A KALMAN FILTER WITH INDIVIDUAL CONTROL FACTORS FOR ECHO CANCELLATION

Constantin Paleologu[†], *Jacob Benesty*[‡], *Silviu Ciochină*[†], *and Steven L. Grant*[§]

[†] University Politehnica of Bucharest, Romania, e-mail: {pale, silviu}@comm.pub.ro
[‡] INRS-EMT, University of Quebec, Montreal, Canada, e-mail: benesty@emt.inrs.ca
[§] Missouri University of Science and Technology, Rolla, USA, e-mail: sgrant@mst.edu

ABSTRACT

In echo cancellation, the main goal is to recover the near-end signal from the error signal of the adaptive filter, which identifies the echo path. In this context, the Kalman filter represents a very appealing choice, since its basic criterion follows the minimization of the system misalignment (instead of the usual error-based cost function). In this paper, we propose a Kalman filter with individual control factors, in terms of using a different level of uncertainty for each coefficient of the filter. As compared to the basic Kalman filter (which imposes the same uncertainty for all the coefficients of the impulse response), the proposed algorithm achieves better performance, especially in terms of the steady-state misalignment.

Index Terms— Echo cancellation, Kalman filter, system identification, adaptive filters.

1. INTRODUCTION

The Kalman filter can be used in many system identification problems [1], [2]. Based on the Bayesian approach, this algorithm recursively estimates a set of unknown variables from a set of (noisy) observations acquired over time. More recently, the Kalman filter was also involved in the context of echo cancellation [3]–[10]. Similar to a system identification problem, the main goal in echo cancellation is to estimate an unknown system (i.e., the echo path) from the microphone signal that contains the echo signal corrupted by different types of "noise" (e.g., the background noise and the near-end speech) [11], [12].

Most of the algorithms used in echo cancellation are based on the minimization of a cost function that depends on the error signal of the adaptive filter. However, in a system identification context, where the output of the unknown system is usually corrupted by another signal (i.e., the system noise), the goal of the adaptive filter is not to make the error signal goes to zero. The objective instead is to recover this system noise from the error signal of the adaptive filter (after this one converges to the true solution). This is mandatory in echo cancellation, where the "corrupting signal" comes from the near-end and should be delivered (as accurately as possible) to the far-end, through the error signal. Consequently, it makes more sense to minimize the system misalignment, like in the Kalman filtering, instead of the classical error-based cost function.

The Kalman filter uses a specific parameter that captures the uncertainties in the system to be identified and acts as a control factor. In the basic approach, this parameter has the same value for all the coefficients of the impulse response. In this paper, we consider a more interesting and realistic case by using a different level of uncertainty for each coefficient, thus resulting a Kalman filter with individual control factors.

The paper is organized as follows. In Section 2, the state variable model for echo cancellation and the basic Kalman filter are presented. The proposed Kalman filter with individual control factors is developed in Section 3. Simulation results presented in Section 4 (in the context of echo cancellation) indicate the good performance of the proposed algorithm. Finally, Section 5 concludes this paper and outlines the relation to prior work.

2. THE KALMAN FILTER FOR ECHO CANCELLATION

In a system identification problem, the reference (or desired) signal at the discrete-time index n is defined as

$$d(n) = \mathbf{x}^T(n)\mathbf{h}(n) + v(n), \quad (1)$$

where $\mathbf{x}(n) = \left[\begin{array}{cccc} x(n) & x(n-1) & \cdots & x(n-L+1) \end{array} \right]^T$ is a vector containing the L most recent time samples of the deterministic input signal, $x(n)$, superscript ^{T} denotes transpose of a vector or a matrix, $\mathbf{h}(n) = \left[\begin{array}{cccc} h_0(n) & h_1(n) & \cdots & h_{L-1}(n) \end{array} \right]^T$ is the impulse response (of length L) of the system that we need to identify, and $v(n)$ is a zero-mean stationary white Gaussian noise signal. In the context of echo cancellation, $d(n)$ is the microphone signal, $x(n)$ is the far-end (or loudspeaker) signal, and $\mathbf{h}(n)$ denotes the impulse response of the echo path (from

This work was supported under the Grant UEFISCDI PN-II-ID-PCE-2011-3-0097.

the loudspeaker to the microphone) [11], [12]. The objective is to estimate or identify $\mathbf{h}(n)$ with an adaptive filter $\widehat{\mathbf{h}}(n) = \left[\begin{array}{cccc} \widehat{h}_0(n) & \widehat{h}_1(n) & \cdots & \widehat{h}_{L-1}(n) \end{array} \right]^T$.

In order to model the system impulse response as a state equation, we assume that $\mathbf{h}(n)$ is a zero-mean random vector, which follows a simplified first-order Markov model, i.e.,

$$\mathbf{h}(n) = \mathbf{h}(n-1) + \mathbf{w}(n), \quad (2)$$

where $\mathbf{w}(n)$ is a zero-mean white Gaussian noise signal vector, which is uncorrelated with $\mathbf{h}(n-1)$ and $v(n)$. The correlation matrix of $\mathbf{w}(n)$ is assumed to be $\mathbf{R}_{\mathbf{w}}(n) = \sigma_w^2(n)\mathbf{I}_L$, where $\mathbf{I}_L$ is the $L \times L$ identity matrix. The variance, $\sigma_w^2(n)$, captures the uncertainties in $\mathbf{h}(n)$.

Relation (1) represents the observation equation, while expression (2) is called the state equation. Consequently, the echo cancellation problem may be restated based on these two fundamental equations and the objective is to find the optimal recursive estimator of $\mathbf{h}(n)$ denoted by $\widehat{\mathbf{h}}(n)$.

Next, the Kalman filter can be derived based on this simplified model defined by (1) and (2). In the context of the linear sequential Bayesian approach, the optimal estimate of the state vector, $\mathbf{h}(n)$, has the form [13]:

$$\widehat{\mathbf{h}}(n) = \widehat{\mathbf{h}}(n-1) + \mathbf{k}(n)e(n), \quad (3)$$

where $\mathbf{k}(n)$ is the Kalman gain vector and

$$e(n) = d(n) - \mathbf{x}^T(n)\widehat{\mathbf{h}}(n-1) \quad (4)$$

is the a priori error signal between the microphone signal and the estimate of the echo signal. Also, we can define the state estimation error or a posteriori misalignment as

$$\boldsymbol{\mu}(n) = \mathbf{h}(n) - \widehat{\mathbf{h}}(n). \quad (5)$$

The correlation matrix of $\boldsymbol{\mu}(n)$ is

$$\mathbf{R}_{\boldsymbol{\mu}}(n) = E\left[\boldsymbol{\mu}(n)\boldsymbol{\mu}^T(n)\right], \quad (6)$$

where $E[\cdot]$ denotes mathematical expectation. We can also define the a priori misalignment as

$$\begin{aligned} \mathbf{m}(n) &= \mathbf{h}(n) - \widehat{\mathbf{h}}(n-1) \\ &= \boldsymbol{\mu}(n-1) + \mathbf{w}(n), \end{aligned} \quad (7)$$

for which its correlation matrix is

$$\begin{aligned} \mathbf{R}_{\mathbf{m}}(n) &= E\left[\mathbf{m}(n)\mathbf{m}^T(n)\right] \\ &= \mathbf{R}_{\boldsymbol{\mu}}(n-1) + \mathbf{R}_{\mathbf{w}}(n). \end{aligned} \quad (8)$$

The Kalman gain vector is obtained by minimizing the criterion:

$$J(n) = \frac{1}{L}\text{tr}\left[\mathbf{R}_{\boldsymbol{\mu}}(n)\right] \quad (9)$$

with respect to $\mathbf{k}(n)$. From this minimization, we find that

$$\mathbf{k}(n) = \frac{\mathbf{R}_{\mathbf{m}}(n)\mathbf{x}(n)}{\mathbf{x}^T(n)\mathbf{R}_{\mathbf{m}}(n)\mathbf{x}(n) + \sigma_v^2(n)}, \quad (10)$$

where $\sigma_v^2(n) = E\left[v^2(n)\right]$. Also, it results that

$$\mathbf{R}_{\boldsymbol{\mu}}(n) = \left[\mathbf{I}_L - \mathbf{k}(n)\mathbf{x}^T(n)\right]\mathbf{R}_{\mathbf{m}}(n). \quad (11)$$

Summarizing, the following equations define the well-known Kalman filter [14]:

$$\begin{aligned} \mathbf{R}_{\mathbf{m}}(n) &= \mathbf{R}_{\boldsymbol{\mu}}(n-1) + \mathbf{R}_{\mathbf{w}}(n), && (12) \\ \mathbf{k}(n) &= \frac{\mathbf{R}_{\mathbf{m}}(n)\mathbf{x}(n)}{\mathbf{x}^T(n)\mathbf{R}_{\mathbf{m}}(n)\mathbf{x}(n) + \sigma_v^2(n)}, && (13) \\ e(n) &= d(n) - \mathbf{x}^T(n)\widehat{\mathbf{h}}(n-1), && (14) \\ \widehat{\mathbf{h}}(n) &= \widehat{\mathbf{h}}(n-1) + \mathbf{k}(n)e(n), && (15) \\ \mathbf{R}_{\boldsymbol{\mu}}(n) &= \left[\mathbf{I}_L - \mathbf{k}(n)\mathbf{x}^T(n)\right]\mathbf{R}_{\mathbf{m}}(n). && (16) \end{aligned}$$

The matrix $\mathbf{R}_{\mathbf{w}}(n)$ plays a major role in the performance of the algorithm, in terms of the values of $\sigma_w^2(n)$ [7], [9]. Small values of $\sigma_w^2(n)$ imply a good misalignment but a poor tracking, while large values of $\sigma_w^2(n)$ (i.e., high uncertainties in the echo path) imply a good tracking but a high misalignment. Consequently, this parameter highly influences the tracking abilities and the convergence of the Kalman filter.

3. KALMAN FILTER WITH INDIVIDUAL CONTROL FACTORS

Let us refer again to (2), where $\mathbf{w}(n)$ is a zero-mean white Gaussian noise signal vector. In all previous studies in echo cancellation, it is assumed that the correlation matrix of $\mathbf{w}(n)$ is

$$\mathbf{R}_{\mathbf{w}}(n) = \sigma_w^2(n)\mathbf{I}_L, \quad (17)$$

i.e., all the coefficients have the same level of uncertainty.

In the proposed approach, let us consider a more interesting and realistic case:

$$\mathbf{h}(n) = \mathbf{h}(n-1) + \widetilde{\mathbf{w}}(n), \quad (18)$$

where $\widetilde{\mathbf{w}}(n) = \left[\begin{array}{cccc} \widetilde{w}_0(n) & \widetilde{w}_1(n) & \cdots & \widetilde{w}_{L-1}(n) \end{array} \right]^T$ and

$$E\left[\widetilde{w}_k(n)\widetilde{w}_l(n)\right] = \begin{cases} \sigma_{\widetilde{w}_l}^2, & k = l \\ 0, & k \neq l \end{cases}, \quad (19)$$

with $k, l = 0, 1, \ldots, L-1$. In other words, we consider independent fluctuations for each coefficient. Consequently,

$$\begin{aligned} \mathbf{R}_{\widetilde{\mathbf{w}}}(n) &= E\left[\widetilde{\mathbf{w}}(n)\widetilde{\mathbf{w}}^T(n)\right] \\ &= \text{diag}\left[\sigma_{\widetilde{w}_0}^2(n), \sigma_{\widetilde{w}_1}^2(n), \ldots, \sigma_{\widetilde{w}_{L-1}}^2(n)\right], \end{aligned} \quad (20)$$

where diag$[\cdot]$ denotes a diagonal matrix.

Based on the state equation (18), we have

$$\widetilde{w}_l(n) = h_l(n) - h_l(n-1),\ l = 0, 1, \ldots, L-1, \quad (21)$$

so that

$$\begin{aligned} \sigma^2_{\widetilde{w}_l}(n) &= E\left[\widetilde{w}_l^2(n)\right] \\ &= E\left\{\left[h_l(n) - h_l(n-1)\right]^2\right\}. \end{aligned} \quad (22)$$

The value of $\sigma^2_{\widetilde{w}_l}(n)$ expresses the level of uncertainty in $h_l(n)$, with $l = 0, 1, \ldots, L-1$.

The form of the state equation (18) leads to a Kalman filter with individual control factors (ICF-KF). The proposed algorithm uses

$$\mathbf{R_m}(n) = \mathbf{R}_{\boldsymbol{\mu}}(n-1) + \mathbf{R}_{\widetilde{\mathbf{w}}}(n) \quad (23)$$

instead of (12), while the rest of the equations are the same as in the Kalman filter. Note that the ICF-KF takes into account the variation of each coefficient from one iteration to another (i.e., its level of uncertainty).

Following (22), a natural way to estimate $\sigma^2_{\widetilde{w}_l}(n)$ is

$$\begin{aligned} \widehat{\sigma}^2_{\widetilde{w}_l}(n) &= \lambda \widehat{\sigma}^2_{\widetilde{w}_l}(n-1) \\ &+ (1-\lambda)\left[\widehat{h}_l(n-1) - \widehat{h}_l(n-2)\right]^2, \\ & l = 0, 1, \ldots, L-1, \end{aligned} \quad (24)$$

with $\lambda = 1 - 1/(\kappa L)$ and $\kappa \geq 1$; thus, the correlation matrix from (20) becomes

$$\widehat{\mathbf{R}}_{\widetilde{\mathbf{w}}}(n) = \text{diag}\left[\widehat{\sigma}^2_{\widetilde{w}_0}(n), \widehat{\sigma}^2_{\widetilde{w}_1}(n), \ldots, \widehat{\sigma}^2_{\widetilde{w}_{L-1}}(n)\right]. \quad (25)$$

Of course, we can get back to (17) by using the mean of the parameters estimated in (24), i.e.,

$$\widehat{\sigma}^2_{\widetilde{w}}(n) = \frac{1}{L}\sum_{l=0}^{L-1} \widehat{\sigma}^2_{\widetilde{w}_l}(n). \quad (26)$$

Furthermore, using $\lambda = 0$ in (24) (i.e., without temporal averaging), the mean value from (26) becomes

$$\begin{aligned} \widehat{\sigma}^2_{\widetilde{w}}(n) &= \frac{1}{L}\sum_{l=0}^{L-1}\left[\widehat{h}_l(n-1) - \widehat{h}_l(n-2)\right]^2 \\ &= \frac{1}{L}\left\|\widehat{\mathbf{h}}(n-1) - \widehat{\mathbf{h}}(n-2)\right\|_2^2, \end{aligned} \quad (27)$$

where $\|\cdot\|_2$ is the ℓ_2 norm.

When the algorithm starts to converge, the misalignment of the individual coefficients tend to become uncorrelated. On the other hand, we could encounter large variations for some coefficients (like in a tracking situation, when the echo path abruptly changes), which could bias the overall behavior of the algorithm. In order to prevent this issue, we propose to additionally control the estimation from (24), by imposing a maximum level of uncertainties given in (27), i.e.,

$$\underline{\widehat{\sigma}}^2_{\widetilde{w}_l}(n) = \min\{\widehat{\sigma}^2_{\widetilde{w}_l}(n), \widehat{\sigma}^2_{\widetilde{w}}(n)\},\ l = 0, 1, \ldots, L-1. \quad (28)$$

Finally, we should note that another important parameter to be found is the system noise power, $\sigma_v^2(n)$. Usually, it can be estimated during silences of the near-end talker, i.e., in the single-talk scenario [15]. Nevertheless, the most critical situation in echo cancellation is the double-talk case, when the near-end signal is a combination of the background noise and the near-end speech. In this scenario, the parameter $\sigma_v^2(n)$ can be estimated as discussed in [9], [16].

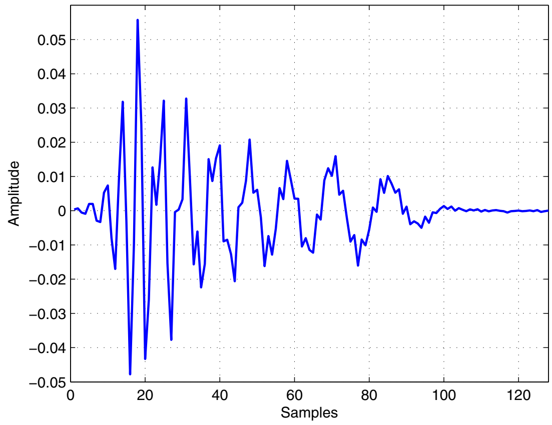

Fig. 1. Impulse response used in simulations (the fourth echo path from G168 Recommendation).

4. SIMULATION RESULTS

In the experiments, we consider a network echo cancellation scenario, in the framework of G168 Recommendation [17]. The echo path is depicted in Fig. 1; it is the fourth impulse response (of length $L = 128$) from the above recommendation. The sampling rate is 8 kHz. All adaptive filters used in the experiments have the same length as the echo path. The far-end signal (i.e., the input signal) is either a white Gaussian signal, an AR(1) process generated by filtering a white Gaussian noise through a first-order system $1/(1 - 0.8z^{-1})$, or a speech signal. The output of the echo path is corrupted by an independent white Gaussian noise with 20 dB signal-to-noise ratio (SNR). We assume that the variance of the noise, $\sigma_v^2(n)$, is available in most of the simulations (excepting the last one, performed in a double-talk scenario). An echo path change scenario is simulated in all the single-talk experiments (in order to evaluate the tracking capabilities of the algorithms), by shifting the impulse response to the right by 12 samples in the middle of each simulation. The performance measure used in simulations is the normalized misalignment (in dB) evaluated

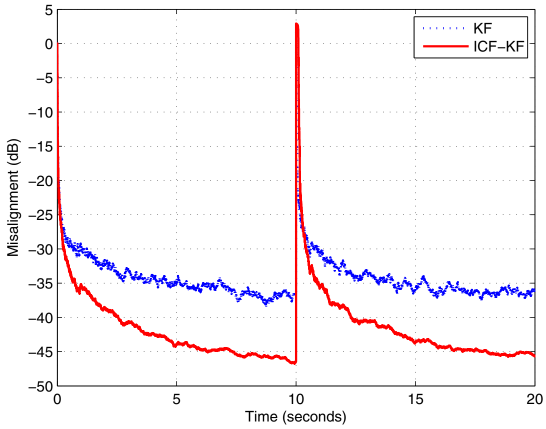

Fig. 2. Misalignment of the KF and ICF-KF. The input signal is white and Gaussian, $L = 128$, and SNR = 20 dB. Echo path changes at time 10 seconds.

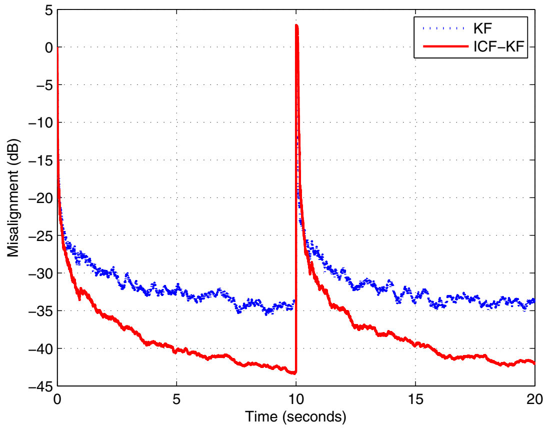

Fig. 3. Misalignment of the KF and ICF-KF. The input signal is an AR(1) process. Other conditions as in Fig. 2.

as

$$\text{Mis}(n) = 20\log_{10}\frac{\left\|\widehat{\mathbf{h}}(n) - \mathbf{h}(n)\right\|_2}{\|\mathbf{h}(n)\|_2}. \quad (29)$$

In all the experiments, the proposed ICF-KF is compared to the Kalman filter (KF) presented in [10].

In Fig. 2, the input signal is a white Gaussian noise. As we can see from this figure, both algorithms have similar initial convergence rate and tracking reaction, but the ICF-KF achieves a lower misalignment as compared to KF.

Figures 3 and 4 compare again the KF and ICF-KF this time with AR(1) and speech signals as inputs. The ICF-KF still performs much better than the KF, in terms of the final misalignment.

Finally, a double-talk scenario is evaluated in Fig. 5, without using any double-talk detector (DTD). The variance of the near-end signal is evaluated as in [9]. It can be noticed that both algorithms show good robustness against the near-end speech, while the ICF-KF achieves a lower misalignment.

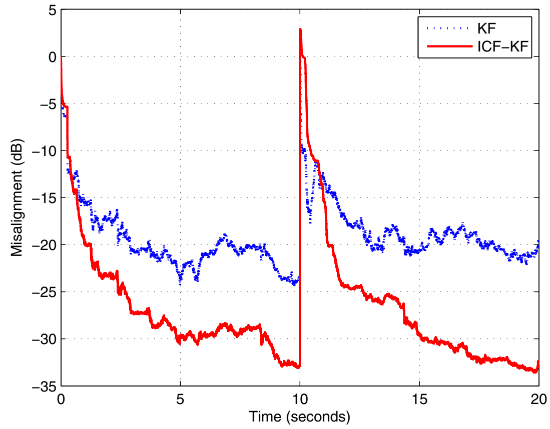

Fig. 4. Misalignment of the KF and ICF-KF. The input signal is speech. Other conditions as in Fig. 2.

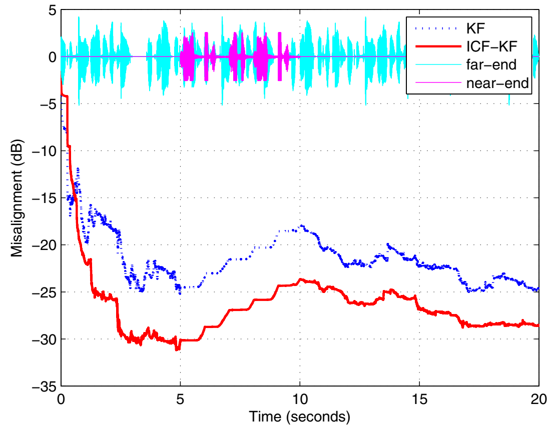

Fig. 5. Misalignment of the KF and ICF-KF. Double-talk scenario, without using any DTD.

5. DISCUSSION

In this paper, we have proposed a Kalman filter with individual control factors, in the context of echo cancellation. In prior work on Kalman filter for echo cancellation, there is a specific parameter that captures the uncertainties in the system and has the same value for all the coefficients of the impulse response. The proposed ICF-KF considers independent fluctuations in the state equation and, consequently, it uses a different level of uncertainty for each coefficient of the filter. Simulation results indicate the good performance of the proposed algorithm.

6. REFERENCES

[1] R. E. Kalman, “A new approach to linear filtering and prediction problems,” *J. Basic Engineering*, vol. 82, pp. 35–45, Mar. 1960.

[2] R. Faragher, “Understanding the basis of the Kalman filter via a simple and intuitive derivation,” *IEEE Signal Processing Magazine*, vol. 29, pp. 128–132, Sept. 2012.

[3] G. Enzner, *A Model-Based Optimum Filtering Approach to Acoustic Echo Control: Theory and Practice*. Dissertation, RWTH Aachen, Aachener Beitrge zu digitalen Nachrichtensystemen, Vary P. (ed.), Wissenschaftsverlag Mainz, Aachen, June 2006.

[4] G. Enzner and P. Vary, “Frequency-domain adaptive Kalman filter for acoustic echo control in hands-free telephones,” *Signal Processing*, vol. 86, pp. 1140–1156, 2006.

[5] S. Malik and G. Enzner, “Model-based vs. traditional frequency-domain adaptive filtering in the presence of continuous double-talk and acoustic echo path variability,” in *Proc. IWAENC*, 2008.

[6] G. Enzner, “Model-based interrelations of adaptive filter algorithms in acoustic echo control,” in *Proc. IEEE PacRim*, 2009, pp. 909–913.

[7] G. Enzner, “Bayesian inference model for applications of time-varying acoustic system identification,” in *Proc. EUSIPCO*, 2010, pp. 2126–2130.

[8] S. Malik and G. Enzner, “State-space frequency-domain adaptive filtering for nonlinear acoustic echo cancellation,” *IEEE Trans. Audio, Speech, Language Processing*, vol. 20, pp. 2065–2079, Sept. 2012.

[9] C. Paleologu, J. Benesty, and S. Ciochină, “Study of the general Kalman filter for echo cancellation,” *IEEE Trans. Audio, Speech, Language Processing*, vol. 21, pp. 1539–1549, Aug. 2013.

[10] C. Paleologu, J. Benesty, and S. Ciochină, “Study of the optimal and simplified Kalman filters for echo cancellation,” in *Proc. IEEE ICASSP*, 2013, pp. 580–584.

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

[12] C. Paleologu, J. Benesty, and S. Ciochină, *Sparse Adaptive Filters for Echo Cancellation*. Morgan & Claypool Publishers, Synthesis Lectures on Speech and Audio Processing, 2010.

[13] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Englewood Cliffs, NJ: Prentice Hall, 1993.

[14] A. H. Sayed and T. Kailath “A state-space approach to adaptive RLS filtering,” *IEEE Signal Processing Magazine*, vol. 11, pp. 18–60, July 1994.

[15] J. Benesty, H. Rey, L. Rey Vega, and S. Tressens, “A non-parametric VSS NLMS algorithm,” *IEEE Signal Processing Lett.*, vol. 13, pp. 581–584, Oct. 2006.

[16] C. Paleologu, J. Benesty, S. Ciochină, and V. Popescu, “Robust general Kalman filter for echo cancellation,” in *Proc. EUSIPCO*, 2013, 5p.

[17] *Digital Network Echo Cancellers*, ITU-T Rec. G.168, 2002.