LOOK-AHEAD NEAR-END COMPENSATION IN ACOUSTIC ECHO CANCELLATION

Luis Weruaga

Department of Electrical and Computer Engineering
Khalifa University, Sharjah, United Arab Emirates
E-mail: weruaga@ieee.org

ABSTRACT

This paper presents a novel method for the compensation of near-end interferences in acoustic echo cancellation. The technique is derived from a rigorous analysis dealing with the statistical impact of the update over the "look-ahead" error. The outcome is a well-conditioned problem, whose numerical solution is shaped from maximum-likelihood principles. Results on a realistic simulated scenario show the excellent performance and robustness of the proposed method.

***Index Terms*—** Acoustic cancellation, near-end signal compensation, autoregressive model, maximum likelihood.

1. INTRODUCTION

Acoustic echo cancellation (AEC) in non-stationary noisy environments [1–3] is considered to be a challenging problem not fully solved. Due to the long nature of acoustic paths, the adaptation of the canceller coefficients in the frequency domain (FD) is preferred, not only because of its computational savings, but also for its faster speed of convergence. In addition to that, the orthogonality properties of the frequency-domain basis permits the independent adaptation of each spectral weight, therewith providing frequency-selective immunity against interferences and background noise that may be appear in the near end [4].

This important aspect, namely robustness against near-end signals, has been a problem thoroughly studied during the last decade [2–7]. A popular approach thereto is the use of double-talk detection (DTD) [2, 3], in order to literally stop the filter update when near-end activity is detected. A more elegant methodology though is the use of a frequency-tailored step size [4] to perform a sort of "soft" DTD on every spectral bin independently. Despite the existing consensus on the theoretical framework [8] of this second approach, estimating the terms involved in the optimal FD step size is extremely challenging, as that is actually an ill-posed problem. In the practice, the existing works [5–7] perform in real non-stationary acoustic scenarios far from optimally.

This paper discloses a novel estimation framework that proves the theoretical solution [4, 8] to be successful in the practice. The adoption of a "look-ahead" strategy as mechanism to obtain well-conditioned solutions is the main novel aspect of the method. Furthermore, the large number of parameters to estimate is substantially reduced by means of maximum-likelihood noise-compensated autoregressive estimation (ML-NCAR), a topic studied in depth by the author [9, 10]. This aspect is also key in preventing overfitted solutions. The paper contents follow: the optimal FD-AEC fundamentals are presented in Sec. 2, the analytical framework and the proposed numerical method are exposed in Sec. 3 and Sec. 4 respectively, the performance evaluation is included in Sec. 5, finally, the conclusions close the paper.

2. OPTIMAL FREQUENCY ADAPTIVE FILTER

A linear adaptive canceller consists of an L-tap transversal filter (FIR) with output given by

$$y(n) = \sum_{\ell=0}^{L-1} w_\ell x(n-\ell) \tag{1}$$

where $x(n)$ is the far-end signal, n denotes (discrete) time, and w_ℓ are the filter coefficients. The frequency-domain (FD) convolution is an efficient way to obtain the filter output

$$\mathbf{y}_m = \mathbf{q} \circ \left(\mathbf{F}^{-1}(\mathbf{W}_m \circ \mathbf{X}_m)\right) \tag{2}$$

where subscript m is the block index (hence related to time), the operation $\circ$ denotes element-by-element multiplication,[1] $\mathbf{F}$ is the N-point DFT (discrete Fourier transform) matrix, and the N-point column vectors $\mathbf{W}_m$ and $\mathbf{X}_m$ are built as follows

$$\mathbf{W}_m = \mathbf{F}\left[w_{m,0} \ \cdots \ w_{m,L-1} \ 0 \cdots 0\right]^T \tag{3}$$

$$\mathbf{X}_m = \mathbf{F}\left[x(mM-N+1) \ \cdots \ x(mM)\right]^T. \tag{4}$$

The operation (2) delivers in reality $M = N-L+1$ valid output samples of the transversal filter, hence the window vector is defined accordingly $\mathbf{q} = [\mathbf{0}_{N-M} \ \mathbf{1}_M]^T$. Without loss of generality, N is considered to be power of 2.

The error between the reference $z(n)$ and the canceller output $y(n)$ in a block is thus obtained as

$$\mathbf{e}_m = \left[0 \cdots 0 \ z((m-1)M+1) \ \cdots \ z(mM)\right]^T - \mathbf{y}_m. \tag{5}$$

[1]The Hadamard vector product $\circ$ is often replaced in the literature by the product of one vector with a diagonal matrix built with the other vector.

At the expense of an output delay equal to the block size M, obtaining the filter output as in (2) is computationally more efficient than the direct FIR evaluation (1). Moreover, updating the FD weight (3) explicitly is preferable (instead of w_ℓ) as it brings computational savings as well as attractive convergence benefits. This popular FD update corresponds to

$$\mathbf{W}_{m+1} = \mathbf{W}_m + \mathbf{F}\left(\mathbf{p} \circ \left(\mathbf{F}^{-1}(\boldsymbol{\mu}_m \circ \mathbf{U}_m)\right)\right) \quad (6)$$

where $\mathbf{U}_m$ is the estimated weight mismatch, $\boldsymbol{\mu}_m$ is the FD step size vector, and $\mathbf{p} = [\mathbf{1}_L \ \mathbf{0}_{N-L}]^T$ makes the overall weight update valid only in its first L time coefficients. The estimate of the weight mismatch $\mathbf{U}_n$ corresponds to

$$\mathbf{U}_m = \mathbf{R}_m^{-1}\left(\mathbf{X}_m^* \circ (\mathbf{F}\,\mathbf{e}_m)\right) \quad (7)$$

where * denotes complex conjugate, and $\mathbf{R}_n = \mathrm{E}\{\mathbf{X}_n\mathbf{X}_n^H\}$ is the autocorrelation matrix, which results to be diagonal. Here and in the sequel, $\mathrm{E}\{\cdot\}$ denotes statistical expectation, and ^{H} the hermitian operator (transpose and complex conjugate).

In the practice, the reference signal $z(n) = d(n) + v(n)$ is composed of the actual echo $d(n)$ and the near-end signal $v(n)$. The reference signal can be thus described spectrally as $Z_k = D_k + V_k$, where D_k and V_k corresponds to the spectral samples of $d(n)$ and $v(n)$ respectively, and subindex k is the frequency bin. However, the near-end samples V_k do not represent a valid reference in the update, hence corrupting the update (6) of the filter. An elegant way to overcome this drawback is namely to select an appropriate value for the step $\boldsymbol{\mu}_m$ at each block update: the optimal FD step was found in [4, 8] to be[2]

$$\mu_{m,k} = \frac{\mathrm{E}\left\{\left|D_{m,k} - Y_{m,k}\right|^2\right\}}{\mathrm{E}\left\{\left|V_{m,k}\right|^2\right\} + \mathrm{E}\left\{\left|D_{m,k} - Y_{m,k}\right|^2\right\}} \quad (8)$$

which in plain words represents the percentage of residual echo inside the error. The numerical evaluation of the optimal step (8) has been matter of study during the last decade [5–7]. This estimation task is a challenging problem as neither residual echo nor near-end samples are directly available.

3. FULL-WEIGHT UPDATE IMPACT

As the full-weight update yields minimum *a posteriori* square error [11], the canceller weights get corrupted with the near-end signal $v(n)$ present in the reference $z(n)$. Assessing objectively the quality of the update requires *testing* data statistically independent from the *training* data. A tentative choice is the data from the next $(m+1)$th block. Therefore, we consider the following three errors

$$E_k = Z_{m,k} - W_{m,k}X_{m,k} \quad (9a)$$

$$E'_k = Z_{m+1,k} - W_{m,k}X_{m+1,k} \quad (9b)$$

$$E''_k = Z_{m+1,k} - (W_{m,k} + U_{m,k})\,X_{m+1,k} \quad (9c)$$

[2]Another expression of the optimal frequency-domain step size has been recently proposed [7], but it does not differ in essence from (8).

where E'_k and E''_k correspond to the "look-ahead" errors resulting from the weight update for two extreme cases, namely, a frozen ($\boldsymbol{\mu}_m = \mathbf{0}_N$) and a full ($\boldsymbol{\mu}_m = \mathbf{1}_N$) update. In order to consider valid the previous statement (9), the look-ahead error E''_k must be obtained with the windowless update

$$\mathbf{W}_{m+1} = \mathbf{W}_m + \boldsymbol{\mu}_m \circ \mathbf{U}_m \quad (10)$$

which converges, although somewhat slower than (6) because the number of parameters nearly doubles ($N \approx 2L$). This way of proceeding, namely, excluding the $\mathbf{p}$-windowing operation in the update, prevents the characteristic and undesired spectral distortion of a rectangular window operation.[3]

The aim of this analysis is to obtain a meaningful expression of the expected power spectrum of each one of the errors (9). Let $G_{m,k} = H_k - W_{m,k}$ be the canceller misalignment, where H_k is the frequency response of the acoustic echo path, considered constant during consecutive segments. Given that the segment length N is usually large, the signals from adjacent segments can be thus considered uncorrelated; the power spectral density (PSD) of error E_k and look-ahead error E'_k result thus after simplifications in

$$\mathcal{E}_k = \mathcal{V}_{m,k} + \mathcal{G}_{m,k}\mathcal{X}_{m,k} \quad (11a)$$

$$\mathcal{E}'_k = \mathcal{V}_{m+1,k} + \mathcal{G}_{m,k}\mathcal{X}_{m+1,k} \quad (11b)$$

where $\mathcal{E} = \mathrm{E}\{|E|^2\}$, $\mathcal{V} = \mathrm{E}\{|V|^2\}$, $\mathcal{X} = \mathrm{E}\{|X|^2\}$, and $\mathcal{G} = \mathrm{E}\{|G|^2\}$. On the other hand, the PSD of the look-ahead error E''_k happens to result in

$$\mathcal{E}''_k = \mathcal{V}_{m+1,k} + e^{-\frac{1}{2}}\mathcal{G}_{m,k}\mathcal{X}_{m+1,k} + \frac{\mathcal{X}_{m+1,k}}{\mathcal{X}_{m,k}}\mathcal{V}_{m,k} \quad (12)$$

where the last term in (12) corresponds to the weight corruption caused by the presence of near-end activity. On the positive side, the canceller mismatch must experience a decrease by a factor of $\exp(-1/2)$ in accordance to the following argument: the fastest convergence of the windowless update (10) in a noiseless situation (for $\mathcal{V}_{m,k} = 0$) follows the rule

$$\mathcal{G}_{m+1,k} = \left(1 - \frac{1}{N}\right)^M \mathcal{G}_{m,k} \simeq e^{-\frac{1}{2}}\mathcal{G}_{m,k} \quad (13)$$

where the last simplification results from the asymptotic limit for large M, and given that $M \simeq N/2$.

4. ESTIMATION OF THE OPTIMAL STEP

We can rewrite equations (11) and (12) in the following system of linear equations subject to inequality constraints

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & \rho_k \\ \rho_k & 1 & 0.6\rho_k \end{bmatrix} \begin{bmatrix} \mathcal{V}_{m,k} \\ \mathcal{V}_{m+1,k} \\ \mathcal{X}_{m,k}\mathcal{G}_{m,k} \end{bmatrix} = \begin{bmatrix} \mathcal{E}_k \\ \mathcal{E}'_k \\ \mathcal{E}''_k \end{bmatrix} \quad (14)$$

subject to $\mathcal{V}_{m,k}, \mathcal{V}_{m+1,k}, \mathcal{G}_{m,k} \geqslant 0$

[3]In consequence, we do not need to make use of the following approximation used in the literature [4–7], $\mathbf{F}(\mathbf{p} \circ (\mathbf{F}^{-1})) \approx (1/2)\mathbf{I}$.

where

$$\rho_k = \frac{\mathcal{X}_{m+1,k}}{\mathcal{X}_{m,k}}. \tag{15}$$

The condition number of the system matrix in (14) results in $\kappa(\rho_k) \approx \rho_k + \rho_k^{-1} + 3/2$. This condition number is minimum at $\rho_k = 1$, that is, for $\mathcal{X}_{m+1,k} = \mathcal{X}_{m,k}$.[4]

4.1. Reducing the Parametric Space

Instead of considering $\mathcal{V}_{m,k}$, $\mathcal{V}_{m+1,k}$ and $\mathcal{G}_{m,k}$ parameters of the solution equation (14), we assume them to follow a parametric autoregressive (AR) model. We thus define the AR transfer function

$$A(e^{j\omega}) = \sum_{\ell=0}^{P} \alpha_\ell \, e^{-j\omega\ell} \tag{16}$$

where α_ℓ are the AR coefficients and P is the AR order, to describe the near-end signal at the mth segment as $\mathcal{V}_{m,k} \equiv 1/|A(w_N^k)|^2$, where $w_N^k = \exp(jk2\pi/N)$. Note that we will use interchangeably A_k to denote $A(w_N^k)$. The realization of the near-end term $\mathcal{V}_{m,k}$ in error E_k and look-ahead error E_k'' are stochastically independent, hence the estimation of the AR model (16) from each error yields (slightly) different results: we will thus define the AR model $A''(e^{j\omega})$ of parameters α_ℓ'' to account for the near-end signal observed in E_k''; obviously, there is no need for an $A'(e^{j\omega})$.

Accordingly, we define the AR models $B(e^{j\omega})$, $B'(e^{j\omega})$, $B''(e^{j\omega})$ of order Q and parameters β_ℓ, β_ℓ', β_ℓ'' to model the canceller weight mismatch $\mathcal{G}_{m,k}$ in each error respectively. Finally, we define only one P-order AR model $C(e^{j\omega})$ to model $\mathcal{V}_{m+1,k}$ in both E_k' and E_k''. The use of an AR model is justified because:

1. many audio signals, and in especial speech, can be physically described with an autoregressive model,
2. since $P, Q \lll N$, the number of unknowns in the problem reduces substantially, and
3. as the AR power spectrum is always positive, the constraints in (14) are implicitly met.

4.2. Maximum-Likelihood Estimation

We may take error E_k to be a statistically-independent complex (bivariate) Gaussian variable of zero mean and variance $\mathcal{E}_k$; hence its probability density function (PDF) is

$$p(E_k) = \frac{1}{\pi\mathcal{E}_k} \exp\!\big(-|E_k|^2/\mathcal{E}_k\big)\, p(\boldsymbol{\alpha})\, p(\boldsymbol{\beta}) \tag{17}$$

where $\boldsymbol{\alpha} = \{\alpha_0, \cdots, \alpha_P\}$, $\boldsymbol{\beta} = \{\beta_0, \cdots, \beta_P\}$, and $p(\boldsymbol{\alpha})$ and $p(\boldsymbol{\beta})$ are prior probabilities. The PDF $p(\boldsymbol{\alpha})$ is defined as

$$p(\boldsymbol{\alpha}) = \tfrac{|\boldsymbol{J}_\alpha|^{1/2}}{(2\pi)^{P/2}} \exp\!\big(-\tfrac{1}{2}(\boldsymbol{\alpha}-\boldsymbol{\alpha}_o)\boldsymbol{J}_\alpha(\boldsymbol{\alpha}-\boldsymbol{\alpha}_o)^H\big) \tag{18}$$

[4] The system becomes ill-conditioned in the trivial case of $\mathcal{X}_{m,k} \simeq 0$ or $\mathcal{X}_{m+1,k} \simeq 0$, that is, with the lack of training or testing data respectively.

where $\boldsymbol{\alpha}_o$ represents the actual (unknown) AR parameters, set as the average $\boldsymbol{\alpha}_o = (\boldsymbol{\alpha} + \boldsymbol{\alpha}'')/2$, and $\boldsymbol{J}_\alpha$ is the Fisher information matrix (FIM) [10]. According to (14)

$$\mathcal{E}_k = \frac{1}{|A_k|^2} + \frac{\mathcal{X}_{m,k}}{|B_k|^2}. \tag{19}$$

Likewise we can write

$$p(E_k'') = \frac{1}{\pi\mathcal{E}_k''} \exp\!\big(-|E_k''|^2/\mathcal{E}_k''\big)\, p(\boldsymbol{\alpha}'')\, p(\boldsymbol{\beta}'') \tag{20}$$

where $\mathcal{E}_k''$ can be deduced from (14) as

$$\mathcal{E}_k'' = \frac{\mathcal{X}_{m+1,k}}{\mathcal{X}_{m,k}} \frac{1}{|A_k''|^2} + \frac{0.6\mathcal{X}_{m+1,k}}{|B_k''|^2} + \frac{1}{|C_k|^2}. \tag{21}$$

In consequence, the original problem (14) is revamped as the maximization (of the logarithm) of the likelihood with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}''$, that is, to solve the following problem

$$\max_{\boldsymbol{\alpha},\boldsymbol{\alpha}''} \sum_{k=0}^{N-1} \log p(E_k) + \log p(E_k''). \tag{22}$$

As E_k and E_k'' are probabilistically independent, the joint likelihood results in the product of each individual likelihood, hence (22). Problem (22) corresponds to a noise-compensated AR analysis (NCAR). The ML solution to the NCAR problem has been studied recently in [9] by the author, proposing a method capable of nearly attaining the lower estimation variance [10].

The null of the gradient of the functional with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}''$ yields the following quasi-Newton method

$$\boldsymbol{\alpha}_o^{(\xi)} = \tfrac{1}{2}\big(\boldsymbol{\alpha}^{(\xi)} + \boldsymbol{\alpha}''^{(\xi)}\big) \tag{23a}$$

$$\boldsymbol{\alpha}^{(\xi+1)} \left(\mathbf{H}_\alpha^{(\xi)} + \lambda \boldsymbol{J}_\alpha^{(\xi)}\right) = \mathbf{g}_\alpha^{(\xi)} + \lambda \boldsymbol{\alpha}_o^{(\xi)} \boldsymbol{J}_\alpha^{(\xi)} \tag{23b}$$

$$\boldsymbol{\alpha}''^{(\xi+1)} \left(\mathbf{H}_\alpha''^{(\xi)} + \lambda \boldsymbol{J}_\alpha^{(\xi)}\right) = \mathbf{g}_\alpha''^{(\xi)} + \lambda \boldsymbol{\alpha}_o^{(\xi)} \boldsymbol{J}_\alpha^{(\xi)} \tag{23c}$$

where λ is the tradeoff regularization hyper-parameter, and superscript ξ denotes iteration. The approximate Hessian Toeplitz matrix $\mathbf{H}_\alpha$ is built in its ith diagonal as

$$h_{\alpha,i}^{(\xi)} = \textstyle\sum_{k=0}^{N-1} \big(\psi_k^{(\xi)}\big)^2 |E_k|^2 w_N^{ki} \tag{24}$$

while the ith component of vector $\mathbf{g}_\alpha$ is given by

$$g_{\alpha,i}^{(\xi)} = \sum_{k=0}^{N-1} \big(\psi_k^{(\xi)}\big)^2 \left(\frac{1}{|A_k^{(\xi)}|^2} + \frac{\mathcal{X}_{m,k}}{|B_k^{(\xi)}|^2}\right) A_k^{(\xi)} w_N^{ki}. \tag{25}$$

Here $\psi_k^{(\xi)}$ is the spectral weight, built as

$$\psi_k^{(\xi)} = \frac{\big|B_k^{(\xi)}\big|^2}{\big|B_k^{(\xi)}\big|^2 + \mathcal{X}_{m,k}\big|A_k^{(\xi)}\big|^2} \tag{26}$$

Matrix $\mathbf{H}_\alpha''$ and $\mathbf{g}_\alpha''$ are obtained accordingly. Note that (24) and (25) correspond to an inverse DFT, and they can be thus evaluated efficiently. Deriving the likelihood terms required to estimate $V_{m+1,k}$ and $G_{m,k}$ is not difficult, and due to space constraints it is thus left as exercise. The solution upon convergence is used to evaluate directly the optimal weight (8).

5. SIMULATION RESULTS

The proposed method for near-end-compensated echo cancellation was validated on realistic simulated scenarios. The speech signals used in the experiments, sampled at 16 kHz, belong to a private database.

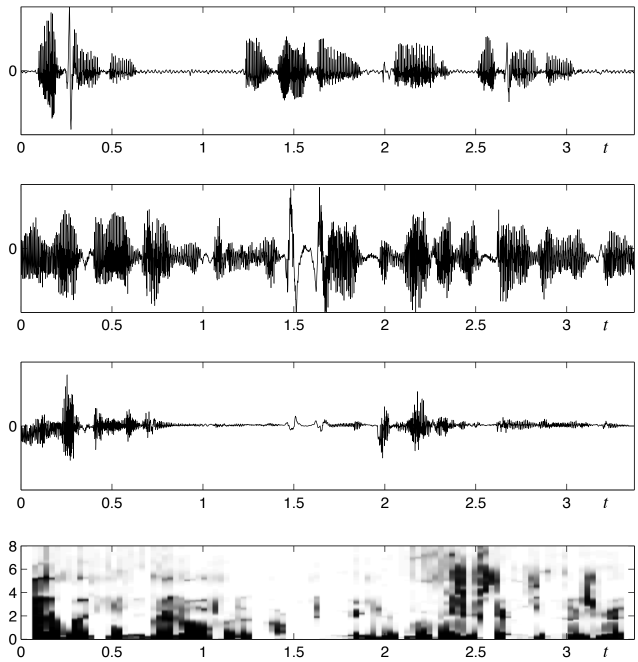

Fig. 1. From top to bottom: near-end signal, echo signal, residual echo (proposed method), and resulting FD step size (darker is larger). Frequency axis in kHz, time in seconds.

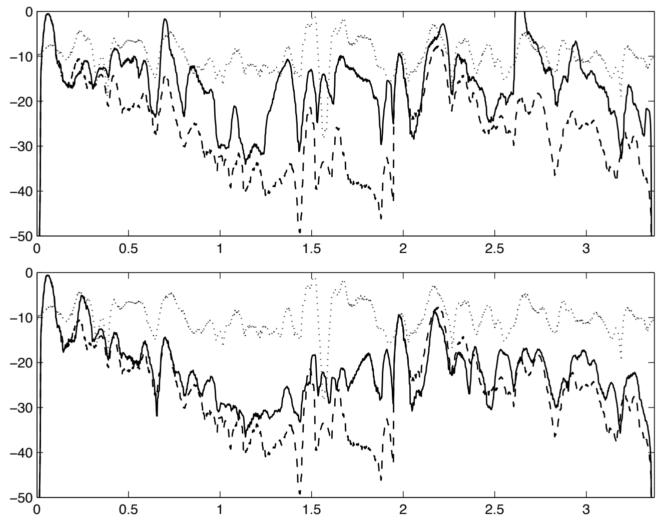

Fig. 2. Performance of [7] (top) and proposed method (bottom). Energy level (in dB) of echo (dotted), residual echo in *genius* method (dashed), and that of each method (solid).

The impulse response of the acoustic path was simulated according to a room of moderate dimensions (3 meters in each dimension), aimed to resemble a car chamber;[5] the length in samples of the response is $L = 511$, hence $N = 1024$. Fig. 1 depicts the near-end signal (which acts as disturbance in the system identification) and the echo signal. In order to test the tracking performance, the echo path changes abruptly at time $t = 2$. Three methods were tested in this scenario: the *genius* method (the optimal step (8) is obtained with the actual weight mismatch and near-end signal), the method presented recently in [7] (the weight mismatch and near-end information is obtained by averaging the instantaneous spectra of previous blocks), and our proposed method ($P = 14$, $Q = 4$). The methods count with the same (all zero) initialization.

Fig. 2 shows the average performance over 100 simulations. Each simulation has a different acoustic path (position of microphone and loudspeaker were randomly positioned inside the virtual acoustic chamber). The competitive method [7] was implicitly designed for stationary environments. In this realistic scenario, however, that solution yields poor estimation of the optimal weight, suffering from the presence of the near-end signal (note the deterioration with near-end signal activity). On the contrary, our proposed method achieves low residual echo level, as the weights are softly updated based on the accurate detection of near-end spectral activity. Its tracking performance is also excellent. This performance exhibited in all other simulation experiments we conducted.

Fig. 1 further shows the residual echo as well as the FD step size resulting from the proposed method. The residual echo manifests steady decrease in amplitude, only becoming significant at time $t = 2$ due to the abrupt echo path change. The method's ability to detect near-end activity is remarkable as shown by the white areas in the FD step size, which faithfully match the presence of the near-end signal. In summary, the proposed method can be seen as a frequency-dependent soft DTD. Due to time and space constraints comparison against the state of art in hard DTD [2, 3] is planned for the near future.

6. CONCLUSIONS

The estimation of near-end and residual echo components is indispensable as much as challenging a task in acoustic echo cancellation. Such an ill-posed problem (estimating two terms from a single reference) is approached in this paper with a "look-ahead" strategy that results in a well-conditioned scenario involving an extra term to estimate (from three valid references). Furthermore, the use of autoregressive models helps reduce the large number of parameters in the problem, hence improving robustness against overfitting. Results on realistic simulated scenarios prove the validity of this promising novel strategy.

[5] From empirical analysis, the acoustic response (even with random ones) has no notorious effect in the performance of the method.

7. REFERENCES

[1] G. Enzner and P. Vary, "Frequency-domain adaptive Kalman filter for acoustic echo control in hands-free telephones," *Signal Processing*, vol. 86, no. 6, pp. 1140–56, Jun. 2006.

[2] T. Gänsler and J. Benesty, "The fast normalized cross-correlation double-talk detector," *Signal Processing*, vol. 86, no. 6, pp. 1124–1139, Jun. 2006.

[3] T. van Waterschoot, G. Rombouts, P. Verhoeve, and M. Moonen, "Double-talk-robust prediction error identification algorithms for acoustic echo cancellation," *IEEE Trans. Signal Process.*, vol. 55, no. 3, pp. 846–858, Mar. 2007.

[4] B. H. Nitsch, "A frequency-selective stepfactor control for an adaptive filter algorithm working in the frequency domain," *Signal Process.*, vol. 80, no. 9, pp. 1733–1745, Sep. 2000.

[5] G. Enzner and P. Vary, "Robust and elegant, purely statistical adaptation of acoustic echo canceller and post-filter," *Proc. Intl. Wksp. Acoust. Echo Noise Control (IWAENC)* 2003, pp. 43–46.

[6] Y. Zhou and X. Li, "A variable step-size for frequency-domain acoustic echo cancellation," *Proc. IEEE Wksp. Appl. Signal Process. Audio Acoust. (WASPAA)* 2007, pp. 21–24.

[7] K. Shi and X. Ma, "A frequency domain step-size control method for LMS algorithms," *IEEE Signal Process. Lett.*, vol. 17, no. 2, pp. 125–128, Feb. 2010.

[8] P. Vary and R. Martin, *Digital speech transmission: Enhancement, coding and error concealment*. John Wiley & Sons, 2006.

[9] L. Weruaga and L. Dimitrov, "The spectral nature of maximum-likelihood noise-compensated autoregressive analysis," *IEEE Trans. Audio, Speech, Language Process.*, vol. 21, no. 8, pp. 1760–1765, Aug. 2013.

[10] L. Weruaga and M. Melko, "Asymptotic Cramér–Rao bound for noise-compensated autoregressive estimation," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 59, no. 9, pp. 2017–2024, Sep. 2012.

[11] J. Benesty, H. Rey, L. Rey-Vega, and S. Tressens, "A nonparametric VSS NLMS algorithm," *IEEE Signal Process. Lett.*, vol. 13, no. 10, pp. 581–584, Oct. 2006.