Proceedings of the Ninth International Conference on Machine Learning and Cybernetics, Qingdao, 11-14 July 2010

A FREQUENCY-DOMAIN DOUBLE-TALK DETECTOR COMBINED WITH THE REGULARIZED GMDF ADAPTIVE FILTER

HSU-CHANG HUANG[1], JUNGHSI LEE[2]

Department of Electrical Engineering, Yuan-Ze University, Chung-Li, Taoyuan, 32026, TAIWAN
E-MAIL: s958501@mail.yzu.edu.tw, eejlee@saturn.yzu.edu.tw

Abstract:

Adaptive filter and double-talk detector (DTD) are two essential parts in acoustic echo cancellation of hands-free communications. The generalized frequency-domain multidelay adaptive filter (GMDF) is an attractive choice due to its convergence properties and computational parsimony. Very recently we presented a robust time-domain DTD equipped with a near-end voice detector, a double-to-single talk detector and two auxiliary filters [7]. In this paper, we expand the work of [7] to introduce a frequency-domain DTD combined with the regularized GMDF adaptive filter. Extensive simulations demonstrated the proposed system was capable of differentiating echo path changes from double-talk situation, and performed better than other competitive schemes.

Keywords:

Acoustic echo cancellation; double-talk detector; adaptive filter; NLMS; MDF.

1. Introduction

Acoustic echoes are major sources of annoyance in hands-free communications such as mobile radiotelephone, speakerphone, teleconferencing and video conferencing. In these applications, acoustic echo canceller (AEC) is employed to reduce the annoying acoustic echo by means of estimating the loudspeaker-enclosure-microphone system with adaptive filter. The acoustic echo cancellation process becomes complicated whenever the near-end speech and far-end speech occur simultaneously, the so-called double talk (DT) mode. In this DT stage, the adaptation of the adaptive filter will be severely disturbed by the near-end signal. Therefore, a dependable double-talk detector is required to decide whether it is in DT mode. If so, the AEC has to either slow down or freeze the adaptation of the filter to prevent it from divergence [1-4, 7-9].

In hands-free communications, the movement of objects or people changes the acoustic echo paths and introduces a more difficult problem for double-talk detector (DTD). The DTD might see the double-talk situation as change of echo path. In that case, the adaptive filter will keep updating and result in the divergence of the system. On the other hand, the DTD may declare echo-path changes as DT mode. Therefore, researchers have been looking for a DTD which is able to efficiently differentiate echo-path changes from DT mode [1], [4], [7] for the past 10 years.

Recently we presented a time-domain DTD equipped with a near end voice detector (NEVD), a double-to-single detector (DSD) and two auxiliary filters [7]. This DTD system performs well in transition of double-talk and single talk and can distinguish between echo-path changes and double-talk mode. The auxiliary filters give the AEC a better chance of saving good estimate of echo path during the transition from single-talk to double-talk. This DTD system [7] was shown to outperform Park's method [8].

The generalized multidelay frequency domain adaptive filter is attractive in the application of AEC. The GMDF has very low computational complexity due to the usage of fast Fourier transform (FFT). Our recent paper [6] derived a very relaxed fixed common step-size bound for GMDF which made it more practical. That is, we now could choose suitable fixed common step-size to maintain good tracking and convergence performance as well. Just like the regularization process considered for the normalized least-mean-square (NLMS) algorithm, the GMDF could perform better by properly regularized in the frequency domain. Based on the step-size bound analysis in [6], we developed a regularized GMDF algorithm [5] recently.

In this paper, we expand the idea of [7] to introduce a frequency-domain DTD combined with the regularized GMDF adaptive filter. The proposed DTD system performed well during double-talk stage in an echo-path changing scenario. The superiority of our method was verified by extensive simulations.

2. Summary of the regularized GMDF algorithm [5]

We summarize the regularized GMDF adaptive filter in this section. Consider the GMDF with L sub-filters,

978-1-4244-6525-5/10/$26.00 ©2010 IEEE

each of order N, and FFT size is $2N$. Without loss of generality, we assume that $M = NL$ where M is the order of the GMDF adaptive filter. The GMDF uses a positive integer α to control the overlap between the successive input blocks. Consequently, it updates the coefficients every $R = N/\alpha$ samples. In the k^{th} block iteration, defining reference input vector $\mathbf{x}_k$ and desired response vector $\mathbf{y}_k$, respectively, as

$$\mathbf{x}_k = \left[x(kR), x(kR+1), \cdots, x(kR+N-1)\right]^T, \quad (1)$$

$$\mathbf{y}_k = \left[y(kR), y(kR+1), \cdots, y(kR+N-1)\right]^T. \quad (2)$$

Frequency-domain input vector for l^{th} sub-filter, denoted as $\mathbf{X}_{l,k}$, $l = 1,2,\cdots,L$ is computed as

$$\mathbf{X}_{l,k} = FFT\left[\mathbf{x}_{k-l\alpha}^T, \mathbf{x}_{k-(l-1)\alpha}^T\right]^T. \quad (3)$$

The corresponding frequency-domain coefficient vector $\mathbf{H}_{l,k}$ is defined accordingly as

$$\mathbf{H}_{l,k} = FFT\left[\mathbf{h}_{l,k}^T, 0_{N\times 1}^T\right]^T, \quad (4)$$

where $\mathbf{h}_{l,k}$ is the l^{th} sub-filter's time-domain coefficient vector. Output vector of filter $\hat{\mathbf{d}}_k$ is calculated as

$$\hat{\mathbf{d}}_k = \text{last N points of } FFT^{-1}\left[\sum_{l=1}^{L} \mathbf{H}_{l,k} \otimes \mathbf{X}_{l,k}\right], \quad (5)$$

where $\otimes$ denotes element-wise multiplication. Frequency-domain error vector $\mathbf{E}_k$ is obtained as follows

$$\mathbf{e}_k = \mathbf{y}_k - \hat{\mathbf{d}}_k, \quad (6)$$

$$\mathbf{E}_k = FFT\left[0_{N\times 1}^T, \mathbf{e}_k^T\right]^T. \quad (7)$$

The frequency power of the l^{th} subfilter at k^{th} block iteration is calculated as

$$\mathbf{Z}_{l,k} = \beta \mathbf{Z}_{l,k-1} + (1-\beta)\overline{\mathbf{X}}_{l,k} \otimes \mathbf{X}_{l,k}, \quad (8)$$

where $\overline{\mathbf{X}}_{l,k}$ denotes the complex conjugate of $\mathbf{X}_{l,k}$, and β is a forgetting factor. The coefficient vector $\mathbf{H}_{l,k}$ is updated as

$$\mathbf{H}_{l,k+1} = \mathbf{H}_{l,k} + \frac{2\mu_{GMDF}}{M}\mathbf{\Phi}_{l,k}, \quad (9)$$

where μ_{GMDF} is a fixed common step-size parameter of the GMDF filter. In (9), $\mathbf{\Phi}_{l,k}$, the new information for updating, is obtained as

$$\mathbf{\Phi}_{l,k} = FFT\left[\phi_{l,k}^T, 0_{N\times 1}^T\right]^T, \quad (10)$$

and

$$\phi_{l,k} = \text{first part of } FFT^{-1}\left[\left(\mathbf{E}_k \otimes \overline{\mathbf{X}}_{l,k}\right) \odot \left(\mathbf{Z}_{l,k} + \delta \cdot \mathbf{1}_{2N\times 1}\right)\right], \quad (11)$$

where $\odot$ denotes element-wise division and δ denotes the regularization parameter. The regularization parameter δ has to be greater than $\sigma_x^2 \cdot \mu_{GMDF}$. It was shown in [5] that a workable, more conservative choice is

$$\delta > 3\mu_{GMDF}\sigma_x^2. \quad (12)$$

3. The proposed DTD

The block diagram of an acoustic echo canceller equipped with the proposed DTD in frequency-domain is illustrated in Fig. 1. Our DTD includes a near-end voice detector (NEVD) and two auxiliary filters. The role of NEVD is to control the adaptive filter in working or halting status. The auxiliary filters are to save good estimates of echo path so as to prevent the AEC from divergence.

3.1. NEVD indicator ξ_k

Since GMDF algorithm updates the coefficients every R samples at k^{th} block iteration, we obtain the frequency domain desired response vector $\mathbf{Y}_k$, and filter output vector $\hat{\mathbf{D}}_k$, with FFT size R as follows

$$\mathbf{Y}_k = FFT\left[\textit{last } R \textit{ point of } \mathbf{y}_k\right]^T, \quad (13)$$

$$\hat{\mathbf{D}}_k = FFT\left[\textit{last } R \textit{ point of } \hat{\mathbf{d}}_k\right]^T. \quad (14)$$

Let $\sigma_{\mathbf{Y},k}^2$ and $\sigma_{\hat{\mathbf{D}},k}^2$ represent the estimated power of $y(n)$ and $\hat{d}(n)$, respectively, in frequency domain at k^{th} iteration, and being calculated respectively as

$$\sigma_{\mathbf{Y},k}^2 = \beta\sigma_{\mathbf{Y},k-1}^2 + (1-\beta)\overline{\mathbf{Y}}_k \otimes \mathbf{Y}_k, \quad (15)$$

$$\sigma_{\hat{\mathbf{D}},k}^2 = \beta\sigma_{\hat{\mathbf{D}},k-1}^2 + (1-\beta)\overline{\hat{\mathbf{D}}}_k \otimes \hat{\mathbf{D}}_k. \quad (16)$$

Finally, the $R\times 1$ NEVD indicator vector ξ_k is defined as

$$\xi_k = \sqrt{\sigma_{\hat{\mathbf{D}},k}^2 \odot \sigma_{\mathbf{Y},k}^2}, \quad (17)$$

where $\sqrt{\ }$ denotes element-wise square root.

To reduce the possibility of erroneous decision, we use L consecutive indicator vectors to determine whether the system is in DT mode. Defining the decision variable q_l at k^{th} iteration as

$$q_l = Avg\{\xi_{k-l+1}\}, \; l = 1,2,\ldots,L, \quad (18)$$

where $Avg\{\cdot\}$ denotes average operation on the vector's

elements. Letting indicator $b_l = 0$ if $q_l < T$, or $b_l = 1$ if $q_l \geq T$, where T is a parameter obtained via extensive experiments. Defining the DT verification vector $\mathbf{B}_k$ as

$$\mathbf{B}_k = [b_1 \;\; b_2 \;\; \cdots \;\; b_L]. \tag{19}$$

Again, choosing a threshold T_F, if $Avg\{\mathbf{B}_k\} < T_F$, then setting $S_{DT,k} = 1$ and declaring the system is in DT mode. Otherwise, the system sets $S_{DT,k} = 0$.

3.2. Auxiliary filters

Our AEC uses two auxiliary filters: AF_1 and AF_2, both filters are employed only to store estimated echo path for later use. The filter output $\hat{d}(n)$ is produced by the regularized GMDF at all times. Auxiliary filters work as follows. At k^{th} block iteration, if $S_{DT,k} = 0$, it is NOT in the DT mode, the GMDF keeps updating the coefficients and the counter C_{AF} increases by one. When C_{AF} reaches a preset T_{AF}, coefficients of AF_1 are copied into AF_2, followed by copying coefficients of the GMDF to AF_1, then reset counter C_{AF} to zero. If $S_{DT,k} = 1$, it is in the DT mode, AEC freezes adaptation of the GMDF, copies coefficients of AF_2 to AF_1 and to the GMDF as well. The role of AF_1 is to have a buffer filter so that the AEC would have a higher probability of storing good acoustic echo path estimates during the transition from single talk to double talk.

4. Computational Complexity

We examine the computational complexity of the proposed DTD with regularized GMDF algorithm. The regularized GMDF algorithm requires five $2N-$point FFTs for each block iteration. The DTD requires $2R$ $-$point FFTs and $10R$ real multiplications (RM) to calculate NEVD indicator vector ξ_k. A $2N-$point FFT requires $2N\log_2 N$ RM. Therefore, the GMDF requires approximately $\alpha\{(6+4L)\log_2 N + 8L + 6\}$ RM and the DTD needs $2\log_2(R/2)+10$ RM for each data sample processed. In summary, the proposed system consumes $\alpha\{(6+4L)\log_2 N + 8L + 6\} + 2\log_2(N/2\alpha)+10$ RM for each sample. In comparison, the time-domain DTD with NLMS filter of [7] requires about $2LN+8$ RM each sample. We illustrate this comparison by means of the following examples. Given $N = 128$, $L = 4$, and $\alpha = 4$, the RM required by [7] is 1.32 times of our proposed system, and for $N = 512$, $L = 2$, and $\alpha = 1$, [7] requires 12.0 times RM of ours. Park's DTD employs NLMS algorithm, therefore, it exhibits similar computational complexity as that of [7].

5. Simulation results

We now investigate the performance of our DTD combined with the regularized GMDF via simulation. The regularized GMDF was employed to estimate a 512-tap acoustic echo path, $\mathbf{h}_{opt}(n)$, measured in a small office at 8000 Hz sampling rate. The excitation signal was a 16-second Chinese speech. White Gaussian noise with SNR 39dB was added to the acoustic echo. Double talk situation and sudden echo path change were considered in the simulation. We compared the performance of our DTD to two competitive algorithms: Park's DTD [8] and the time-domain DTD presented in [7].

The regularized GMDF chose $\mu_{GMDF} = 0.6 \times R$, $L = 4$, $\alpha = 4$, and NLMS was run with $\mu = 0.6$. Choosing step sizes this way gives both adaptive filters the same convergence properties when input signal is a white Gaussian process [6]. Parameters of our proposed DTD were chosen as $T_{AF} = 32$, $T = 0.65$, and $T_F = 0.75$. Parameters of [7] were chosen to be $\xi(n) = 0.65$ and $\zeta(k) = 0.95$. Parameters of Park's DTD were chosen as $\rho_{e,y}(n) = 0.35$ and $\rho_{\hat{d},y}(n) = 0.8$. It should be noted that these settings were suggested in [7] and [8]. The echo return loss enhancement (ERLE) and the normalized squared coefficient error (NSCE), used to evaluate the performance, were defined as

$$ERLE(n) = 10\log_{10}\frac{(y(n)-s(n))^2}{(e(n)-s(n))^2}, \tag{20}$$

and

$$NSCE(n) = 10\log_{10}\frac{\|\mathbf{h}_{opt}(n)-\mathbf{h}(n)\|^2}{\|\mathbf{h}_{opt}(n)\|^2}, \tag{21}$$

respectively. Note that we have included $s(n)$ in (20) so as to better evaluate the performance of AEC during DT stages.

The first experiment had DT periods from 11 to 14 seconds, and the acoustic echo path $\mathbf{h}_{opt}(n)$ was circularly shifted by 200 samples at time 5.3. Far-end speech $x(n)$ was shown in Fig. 2. Simulation results are illustrated in

Figs. 3-5. Fig. 3 shows that the Park's method made a few wrong decisions to declare DT mode when echo-path changed. Fortunately, this did not cause any extra ERLE loss comparing to the time-domain DTD [7]. During DT stage, Park's method declared single talk status from 13.84 to 14 seconds. Note that there was a slight decision delay occurring at the beginning of DT stage for all three methods. As a result, the error increased dramatically at the beginning of DT mode and all methods had an NSCE sudden drop by about 20dB as demonstrated in Fig. 4. In spite of this, the proposed AEC and the time-domain DTD [7] efficiently recovered back to good level of NSCE and ERLE in less than 1.0 second while Park's DTD performed badly during DT stage. Fig. 5 illustrates that our method is about 30dB ERLE better than the other two methods in the beginning of DT mode. Figs. 4 and 5 clearly show that our AEC has the best performance in tracking and convergence.

We conducted another experiment involving abrupt echo path change: $\mathbf{h}_{opt}(n)$ was switched to $-\mathbf{h}_{opt}(n)$ at time 5.3. The DT phase is set from 11 to 14 seconds. Simulation results are illustrated in Figs. 6-8. Fig. 6 shows that Park's scheme made several incorrect decisions to declare DT mode and was dead there during 5.3 to 9.75 second. The damage was obvious: adaptive filter was frozen in that period. Our method and the time-domain DTD [7] performed similarly to that they behaved in the first experiment. We believe the good properties of our frequency-domain DTD and the time-domain DTD [7] are due to the employment of two auxiliary filters.

6. Conclusions

In this paper, we presented a new frequency-domain DTD combined with the regularized GMDF algorithm. The DTD comprises an NEVD indicator and two auxiliary filters. The auxiliary filters are only to store estimated echo path for later use. Simulation results demonstrated that our DTD was capable of distinguishing echo path changes from double-talk situation, and outperformed other competitive algorithms. We also made a comparison study of these algorithms in computational complexity. It was obvious that our scheme prevailed in this category as well.

References

[1] P. Ahgren and A. Jakobsson, "A study of doubletalk detection performance in the presence of acoustic echo path changes," Proceedings of 2005 IEEE ICASSP, Vol. III, pp. 141-144, 2005.

[2] J. Benesty and T. Gansler, "A multidelay double-talk detector combined with the MDF adaptive filter," EURASIP Journal on Applied Signal Processing, pp. 1056-1063, 2003.

[3] J. Benesty, D. R. Morgan, and J. H. Cho, "A new class of doubletalk detectors based on cross-correlation." IEEE Transactions on Speech and Audio Processing, Vol. 8, No. 2, pp. 168-172, Mar. 2000.

[4] J. C. Jeng and S. F. Hsieh, "Decision of double-talk and time-variant echo path for acoustic echo cancellation," IEEE Signal Processing Letters, Vol. 10, No. 11, pp. 317-319, Nov. 2003.

[5] J. Lee and H. C. Huang, "A New Frequency-domain Regularization for the GMDF Algorithm," proceedings of IMECS, pp 1495-1498, Mar. 19-21, 2008.

[6] J. Lee and H. C. Huang, "On the Fixed Common Step-Size of the Frequency-Domain Normalized Generalized Multidelay Adaptive Filter," IAENG International Journal of Computer Science, Vol. 35, Issue 1, pp 140-144, Mar. 2008.

[7] J. Lee and H. C. Huang, "A robust double-talk detector for acoustic echo cancellation," proceedings of IMECS 2010, pp 1239-1242, Mar. 2010.

[8] S. J. Park, C. G. Cho, C. Lee, and D. H. Youn, "Integrated echo and noise canceler for hands-free applications," IEEE Transactions on Circuits and Systems II, Vol. 49, Issue 3, pp. 186-195, March 2002.

[9] H. Ye and B. X. Wu, "A new double-talk detection algorithm based on the orthogonality theorem," IEEE Transactions on Communications, Vol. 39, pp. 1542-1545, Nov. 1991.

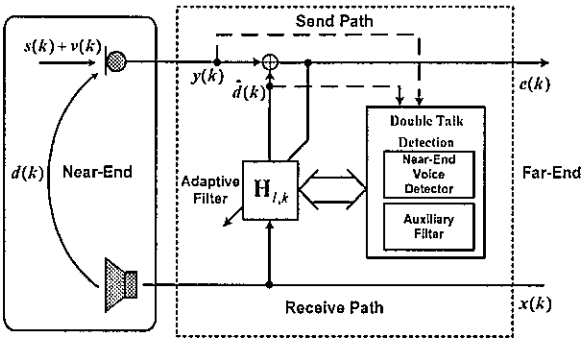

Fig. 1, Basic structure of acoustic echo cancellation with double-talk detection.

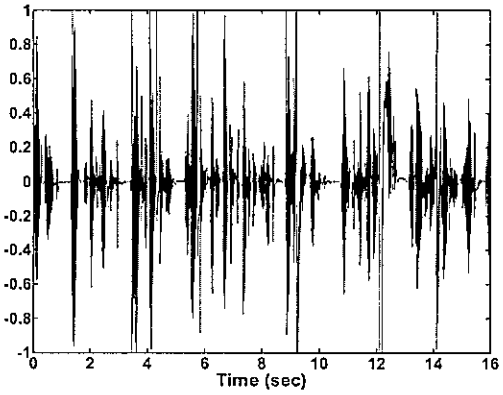

Fig. 2, Far-end speech used in the simulation

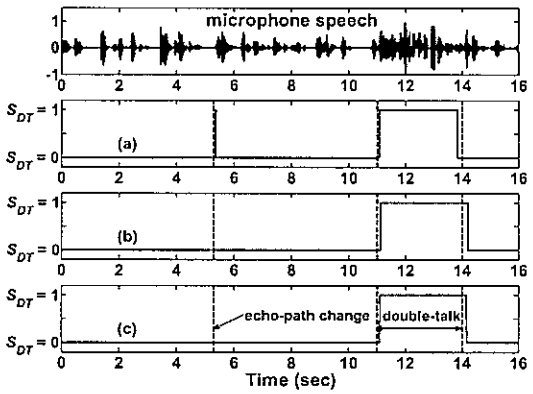

Fig. 3, DTD decision results of (a) Park's method, (b) time-domain DTD [7], and (c) our proposed method.

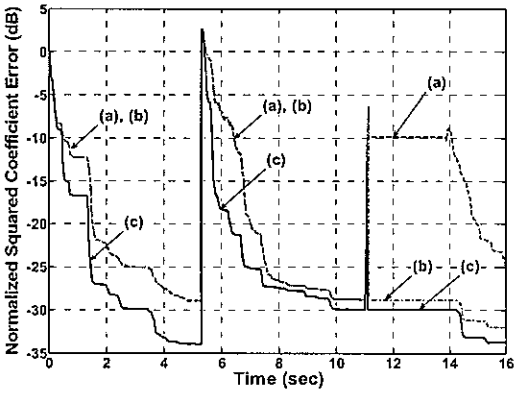

Fig. 4, NSCE curves of (a) Park's method, (b) time-domain DTD [7], and (c) our proposed method.

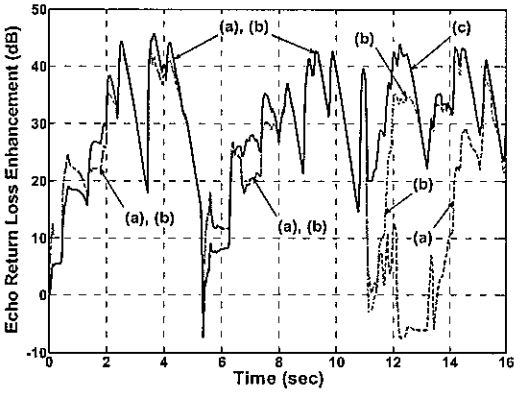

Fig. 5, ERLE curves of (a) Park's method, (b) time-domain DTD [7], and (c) our proposed method.

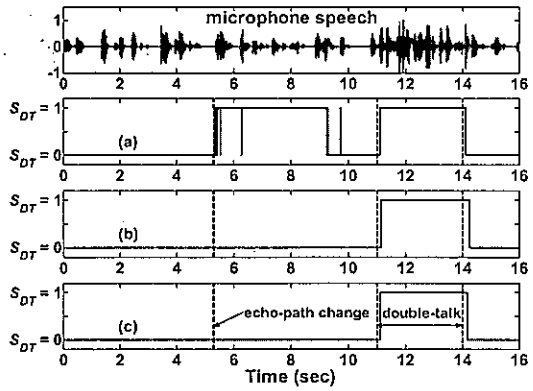

Fig. 6, DTD decision results of (a) Park's method, (b) time-domain DTD [7], and (c) our proposed method.

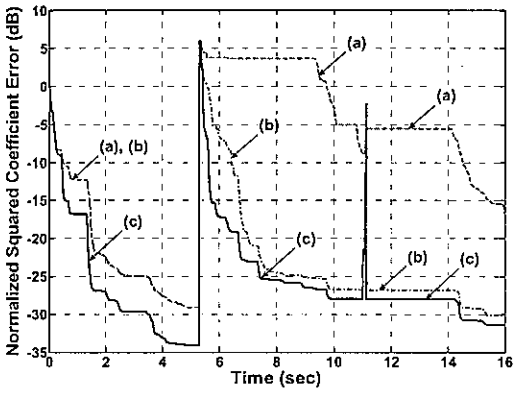

Fig. 7, NSCE curves of (a) Park's method, (b) time-domain DTD [7], and (c) our proposed method.

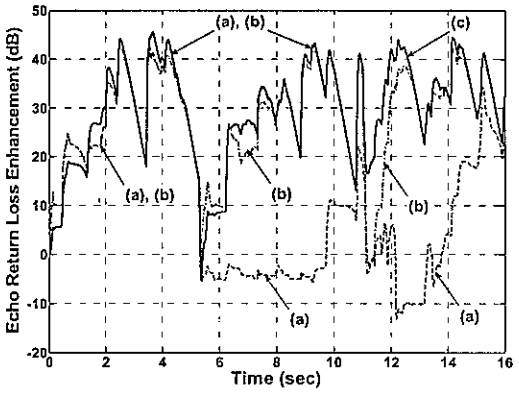

Fig. 8, ERLE curves of (a) Park's method, (b) time-domain DTD [7], and (c) our proposed method.