SPECTRAL DOMAIN B-SPLINE IDENTIFICATION IN ACOUSTIC ECHO CANCELLATION

Y. Zakharov and T. Tozer

Department of Electronics, University of York, York YO10 5DD, UK, yz1,tct@ohm.york.ac.uk

ABSTRACT

Spectral domain B-spline identification is proposed for acoustic echo cancellation. Two approaches are considered. The first is based on solution of normal equations; we describe an efficient technique for such a solution, which benefits from the sparseness of the system matrix due to B-splines. The second approach is based on using local splines, enabling further simplification. We also show how the proposed techniques can be used for efficient double-talk detection. The echo cancellation performance and complexity of the proposed techniques are compared with that of a low-complexity cross-spectral technique and the affine projection (AP) algorithm possessing high cancellation performance. The B-spline identification allows cancellation performance comparable with that of the AP algorithm and complexity close to that of the cross-spectral algorithm.

1. INTRODUCTION

Acoustic echo cancellation is based on identification of the acoustic impulse or frequency response, modeling the echo, and subtracting the echo model from an input (microphone) signal. A variety of methods have been proposed for solution of the problem in both time and spectral domains. Most often, time-domain FIR filtering, which allows a low processing delay, is used to model the echo. For estimation of FIR filter taps, the classical NLMS algorithm is often used in practice, providing low complexity, but also possessing low convergence speed. A faster convergence is achieved by the affine projection (AP) algorithm [1]; however, this is complex for implementation. The Fast AP algorithm has been proposed [2], but it demonstrates numerical instability and it is sensitive to noise. Good cancellation performance is achieved by using the least squares (LS) block approach [3]. However, even with the Toeplitz approximation of normal equations in the LS problem and use of the computationally efficient Levinson algorithm, the complexity of the echo canceler is still high.

Spectral-domain echo cancelers are capable of reducing the computational load, but they lead to a high processing delay [4]. To achieve a low delay and low complexity, time-domain FIR filtering and spectral-domain block identification are used in the technique proposed in [5]. However, this technique demonstrates instability and its performance significantly degrades in the presence of noise and double-talk.

In this paper, we adopt the approach from [5]. However, we propose B-spline identification of the acoustic frequency response; specifically, we use cubic B-splines. Two approaches for solving the LS problem are considered. The first is based on solution of normal equations. We show how the recently introduced dichotomous coordinate descent (DCD) algorithm [6] can be used for this purpose, by benefiting from the sparseness of the system matrix due to B-splines, as well as multiplication-free iterations. The second approach is based on using local splines, enabling further simplification, but resulting in somewhat higher approximation error. We also show how the spectral domain identification can be used for efficient double-talk detection. The performance and complexity of the proposed techniques are compared by simulation with that of the cross-spectral technique from [5] and the AP algorithm.

2. SPECTRAL-DOMAIN SPLINE-IDENTIFICATION

In application to acoustic echo cancellation, the identification problem can be described as follows. The microphone signal is

$$y(t) = u(t) + z(t) \tag{1}$$

where t is discrete time, $u(t) = \sum_{\tau=0}^{L-1} x(t-\tau)h(\tau)$ is the echo signal, $z(t)$ is the near-end signal and/or white Gaussian noise, $x(t)$ is the excitation (far-end) signal, $h(\tau)$ is the acoustic impulse response to be estimated, and L is the length of $h(\tau)$. In the spectral domain, this can be represented as

$$Y(\omega_k) = X(\omega_k)H(\omega_k) + Z(\omega_k) \tag{2}$$

where $X(\omega_k)$ is the spectrum of the excitation signal $x(t)$, $Z(\omega_k)$ is the spectrum of the noise and near-end signal $z(t)$, and $\omega_k \in \Omega$ form a frequency grid in the frequency bandwidth of interest $\Omega = [\omega_l, \omega_u]$. The spectra of microphone and excitation signals are calculated over a block of N samples by using FFT as

$$Y(\omega_k) = \frac{1}{N}\sum_{i=0}^{N-1} w(i)y(i+t-N+1)e^{-j\frac{2\pi ki}{N}}, \tag{3}$$

$$X(\omega_k) = \frac{1}{N}\sum_{i=0}^{N-1} w(i)x(i+t-N+1)e^{-j\frac{2\pi ki}{N}} \tag{4}$$

where $w(i)$ is a window (e.g., the Hamming window). Since $y(t)$ and $x(t)$ are real-valued, we are only interested in the first $N/2$ frequency bins of the FFTs, $k = 0, \ldots, N/2-1$.

The frequency response $H(\omega_k)$ is approximated by a series

$$\hat{H}(\omega_k) = \sum_{p=1}^{N_\varphi} c_p \varphi_p(\omega_k) \tag{5}$$

where $\{\varphi_p(\omega_k)\}$ are N_φ basis functions. Minimisation of the error

$$\varepsilon^2 = \sum_{\omega_k \in \Omega} \left| Y(\omega_k) - X(\omega_k)\hat{H}(\omega_k) \right|^2 \tag{6}$$

results in expansion coefficients $\mathbf{c} = [c_1, \ldots, c_{N_\varphi}]^T$ being the solution of normal equations

$$\mathbf{R}\mathbf{c} = \boldsymbol{\xi} \tag{7}$$

where the vector $\boldsymbol{\xi}$ contains elements

$$\xi_q = \sum_{\omega_k \in \Omega} S_{YX}(\omega_k)\varphi_q(\omega_k), \;\; q = 1, \ldots, N_\varphi \tag{8}$$

the matrix $\mathbf{R}$ contains elements

$$r_{qp} = \sum_{\omega_k \in \Omega} S_{XX}(\omega_k)\varphi_q(\omega_k)\varphi_p^*(\omega_k), \;\; q, p = 1, \ldots, N_\varphi, \tag{9}$$

$S_{YX} = Y(\omega_k)X^*(\omega_k)$ and $S_{XX} = X(\omega_k)X^*(\omega_k)$ are respectively the cross-spectrum and auto-spectrum, and $(\cdot)^*$ denotes complex conjugate. To improve the convergence when solving the system (7), the matrix $\mathbf{R}$ is regularised as $\mathbf{R} \Rightarrow \mathbf{R} + \delta\mathbf{I}$, where $\delta > 0$ is a regularisation parameter and $\mathbf{I}$ is the identity matrix.

If basis functions are complex harmonics, the echo path is modeled as a FIR filter with N_φ filter taps, $\mathbf{R}$ is the auto-correlation matrix of the excitation signal, and $\boldsymbol{\xi}$ is the cross-correlation vector of the excitation and microphone signals. This case corresponds to the block LS approach adopted in [3]. Unfortunately, for harmonic basis functions, the matrix $\mathbf{R}$ in general is not sparse, which makes solving the system (7) with large N_φ a complicated problem.

We propose polynomial spline-approximation. Splines provide a low approximation error with a low degree of the polynomial; cubic splines are considered to provide the best trade off between accuracy and complexity [7, 8, 9]. Among spline basis functions, B-splines possess features attractive for implementation [7, 8], e.g., they have the minimum support which leads to simple calculation of the sparse matrix $\mathbf{R}$ and vector $\boldsymbol{\xi}$, and simplify (due to the sparseness) the solution of equations (7). Below we use cubic B-splines

$$\varphi_p(\omega_k) = b_3(\omega - \omega_l - (p-2)\Delta\omega) \tag{10}$$

where $\Delta\omega = (\omega_u - \omega_l)/(N_\varphi - 3)$ and

$$b_3(\omega) = \begin{cases} \frac{1}{6}\left(2 - \frac{|\omega|}{\Delta\omega}\right)^3 - \frac{2}{3}\left(1 - \frac{|\omega|}{\Delta\omega}\right)^3, & 0 \le |\omega| < \Delta\omega \\ \frac{1}{6}\left(2 - \frac{|\omega|}{\Delta\omega}\right)^3, & \Delta\omega \le |\omega| < 2\Delta\omega \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

It is convenient to choose $\Delta\omega$ as a multiple of the FFT bin: $\Delta\omega = (2\pi/N)D$, where D is an integer. Then samples of basis functions are discrete shifts of $4D$ samples of $b_3(\omega)$ by D FFT frequency bins. Fig.1 shows a few basis B-splines in a part of the frequency range Ω in the case of $N = 8192$, $D = 7$, and the sampling frequency $F_s = 8$ kHz; circles indicate B-spline values at the FFT frequencies.

The block estimate of the impulse response is obtained by the inverse Fourier transform and truncation:

$$\hat{h}_b(\tau) = \frac{1}{N}\sum_{k=0}^{N-1} \hat{H}(\omega_k)e^{j\omega_k\tau}, \;\; \tau = 0, 1, \ldots, L-1. \tag{12}$$

The final estimate of the room impulse response $h(\tau)$ is updated as

$$\hat{h}_t(\tau) = (1-\alpha_t)\hat{h}_{t-M}(\tau) + \alpha_t\hat{h}_b(\tau), \;\; \tau = 0, \ldots, L-1, \tag{13}$$

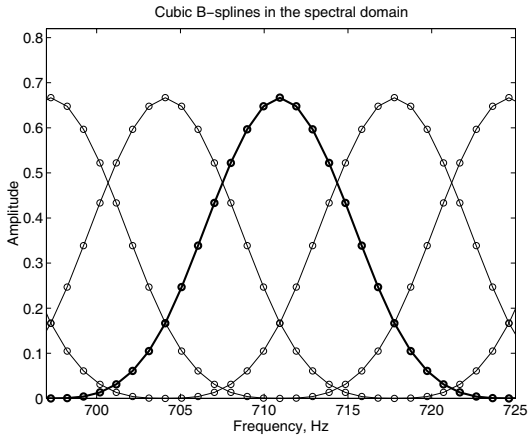

Fig. 1. *Basis B-splines;* $D = 7$, $N = 8192$, $F_s = 8$ *kHz.*

where M is the number of samples between two blocks and α_t is a time-varying forgetting factor determined as described in section 5. The estimate $\hat{h}_t(\tau)$ is used for FIR filtering to model the echo; it remains constant over M samples until the next block processing.

3. DCD ALGORITHM

A significant computational problem with such identification is in the solution of the system of equations (7). The use of conventional techniques like the Cholesky decomposition leads to high complexity (see section 7).

We use the recently introduced DCD algorithm [6]. It is based on binary representation of elements of the solution vector with M_b bits within an amplitude range $[-H, H]$. It starts an iterative approximation of the solution vector $\mathbf{c}$ from the most significant bit. Once the most significant bit has been found for all vector elements, the algorithm starts updating the next less significant bit, and so on. If an update happens (such an iteration is called 'successful'), the vector $\boldsymbol{\xi}$ is also updated. Parameters of the DCD algorithm are the number of bits M_b representing elements of the vector $\mathbf{c}$ within the amplitude range $[-H, H]$ and the maximum number of iterations N_{it}.

As $\mathbf{R}$ is real, the system can be solved separately for real $\boldsymbol{\xi}^{(r)}$ and imaginary $\boldsymbol{\xi}^{(i)}$ parts of the vector $\boldsymbol{\xi} = \boldsymbol{\xi}^{(r)} + j\boldsymbol{\xi}^{(i)}$ as $\mathbf{R}\mathbf{c}^{(r)} = \boldsymbol{\xi}^{(r)}$ and $\mathbf{R}\mathbf{c}^{(i)} = \boldsymbol{\xi}^{(i)}$, respectively. Then the solution is represented as $\mathbf{c} = \mathbf{c}^{(r)} + j\mathbf{c}^{(i)}$.

In application to the cubic B-spline identification, the DCD algorithm can be described as follows (for the real part of the system).

Initialization: $\mathbf{c}^{(r)} = \mathbf{0}$, the step-size $d = H$.

For every m^{th} bit, $m = 1, \ldots, M_b$, the step-size is reduced as $d = d/2$, the iteration counter $it = 0$, and iterations start:

(1) Indicator of 'successful' iterations is reset, $Flag = 0$, and the iteration counter is incremented, $it = it + 1$;

(2) For $p = 0, \ldots, N_\varphi - 1$: if $|\xi_p^{(r)}| > (d/2)r_{pp}$, then the iteration is 'successful', $Flag = 1$, $c_p^{(r)} = c_p^{(r)} + d$, elements of the vector $\boldsymbol{\xi}$ with indexes $q \in [max(1, p-n), min(N_\varphi, p+n)]$ are updated as $\xi_q^{(r)} = \xi_q^{(r)} - sgn(\xi_p^{(r)})dr_{pq}$.

(3) If $it = N_{it}$, then the algorithm stops.

(4) If $Flag = 1$, then steps (1), (2), and (3) repeat; otherwise iterations start for the next less significant bit ($m = m + 1$) with a reduced step-size ($d = d/2$).

The DCD algorithm guarantees convergence to the true solution if elements of the true solution vector $\mathbf{c}$ are within the interval $[-H, H]$. If H is a power of two, then multiplications by factors of power of two are only used; these can be replaced by bit shifts. Thus, the DCD algorithm can be implemented without explicit multiplications, which may be useful in hardware implementation.

4. LOCAL SPLINE-APPROXIMATION

Although the DCD algorithm allows efficient solution of normal equations, it may still require a considerable computational load. The use of local splines allows us to avoid such calculations with a slightly higher approximation error [9]. The spline coefficients c_n can be calculated as follows

$$c_q = a_0\xi_q + a_1(\xi_{q-1} + \xi_{q+1}) + a_2(\xi_{q-2} + \xi_{q+2}), \quad (14)$$

$$\xi_q = \frac{\sum_{\omega_k \in \Omega_q} S_{YX}(\omega_k)}{\sum_{\omega_k \in \Omega_q} S_{XX}(\omega_k)}, \quad (15)$$

where $\Omega_q = [\omega_l + \Delta\omega(q-2) - \Delta\omega/2, \omega_l + \Delta\omega(q-2) + \Delta\omega/2]$. The weights should satisfy the condition $a_0 + 2a_1 + 2a_2 = 1$ and they are tuned to obtain the best cancellation performance. In our simulation, the weights are $a_0 = 1.94$, $a_1 = -0.58$, $a_2 = 0.11$.

5. DOUBLE-TALK DETECTION

Relationship between the error ε^2 in (6) and the energy of the microphone signal $E_y = \sum_{\omega_k \in \Omega} |Y(\omega_k)|^2$ characterizes accuracy of the identification. If the accuracy is low, i.e. ε^2 is close to E_y (e.g. in double-talk situation), then the block impulse response estimate is ignored by setting $\alpha_t = 0$. If ε^2 is much smaller than E_y, we can update the impulse response estimate by adding the block estimate with the weight α_t close to 1. This can be considered as a spectral domain implementation of the "two-path" approach [4]. Note that main computations for calculation of ε^2 and E_y have already been done and implementation of this approach requires a small extra computational load. In the simulation below, we use the following mapping of the relationship between ε^2 and E_y to the forgetting factor α_t:

$$\alpha_t = \begin{cases} \alpha(1), & \varepsilon^2 < \rho(1)E_y \\ \alpha(2), & \rho(1)E_y \le \varepsilon^2 < \rho(2)E_y \\ \alpha(3), & \rho(2)E_y \le \varepsilon^2 < \rho(3)E_y \\ 0, & \varepsilon^2 \ge \rho(3)E_y. \end{cases} \quad (16)$$

The vectors $\boldsymbol{\alpha} = [\alpha(1),\ \alpha(2),\ \alpha(3)]$ and $\boldsymbol{\rho} = [\rho(1),\ \rho(2),\ \rho(3)]$ are chosen to obtain the best cancellation performance. Other dependencies α_t on ε^2 and E_y can also be used. We have found the mapping (16) efficient and simple for implementation.

6. NUMERICAL RESULTS

We simulate acoustic echo cancellation in the following scenarios. The acoustic impulse response $\mathbf{h} = [h(0), \ldots, h(L-1)]$ has a length $L = 512$. The excitation signal is 11-sec female speech sampled at $F_s = 8$ kHz with a 16-bit resolution. Experimental plots have been obtained by averaging the misalignment

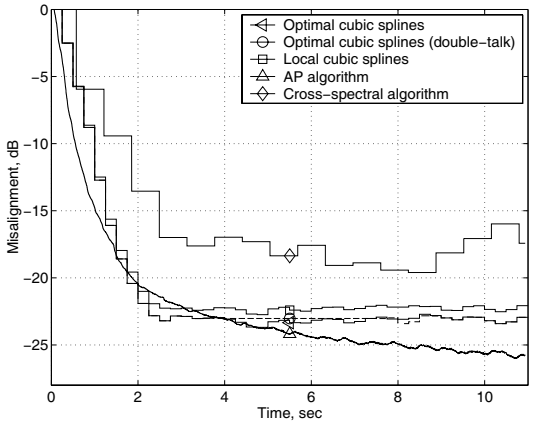

Fig. 2. *Identification performance; SNR=30dB*

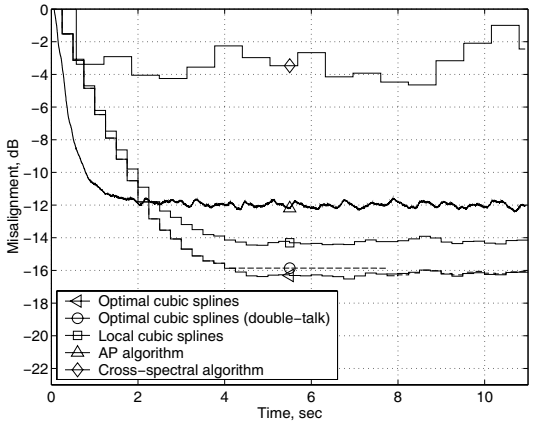

Fig. 3. *Identification performance; SNR=15dB*

$||\mathbf{h} - \hat{\mathbf{h}}_t||^2/||\mathbf{h}||^2$ in 20 trials. In each trial, new excitation speech and noise signals are used. Echo attenuation (ERLE) is calculated over intervals between 2nd and 11th seconds and averaged over the 20 trials. In double-talk scenarios, the near-end speech is applied between 4th and 7th seconds with a power equal to that of the echo signal.

We compare: (1) optimal splines with ideal matrix inversion; (2) optimal splines with the DCD algorithm; (3) local splines; (4) the AP algorithm; and (5) the cross-spectral algorithm. Parameters of the spline identification are: $N = 8192$, $M = 2000$, $D = 7$, $N_\varphi = 585$, $\delta = 0.5$, $\boldsymbol{\alpha} = [0.4,\ 0.1,\ 0.05]$, $\boldsymbol{\rho} = [0.015,\ 0.1,\ 0.25]$, $\omega_l = 0$, $\omega_u = \pi F_s$. Parameters of the DCD algorithm are: $H = 1$, $M_b = 8$, $N_{it} = 20$. In the AP algorithm, the AP order is $N_{AP} = 8$, the step-size is 0.125, and the regularisation parameter $\delta = 10^8$. In the cross-spectral algorithm, the average of cross-spectrum and auto-spectrum is performed over intervals of 0.64 sec as in [5].

Fig.2 shows misalignment for scenarios with noise 30 dB down from the echo ($SNR = 30$ dB). The optimal splines with ideal matrix inversion and with the DCD algorithm demostrate equal performance, therefore we show one plot only. Local splines provide a slightly higher (by about 1 dB) steady-state misalignment than the optimal splines. The convergence speed of spline algo-

Table 1. Echo attenuation in noisy environments

Algorithm	SNR=15dB	SNR=30dB
Opt. splines	20.8dB	29.7dB
Opt. splines-DCD	20.8dB	29.7dB
Local splines	19.9dB	28.7dB
AP algorithm	19.1dB	33.8dB
Cross-spectral	11.5dB	22.0dB

Table 2. Echo attenuation in noisy and double-talk environments

Algorithm	SNR=15dB	SNR=30dB
Opt. splines	20.1dB	28.4dB
Opt. splines-DCD	20.1dB	28.4dB
Local splines	19.2dB	27.6dB

rithms is close to that of the AP algorithm. The cross-spectral algorithm has a poorer performance. For $SNR = 15$ dB (Fig.3), optimal splines provide about 2 dB, 4 dB, and 12 dB improvement of the steady-state misalignment over local splines, the AP algorithm, and the cross-spectral algorithm, respectively. The convergence of spline algorithms is slower than that of the AP algorithm. This is due to a smaller forgetting factor α_t (16) as the error ε^2 in the higher noise is increased; the convergence speed can be increased by increasing $\alpha(2)$, but at the expense of the steady-state performance. The echo attenuation (ERLE) performance for noisy scenarios is shown in Table 1. For double-talk and noisy scenarios, the echo attenuation is shown in Table 2. It can be seen that the double-talk does not affect significantly the performance of the proposed echo cancelers, which demonstrates efficiency of the double-talk detection algorithm described in section 5.

We can conclude that the cancellation performance of the spline algorithms is comparable with that of the AP algorithm and significantly better than that of the cross-spectral algorithm.

7. COMPLEXITY

The complexity of the optimal cubic spline identification can be approximately represented as $P_{opt} \approx 2P_w + 3P_{FFT} + P_{S_{XX}} + P_{S_{XY}} + P_{\mathbf{R}} + P_{\boldsymbol{\xi}} + P_{\mathbf{c}} + P_{\hat{H}}$. The windowing in (3) or (4) requires $P_w = N$ MACs (multiply-accumulate operations). FFTs in (3) and (4) and inverse FFT in (12) each requires about $P_{FFT} = Nlog_2(N)$ MACs. Cross-spectrum and auto-spectrum calculation for $N/2$ frequency bins require $P_{S_{XY}} = 2N$ and $P_{S_{XX}} = N$ MACs, respectively. Due to the symmetry of the matrix $\mathbf{R}$, the fact that it is 7-diagonal, and the cubic B-splines have a support of $4D$ samples, the computational load for calculating $\mathbf{R}$ is $P_{\mathbf{R}} = 10N$ MACs. Similarly, calculation of $\boldsymbol{\xi}$ requires $P_{\boldsymbol{\xi}} = 4N$ MACs. The frequency response estimate (5) requires $P_{\hat{H}} = 4N$ MACs. Solution of the system (7) by using such an efficient technique as the Cholesky algorithm would require $P_{\mathbf{c}} = N_\varphi^3/3$ MACs; for $N_\varphi = 585$, it would be too complex for real-time, $P_{\mathbf{c}} \approx 7 \cdot 10^7$ MACs. For the DCD algorithm, in all simulation trials $P_{\mathbf{c}}$ has not exceeded $1.8 \cdot 10^5$ operations, which is significantly smaller. Then the total computational load is $P_{opt} \approx 6.9 \cdot 10^5$ MACs or $P_{opt}/M \approx 340$ MACs/sample. For local splines, similar analysis results in a total of 190 MACs/sample. The cross-spectral technique requires 60 MACs/sample. The AP algorithm (together with the FIR filtering) requires about 4600 MACs/sample plus an extra complexity for inversion of a $N_{AP} \times N_{AP}$ matrix.

Thus, the optimal spline identification with the DCD algorithm, local splines, the cross-spectral and AP algorithms require 340, 190, 90, and 4600 MACs/sample, respectively. By taking into account the FIR filtering and normalizing to the number of FIR taps L, these can be represented as 1.7, 1.4, 1.1, and 9.0 MACs/sample/tap, respectively. Note that the NLMS algorithm requires 2 MACs/sample/tap; thus, all the spectral-domain algorithms have smaller complexity than the NLMS algorithm.

8. CONCLUSIONS

We have proposed new acoustic echo cancellation algorithms. These are based on time-domain FIR filtering and spectral domain cubic spline identification of the acoustic frequency response. We have considered optimal splines with ideal matrix inversion, optimal splines with the DCD algorithm, and local splines. Ideal matrix inversion and the DCD algorithm result in identical echo cancellation performance, while the DCD algorithm allows significant reduction in the complexity. The DCD-based optimal B-spline identification requires 1.7 MACs/sample/tap, while local splines require 1.4 MACs/sample/tap. The cross-spectral technique, though having a smaller complexity (1.1 MACs/sample/tap), provides poorer cancellation performance, especially at low SNRs. The affine projection algorithm provides a better echo attenuation at high SNRs and comparable at low SNRs; however, it is complex for implementation. The proposed techniques have also been shown to provide efficient double-talk detection.

9. REFERENCES

[1] A. H. Sayed, *Fundamentals of adaptive filtering*, A John Wiley & Sons, Inc., 2003.

[2] S. L. Gay and S. Tavathia, "The fast affine projection algorithm," in *Proceedings ICASSP'95*, 1995, pp. 3023–3026.

[3] E. Woudenberg, F. K. Soong, and B. H. Juang, "A block least squares approach to acoustic echo cancellation," in *Proc. ICASSP'99, Phoenix, USA*, March 1999, vol. 2, pp. 869–872.

[4] S. L. Gay and J. Benesty, *Acoustic signal processing for telecommunication*, Kluwer Academic Publishers, 2001.

[5] T. Okuno, M. Fukushima, and M. Tohyama, "Adaptive cross-spectral technique for acoustic echo cancellation," *IEICE Trans. Fundamentals*, vol. E82-A, no. 4, pp. 634–639, April 1999.

[6] Y. V. Zakharov and T. C. Tozer, "Multiplication-free iterative algorithm for LS problem," *Electronics Letters*, vol. 40, no. 9, pp. 567–569, April 2004.

[7] M. Unser, A. Aldroubi, and M. Eden, "B-Spline signal processing: Part I - Theory," *IEEE Trans. Signal Processing*, vol. 41, no. 2, pp. 821–833, Feb. 1993.

[8] M. Unser, A. Aldroubi, and M. Eden, "B-Spline signal processing: Part II - Efficient design and applications," *IEEE Trans. Signal Processing*, vol. 41, no. 2, pp. 834–848, Feb. 1993.

[9] Y. V. Zakharov, T. C. Tozer, and J. F. Adlard, "Polynomial spline-approximation of Clarke's model," *IEEE Trans. Signal Processing*, vol. 52, no. 5, pp. 1198–1208, May 2004.