

ACCURATE ADAPTIVE FILTERING IN SQUARE-ROOT HANN WINDOWED SHORT-TIME FOURIER TRANSFORM DOMAIN

Suehiro Shimauchi and Hitoshi Ohmuro

NTT Media Intelligence Laboratories
NTT Corporation, Tokyo, Japan

ABSTRACT

A novel short-time Fourier transform (STFT) domain adaptive filtering scheme is proposed that can be easily combined with nonlinear post filters such as residual echo or noise reduction in acoustic echo cancellation. Unlike normal STFT subband adaptive filters, which suffers from aliasing artifacts due to its poor prototype filter, our scheme achieves good accuracy by exploiting the relationship between the linear convolution and the poor prototype filter, i.e., the STFT window function. The effectiveness of our scheme was confirmed through the results of simulations conducted to compare it with conventional methods.

Index Terms— Adaptive filters, short-time Fourier transform, square-root Hann window, acoustic echo cancellation

1. INTRODUCTION

Adaptive filtering technique is widely used in various system identification applications such as acoustic echo cancellation (AEC) [1]. In AEC applications, an adaptive filter identifies the acoustic echo path between the loudspeaker and microphone in order to estimate the echo signal, and then subtracts the estimated echo from the microphone signal to achieve the echo cancellation. The adaptive filter for AEC is often combined with a nonlinear post filter [2, 3], which reduces the residual echo and/or noise. The seamless combination of an adaptive filter and a post filter has been investigated [4, 5].

Typical post filters are processed in the short-time Fourier transform (STFT) domain. Therefore, STFT-domain adaptive filters are a reasonable choice for efficient implementation. The STFT can be regarded as a discrete Fourier transform (DFT) filter bank whose prototype filter is a DFT window function, e.g., a Hann window. The DFT window length is much shorter than those of the prototype filters designed for general subband adaptive filters [6]. Therefore, when the STFT subband adaptive filter is straightforwardly developed, its accuracy is limited by severe aliasing artifacts of the decimated subband signals due to the poor prototype filter. To improve the performance of the STFT adaptive filter, Avargel and Cohen [7] introduced crossband filters to connect each subband reference signal with neighbor subbands to compensate for the aliasing artifacts. Krini and Schmidt [5] introduced additional filter coefficients for subsampled subband reference signals obtained by an interpolation technique. Instead of achieving the STFT adaptive filter, Lu and Champagne [4] proposed a combination of the subband adaptive

filter and the subband post filter, both of which are processed with a sufficiently long prototype filter based on the interpolation of a quadrature mirror filter (QMF).

On the other hand, frequency domain (FD) adaptive filters [6, 8], whose structures are similar to those of the subband adaptive filters, are known to be free from the above-mentioned aliasing problem. This is because they are designed to guarantee the linear convolution property. This property is achieved by converting the system output estimates into a time domain, where the linear convolution components and the circular convolution artifacts are obtained separately. Thus the errors between the desired outputs and its estimates can be evaluated without any effects from circular convolution artifacts. However, due to the fullband error evaluation, the FD adaptive filter is not necessarily combined efficiently with the post filter in the STFT domain.

In this paper, we propose a novel scheme for the STFT adaptive filtering, which can be easily combined with the STFT post filters. The prior STFT adaptive filters proposed by Avargel and Cohen [7] and Krini and Schmidt [5] can be regarded as modified versions of the subband adaptive filters, which are derived by considering how to reduce the aliasing artifacts. Our approach is based on FD adaptive filters [8], which are derived by considering how to reduce the circular convolution artifacts. However, unlike the case with conventional FD adaptive filters, our adaptive filtering scheme directly evaluates the subband errors in the STFT domain. Eneman and Moonen [9] and Merched and Sayed [10] investigated the relationship between subband and FD adaptive filters. It can be understood from their results that the accuracy of subband adaptive filters can be improved if their structures are modified so that the fullband errors are evaluated as the FD adaptive filters do. However, on the contrary, our interest here is in investigating how to maintain the good accuracy of the FD adaptive filters when the subband errors are directly evaluated. To achieve this, we focus on a particular STFT case where the square-root Hann (SR-H) window is chosen for frequency analysis and synthesis, which is a popular choice in speech and audio processing [11].

2. DIRECT LINEAR CONVOLUTION IN FREQUENCY DOMAIN

We review how the FD adaptive filters guarantee the linear convolution property and consider how to directly obtain the accurate estimates of system outputs in the frequency domain.

The unknown system is assumed to be modeled as a linear finite impulse response (FIR) filter. The signal $d(n)$ is output for the input $x(n)$, where n indicates the discrete time index. In the block manner, we observe the signal block $\mathbf{y}(m)$ as a mixture of the system output $\mathbf{d}(m)$ and additive noise $\mathbf{v}(m)$:

$$\mathbf{y}(m) = \mathbf{d}(m) + \mathbf{v}(m), \quad (1)$$

where $\mathbf{y}(m) = [y(mN - 2N + 1), \dots, y(mN)]^T$, $\mathbf{d}(m) = [d(mN - 2N + 1), \dots, d(mN)]^T$, $\mathbf{v}(m) = [v(mN - 2N + 1), \dots, v(mN)]^T$. Here, two successive N -sample frames are combined as a signal block for every frame index m . This is suitable for the 50% overlap STFT processing. The system output $\mathbf{d}(m)$ is generated by the linear convolution of the system input $x(n)$ and system impulse response $\mathbf{h} = [h_0, h_1, \dots, h_{2NL-1}]^T$ as follows,

$$\mathbf{d}(m) = \mathbf{H} \begin{bmatrix} x(mN-2N(L+1)+1) \\ \vdots \\ x(mN) \end{bmatrix}, \quad (2)$$

where $\mathbf{H}$ is the impulse response matrix,

$$\mathbf{H} = \begin{bmatrix} 0 & h_{2NL-1} & \dots & h_1 & h_0 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & h_{2NL-1} & \dots & h_1 & h_0 & 0 \\ 0 & \dots & 0 & 0 & h_{2NL-1} & \dots & h_1 & h_0 \end{bmatrix}, \quad (3)$$

and the length of the impulse response is assumed to be not longer than $2NL$. In the multidelay block frequency domain (MDF) model [8], the convolution is calculated in the frequency domain by separating the impulse response into L block sections. The outputs of frequency domain filtering correspond to the following $4N$ -sample signal block in the time domain:

$$\begin{bmatrix} \bar{\mathbf{d}}^l(m) \\ \mathbf{d}^l(m) \end{bmatrix} = \sum_{l=0}^{L-1} \begin{bmatrix} \bar{\mathbf{d}}^l(m) \\ \mathbf{d}^l(m) \end{bmatrix}, \quad (4)$$

where the last $2N$ -sample block is the linear convolution output $\mathbf{d}(m)$, and the first $2N$ -sample block is the circular convolution output $\bar{\mathbf{d}}(m)$, which is not observed in the real world. To analyze this in more detail, we focus on the l -th block section, which can be expressed as

$$\begin{bmatrix} \bar{\mathbf{d}}^l(m) \\ \mathbf{d}^l(m) \end{bmatrix} = C \left(\begin{bmatrix} \mathbf{h}^l \\ \mathbf{0}_{2N} \end{bmatrix} \right) \begin{bmatrix} \mathbf{x}^{l+1}(m) \\ \mathbf{x}^l(m) \end{bmatrix}, \quad (5)$$

where

$$\mathbf{h}^l = [h_{2Nl}, \dots, h_{2Nl+2N-1}]^T, \quad (6)$$

$$\mathbf{x}^l(m) = [x(mN - 2N(l+1) + 1), \dots, x(mN - 2Nl)]^T. \quad (7)$$

Here, $\mathbf{0}_{2N}$ is the $2N$ -elements zero vector, and $C(\mathbf{a})$ indicates a circulant matrix generated from P -dimensional vector $\mathbf{a} = [a_0, \dots, a_{P-1}]^T$ as

$$C(\mathbf{a}) = \begin{bmatrix} a_0 & a_{P-1} & \dots & a_1 \\ a_1 & a_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_{P-1} \\ a_{P-1} & \dots & a_1 & a_0 \end{bmatrix}. \quad (8)$$

Now the signal block can be decomposed as

$$\begin{bmatrix} \bar{\mathbf{d}}^l(m) \\ \mathbf{d}^l(m) \end{bmatrix} = \begin{bmatrix} \mathbf{d}_{(+)}^l(m) \\ \mathbf{d}_{(+)}^l(m) \end{bmatrix} + \begin{bmatrix} -\mathbf{d}_{(-)}^l(m) \\ \mathbf{d}_{(-)}^l(m) \end{bmatrix}, \quad (9)$$

where

$$\mathbf{d}_{(+)}^l(m) = C(\mathbf{h}^l) \mathbf{x}_{(+)}^l(m)/2, \quad (10)$$

$$\mathbf{d}_{(-)}^l(m) = T(\mathbf{h}^l) \mathbf{x}_{(-)}^l(m)/2, \quad (11)$$

$$\mathbf{x}_{(+)}^l(m) = \mathbf{x}^l(m) + \mathbf{x}^{l+1}(m), \quad (12)$$

$$\mathbf{x}_{(-)}^l(m) = \mathbf{x}^l(m) - \mathbf{x}^{l+1}(m). \quad (13)$$

and $T(\mathbf{a})$ indicates a non-circulant Toeplitz matrix:

$$T(\mathbf{a}) = \begin{bmatrix} a_0 & -a_{P-1} & \dots & -a_1 \\ a_1 & a_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & -a_{P-1} \\ a_{P-1} & \dots & a_1 & a_0 \end{bmatrix}. \quad (14)$$

In the MDF case, the $4N$ -point DFT is directly applied to (5). Thus, the results are affected by the circular convolution artifact $\bar{\mathbf{d}}^l(m)$, which can not be easily removed in the frequency domain. Here we consider calculating (10) and (11) in the frequency domain by applying $2N$ -point DFT so that the frequency domain counterpart of $\mathbf{d}^l(m)$ can be directly obtained. In (10), the circulant matrix $C(\mathbf{h}^l)$ is diagonalized by the DFT operation. Thus, it can be efficiently calculated in the frequency domain. On the other hand, (11) has the non-circulant matrix. Therefore, we consider modifying the following equation:

$$\begin{bmatrix} -\mathbf{d}_{(-)}^l(m) \\ \mathbf{d}_{(-)}^l(m) \end{bmatrix} = \frac{1}{4} C \left(\begin{bmatrix} \mathbf{h}^l \\ -\mathbf{h}^l \end{bmatrix} \right) \begin{bmatrix} -\mathbf{x}_{(-)}^l(m) \\ \mathbf{x}_{(-)}^l(m) \end{bmatrix}. \quad (15)$$

Equation (11) is obtained by decimating (15). We apply a modulation window to (15) as follows. The left hand side of (15) becomes,

$$D(\mathbf{w}_{4N}) \begin{bmatrix} -\mathbf{d}_{(-)}^l(m) \\ \mathbf{d}_{(-)}^l(m) \end{bmatrix} = \begin{bmatrix} D(\mathbf{w}_{2N}) \mathbf{d}_{(-)}^l(m) \\ D(\mathbf{w}_{2N}) \mathbf{d}_{(-)}^l(m) \end{bmatrix}. \quad (16)$$

The right hand side of (15) becomes,

$$\begin{aligned} & \frac{1}{4} D(\mathbf{w}_{4N}) C \left(\begin{bmatrix} \mathbf{h}^l \\ -\mathbf{h}^l \end{bmatrix} \right) \begin{bmatrix} -\mathbf{x}_{(-)}^l(m) \\ \mathbf{x}_{(-)}^l(m) \end{bmatrix} \\ &= \frac{1}{4} C \left(\begin{bmatrix} D^H(\mathbf{w}_{2N}) \mathbf{h}^l \\ D^H(\mathbf{w}_{2N}) \mathbf{h}^l \end{bmatrix} \right) \begin{bmatrix} D(\mathbf{w}_{2N}) \mathbf{x}_{(-)}^l(m) \\ D(\mathbf{w}_{2N}) \mathbf{x}_{(-)}^l(m) \end{bmatrix} \end{aligned} \quad (17)$$

where

$$\mathbf{w}_{4N} = [e^{-j2\pi \frac{0}{4N}}, \dots, e^{-j2\pi \frac{4N-1}{4N}}]^T, \quad (18)$$

$$\mathbf{w}_{2N} = [e^{-j2\pi \frac{0}{4N}}, \dots, e^{-j2\pi \frac{2N-1}{4N}}]^T, \quad (19)$$

$$D(\mathbf{a}) = \text{diag}(a_0, \dots, a_{P-1}), \quad (20)$$

and H indicates a conjugate transpose. Consequently, we obtain a half-wave modulated version of (11) as,

$$\begin{aligned} & D(\mathbf{w}_{2N}) \mathbf{d}_{(-)}^l(m) \\ &= C(D^H(\mathbf{w}_{2N}) \mathbf{h}^l) D(\mathbf{w}_{2N}) \mathbf{x}_{(-)}^l(m)/2. \end{aligned} \quad (21)$$

Fortunately, the non-circulant matrix in (11) turns into a circulant matrix in (21), and the imaginary part of the modulation window $\mathbf{w}_{2N}$ is exactly the same as the SR-H window with the opposite sign.

3. PROPOSED ADAPTIVE FILTERING SCHEME

3.1. Adaptive filter structure

According to the discussion in the previous section, we derive a new frequency domain adaptive filter structure, as shown in Fig. 1. The observed signal block $\mathbf{y}(m)$ is converted into the frequency domain by the STFT as,

$$\underline{\mathbf{y}}(m) = \mathbf{F}D(\mathbf{s}_{2N})\mathbf{y}(m), \quad (22)$$

where $\mathbf{s}_{2N}$ is the SR-H window:

$$\mathbf{s}_{2N} = [\sin\left(2\pi\frac{0}{4N}\right), \dots, \sin\left(2\pi\frac{2N-1}{4N}\right)]^T. \quad (23)$$

Here, $\mathbf{F}$ indicates the $2N$ -point DFT matrix, and the frequency domain signal blocks are denoted with an underline. The reference signal blocks, $\mathbf{x}_{(+)}^0(m)$ and $\mathbf{x}_{(-)}^0(m)$, are converted as,

$$\underline{\mathbf{x}}_{(+)}^0(m) = \mathbf{F}\mathbf{x}_{(+)}^0(m)/2, \quad (24)$$

$$\underline{\mathbf{x}}_{(-)}^0(m) = \mathbf{F}D(\mathbf{w}_{2N})\mathbf{x}_{(-)}^0(m)/2, \quad (25)$$

where $\mathbf{x}_{(-)}^0(m)$ is converted after modulation referring to (21). The estimated outputs of the sum and difference parts for l -th filter block can be obtained as,

$$\hat{\underline{\mathbf{d}}}_{(+)}^l(m) = D(\hat{\underline{\mathbf{h}}}_{(+)}^l(m))\underline{\mathbf{x}}_{(+)}^l(m), \quad (26)$$

$$\hat{\underline{\mathbf{d}}}_{(-)}^l(m) = D(\hat{\underline{\mathbf{h}}}_{(-)}^l(m))\underline{\mathbf{x}}_{(-)}^l(m), \quad (27)$$

where $\hat{\underline{\mathbf{h}}}_{(+)}^l(m)$ and $\hat{\underline{\mathbf{h}}}_{(-)}^l(m)$ respectively correspond to the estimates of the DFTs of $\mathbf{h}^l$ and $D(\mathbf{w}_{2N})\mathbf{h}^l$. Here, the l -th converted reference signal blocks can be obtained as,

$$\underline{\mathbf{x}}_{(+)}^l(m) = \underline{\mathbf{x}}_{(+)}^0(m-2l), \quad (28)$$

$$\underline{\mathbf{x}}_{(-)}^l(m) = \underline{\mathbf{x}}_{(-)}^0(m-2l). \quad (29)$$

By taking into account that the SR-H window is now applied to the desired system output in (22), the frequency-domain estimate of the system output is generated as

$$\hat{\underline{\mathbf{d}}}(m) = C(\underline{\mathbf{s}}_{2N}) \sum_{l=0}^{L-1} \hat{\underline{\mathbf{d}}}_{(+)}^l(m) + C(\underline{\mathbf{g}}_{2N}) \sum_{l=0}^{L-1} \hat{\underline{\mathbf{d}}}_{(-)}^l(m), \quad (30)$$

where

$$\underline{\mathbf{s}}_{2N} = \mathbf{F}\mathbf{s}_{2N}, \quad (31)$$

$$\underline{\mathbf{g}}_{2N} = [-j/2, j/2, 0, \dots, 0]^T, \quad (32)$$

and $C(\underline{\mathbf{s}}_{2N})$ applies the SR-H window to $\hat{\underline{\mathbf{d}}}_{(+)}^l(m)$, and $C(\underline{\mathbf{g}}_{2N})$ extracts the SR-H windowed components from $\hat{\underline{\mathbf{d}}}_{(-)}^l(m)$ in the frequency domain. While most elements of $\underline{\mathbf{g}}_{2N}$ are zero as is, most elements of $\underline{\mathbf{s}}_{2N}$ are very small but are not exactly zero. For computational efficiency, we use an approximated version of $\underline{\mathbf{s}}_{2N}$ by making the small elements zero while keeping the K non-zero elements. The

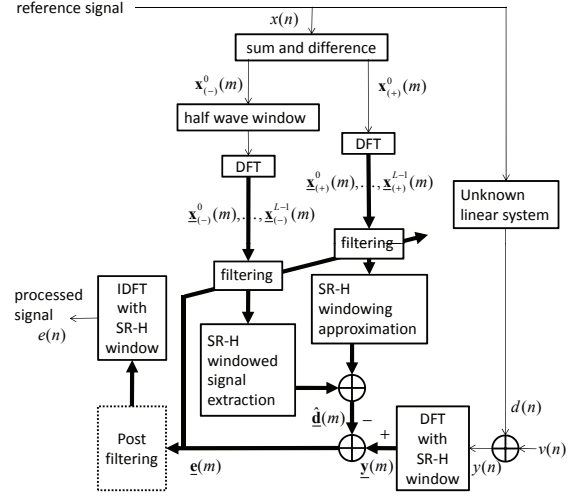


Fig. 1. Proposed adaptive filter structure.

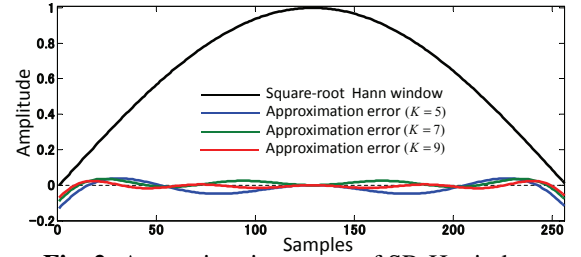


Fig. 2. Approximation errors of SR-H window.

output error is directly calculated in the frequency domain,

$$\underline{\mathbf{e}}(m) = \underline{\mathbf{y}}(m) - \hat{\underline{\mathbf{d}}}(m). \quad (33)$$

After applying some post filtering to $\underline{\mathbf{e}}(m)$, the processed signal in the time domain, $e(n)$, is obtained through the inverse DFT (IDFT) with the SR-H window,

$$\mathbf{e}(m) = D(\mathbf{w}_{2N})\mathbf{F}^{-1}\underline{\mathbf{e}}(m). \quad (34)$$

3.2. Adaptive algorithm

The frequency domain filter coefficient blocks $\hat{\underline{\mathbf{h}}}_{(+)}^l(m)$ and $\hat{\underline{\mathbf{h}}}_{(-)}^l(m)$ can be updated as,

$$\hat{\underline{\mathbf{h}}}_{(+)}^l(m+1) = \hat{\underline{\mathbf{h}}}_{(+)}^l(m) + D(\underline{\mu}_{(+)}(m))D^H(\underline{\mathbf{x}}_{(+)}^l(m))\underline{\mathbf{e}}_{(+)}(m), \quad (35)$$

$$\hat{\underline{\mathbf{h}}}_{(-)}^l(m+1) = \hat{\underline{\mathbf{h}}}_{(-)}^l(m) + D(\underline{\mu}_{(-)}(m))D^H(\underline{\mathbf{x}}_{(-)}^l(m))\underline{\mathbf{e}}_{(-)}(m), \quad (36)$$

where

$$D(\underline{\mu}_{(+)}(m)) = \mu D^{-1}(\mathbf{r}_{(+)}(m)), \quad (37)$$

$$D(\underline{\mu}_{(-)}(m)) = \mu D^{-1}(\mathbf{r}_{(-)}(m)), \quad (38)$$

$$\mathbf{r}_{(+)}(m) = \alpha \mathbf{r}_{(+)}(m-1) + (1-\alpha) \sum_{l=0}^{L-1} D^H(\underline{\mathbf{x}}_{(+)}^l(m))\underline{\mathbf{x}}_{(+)}^l(m), \quad (39)$$

$$\mathbf{r}_{(-)}(m) = \alpha \mathbf{r}_{(-)}(m-1) + (1-\alpha) \sum_{l=0}^{L-1} D^H(\underline{\mathbf{x}}_{(-)}^l(m))\underline{\mathbf{x}}_{(-)}^l(m), \quad (40)$$

and μ is a step size. The errors $\underline{\mathbf{e}}_{(+)}(m)$ and $\underline{\mathbf{e}}_{(-)}(m)$ are modified versions of $\underline{\mathbf{e}}(m)$ that enable the algorithm to take into ac-

count the combination of the sum and difference signals with different window modulation effects and the approximation error of $\underline{s}_{2N}$. The modified errors are calculated as,

$$\underline{e}_{(+)}(m) = D(\mathbf{q}_{2N}) D(\mathbf{r}_{2N}) \underline{e}(m), \quad (41)$$

$$\underline{e}_{(-)}(m) = j D(\mathbf{r}_{2N}) \underline{e}(m), \quad (42)$$

where

$$\mathbf{r}_{2N} = [1, -0.5, 0, 0, \dots, 0, 0, -0.5]^T, \quad (43)$$

$$\mathbf{q}_{2N} = [0.6366, -0.6366, 0, 0, \dots, 0, 0]^T. \quad (44)$$

As shown in Fig. 2, the approximation of $\underline{s}_{2N}$ with K non-zero elements causes errors mainly around both edge sides in the time domain, where the approximated window is normalized in order to have 1 at the center point. Therefore, the frequency domain filter $\mathbf{r}_{2N}$, which corresponds to the Hann window in the time domain, is applied to $\underline{e}(m)$ in order to reduce the influence of the window approximation error on the system identification. The SR-H window can be regarded as a kind of half-wave modulation, which shifts by 0.5 point in the DFT domain. The simple two-tap filter $\mathbf{r}_{2N}$ is introduced to compensate for this modulation effect since the filter for the sum signal, $\hat{\mathbf{h}}_{(+)}^l(m)$, outputs non-modulated estimates as is. The imaginary unit j is multiplied in (42) since the imaginary part of the output of $\hat{\mathbf{h}}_{(-)}^l(m)$ corresponds to the SR-H windowed signal.

Because $\hat{\mathbf{h}}_{(-)}^l(m)$ should be a modulated version of $\hat{\mathbf{h}}_{(+)}^l(m)$, the filter coefficients can be constrained to satisfy this. Although this constraint requires four extra operations of $2N$ -point DFT or IDFT for each l block, it is applied to only one filter block at each iteration, as in the MDF case [8].

4. SIMULATIONS

We carried out some simulations assuming the AEC applications to compare the performance of the proposed scheme with those of the normal STFT subband adaptive filter, its interpolated version [5], and the MDF adaptive filter [8]. Through all simulations, the frame size $N = 128$, the target linear impulse response was a real room impulse response with 8-kHz sampling frequency that was truncated by 1024 samples, and its property was changed at 4 s as an echo path change. For the proposed method, the number of filter blocks was $L = 4$, which corresponds to 1024-tap impulse response. The SR-H window was approximated with $K = 9$. For the normal subband method, $2L$ filter blocks were used for the equivalent filter length. For the interpolated subband method, $4L$ filter blocks were used for the equivalent filter length. The reference signal blocks were interpolated by the ideal filter. For the MDF method, L filter blocks were used for the equivalent filter length. For all methods, the filter coefficients were updated every N -sample frame. All step sizes were chosen as $\mu = 0.5$. White noise and male speech were used as the reference signals. No additive noise was mixed into the system output so as to concentrate on the accuracy limitation due to the structure difference. Figures 3 and 4 plot the mean squared error (MSE) convergence for the white noise input and the male speech input, respectively. In both cases,

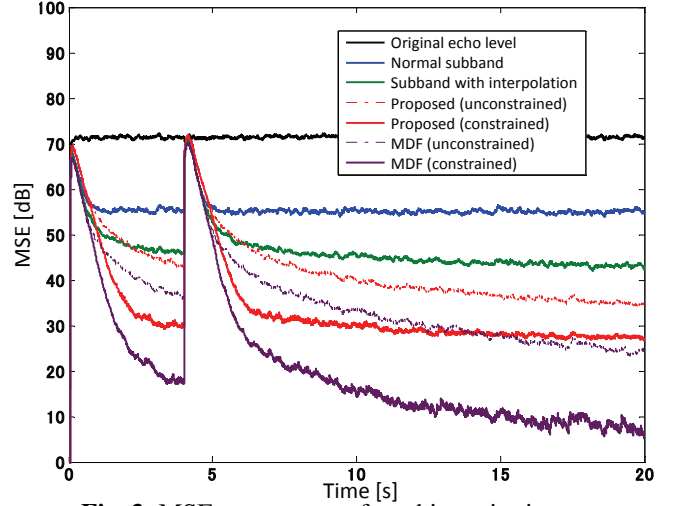


Fig. 3. MSE convergence for white noise input.

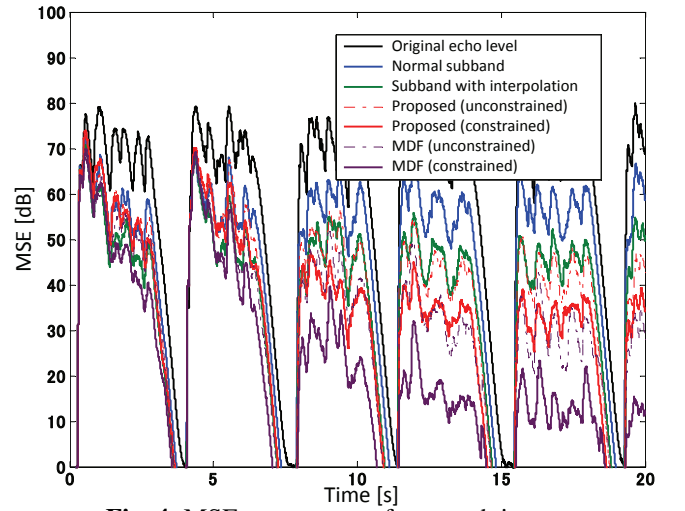


Fig. 4. MSE convergence for speech input.

although the MDF method, in whose structure the fullband errors are evaluated, was outperformed, the proposed method achieved the best performance among the methods that evaluate subband errors, especially when the filter coefficient constraint was applied like it was in the MDF case.

5. CONCLUSIONS

A novel STFT adaptive filtering scheme was proposed that can be easily combined with post filters such as residual echo and/or noise reduction in acoustic echo cancellation applications. Unlike the normal STFT subband adaptive filter, which suffers from aliasing artifacts due to the poor prototype filter, i.e., the DFT window function, our scheme achieved good accuracy by exploiting the relationship between the linear convolution and the SR-H window. Simulations were conducted to compare the proposed scheme with conventional methods. The scheme was confirmed to be effective from the results. This paper focused on the accuracy of the scheme. More computationally efficient implementation is a remaining issue to be investigated in the future.

6. REFERENCES

- [1] C. Breining, P. Dreiscitel, E. Hansler, A. Mader, B. Nitsch, H. Puder, T. Schertler, G. Schmidt, and J. Tilp, "Acoustic echo control. an application of very-high-order adaptive filters," *Signal Processing Magazine, IEEE*, vol. 16, no. 4, pp. 42–69, 1999.
- [2] W.L.B. Jeannes, P. Scalart, G. Faucon, and C. Beaugeant, "Combined noise and echo reduction in hands-free systems: a survey," *IEEE Trans. on Speech and Audio Processing*, vol. 9, no. 8, pp. 808–820, 2001.
- [3] S. Gustafsson, R. Martin, P. Jax, and P. Vary, "A psychoacoustic approach to combined acoustic echo cancellation and noise reduction," *IEEE Trans. on Speech and Audio Processing*, vol. 10, no. 5, pp. 245–256, 2002.
- [4] X. Lu and B. Champagne, "Acoustic echo cancellation with post-filtering in subband," in *Proc. WASPAA 2003*, 2003, pp. 29–32.
- [5] M. Krini and G. Schmidt, "Method for temporal interpolation of short-term spectra and its application to adaptive system identification," in *Proc. ICASSP 2012*, 2012, pp. 45–48.
- [6] J.J. Shynk, "Frequency-domain and multirate adaptive filtering," *Signal Processing Magazine, IEEE*, vol. 9, no. 1, pp. 14–37, 1992.
- [7] Y. Avargel and I. Cohen, "System identification in the short-time fourier transform domain with crossband filtering," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 15, no. 4, pp. 1305–1319, 2007.
- [8] J.-S. Soo and K.K. Pang, "Multidelay block frequency domain adaptive filter," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 38, no. 2, pp. 373–376, 1990.
- [9] K. Eneman and M. Moonen, "Hybrid subband/frequency-domain adaptive systems," *Signal Processing*, vol. 81, no. 1, pp. 117 – 136, 2001.
- [10] R. Merched and A.H. Sayed, "An embedding approach to frequency-domain and subband adaptive filtering," *IEEE Trans. on Signal Processing*, vol. 48, no. 9, pp. 2607–2619, 2000.
- [11] R. Martin and R.V. Cox, "New speech enhancement techniques for low bit rate speech coding," in *Proc. 1999 IEEE Workshop on Speech Coding Proceedings*, 1999, pp. 165–167.