag{12}$$

In effect, vector $\mathbf{z}(n)$ corresponds to the first (i.e. $l = 0$) polyphase component of the complete vector $\mathbf{s}(n)$, while the remaining polyphase components represent the

[2] That is, $N - q$ additions and q multiplications if the normalization factors α_j in (2) are included.

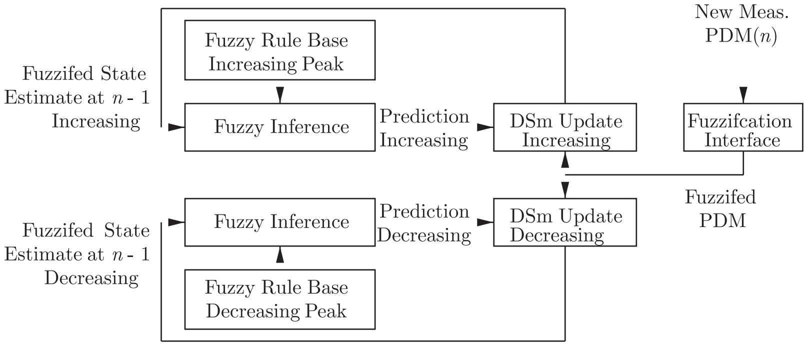

Fig. 4. Block diagram: peak tendency estimator.

transformed input vectors time delayed by $l = 1, \ldots, M-1$ sample(s), i.e. $\mathbf{z}(n-1)$ down to $\mathbf{z}(n-M+1)$.

Unlike the recently proposed partial block wavelet transform-based algorithm in [16] which calculates a transformed input coefficient every M samples, the proposed approach here can produce a new transformed input sample $s(n)$ every iteration. This makes it possible to drive the partial Haar adaptive filter with a specific polyphase component of $\mathbf{s}(n)$, which in turn results in its convergence to one of M different steady-state solutions (see Fig. 2, for example), each with its unique peak magnitude.

In what follows, the term *context* will be used to denote the polyphase component of $\mathbf{s}(n)$, corresponding to shift index $l \in \{0, \ldots, M-1\}$, that is being used as input to the partial Haar adaptive filter. Storing all N samples of $\mathbf{s}(n)$ makes M contexts and their resulting steady-state solutions readily available. At any given time, the partial Haar adaptive filter can be driven to *one*[3] of these M solutions, increasing the system's flexibility by making it possible to avoid suboptimal cases (such as illustrated by the right-most curve in Fig. 3).

3.2. *Peak tendency estimation*

Without further *a priori* knowledge (e.g. a pre-determined threshold), the magnitude of a detected peak *alone* is not a sufficient indicator of how well the bulk delay estimation process is progressing for a given context. A method is needed that includes a relative measure of the peak *quality* with regard to the detection process, in addition to a means for tracking this measure's behaviour over time. To categorize the performance of a context, i.e. if it is suitable or not for quick and correct peak delay estimation, a PTE is therefore presented below. This estimator is based on the work of [21], where an original approach for target behaviour tendency estimation (i.e. receding or approaching) is developed using the DSmT of plausible and paradoxical reasoning [18].

To mitigate the local effects of coefficient noise on the peak location of the partial Haar adaptive filter $\mathbf{v}(n)$, a peak discernibility measure (PDM) is proposed here which is calculated as follows: partition vector $\mathbf{v}(n)$ into three contiguous groups of similar size and find the maximum peak magnitude for each of the three groups. Let $c_{\min}(n)$ and $c_{\max}(n)$, respectively, denote the minimum and maximum among the three maxima so obtained. The PDM at time n is defined as[4]

$$\mathrm{PDM}(n) = 1 - c_{\min}(n)/c_{\max}(n). \tag{13}$$

When $c_{\max}(n)$ is large compared to $c_{\min}(n)$, then the PDM approaches 1 (a large localized peak). However, if the two are comparable (which is usually the case for coefficient noise), then the PDM approaches 0. In Appendix A, we show that the PDM (13) is indeed related to the probability of correctly locating the peak of $\mathbf{v}(n)$, and thus provides an indication of the quality of the peak detection process.

The proposed PTE takes PDM(n) as input at time n and uses a *combination* of DSmT and fuzzy inference similar to [21]. Unlike traditional logic, fuzzy inferences can be made even when rules are only partially satisfied. This is in stark contrast to crisp logic, where a rule is satisfied only if a premise of interest matches an implication's antecedent exactly. This is a powerful extension for dealing with fuzzy sets as linguistic descriptors that are difficult to clearly define, and possess a degree of uncertainty (interested readers may consult [17,22–25] for a good introduction to fuzzy systems). This inference mechanism makes fuzzy systems well-suited for complex or nonlinear control applications, such as step-size control in adaptive filtering algorithms [26–28]. However, unlike such approaches, the method described here is solely used to track and categorize a peak's discernibility as increasing or decreasing (where this information is used in the algorithm described in Section 3.3).

The adopted PTE system is represented in block diagram form in Fig. 4. It consists of two state prediction filters running in parallel on the sequence of fuzzified inputs PDM(n), with each filter using a different model of possible PDM behaviour. The top model corresponds to a PDM that increases over time and eventually results in the

[3] An obvious, but much more expensive solution is to use M partial Haar adaptive filters in parallel, each operating with a different context and choosing the peak delay to correspond to the maximum peak among the global maxima for each filter.

[4] Set PDM(n) = 0, if $c_{\max}(n) = 0$.

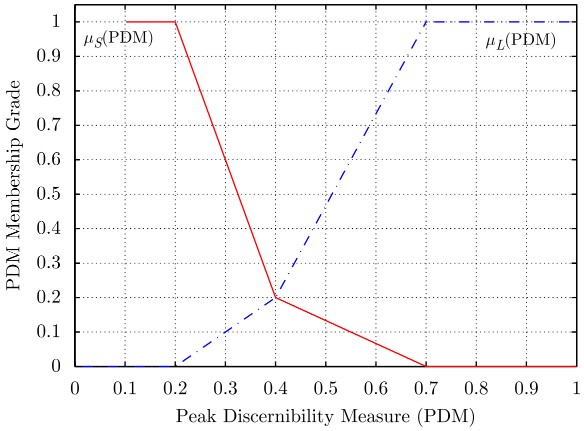

Fig. 5. Fuzzy interface including small and large membership functions.

proper detection of a dispersive region. The bottom model corresponds to a decreasing PDM which can be used to detect coefficient noise. The basic components of the PTE system are briefly described below (see [18,21,29] for additional details on DSmT and its applications to behaviour tendency estimation).

Two fuzzy sets, labelled S and L and defined in terms of the membership functions shown in Fig. 5, are used to characterize a small and large discernibility, respectively. At time n, a new (row) vector of fuzzy PDM values is obtained as $\boldsymbol{v}_n = [v_n(S), v_n(S\cup L), v_n(L), v_n(S\cap L)]$ where

$$v_n(L) = \mu_L(\mathrm{PDM}(n)), \tag{14}$$

$$v_n(S) = \mu_S(\mathrm{PDM}(n)), \tag{15}$$

$$v_n(S\cup L) = 1 - v_n(S) - v_n(L) \tag{16}$$

and $v_n(S\cap L) = 0$. The terms $v_n(\cdot)$ denote the basic belief masses (bbm) used within the DSmT framework.

The peak tendency at time n is internally characterized by a pair of fuzzy PDM state vectors corresponding to the two models, i.e. $\boldsymbol{\mu}_n^{\mathrm{inc}}$ and $\boldsymbol{\mu}_n^{\mathrm{dec}}$, where ($\mathscr{M} = \mathrm{inc}$ or dec)

$$\boldsymbol{\mu}_n^{\mathscr{M}} = [\mu_n^{\mathscr{M}}(S), \mu_n^{\mathscr{M}}(S\cup L), \mu_n^{\mathscr{M}}(L), \mu_n^{\mathscr{M}}(S\cap L)], \tag{17}$$

with $\mu_n^{\mathscr{M}}(\cdot)$ representing the corresponding bbm values. The state vector of each model is updated recursively in two steps: (1) the previous state vector $\boldsymbol{\mu}_{n-1}^{\mathscr{M}}$ is processed by a fuzzy inference engine to obtain a prediction of the current state vector, denoted $\boldsymbol{\mu}_{n|n-1}^{\mathscr{M}}$; (2) the predicted state vector $\boldsymbol{\mu}_{n|n-1}^{\mathscr{M}}$ is merged with the current observation vector $\boldsymbol{v}_n$ via the DSmT rule of combination to obtain the current state vector $\boldsymbol{\mu}_n^{\mathscr{M}}$.

Table 1
Fuzzy rule bases.

Rule no.	Increasing PDM
1	If PDM$(n-1) = S$ then PDM$(n) = S$
2	If PDM$(n-1) = S$ then PDM$(n) = L$
3	If PDM$(n-1) = L$ then PDM$(n) = L$
	Decreasing PDM
1	If PDM$(n-1) = L$ then PDM$(n) = L$
2	If PDM$(n-1) = L$ then PDM$(n) = S$
3	If PDM$(n-1) = S$ then PDM$(n) = S$

Table 2
Fuzzy graphs corresponding to two models of peak discernibility.

$n-1 \to n$	S	$S\cup L$	L	$S\cap L$
(a) Increasing discernibility, $\mathbf{G}^{\mathrm{inc}}$				
S	1	0	1	0
$S\cup L$	0	0	0	0
L	0.2	0	1	0
$S\cap L$	0	0	0	0
(b) Decreasing discernibility, $\mathbf{G}^{\mathrm{dec}}$				
S	1	0	0.2	0
$S\cup L$	0	0	0	0
L	1	0	1	0
$S\cap L$	0	0	0	0

The respective rule-bases for each model in Table 1 consist of if–then statements with antecedents A_i and consequents C_i (e.g. for rule 1 of the increasing PDM model, $A_1 = C_1 = S$). These rules are used together with the membership functions in Fig. 5 to construct a 4×4 fuzzy graph matrix $\mathbf{G}^{\mathscr{M}}$ for each of the two models, as given in Table 2, which quantifies the degree of association between two fuzzy sets for a specific model governed by a corresponding rule-base. The graph entries for a transition $A \to C$ are computed with the maximum value of the fuzzy intersection function between A and all A_j such that $C_j = C$ [22]:

$$g_{A\to C} = \max_{j,x\in[0,1]} \min(\mu_A(x), \mu_{A_j}(x)), \tag{18}$$

where A and C take values in the DSmT set of composite propositions $\mathscr{D} = \{S, L, S\cap L, S\cup L\}$. As long as the membership functions μ_A and μ_{A_j} overlap, the degree of association is non-zero. The predicted state vector for each model is obtained as

$$\boldsymbol{\mu}_{n|n-1}^{\mathscr{M}} = \boldsymbol{\mu}_{n-1}^{\mathscr{M}} \circ \mathbf{G}^{\mathscr{M}}, \tag{19}$$

where $\circ$ denotes the matrix form of Zadeh's max–min rule of composition [17]. Each vector $\boldsymbol{\mu}_{n|n-1}^{\mathscr{M}}$ is then normalized so that its entries sum up to 1.

At time n, the new input vector $\boldsymbol{v}_n$ is separately combined with each of the predicted state vectors $\boldsymbol{\mu}_{n|n-1}^{\mathscr{M}}$ using the DSmT rule of combination

$$\mu_n^{\mathscr{M}}(C) = \sum_{\substack{A,B\in\mathscr{D},\\ A\cap B = C}} \mu_{n|n-1}^{\mathscr{M}}(A) v_n^{\mathscr{M}}(B), \quad C \in \mathscr{D} \tag{20}$$

to obtain the updated fuzzy state vectors $\boldsymbol{\mu}_n^{\mathscr{M}}$.

Finally, each model's corresponding set of updated bbms ($\boldsymbol{\mu}_n^{\mathscr{M}}$) are converted into the Pignistic probabilities [18]

$$P^{\mathscr{M}}\{S\} = \mu_n^{\mathscr{M}}(S) + 0.5\mu_n^{\mathscr{M}}(S\cup L) + 0.5\mu_n^{\mathscr{M}}(S\cap L), \tag{21}$$

$$P^{\mathscr{M}}\{L\} = \mu_n^{\mathscr{M}}(L) + 0.5\mu_n^{\mathscr{M}}(S\cup L) + 0.5\mu_n^{\mathscr{M}}(S\cap L). \tag{22}$$

The correct behavioural model corresponds to the model with the smallest Pignistic entropy, as given by

$$H^{\mathscr{M}}_{\text{pig}} = -\sum_{A\in\{S,L\}} P^{\mathscr{M}}\{A\}\ln(P^{\mathscr{M}}\{A\}). \tag{23}$$

3.3. Suboptimal-context escape (SCE)

The SCE approach that we propose here monitors the correct peak model from the PTE and decides if the current context should be changed (i.e. if a different polyphase component of the input RPHT should be used) in the event that a peak's discernibility is poor. Deciding when to make a context switch requires some care. While the partial Haar adaptive filter needs time to reach steady state, it is usually the case that the time required to correctly locate a peak is far less than the convergence time in a proper operating context. It is thus inefficient to wait until the Haar-domain adaptive filter has converged to decide whether to remain in the current operating context or to attempt a new context.

The proposed solution consists of a schedule of $M = N/q$ non-decreasing trial periods $\boldsymbol{\tau} = \{\tau_1, \ldots, \tau_M\}$ ($\tau_1 \leqslant \tau_2 \leqslant \cdots \leqslant \tau_M$) to sequentially test each of the M possible contexts. The selection of a proper schedule is based on prior knowledge about the amount of time required by the peak delay estimator to correctly locate a peak in different contexts. The first trial period τ_1 should be related to the number of samples required to properly detect the global peak in an optimal context. The final trial period τ_M reflects the amount of time willing to be spent operating in a worst-case context. The remaining trial periods can be uniformly spaced between these two limits. Intuitively, a schedule of non-decreasing trial periods ensures that the likelihood of selecting (yet-unattempted) contexts that can successfully lead to a correct estimate of the peak delay increases. At the same time, beginning the schedule with shorter waiting times escapes any suboptimal contexts earlier in the peak delay estimation process. Further analytical motivations in terms of the mean time to correctly locate the bulk delay are provided in Appendix B.

In our implementation of this scheme, two counters C_{inc} and C_{dec} are used to maintain the number of samples that the peak tendency has either been increasing or decreasing during the current trial period τ_k. If $C_{\text{dec}} > \tau_k$ and jitter occurs in the peak delay estimate, then the filter and counters are reset to zero and a new context is attempted for the next trial period τ_{k+1}. If instead $C_{\text{inc}} > \tau_k$, i.e. peak delay is correctly estimated, then the counters are reset and the trial period is set back to its initial value τ_1. If all contexts have been attempted unsuccessfully, the process is repeated beginning with τ_2.

3.4. Improved tracking (IT)

Abrupt changes in an echo-path impulse response can be associated with a change in the bulk delay, or a phase roll, whereby the impulse response coefficients change signs. The tracking of a dispersive region after an abrupt change in the echo path can be seen as a competition between filter coefficient magnitudes. Indeed, when the partial Haar adaptive filter is initialized to zero, a peak's magnitude only has to compete with the low-magnitude coefficient noise of the surrounding taps, which makes its detection easier. However, in the event of an abrupt change in the echo-path impulse response (as can occur on IP networks for example), a new peak might have to compete with the decreasing magnitude of an old peak. This can cause the peak delay estimator to take much longer to find the *new* peak.

The proposed IT approach takes advantage of the following observation: After an abrupt change in the echo-path impulse response occurs, one finds that although the location of the new peak as well as its steady-state magnitude is unknown, the new steady-state magnitude of an *old* peak is approximately zero. Thus, if a significant decrease in the current peak's magnitude is detected (which usually signals a change in the echo-path impulse response), the entire partial Haar filter is reset to $\mathbf{v}(n) = \mathbf{0}$. This is a feasible solution because this filter is not directly being used to cancel echo. This allows the new peak to solely compete with the low-magnitude coefficient noise of its neighbouring taps instead of the decreasing magnitude of the previous peak. As a result, the performance gains obtained in [14] for the stationary case can be extended to cases where abrupt changes in the echo-path impulse response occur.

A reset is only deemed necessary when a decrease in magnitude is detected for a peak whose tendency has been in an *increasing* state during more than T_{inc} (not necessarily consecutive) iterations. To prevent the effects of false resets, the location of the peak *before* resetting the partial Haar adaptive filter, is stored and fed to the bulk delay unit. Only once the new peak tendency has been categorized as increasing for a total of T_{inc} samples, is the location of the new peak used to center the short time-domain filter. This way, if a reset is in fact necessary, the bulk delay unit will offset the short time-domain filter to the *new* peak location once it has been in an increasing state for a total of T_{inc} samples. If, however, a reset is unnecessary, then the old peak location is still used to center the short time-domain filter and the partial Haar adaptive filter will readapt to the old solution without affecting the time-domain adaptation. To avoid resetting the partial Haar adaptive filter too often, no new reset operation is attempted for at least T_{RS} iterations following a previous reset. In practice, T_{RS} and T_{inc} can be selected as $T_{\text{RS}} \approx 0.25T_{\text{inc}} \approx \tau_1$. The above IT approach can be easily integrated with the SCE approach in Section 3.3.

3.5. Multiple echo (ME) cancellation

The solution proposed here for ME cancellation with the PHDAF is based on a coordinator-multi-agent architecture similar to [30]. The following assumptions are made:

- The length of the dispersive regions is bounded by a known integer L in the time domain.

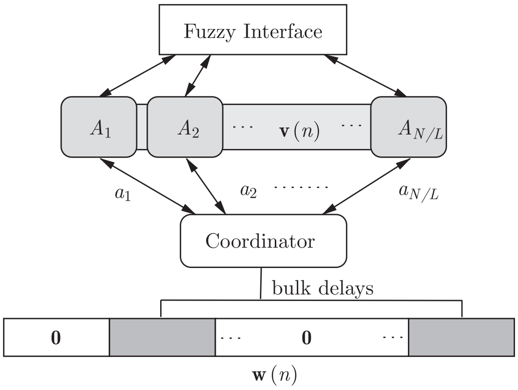

Fig. 6. Multiple echo cancellation system architecture.

- The delay between successive echoes (or reflections) is larger than L, i.e. dispersive regions do not overlap.
- The number of dispersive regions is upper bounded by a small integer n_D.

These standard assumptions ensure that some advantages can be gained by using an adaptive solution tailored to the ME case; otherwise, the use of a full time-domain adaptive filter may be more appropriate. For instance, most network hybrid impulse responses range from 3 to 12 ms, or a corresponding length of 24–96 samples at a sampling rate of 8 kHz [20], which is typically much smaller than the maximum possible bulk delay.

A block diagram of the proposed multiple echo-PHDAF (ME-PHDAF) is shown in Fig. 6. The partial Haar adaptive filter is partitioned into N/L sections, each of length L/M coefficients. Assigned to each *region* is an agent A_i that runs a distributed form of the SCE algorithm that departs slightly from the case of a single dispersive region in Section 3.3. The length of each region is $\beta L/M$, obtained by extending the corresponding section of the partial Haar filter so that it overlaps with its left neighbour; the overlapping factor is typically set to $\beta = 1.5$. The PDM of agent i at time n is now given by

$$\mathrm{PDM}_i = 1 - \frac{\min(c_{i,\max}, \tilde{c})}{c_{i,\max}}. \tag{24}$$

Each agent sends its $c_{i,\min}$ to the coordinator (see Fig. 6) where $\tilde{c}$ is calculated as $\tilde{c} = \max_i c_{i,\min}$. The use of the min operator in (24) is necessary because $\tilde{c} > c_{i\,\max}$ in some cases, making the original PDM negative. This form of cooperating agents preserves the global characteristic of the PDM.

If an agent requires a context switch, then only its corresponding *section*'s (not region) context should be changed. The polyphase decomposition of the RPHT from Section 3.1 comes in handy here, since each set of transformed input samples spanning a given section is readily available. The use of overlapping regions help to ensure that contiguous agents sharing a peak near their boundary have the same operating context, so that the orthogonality of the corresponding Haar basis vectors is preserved. Each agent is classified as active ($a_i = 1$) or inactive ($a_i = 0$), depending on its peak's behavioural model. An agent becomes active if its peak tendency has been in an increasing state for more than T_{inc} samples; otherwise, it is deactivated. A maximum of n_D agents can be active at any given time, based on a first-in first-out mechanism.

With information about the state of an agent, the central coordinator does one of two things: If one or more agents are active, the coordinator uses the peak locations of those agents to update only certain tap-weights of the time-domain adaptive filter centered on these locations. In effect, the input signal $u(n)$ in Fig. 1 is time-shifted and applied as input to short adaptive filter sections of length L, represented by the shaded areas in Fig. 6. If none of the agents are active (i.e. during initialization of the algorithm), the coordinator uses a single-echo version of the SCE scheme to locate a single dispersive region and proceeds to update the coefficients of a time-domain filter of length L centered on this delay.

3.6. Computational complexity

For convenience, we shall refer to the PHDAF algorithm incorporating the proposed SCE and IT schemes of Sections 3.3 and 3.4 as the improved PHDAF (I-PHDAF), and to the PHDAF incorporating the multi-agent structure in Section 3.5 as the ME-PHDAF. We first note that the partial Haar transform used in these algorithms is computationally efficient as it does not require explicit multiplication operations. Indeed, for a given scale index j, the normalization factor α_j is constant; it can be factored out of the matrix $\mathbf{H}_q$ and absorbed as part of subsequent operations on the transformed output vector.

The bulk of the SCE and IT schemes' computational complexity lies with the PTE. Table 3 shows the number of arithmetic operations per iteration required by the PTE, the SCE and the IT schemes, along with the required number of operations for an NLMS-based realization of the basic PHDAF structure in Fig. 1. We note that the arithmetic operations associated with the PTE are drastically reduced since many bbm terms, such as $v_n(S \cap L)$ and $\mu^{\mathscr{M}}_{n|n-1}(S \cup L)$, are zero throughout the PTS's operation [21]. The logarithmic operations needed in (23) can also be eliminated since only the relative entropy is required. An alternate measure that preserves the desired relationship and requires only a single comparison operation is given by $\tilde{H}^{\mathscr{M}}_{\mathrm{pig}} = \min(P^{\mathscr{M}}\{S\}, P^{\mathscr{M}}\{L\})$. Therefore, as an example, letting $N = 1024$, $q = 256$, and $L = 128$, the percentage increase in complexity when using the I-PHDAF algorithm, versus the basic PHDAF algorithm, is only $\frac{61}{1548} = 3.94\%$.

Table 3
Number of arithmetic operations per iteration—single dispersive region (I-PHDAF).

	PTE	SCE	IT	PHDAF
Add.	17	1	0	$2q + 2L + N/q + 1$
Mult.	18	0	0	$2q + 2L + 6$
Div.	3	0	0	1
Comp.	11	5	6	0

Table 4
Number of arithmetic operations per iteration—multiple dispersive regions (ME-PHDAF).

	Multi-agent	PHDAF
Add.	$18N/L$	$2q+2nDL+N/q+1$
Mult.	$18N/L$	$2q+2nDL+6$
Div.	$3N/L$	1
Comp.	$18N/L-1$	0

The computational complexity of the ME-PHDAF is summarized in Table 4. Note that the complexity of the PHDAF component now includes provisions for the use of n_D short time-domain adaptive filters. The complexity of the multi-agent scheme is on the order of N/L times that of the basic PTE/SCE scheme in Table 3. To reduce the computational load, the ME-PHDAF can be programmed to allow only one agent to estimate its peak tendency every input sample, i.e. agents time-share the PTE. This also requires scaling the trial periods in the schedule $\boldsymbol{\tau}$ and relevant counters by N/L.

4. Computer simulations

4.1. Methodology

The set of hybrid impulse responses used in the following simulations are taken from Annex D of the ITU-T G.168 Recommendation for digital network echo cancellers [20]. There are eight impulse responses $m_i(n)$ $(i=1,2,\ldots,8)$ with lengths L_i that range from 64 to 128 at a sampling rate of 8 kHz; it is assumed that $m_i(n)=0$ for $n<0$ and for $n\geqslant L_i$. These impulse responses are time shifted as needed and scaled to obtain an echo return loss of 15 dB. The input samples $u(n)$ are taken from a white zero-mean Gaussian process with unit variance. The error signal $d(n)$ is obtained by filtering $u(n)$ with the selected impulse response and adding a white Gaussian measurement noise $v(n)$, uncorrelated with $u(n)$ and of zero-mean and variance $\sigma_v^2=10^{-\mathrm{SNR}/10}$. The SNR is set to 30 dB, unless stated otherwise.

The length of the input data vector (i.e. maximum length of unknown echo path) is set to $N=1024$. The PHDAF and proposed I-PHDAF and ME-PHDAF algorithms use a partial Haar transform of size $q=256$ and the length of the dispersive region(s), i.e. short time-domain filter(s), is assumed to be $L=128$. The partial Haar and short time-domain adaptive filters utilize the NLMS algorithm with a step-size $\mu=1$.[5] In the proposed PHDAF algorithms, the initial context is set to 1, which corresponds to no shifting of the partial Haar basis vectors. A conventional NLMS algorithm (i.e. length N) is also used as a reference for comparison.

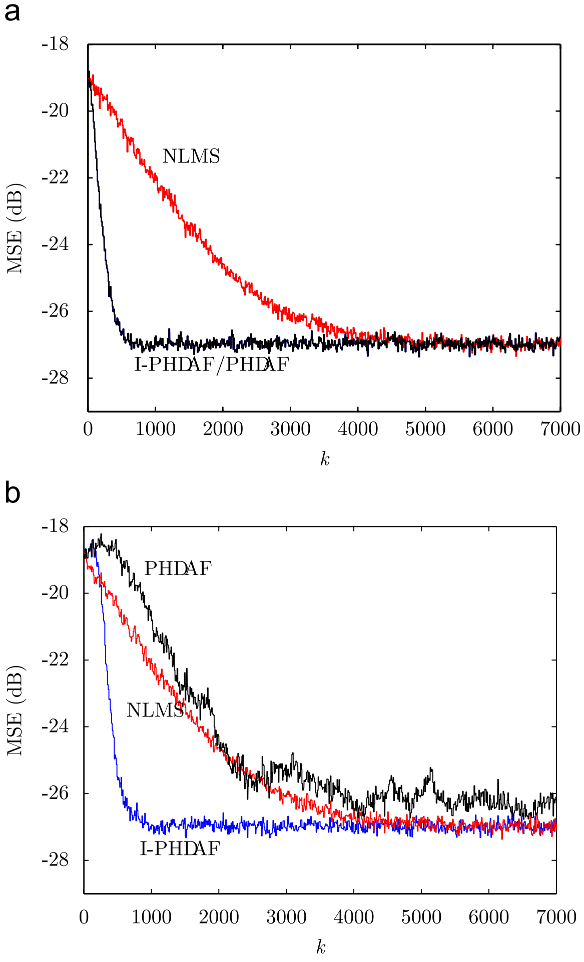

Fig. 7. Learning curves for ITU-T hybrid response $m_5(n)$ using a: (a) best and (b) worst-case bulk delay for the initial context used (SNR = 30 dB).

4.2. Context escaping

We use the following schedule of trial periods to test the different contexts: $\boldsymbol{\tau}=\{150,250,300,400\}$. Fig. 7(a) and (b) show the learning curves corresponding to an echo-path impulse response using ITU-T G.168 hybrid model $m_5(n-l)$ under the best and worst bulk delays (with respect to the initial context), respectively. The curves represent an ensemble of 200 runs for each simulation.

For the best-case bulk delay (Fig. 7(a)), both the I-PHDAF and PHDAF show identical learning curves reaching steady state at around $k=750$ compared to the NLMS which converges at around $k=5000$. This is due to the fact that the NLMS adapts a far larger number of coefficients (1024 compared to 128). In fact, this result (i.e. when the best bulk delay is used for the initial context) may be viewed as a best-case performance for the proposed I-PHDAF (and also PHDAF) since it leads to the fastest possible detection of the global peak. Even if all the contexts are adapted in parallel, this would not lead to a faster convergence at the considered SNR level of 30 dB. For a worst-case bulk delay (Fig. 7(b)), the PHDAF never

[5] In the partial Haar NLMS update equation (4), $\|\mathbf{u}(n)\|^2$ is used instead of $\|\mathbf{z}(n)\|^2$ for the normalization step; this was found to make the algorithm more robust to the rank deficiency of the transform matrix $\mathbf{H}_q$.

Table 5
Comparison of mean times and standard deviations to correctly estimate the peak delay for different SNRs.

SNR (dB)	I-PHDAF		PHDAF	
	Mean	Std.	Mean	Std.
30	91.5	75.4	121.1	203.0
20	107.7	86.4	214.5	664.0
15	167.4	138.3	362.7	1067.7
10	421.4	387.1	531.7	1177.2

seems to reach steady-state, while the proposed I-PHDAF converges much faster, nearly as well as in the optimal case, requiring about $k = 1000$ samples to converge. In effect, the I-PHDAF adds flexibility to the basic PHDAF and prevents it from "getting trapped" in suboptimal contexts. At the same time, the I-PHDAF is fairly stable, i.e. can remain "locked" on an optimal context. This behaviour may significantly increase the convergence speed of the echo canceller, as observed.

In addition to learning curves, the mean time for each echo canceller to correctly estimate the location of a dispersive region was compared for different SNRs. Each row in Table 5 consists of the average and standard deviation over 500 independent runs. For each run, one of the eight ITU-T G.168 hybrid impulse responses is selected with equal probability, and a random bulk delay uniformly distributed in the interval $[0, 895]$ is included. In all cases, the proposed I-PHDAF algorithm finds the dispersive region faster (the mean time is smaller) and more consistently (the standard deviation is much smaller). Although the PHDAF and the I-PHDAF display comparable mean times to convergence at very low SNR, the standard deviation of the PHDAF is three times larger. This reveals the robustness of using a fixed schedule τ together with a PTE for different values of SNR. Of course, if the SNR does not change much over a specific channel, then schedules can be constructed specifically for those cases.

4.3. Improved tracking

For reasons of brevity, only one generic scenario will be considered here in which the bulk delay of the true echo-path impulse response abruptly changes from an optimal to a suboptimal delay with respect to the initial context. Specifically, the change is from $m_1(n-640)$ to $m_5(n-322)$ at time $k = 5200$. The parameters of the proposed IT approach are chosen as $T_{RS} = 32$ and $T_{inc} = 128$.

The corresponding learning curves, averaged over 200 independent runs are shown in Fig. 8. It can be seen that, initially, both echo cancellers converge optimally. However, after the abrupt change in bulk delay, the I-PHDAF requires much fewer samples to converge. The learning curve of the PHDAF features a plateau-region extending for almost 2000 samples after the change. This corresponds to the amount of time it takes the new peak's magnitude (which is very small since the new bulk delay is suboptimal) to exceed the decreasing magnitude of the old peak.

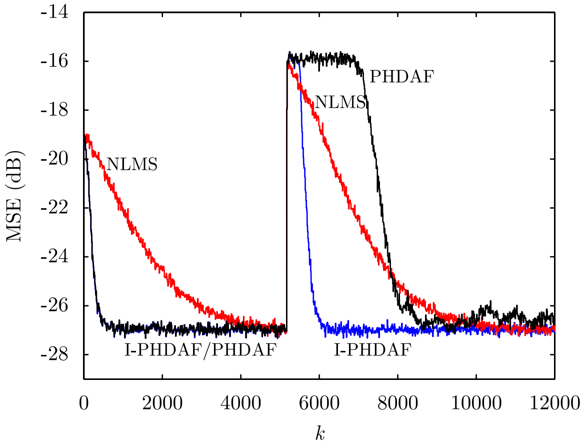

Fig. 8. Tracking behaviour of the I-PHDAF compared to the PHDAF and NLMS algorithm with a best-to-worst-case change in bulk delay (SNR = 30 dB).

4.4. ME cancellation

The ME-PHDAF algorithm of Section 3.5 is tested with $N/L = 8$ agents operating in parallel to classify and locate dispersive regions. The number of possible dispersive regions is set to $n_D = 3$ and the parameter $T_{inc} = 64$ samples.

The learning curves in Fig. 9(a) and (b) first compare the ME-PHDAF's performance to that of the PHDAF for the two cases of *single* echo path in Fig. 7. Each learning curve was averaged over 200 runs. For the best-case bulk delay in Fig. 9(a), the PHDAF is slightly faster than the ME-PHDAF, although both echo canceller's require only around 1000 samples to converge. However, the distributed form of the SCE algorithm allows the ME-PHDAF to converge again much faster in a worst-case bulk delay in Fig. 9(b).

Fig. 10 shows the performance of the ME canceller when the number of dispersive regions in the true echo-path impulse response is 2. The learning curves are averaged over 600 independent runs consisting of bulk delays and dispersive regions randomly selected according to

$$h(n) = g_1 m_i(n - \Delta_1)(n) + g_2 m_j(n - \Delta_2), \qquad (25)$$

where $h(n)$ is the echo-path impulse response, g_1 and g_2 are fixed gains, i and j are independent equiprobable selections from $\{1,2,\ldots,8\}$, $\Delta_1 = 320 + r_1$, $\Delta_2 = 640 + r_2$ and r_1, r_2 are independent equiprobable in $\{0, 1, 2, 3\}$. The IdNLMS filter represents the ideal case where exact prior knowledge about the location of these two dispersive regions is available, and only the corresponding time-domain filter coefficient are adapted while the remaining coefficients are set to zero. This approach, used as a benchmark, provides the best possible performance in this case. From the results in Fig. 10, we note that for $k<500$, the ME-PHDAF's learning curve is slightly larger than the NLMS (only one peak is initially detected), after which its convergence speed increases, reaching the IdNLMS curve's steady-state MSE at around $k = 2500$.

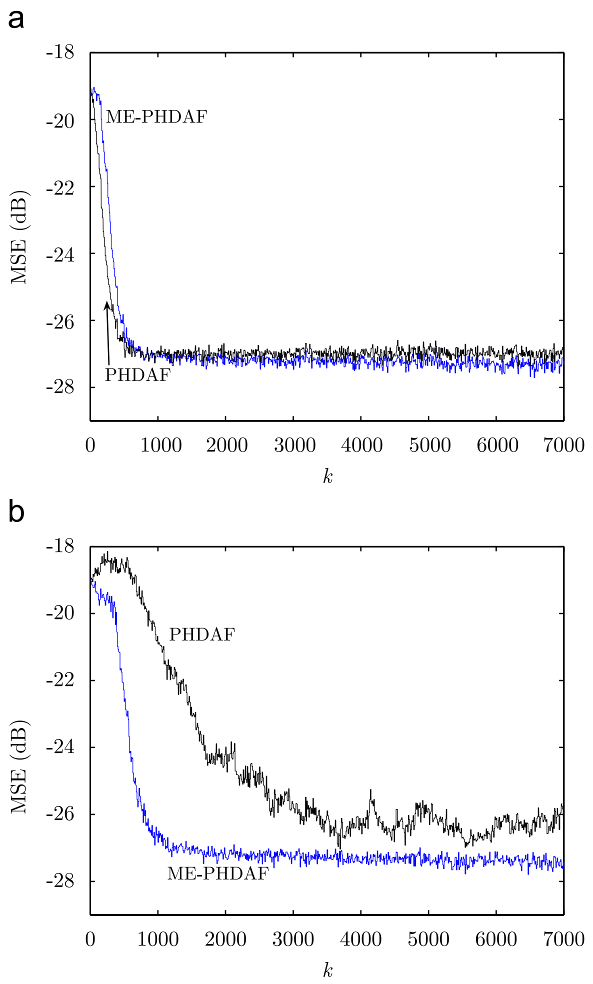

Fig. 9. Learning curves for ITU-T hybrid response $m_5(n)$ learning using a multiple echo canceller and the PHDAF: (a) best-case bulk delay; (b) worst-case bulk delay (SNR = 30 dB).

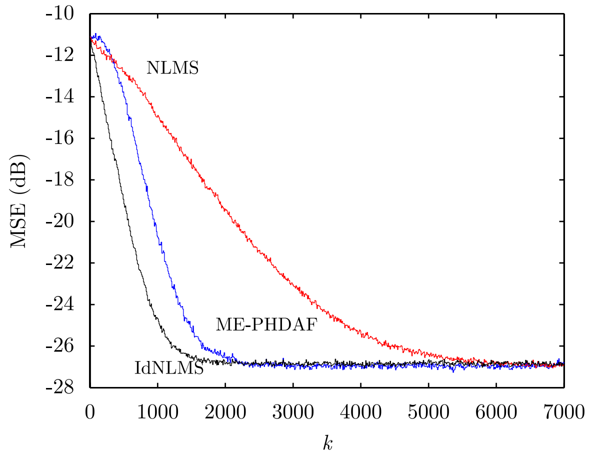

Fig. 10. Learning curves over 600 runs using two random hybrid impulse responses and random bulk delays (SNR = 30 dB).

5. Summary and conclusion

We have proposed feasible solutions to improve the performance of the *partial Haar dual adaptive filter* (PHDAF) in practical applications. These includes: (1) alleviating the dependence of the PHDAF's performance on the bulk delay of the echo-path impulse response; (2) improving the tracking performance of the PHDAF in response to abrupt changes in the echo path; and (3) extending the original PHDAF structure to support the cancellation of MEs.

The proposed algorithmic solutions exploit the polyphase representation of the RPHT and make use of a peak tendency estimator (PTE) based on Dezert–Smarandache theory (DSmT) and fuzzy inference. The PTE monitors the time evolution of the Haar-domain peak coefficient's discernibility, and categorizes it as being in an increasing or a decreasing state. In the proposed IPHDAF algorithm, this information is used to decide if the current operating context (i.e. polyphase component of the Haar transform) should be changed, as well as to detect if an abrupt change in the echo-path impulse response has occurred. To deal with the case of multiple dispersive regions, an ME-PHDAF algorithm is also proposed which uses a distributed (i.e. multi-agent) form of the above PTE-based approach.

The IPHDAF and ME-PHDAF algorithms were evaluated in terms of their mean-square error (MSE) curves as well as their mean time to properly locate dispersive regions for different SNR and channel conditions. Results show that a significantly enhanced performance can be obtained using the proposed algorithms at a minimal increase in computational cost when compared to the original PHDAF algorithm. Finally, we note that the techniques proposed in this paper for the PHDAF are not conceptually limited to the Haar transform and could be applied to other types of wavelets as well.

Appendix A. Theoretical motivation behind the PDM

In this appendix, we further motivate the choice of the PDM (13) on theoretical grounds. We begin by deriving an expression for P_C, defined as the probability of correctly locating the peak magnitude of the partial Haar adaptive filter coefficient vector $\mathbf{v}$, where the dependence on discrete-time n is omitted for convenience. To this end, we generalize the development in [14] to the case of an arbitrary mean vector. Specifically, let $\mathbf{v} = [v_1, \ldots, v_q]$ and assume that the individual weights v_i are statistically independent random variables with mean c_i and probability density functions (pdf) $f_i(u) = \phi(u - c_i)$, derived from a common symmetric pdf $\phi(u)$ with zero-mean, i.e. $\phi(-u) = \phi(u)$. Define $c_{\max} = \max_{i=1,\ldots,q}(|c_i|)$ and let i_0 denote the corresponding index; without loss in generality, assume that $c_{i_0} = c_{\max}$ (positive peak). Then, P_C can be expressed as

$$P_C = \int_{-\infty}^{\infty} \Pr\{|v_i| < |v_{i_0}| \text{ for all } i \neq i_0 | v_{i_0} = u\}\phi(u - c_{\max})\,\mathrm{d}u \tag{A.1}$$

$$= \int_{-\infty}^{\infty} \left\{ \prod_{i \neq i_0} \int_{|u_i| < |u|} \phi(u_i - c_i)\,\mathrm{d}u_i \right\} \phi(u - c_{\max})\,\mathrm{d}u. \tag{A.2}$$

Next, we derive a lower bound and an upper bound for P_C. Let $\mathscr{I}_1, \ldots, \mathscr{I}_K$ define a partition of the set $\mathscr{I} = \{i \in \mathbb{N} : 1 \leqslant i \leqslant q \text{ and } i \neq i_0\}$ and let ρ_k denotes the number of elements in $\mathscr{I}_k$. For each subset $\mathscr{I}_k$, define $b_k = \min_{i \in \mathscr{I}_k}(|c_i|)$ and $d_k = \max_{i \in \mathscr{I}_k}(|c_i|)$. Due to the symmetry of $\phi(v)$, it can be verified that $P_C^l \leqslant P_C \leqslant P_C^u$ where

$$P_C^l = \int_{-\infty}^{\infty} \prod_{k=1}^{K} \left\{ \int_{|u_k|<|u|} \phi(u_k - d_k)\,\mathrm{d}u_k \right\}^{\rho_k} \phi(u - c_{\max})\,\mathrm{d}u \tag{A.3}$$

and P_C^u is given by a similar expression with the numbers d_k replaced by b_k, respectively. We note that this bound on P_C becomes tighter as the size ρ_k of the subsets $\mathscr{I}_k$ is reduced.

Define $c_{\min} = \min_{k=1,\ldots,K}(|d_k|)$, i.e. the minimum among the maximum peaks of each of the intervals $\mathscr{I}_k$. We assume that each subinterval $\mathscr{I}_k$ contains at least one coefficient with peak magnitude less than $c_{\min}$, i.e. $b_k \leqslant c_{\min}$, which is reasonable for larger subintervals. Then we also have $P_C^l \leqslant P_C' \leqslant P_C^u$ where

$$P_C' = \int_{-\infty}^{\infty} \left\{ \int_{|v|<|u|} \phi(v - c_{\min})\,\mathrm{d}v \right\}^{q-1} \phi(u - c_{\max})\,\mathrm{d}u \tag{A.4}$$

now provides an approximation to P_C. It can be shown that: P_C' is an increasing function of $c_{\max}$; P_C' is a decreasing function of $c_{\min}$; $P_C' \approx 1$ in the limit $c_{\max} \gg c_{\min}$; and $P_C' \approx 0$ in the limit $c_{\min} = c_{\max}$ and $q \gg 1$. This behaviour is captured by the ideal PDM

$$\mathrm{PDM}^0 = 1 - c_{\min}/c_{\max}. \tag{A.5}$$

In this respect, Eq. (13) represents a practical attempt at estimating the unknown quantity PDM^0 based on the available data, i.e. the filter coefficients $v_i(n)$ at time n, and the use of only three subintervals.

Appendix B. Motivation behind using a schedule of trial periods

The motivation for using a schedule of trial periods is embedded in system reliability theory [31]. We assume the existence of M systems $Y_1, \ldots, Y_M$ that provide the solution to the same problem. The time for each system to produce the solution is a random variable governed by an exponential cumulative distribution $F(t; \lambda_i) = 1 - \mathrm{e}^{-\lambda_i t}$ for $i = 1, \ldots, M$, with mean $\mu_i = 1/\lambda_i$ and variance $\sigma_i^2 = 1/\lambda_i^2$, where λ_i is the rate parameter for system i. A randomly ordered selection of the M systems is made, as represented by a 1-to-1 permutation π of the integers $\{1, \ldots, M\}$. Then, beginning with system Y_{π_1} up to Y_{π_M}, each system is tested sequentially for a period of time $\tau_1, \ldots, \tau_M$, respectively. Given knowledge of the different rates λ_i, and assuming that the systems operate independently, the problem of interest here is to choose the trial times τ_i, or equivalently the system switching times $s_i = \tau_1 + \cdots + \tau_i$, so that the total time spent is searching a solution is minimized on a statistical basis under the constraint of a fixed total trial time $T = s_M$. The problem is made difficult by the fact that the user does not know the permutation π, i.e. which system is being used in a given trial interval.

Let random variable X denote the total time needed to solve the problem. The conditional CDF of X given a permutation π is shown to be

$$F_X(t|\pi) = 1 - \alpha_k(\pi)\mathrm{e}^{-\lambda_{\pi_k}(t - s_{k-1})}, \quad s_{k-1} \leqslant t \leqslant s_k, \tag{B.1}$$

where we define $\alpha_k(\pi) = \prod_{i=1}^{k-1} \mathrm{e}^{-\lambda_{\pi_i}\tau_i}$ for $k > 1$, $\alpha_1(\pi) = 1$ and $s_0 = 0$. For $t < 0$, we have $F_X(t|\pi) = 0$ while for $t > s_M$, we assume $F_X(t|\pi) = 1$, i.e. the probability of not obtaining a solution within a total time s_M is negligible. The conditional expectation of X can be obtained as

$$E[X|\pi] = \int_0^{\infty} (1 - F_X(t|\pi))\,\mathrm{d}t = \sum_{k=1}^{M} \frac{\alpha_k(\pi)}{\lambda_{\pi_k}}(1 - \mathrm{e}^{-\lambda_{\pi_k}\tau_k}). \tag{B.2}$$

Finally, assuming that the different possible permutations are equiprobable, the expected total solution time is obtained as

$$E[X] = \frac{1}{M!}\sum_{\pi} E[X|\pi], \tag{B.3}$$

where the summation is over all possible permutations π.

While an exhaustive analysis of the above expression for $E[X]$ is beyond the scope of the paper, we can nevertheless obtain interesting insights by considering $M = 2$, in which case (B.3) takes the simplified form

$$\begin{aligned} E[X] = \frac{1}{2}\Bigg[&\frac{1}{\lambda_1} + \frac{1}{\lambda_2}\mathrm{e}^{-\lambda_1\tau_1}\left(1 - \frac{\lambda_2}{\lambda_1} - \mathrm{e}^{-\lambda_2\tau_2}\right) \\ &+ \frac{1}{\lambda_2} + \frac{1}{\lambda_1}\mathrm{e}^{-\lambda_2\tau_1}\left(1 - \frac{\lambda_1}{\lambda_2} - \mathrm{e}^{-\lambda_1\tau_2}\right)\Bigg]. \end{aligned} \tag{B.4}$$

In the case of equal rates, i.e. $\lambda_1 = \lambda_2 = \lambda$, this expression simplifies to $E[X] = (1/\lambda)(1 - \mathrm{e}^{-\lambda T})$ which is independent of the switching time $\tau_1 = s_1$. However, in the case $\lambda_1 \neq \lambda_2$, the situation is different, as revealed by the plots of $E[X]$ versus τ_1 in Fig. B1, obtained under the constraint $\tau_1 + \tau_2 = T$ for different values of the rate parameters λ_1 and λ_2. The value of τ_1 minimizing $E[X]$ in these plots can be obtained by applying the method of Lagrange multipliers and satisfies $(0 < \tau_1 < T)$:

$$\lambda_1\mathrm{e}^{-\lambda_1\tau_1}(1 - \mathrm{e}^{-\lambda_2(T-\tau_1)}) = \lambda_2\mathrm{e}^{-\lambda_2\tau_1}(1 - \mathrm{e}^{-\lambda_1(T-\tau_1)}). \tag{B.5}$$

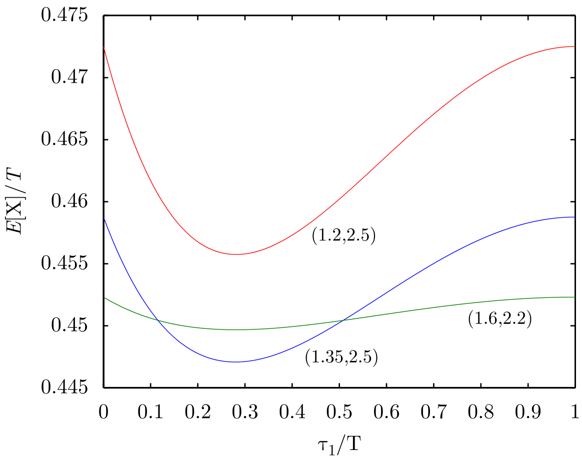

Fig. B1. Plot of $E[X]/T$ vs. τ_1/T for three pairs of $(\lambda_1 T, \lambda_2 T)$.

We generally find that for a wide range of parameter values under consideration, the optimum switching time occurs at $\tau_1 \approx 0.3T$. Clearly, this approach can be generalized to the case of $M > 2$ intervals.

References

[1] S.V. Vaseghi, Advanced Signal Processing and Digital Noise Reduction, Wiley, B.G. Teubner, New York, Stuttgart, 1996.

[2] S. Dimolitsas, J.G. Phipps Jr., A. Wong, Impact of delay on the voice transmission performance of mobile-satellite systems, in: Proceedings of the IEEE 10th International Conference on Digital Satellite Communications, vol. 1, 1995, pp. 96–100.

[3] D.L. Duttweiler, Subsampling to estimate delay with application to echo cancelling, IEEE Trans. Acoust. Speech Signal Process. 31 (1983) 1090–1099.

[4] R.K. Martin, W.A. Sethares, R.C. Williamson, C.R. Johnson Jr., Exploiting sparsity in adaptive filters, IEEE Trans. Signal Process. 50 (8) (2002) 1883–1894.

[5] P.C.-W. Yip, D.M. Etter, An adaptive multiple echo canceller for slowly time-varying echo paths, IEEE Trans. Commun. 38 (1990) 1693–1698.

[6] A. Sugiyama, S. Ikeda, A fast convergence algorithm for adaptive FIR filters with sparse taps, IEICE Trans. Fundamentals E77-A (1994) 681–686.

[7] S. Kawamura, M. Hatori, A tap selection algorithm for adaptive filters, in: Proceedings of the IEEE International Conference on Acoustics, Speech, Signal Processing, 1986, pp. 2979–2982.

[8] D.L. Duttweiler, Proportionate normalized least-mean-squares adaptation in echo cancelers, IEEE Trans. Speech Audio Process. 8 (5) (2000) 508–518.

[9] S.L. Gay, An efficient, fast converging adaptive filter for network echo cancellation, in: Proceedings of the 37th Asilomar Conference on Signals, Systems, and Computers, vol. 1, 1998, pp. 394–398.

[10] J. Benesty, S.L. Gay, An improved PNLMS algorithm, in: Proceedings of the IEEE International Conference on Acoustics, Speech, Signal Processing, vol. 2, 2002, pp. 1881–1884.

[11] J. Cui, P.A. Naylor, D.T. Brown, An improved IPNLMS algorithm for echo cancellation in packet-switched networks, in: Proceedings of the IEEE International Conference on Acoustics, Speech, Signal Processing, vol. 4, 2004, pp. 141–144.

[12] M. Doroslovački, H. Fan, On-line identification of echo-path impulse responses by Haar-wavelet-based adaptive filter, in: Proceedings of the IEEE International Conference on Acoustics, Speech, Signal Processing, vol. 2, 1995, pp. 1065–1068.

[13] K.C. Ho, S.D. Blunt, Rapid identification of a sparse impulse response using an adaptive algorithm in the Haar domain, IEEE Trans. Signal Process. 51 (3) (2003) 628–638.

[14] N.J. Bershad, A. Bist, Fast coupled adaptation for sparse impulse responses using a partial Haar transform, IEEE Trans. Signal Process. 53 (3) (2005) 966–976.

[15] N.J. Bershad, J. Bermudez, J.-Y. Tourneret, Stochastic analysis of the LMS algorithm for system identification with subspace inputs, IEEE Trans. Signal Process. 56 (2008) 1018–1027.

[16] C. Ribas, J. Bermudez, N.J. Bershad, Low-complexity robust sparse channel identification using partial block wavelet transforms—analysis and implementation, in: Proceedings of the IEEE International Conference on Acoustics, Speech, Signal Processing, 2008, pp. 3281–3284.

[17] J.M. Mendel, Fuzzy logic systems for engineering, Proc. IEEE 83 (3) (1995) 345–377.

[18] J. Dezert, Foundations for a new theory of plausible and paradoxical reasoning, Int. J. Inf. Security 9 (2002) 13–57.

[19] C.S. Burrus, R.A. Gopinath, H. Guo, Introduction to Wavelets and Wavelet Transforms, A Primer, first ed., Prentice-Hall, Inc., Englewood Cliffs, NJ, 1998.

[20] Digital network echo cancellers, ITU-T recommendation G.168, Technical Report, International Telecommunication Union, August 2004.

[21] A. Tchamova, T. Semerdjiev, J. Dezert, Estimation of target behavior tendencies using DSmT, in: J. Dezert, F. Smarandache (Eds.), Advances and Applications of DSmT for Information Fusion (Collected Works), vol. I, American Research Press, 2004.

[22] G.J. Klir, B. Yuan, Fuzzy Sets and Fuzzy Logic: Theory and Applications, Prentice-Hall, New Jersey, 1995.

[23] L.-X. Wang, J.M. Mendel, Generating fuzzy rules by learning from examples, IEEE Trans. Syst. Man Cybern. 22 (6) (1992) 1414–1427.

[24] H.T. Nguyen, E.A. Walker, A First Course in Fuzzy Logic, second ed., Chapman & Hall, CRC, London, Boca Raton, FL, 2000.

[25] L.A. Zadeh, The role of fuzzy logic in the management of uncertainty in expert systems, Fuzzy Sets Syst. 11 (1) (1983) 199–227.

[26] L.-X. Wang, J.M. Mendel, Fuzzy adaptive filters, with application to nonlinear channel equalization, IEEE Trans. Fuzzy Syst. 1 (1993) 161–170.

[27] W.-S. Gan, Fuzzy step-size adjustment for the LMS algorithm, Signal Processing 49 (1996) 145–149.

[28] W.-S. Gan, Designing a fuzzy step size LMS algorithm, IEE Proc. Vis. Image Signal Process. 144 (1997) 261–266.

[29] J. Dezert, F. Smarandache, M. Khoshnevisan, Counter-examples to Dempster's rule of combination, in: J. Dezert, F. Smarandache (Eds.), Advances and Applications of DSmT for Information Fusion (Collected Works), American Research Press, 2004.

[30] C. Cruz, D.A. Pelta, A.S. Royo, J.L. Verdegay, Soft computing and cooperative strategies for optimization, in: Proceedings of the IEEE Mid-summer Workshop on Soft Computing in Industrial Applications, Espoo, Finland, 2005.

[31] K.S. Trivedi, Probability and Statistics with Reliability, Queuing, and Computer Science Applications, Wiley, USA, 2002.