LATTICE-LADDER DECORRELATION FILTERS DEVELOPED FOR CO-CHANNEL SPEECH SEPARATION

Kuan-Chieh Yen[1] *Yunxin Zhao*[2]

[1]VerbalTek, Inc., Santa Clara, CA 95051, USA
[2]Dept. of CECS, University of Missouri, Columbia, MO 65211, USA
[1]*kyen@verbaltek.com* [2]*zhao@cecs.missouri.edu*

ABSTRACT

The previously proposed lattice-ladder adaptive decorrelation filtering (LL-ADF) algorithm is further studied and improved in this work, with the aim of developing a more efficient co-channel speech separation system. The effect of the joint linear predictions is first analyzed and the conversions between the lattice coefficients and the prediction and filter vectors are formulated. The implementation issues on the estimation of lattice coefficients are then discussed and the adaptation equations are further refined. Experimental results demonstrate the effectiveness of the algorithm in reducing cross-interference between co-channel speech sources as well as the significant performance improvement over the previous direct-form ADF algorithm. A simplified LL-ADF is also proposed as a compromise between computational cost and system performance.

1. INTRODUCTION

A number of co-channel speech separation algorithms utilizing multi-microphone acquisition and second-order statistics have been proposed in the literature [1][2] and shown effective in separating speech sources from their convolutive mixtures in a variety of applications. In our previous work [3], a co-channel speech separation system was developed based on the adaptive decorrelation filtering (ADF) algorithm [1]. The separation system significantly improved accuracy on co-channel speech recognition as well as intelligibility on clinical subject listening test [4]. The current ADF algorithm has several shortcomings due to the use of direct FIR form in modeling acoustic paths. First of all, it is often difficult to know the required filter length in advance. Secondly, the impulse responses of acoustic paths are usually sparsely spaced, making the direct FIR implementation inefficient. Finally, a slight change in acoustic environment can cause significant changes in FIR coefficients and it can take a long time for the adaptation algorithm to adjust for such a change.

Lattice-ladder structure has been widely used in adaptive signal processing such as linear prediction and noise cancellation [5][6]. Lattice-ladder filters are modular in structure, so that additional stages can be added when necessary without affecting the earlier stages [6]. This provides the potential flexibility to adapt filter length according to environment. Furthermore, when converting a direct-form FIR filter into the lattice-ladder filter, the filter coefficients become more evenly distributed, which enables more efficient filter estimation and adaptation.

In our previous work [7], a lattice-ladder ADF (LL-ADF) algorithm is proposed with the aim of developing a more efficient separation system. It is intended in this work to further study and improve on the previous results. This paper is organized into six sections. In Section 2, co-channel speech separation problem is discussed and the the idea of applying lattice-ladder decorrelation filters in constructing a separation system is introduced. In Section 3, the joint linear predictions defined in [7] is reviewed and their effect on the input signals is analyzed. The lattice-ladder decorrelation filter [7] developed from the joint linear predictions is then described and the conversions between the lattice coefficients and the prediction and filter vectors are formulated. The implementation issues on the estimation algorithm of the lattice coefficients are addressed and the estimation equations are further refined in Section 4. Experimental results are presented in Section 5 and a conclusion is made in Section 6.

2. CO-CHANNEL MODEL AND CO-CHANNEL SPEECH SEPARATION

Assuming that two speech sources exist in a co-channel environment and that two microphones are used to acquire the speech signals. Denoting the speech signal generated by speech source j as $x_j(t)$, $j = 1, 2$ and the signal acquired by microphone i as $y_i(t)$, $i = 1, 2$, the co-channel environment can be modeled in the frequency domain as

$$\begin{aligned} Y_1(f) &= H_{11}(f)X_1(f) + H_{12}(f)X_2(f) \\ Y_2(f) &= H_{22}(f)X_2(f) + H_{21}(f)X_1(f) \end{aligned} \tag{1}$$

where $H_{ij}(f)$ represents the transfer function that models the acoustic path from source j to microphone i. It is shown in [1] that by using a separation system

$$\begin{aligned} V_1(f) &= Y_1(f) - F_{12}(f)Y_2(f) \\ V_2(f) &= Y_2(f) - F_{21}(f)Y_1(f) \end{aligned} \tag{2}$$

which generates output signals $v_i(t)$'s, the speech signals can be separated if the filters F_{ij}'s satisfy $F_{ij} = H_{ij}/H_{jj}$. Since H_{ij}'s are usually unknown and time-varying, F_{ij}'s need to be estimated and tracked.

Assuming that $x_j(t)$'s are zero-mean and uncorrelated, the separated signals should also be zero-mean and uncorrelated. In many situations, decorrelation between $v_i(t)$'s is a simple and effective criterion for estimating F_{ij}'s. Define the length-m vector of a signal $x(t)$ as $\underline{x}_m(t) = [x(t), \cdots, x(t-m+1)]^T$ and choose filters F_{ij}'s to be length-M FIR filters with coefficients $\underline{f}_{ij} = [f_{ij}(0), \cdots, f_{ij}(M-1)]^T$. It was shown in [1] that for $v_i(t)$'s to be uncorrelated, $\underline{f}_{ij}$'s need to satisfy

$$\underline{f}_{ij} = E\left\{\underline{v_j}_M(t)\underline{y_j}_M^T(t)\right\}^{-1} E\left\{\underline{v_j}_M(t)y_i(t)\right\} \tag{3}$$

A lattice-ladder implementation of this separation system is proposed in [7] where two length-M decorrelation filters are applied parallelly. Each decorrelation filter estimates its ideal filter coefficients $\underline{h}^* = E\left\{\underline{v}_M(t)\underline{y}_M^T(t)\right\}^{-1} E\left\{\underline{v}_M(t)d(t)\right\}$ from its three input signals $d(t)$, $y(t)$, and $v(t)$, and generates an output signal $z(t) = d(t) - \underline{h}^T\underline{y}_M(t)$ that is uncorrelated to $v(t)$. By using $y_i(t)$, $y_j(t)$, and $v_j(t)$ in places of $d(t)$, $y(t)$, and $v(t)$, F_{ij} can be implemented and the resulting output is $v_i(t)$. The block diagram of this system is shown in Fig. 1.

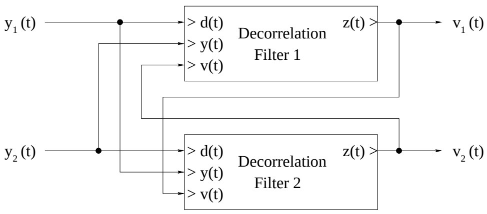

Figure 1. Block diagram of the co-channel speech separation system based on LL-ADF

3. DECORRELATION FILTERING WITH LATTICE-LADDER STRUCTURE

Let the random processes $d(t)$ and $y(t)$ be convolutive mixtures of two zero-mean and uncorrelated signals $x_1(t)$ and $x_2(t)$ and let the random process $v(t)$ be a linear transformation of $x_2(t)$. It is shown in [7] that a length-M decorrelation filter $\underline{h}$ can be estimated using an adaptive filtering process with lattice-ladder structure. The decorrelation filter $\underline{h}$ matches the x_2-component in $y(t)$ to the one in $d(t)$ such that the output signal $z(t) = d(t) - \underline{y}_M^T(t)\underline{h}$ contains only x_1-component and is thus uncorrelated to $v(t)$.

3.1. Joint Linear Predictions

The lattice-ladder structure is derived based on the joint forward and backward linear predictions, where the mth order forward prediction errors are defined as

$$\begin{aligned} e_{pf,m}(t) &= y(t) - \underline{y}_m^T(t-1)\underline{f}_{p,m} \\ e_{qf,m}(t) &= v(t) - \underline{v}_m^T(t-1)\underline{f}_{q,m} \end{aligned} \quad (4)$$

and the mth order backward prediction errors are defined as

$$\begin{aligned} e_{pb,m}(t) &= y(t-m) - \underline{y}_m^T(t)\underline{b}_{p,m} \\ e_{qb,m}(t) &= v(t-m) - \underline{v}_m^T(t)\underline{b}_{q,m} \end{aligned} \quad (5)$$

The ideal values of the mth order prediction vectors are

$$\begin{aligned} \underline{f}_{p,m}^* &= \mathbf{P}_m^{-1}\underline{p}_{f,m} \quad , \quad & \underline{f}_{q,m}^* &= \mathbf{P}_m^{-T}\underline{q}_{f,m} \\ \underline{b}_{p,m}^* &= \mathbf{P}_m^{-1}\underline{p}_{b,m} \quad , \quad & \underline{b}_{q,m}^* &= \mathbf{P}_m^{-T}\underline{q}_{b,m} \end{aligned} \quad (6)$$

in order to satisfy the decorrelation conditions

$$\begin{aligned} & E\left\{\underline{v}_m(t-1)e_{pf,m}(t)\right\} = E\left\{\underline{v}_m(t)e_{pb,m}(t)\right\} \\ & = E\left\{\underline{y}_m(t-1)e_{qf,m}(t)\right\} = E\left\{\underline{y}_m(t)e_{qb,m}(t)\right\} = \underline{0} \end{aligned} \quad (7)$$

Here the mth order cross-correlation matrix $\mathbf{P}_m$ is defined as $E\left\{\underline{v}_m(t)\underline{y}_m^T(t)\right\}$ and the mth order cross-correlation vectors $\underline{p}_{f,m}$, $\underline{p}_{b,m}$, $\underline{q}_{f,m}$, and $\underline{q}_{b,m}$ are defined as $E\left\{\underline{v}_m(t-1)y(t)\right\}$, $E\left\{\underline{v}_m(t)y(t-m)\right\}$, $E\left\{\underline{y}_m(t-1)v(t)\right\}$, and $E\left\{\underline{y}_m(t)v(t-m)\right\}$, respectively. When Eq. (6) is satisfied, the cross-correlations $E_{f,m} = E\left\{e_{pf,m}(t)e_{qf,m}(t)\right\}$ and $E_{b,m} = E\left\{e_{pb,m}(t)e_{qb,m}(t)\right\}$ have the same value, i.e.,

$$E_{f,m}^* = E_{b,m}^* = r_{yv}(0) - \underline{q}_{b,m}^T\mathbf{P}_m^{-1}\underline{p}_{b,m} \equiv \mathcal{E}_m \quad (8)$$

Since the goals of $\underline{f}_{p,m}$ and $\underline{f}_{q,m}$ are to make $E\left\{\underline{v}_m(t-1)e_{pf,m}(t)\right\} = \underline{0}$, and $E\left\{\underline{y}_m(t-1)e_{qf,m}(t)\right\} = \underline{0}$, when $m \to \infty$ and the forward prediction vectors are equal to their ideal values, the forward prediction errors should satisfy

$$E\left\{e_{pf,\infty}(t)v(t-\tau)\right\} = E\left\{e_{qf,\infty}(t)y(t-\tau)\right\} = 0, \quad \tau > 0 \quad (9)$$

By combining Eqs. (4) and (9), it is straightforward to show that

$$E\left\{e_{pf,\infty}(t)e_{qf,\infty}(t-\tau)\right\} = \mathcal{E}_\infty\delta(\tau) \quad (10)$$

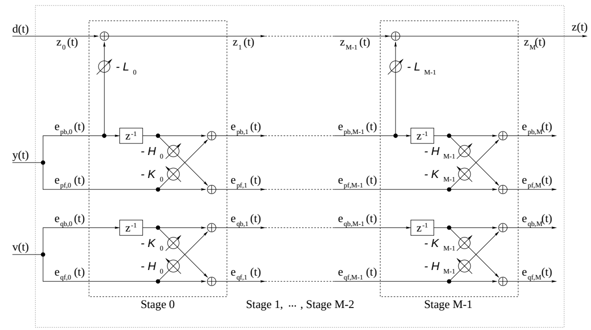

Figure 2. Diagram of the lattice-ladder structure

Therefore, the cross-spectrum between $e_{pf,m}(t)$ and $e_{qf,m}(t)$ is gradually whitened as m increases. Because $e_{pf,m}(t)$ and $e_{qf,m}$ are linear transformations of $y(t)$ and $v(t)$, respectively, this is equivalent to whitening the cross-spectrum of $y(t)$ and $v(t)$. Similarly, it can be shown that

$$E\left\{e_{pb,\infty}(t)e_{qb,\infty}(t-\tau)\right\} = \mathcal{E}_\infty\delta(\tau) \quad (11)$$

and the joint backward linear prediction also whitens the cross-spectrum of $y(t)$ and $v(t)$. Because $v(t)$ contains only x_2-component, the whitening process on the cross-spectrum of $y(t)$ and $v(t)$ is, in fact, the whitening process on $x_2(t)$.

3.2. Lattice-Ladder Structure

To facilitate the construction of lattice ladders, the mth order stage output is defined as

$$z_m(t) = d(t) - \underline{y}_m^T(t)\underline{h}_m \quad (12)$$

To satisfy the decorrelation condition

$$E\left\{\underline{v}_m(t)z_m(t)\right\} = \underline{0} \quad (13)$$

the ideal value of the mth order decorrelation filter $\underline{h}_m$ is

$$\underline{h}_m^* = \mathbf{P}_m^{-1}\underline{s}_m \quad (14)$$

where $\underline{s}_m$ is defined as $E\left\{\underline{v}_m(t)d(t)\right\}$.

It is shown in [7] that the prediction errors and the stage output in each stage can be calculated from the ones in the previous stage as

$$\begin{aligned} e_{pf,m+1}(t) &= e_{pf,m}(t) - \mathcal{H}_m e_{pb,m}(t-1) \\ e_{pb,m+1}(t) &= e_{pb,m}(t-1) - \mathcal{K}_m e_{pf,m}(t) \\ e_{qf,m+1}(t) &= e_{qf,m}(t) - \mathcal{K}_m e_{qb,m}(t-1) \\ e_{qb,m+1}(t) &= e_{qb,m}(t-1) - \mathcal{H}_m e_{qf,m}(t) \\ z_{m+1}(t) &= z_m(t) - \mathcal{L}_m e_{pb,m}(t) \end{aligned} \quad (15)$$

where $\mathcal{H}_m$, $\mathcal{K}_m$, and $\mathcal{L}_m$ are the lattice coefficients in the mth stage with ideal values

$$\begin{aligned} \mathcal{H}_m^* &= \left(r_{yv}(m+1) - \underline{q}_{b,m}^T\mathbf{P}_m^{-1}\underline{p}_{f,m}\right)\mathcal{E}_m^{-1} \\ \mathcal{K}_m^* &= \left(r_{yv}(-m-1) - \underline{q}_{f,m}^T\mathbf{P}_m^{-1}\underline{p}_{b,m}\right)\mathcal{E}_m^{-1} \\ \mathcal{L}_m^* &= \left(r_{dv}(m) - \underline{q}_{b,m}^T\mathbf{P}_m^{-1}\underline{s}_m\right)\mathcal{E}_m^{-1} \end{aligned} \quad (16)$$

From Eqs. (4) and (5), the inputs to the initial stage are $e_{pf,0}(t) = e_{pb,0}(t) = y(t)$, $e_{qf,0}(t) = e_{qb,0}(t) = v(t)$, and $z_0(t) = d(t)$. On the other hand, the output at the final stage, $z_M(t)$, is the desired output $z(t)$. The diagram of the lattice-ladder structure defined by Eq. (15) is shown in Fig. 2.

3.3. Conversion Between the Lattice Coefficients and the Prediction and Filter Vectors

Because the filtering functions performed by $\mathcal{H}_i$, $\mathcal{K}_i$, and $\mathcal{L}_i$, $\forall i \le m$ are equivalent to those performed by $\underline{f}_{p,m+1}$, $\underline{b}_{p,m+1}$, $\underline{f}_{q,m+1}$, $\underline{b}_{q,m+1}$, and $\underline{h}_{m+1}$, there exist a one-on-one conversion between the two sets of parameters. The conversion algorithms are described below.

3.3.1. From the lattice coefficients to the vectors

If the lattice coefficients $\mathcal{H}_i$, $\mathcal{K}_i$, and $\mathcal{L}_i$, $\forall i \leq m$ are available, the vectors $\underline{f}_{p,i}$, $\underline{b}_{p,i}$, $\underline{f}_{q,i}$, $\underline{b}_{q,i}$, and $\underline{h}_i$, $\forall i \leq m+1$ can be computed using the following procedure:

Step 1. Set $i = 1$ and initialize the vectors of order 1 as $\underline{f}_{p,1} = \underline{b}_{q,1} = [\mathcal{H}_0]$, $\underline{b}_{p,1} = \underline{f}_{q,1} = [\mathcal{K}_0]$, and $\underline{h}_1 = [\mathcal{L}_0]$.

Step 2. Computing the vectors of order $i+1$ from the vectors and lattice coefficients of order i as

$$\begin{aligned}
\underline{f}_{p,i+1} = \underline{b}_{q,i+1}^{\uparrow\downarrow} &= \left[\frac{\underline{f}_{p,i}}{0}\right] + \mathcal{H}_i \left[\frac{-\underline{b}_{p,i}}{1}\right] \\
\underline{b}_{p,i+1} = \underline{f}_{q,i+1}^{\uparrow\downarrow} &= \left[\frac{0}{\underline{b}_{p,i}}\right] + \mathcal{K}_i \left[\frac{1}{-\underline{f}_{p,i}}\right] \\
\underline{h}_{i+1} &= \left[\frac{\underline{h}_i}{0}\right] + \mathcal{L}_i \left[\frac{-\underline{b}_{p,i}}{1}\right]
\end{aligned}$$

Here $\uparrow\downarrow$ denotes reverse vector entry order.

Step 3. If $i = m$, stop. Otherwise increase i by 1 and go to Step 2.

3.3.2. From the vectors to the lattice coefficients

If $\underline{f}_{p,m}$ (or $\underline{b}_{q,m}$), $\underline{b}_{p,m}$ (or $\underline{f}_{q,m}$), and $\underline{h}_m$ are available, $\mathcal{H}_i$, $\mathcal{K}_i$, and $\mathcal{L}_i$, $\forall i \leq m-1$ can be computed using the following procedure:

Step 1. Set $i = m$.

Step 2. Set the lattice coefficients of order $i-1$ as $\mathcal{H}_{i-1} = f_{p,i}(i)$, $\mathcal{K}_{i-1} = b_{p,i}(1)$, and $\mathcal{L}_{i-1} = h_i(i)$.

Step 3. For $j = 1, 2, \cdots, i-1$, compute the entries in the prediction and filter vectors of order $i-1$ as

$$\begin{aligned}
f_{p,i-1}(j) = b_{q,i-1}(i-j) &= \tfrac{f_{p,i}(j) + f_{p,i}(i) b_{p,i}(j+1)}{1 - f_{p,i}(i) b_{p,i}(1)} \\
b_{p,i-1}(j) = f_{q,i-1}(i-j) &= \tfrac{b_{p,i}(j+1) + b_{p,i}(1) f_{p,i}(j)}{1 - f_{p,i}(i) b_{p,i}(1)} \\
h_{i-1}(j) = h_i(j) &+ \tfrac{h_i(i)\left[b_{p,i}(j+1) + b_{p,i}(1) f_{p,i}(j)\right]}{1 - f_{p,i}(i) b_{p,i}(1)}
\end{aligned}$$

Step 4. If $i = 1$, stop. Otherwise decrease i by 1 and go to Step 2.

It should be noted that the prediction and filter vectors of lower orders are also computed by the procedure.

4. ESTIMATION OF LATTICE COEFFICIENTS

In the lattice-ladder structure discussed above, the ideal values of the lattice coefficients in each stage need to be estimated. Based on the decorrelation criteria, estimation equations for the lattice coefficients were derived in [7] as

$$\begin{aligned}
\mathcal{H}_m^{(t+1)} &= \mathcal{H}_m^{(t)} + \mu_{\mathcal{H}_m} \left[\Delta_{b,m} e_{pf,m+1}(t) e_{qb,m}(t-1) \right. \\
&\qquad \left. + \Delta_{f,m} e_{pf,m}(t) e_{qb,m+1}(t)\right] \\
\mathcal{K}_m^{(t+1)} &= \mathcal{K}_m^{(t)} + \mu_{\mathcal{K}_m} \left[\Delta_{b,m} e_{qf,m+1}(t) e_{pb,m}(t-1) \right. \\
&\qquad \left. + \Delta_{f,m} e_{qf,m}(t) e_{pb,m+1}(t)\right] \\
\mathcal{L}_m^{(t+1)} &= \mathcal{L}_m^{(t)} + \mu_{\mathcal{L}_m} \Delta_{b,m} e_{qb,m}(t) z_{m+1}(t)
\end{aligned} \tag{17}$$

where $\Delta_{b,m} = sign\{E_{b,m}\}$ and $\Delta_{f,m} = sign\{E_{f,m}\}$. However, two critical implementing issues still need to be addressed: (1) the choice of adaptation gains $\mu_{\mathcal{H}_m}$, $\mu_{\mathcal{K}_m}$, and $\mu_{\mathcal{L}_m}$; and (2) the estimation of $\Delta_{f,m}$ and $\Delta_{b,m}$. Because we use the same lattice coefficient $\mathcal{H}_m$ in the order adaptation of both $e_{pf,m}(t)$ and $e_{qb,m}(t)$ and the same $\mathcal{K}_m$ in that of both $e_{pb,m}(t)$ and $e_{qf,m}(t)$, it can be shown that $\underline{b}_{q,m} = \underline{f}_{p,m}^{\uparrow\downarrow}$, $\underline{b}_{p,m} = \underline{f}_{q,m}^{\uparrow\downarrow}$, and, as a result, $E_{f,m} = E_{b,m}$. Therefore, both $E_{f,m}$ and $E_{b,m}$ can be substituted by E_m, and both $\Delta_{f,m}$ and $\Delta_{b,m}$ can be substituted by Δ_m.

By following the analysis commonly used in the LMS algorithm [8], it can be shown that to guarantee the convergence of the lattice coefficients, the adaptation gains need to satisfy $0 < \mu_{\mathcal{H}_m}, \mu_{\mathcal{K}_m} < \Gamma_m$ and $0 < \mu_{\mathcal{L}_m} < 2\Gamma_m$, with Γ_m being defined as $|E_m|^{-1}$. Based on these bounds, the estimation equations in Eq. (17) can be revised as

$$\begin{aligned}
\mathcal{H}_m^{(t+1)} &= \mathcal{H}_m^{(t)} + \mu_m(t)\left[e_{pf,m+1}(t) e_{qb,m}(t-1) \right. \\
&\qquad \left. + e_{pf,m}(t) e_{qb,m+1}(t)\right] \\
\mathcal{K}_m^{(t+1)} &= \mathcal{K}_m^{(t)} + \mu_m(t)\left[e_{qf,m+1}(t) e_{pb,m}(t-1) \right. \\
&\qquad \left. + e_{qf,m}(t) e_{pb,m+1}(t)\right] \\
\mathcal{L}_m^{(t+1)} &= \mathcal{L}_m^{(t)} + 2\mu_m(t) e_{qb,m}(t) z_{m+1}(t)
\end{aligned} \tag{18}$$

where $\mu_m(t)$ is defined as

$$\mu_m(t) = \frac{\gamma \hat{\Delta}_m(t)}{\left|\hat{E}_m(t)\right| + \alpha\left(\hat{E}_y(t) + \hat{E}_v(t)\right)} \tag{19}$$

with: (1) γ being a constant and $0 < \gamma < 1$; (2) α being a small positive constant; (3) $\hat{\Delta}(t)$ being the estimate of Δ_m at time t using a length-L_Δ window; and (4) $\hat{E}_m(t)$, $\hat{E}_y(t)$, and $\hat{E}_v(t)$ being the estimates of E_m, $E\left\{y^2(t)\right\}$, and $E\left\{v^2(t)\right\}$ at time t using a length-L_E window. The reason for including the second term in the denominator of $\mu_m(t)$ is to prevent instability when $E_m(t)$ changes its sign and may have a small absolute value.

To tolerate estimation errors, γ is usually chosen to be significantly smaller than 1. For highly dynamic signals such as speech, typical choices of γ are around 10^{-4}. The window length L_E needs to be chosen such that reasonably good estimates of the expectations can be obtained without introducing long delays. Typical values for L_E are between 4000 to 10000 samples. Parameters L_Δ and α are unique for the LL-ADF and will be discussed in Section 5.

5. EXPERIMENTS

Several experiments were carried out to evaluate the LL-ADF algorithm. In all experiments, the speech source signals were chosen from the TIMIT database. They were convolved with the impulse responses measured from real acoustic paths and then were combined to generate the co-channel speech signals used in the experiments.

5.1. Choice of Parameters L_Δ and α

The first two experiments investigated the effects of the parameters L_Δ and α on the performance of LL-ADF. A pair of co-channel signals were processed by the LL-ADF (with $M = 200$, $\gamma = 10^{-4}$, and $L_E = 5000$). In the first experiment, α was fixed at 0.25, and L_Δ was set to 1, 2, 5, 7, 8, and 10. In the second experiment, L_Δ was fixed at 5, and α was set to 1, 0.5, 0.25, 0.1, 0.05, 0.02, and 0. The target-to-interference ratio (TIR) improvements (averaged over the two sources) were calculated every 2 s for all cases, and the results are given in Figs. 3 and 4, respectively.

From Fig. 3, using $L_\Delta = 1$ resulted in slow convergence and poor steady-state TIR improvement (TIRI). This is because of inaccurate estimation of Δ_m and high correlation between $\hat{\Delta}_m(t)$ and the prediction errors. Increasing L_Δ to 2 through 7 resulted in faster convergence and better steady-state TIRIs. However, the steady-state performance became less consistent as with further increases of L_Δ due to the longer delay in detecting the change of Δ_m. When $L_\Delta \geq 8$, the delay was so serious that the system became unstable.

The effect of α can be observed from Fig. 4. When $\alpha = 0$, the system converged fast, but the steady-state performance was poor and inconsistent. As α increased, the steady-state performance became better and more consistent, with the penalty of reduced convergence rate. The reason was twofold: (1) using a larger α reduced adaptation gain, and (2) using a larger α made the adaptation gains in different stages more similar. The lattice-ladder structure's whitening effect on the signals was thus reduced.

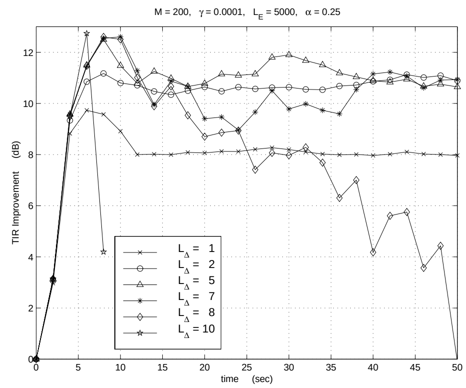

Figure 3. Convergence performance of the LL-ADF algorithm using different values of L_Δ

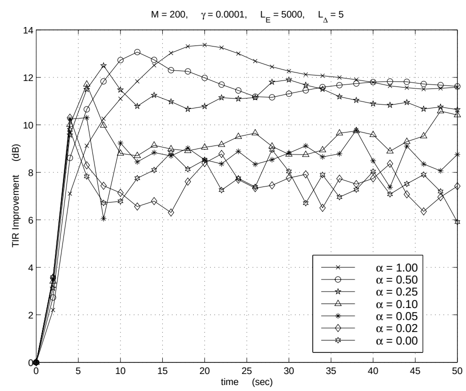

Figure 4. Convergence performance of the LL-ADF algorithm using different values of α

5.2. Steady-State Separation Performance

In the third experiment, a set of co-channel signal pairs was used. The average TIRs were 6.20 dB in y_1 and 5.09 dB in y_2. Upon applying the LL-ADF ($M = 200$, $\gamma = 10^{-4}$, $L_E = 5000$, $L_\Delta = 5$, and $\alpha = 0.25$), the average TIRs were improved to 16.90 dB in v_1 and 17.66 dB in v_2, which was an improvement of more than 10 dB for both source signals.

5.3. LL-ADF, DF-ADF, and Simplified LL-ADF

In the fourth experiment, the same co-channel signals as in Section 5.1 were processed by the LL-ADF, as well as the direct-form ADF (DF-ADF) [3]. The parameters of LL-ADF were the same as in Section 5.2. The filter length of the DF-ADF was also 200. Its adaptation gain was chosen such that the steady-state TIRI would be the same as that of the LL-ADF algorithm. The TIRIs are given in Fig. 5. Compared with the DF-ADF, the LL-ADF provides better dynamic tracking ability without sacrificing steady-state separation performance.

In Eq. (19), it is necessary to keep track of $\hat{E}_m$ and $\hat{\Delta}_m(t)$ in each stage. Since L_E is usually in the range of 4 000 to 10 000, the memory space required for evaluating $\hat{E}_m$ becomes an issue when the number of stages increases. Since the prediction errors are generated from $y(t)$ and $v(t)$, an appropriate substitute for $\left|\hat{E}_m(t)\right|$ is $\frac{1}{2}\left(\hat{E}_y(t) + \hat{E}_v(t)\right)$, and hence $\mu_m(t)$ can be simplified as

$$\mu_{simp,m}(t) = 2\gamma\hat{\Delta}_m(t)/\left(\hat{E}_y(t) + \hat{E}_v(t)\right) \qquad (20)$$

which saves a significant amount of memory space and computational overhead.

To compare the performance of the simplified LL-ADF with that of the original one, the same co-channel signals

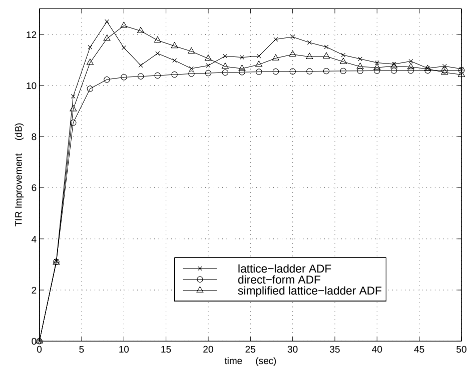

Figure 5. Convergence performance: A comparison among LL-ADF, DF-ADF, and simplified LL-ADF

were also processed by simplified LL-ADF. To achieve the same steady-state TIRI, γ was chosen as 0.08 for $\mu_{simp,m}(t)$, with other parameters remaining the same. The TIRI is also given in Fig. 5. It shows that the simplified LL-ADF converged slightly slower than the LL-ADF, but was still significantly faster than the DF-ADF.

6. CONCLUSION

In this work, the LL-ADF proposed in [7] is further studied. The effect of the joint linear predictions is first analyzed. The conversions between the lattice coefficients and the prediction and filter vectors are then formulated. Finally, the implementation issues on the adaptation algorithm of lattice coefficients are discussed and the algorithm is further refined. Experiments demonstrate the effectiveness of the proposed algorithm in reducing cross-interference between co-channel speech sources as well as the significant performance improvement over the previous direct-form ADF algorithm. A simplified LL-ADF is also proposed as a compromise between computational cost and system performance.

ACKNOWLEDGMENT

This material is based upon work supported by the National Science Foundation under Grant Nos. IIS-99-96042 and EIA-99-11095.

REFERENCES

[1] E. Weinstein, M. Feder, and A. V. Oppenheim, "Multi-Channel Signal Separation by Decorrelation," *IEEE Trans. on SAP*, Vol. 1, No. 4, pp. 405-413, Oct. 1993.

[2] S. Van Gerven and D. Van Compernolle, "Signal Separation by Symmetric Adaptive Decorrelation: Stability, Convergence, and Uniqueness," *IEEE Trans. on SP*, Vol. 43, No. 7, pp. 1602-1612, July 1995.

[3] K. Yen and Y. Zhao, "Adaptive Co-Channel Speech Separation and Recognition," *IEEE Trans. on SAP*, Vol. 7, No. 2, pp. 138-151, March 1999.

[4] Y. Zhao, K. Yen, S. Soli, S. Gao, and A. Vermiglio, "Evaluation of Adaptive Decorrelation Filtering as an Assistive Listening Device," submitted to *JASA*.

[5] L. J. Griffiths, "An Adaptive Lattice Structure for Noise-Cancelling Applications," in *Proc. ICASSP*, pp. 87-90, 1978.

[6] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, 1996.

[7] K. Yen and Y. Zhao, "Lattice-Ladder Structured Adaptive Decorrelation Filtering for Co-Channel Speech Separation," in *Proc. ICASSP*, Vol. I, pp. 388-391, 2000.

[8] B. Widrow and S. D. Sterns, *Adaptive Signal Processing*, Prentice-Hall, 1985.