DEREVERBERATION FOR REVERBERATION-ROBUST MICROPHONE ARRAYS

Takuya Yoshioka, Tomohiro Nakatani

NTT Communication Science Laboratories, NTT Corporation
2-4, Hikari-dai, Seika-cho, Soraku-gun, Kyoto 619-0237, Japan

ABSTRACT

This paper discusses using dereverberation technologies to improve the robustness of beamforming microphone arrays against reverberation. It is known that, while conventional adaptive beamformers are able to effectively cancel the interference from spatially separate noise sources in the absence of reverberation, their efficacy deteriorates in reverberant rooms. To provide a widely applicable solution to this problem, we propose an adaptive multiple point blind equalizer that can shorten all the room impulse responses between the sound sources and microphones. This algorithm reduces the impact of reverberation on the interference cancellation performance when it is used to pre-process the microphone signals prior to beamforming. Experimental results using real conversation data show that a system combining the proposed equalizer and the adaptive beamformer improves noise reduction performance without degrading the speech quality.

***Index Terms*—** Microphone array, adaptive beamformer, reverberation, sound field equalization, weighted prediction error minimization

1. INTRODUCTION

High performance and portability are desired attributes for many consumer products, and this also applies to microphone systems. That is, microphone systems should be able to capture target sound sources with a high signal-to-noise ratio. At the same time, they should be small enough to be placed on tables or even held in the hand. In applications where the capture of the target source is hampered by spatially separate noise sources, adaptive beamformers have been utilized to minimize such degradation [1,2]. Unlike delay-and-sum beamformers, this type of beamformer can cancel out the interference from localized noise sources even when microphone arrays are small.

One major problem that precludes wider use of the adaptive beamformers is that the performance of conventional beamforming algorithms deteriorates in the presence of reverberation. This results from the fact that these algorithms basically rely on the assumption that acoustic waves travel directly from the sources to the microphones. This assumption is not satisfied in reverberant rooms, where the wavefront of each source sound is reflected repeatedly. Although the relative transfer function approach can alleviate this problem [3], this approach leaves reverberation untouched and therefore may not always produce satisfactory sound quality.

This paper discusses the pre-processing of microphone signals with dereverberation technologies to make adaptive beamformers robust against reverberation. According to the review in [4], two major approaches to dereverberation are spectral subtraction [5–7] and blind equalization [8]. The former approach does not recover the phase characteristics, which are crucial for many microphone array systems. On the other hand, the blind equalization approach attempts to recover both the amplitude and phase characteristics. It equalizes or shortens, as far as possible, all the impulse responses between the sound sources and the microphones by using finite impulse response (FIR) filters that have the same number of input and output channels. In other words, this approach self-optimizes the filter coefficients to virtually create "anechoic spots" at all the source positions. To highlight the fact that equalization is performed jointly for all the sources, we use the term "multiple point equalization". However, existing algorithms in this class are based on batch processing. Therefore, they cannot be used for applications where online operation is essential.

This paper presents a principle and an algorithm for adaptive multiple point blind equalization. In Section 2, we present the principle of adaptive multiple point blind equalization and a novel optimization algorithm for the equalization filter coefficients by extending the work of [8]. One desirable property of the proposed algorithm is its ability to handle multiple sound sources, which contrasts with most of the conventional work on equalization-based dereverberation [9–12]. Section 3 evaluates the degree to which the performance of beamforming microphone arrays is improved by pre-processing microphone signals using the proposed algorithm. Finally, we comment on directions for future study in Section 4.

First of all, let us clarify the strategy taken in this work. As stated earlier, this paper seeks an approach that uses dereverberation technologies to pre-process microphone signals for a successive beamforming step. One of the reviewers of this paper has argued that joint optimization of the equalizer and the beamformer would lead to higher dereverberation and noise reduction performance. Indeed, several authors showed that such a joint operation improves both dereverberation and noise reduction performance [13,14]. (Note that these results were obtained for batch processing.) We agree that such tight coupling of the equalizer and the beamformer will achieve the best performance. However, such a joint operation essentially requires us to tailor the dereverberation algorithm to each specific type of beamformer, which impairs flexibility. Rather the goal of the paper is to develop an equalization algorithm appplicable to various types of beamformers. This is why we perform equalization and beamforming in tandem.

2. ADAPTIVE MULTIPLE POINT BLIND EQUALIZATION

Figure 1 shows the overall structure of the proposed microphone system for use with two microphones. First, each mirophone signal passes through a filter-bank analyzer and is decomposed into subband signals. In each subband, the subband signals from all the microphones are jointly dereverberated by an adaptive equalizer. Subsequently, an adaptive beamformer carries out noise reduction for the equalizer output. Any beamforming algorithms can be employed, including the Frost beamformer, the generalized sidelobe canceller, and independent component analysis. Finally, the beamformer outputs from different subbands are combined by a filter-bank synthe-

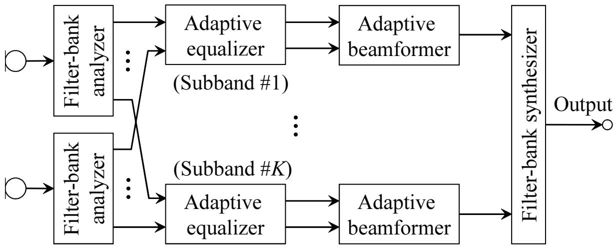

Fig. 1. Schematic of proposed microphone system.

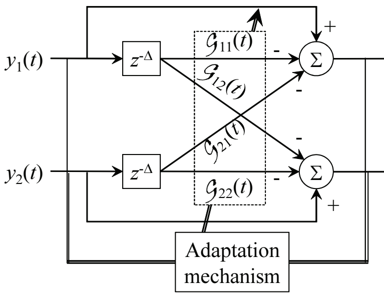

Fig. 2. Block diagram of adaptive multiple point blind equalizer.

sizer that produces a system output. Both equalization and beamforming can be run in parallel for each subband to improve the system throughput. In the following, we exclusively discuss adaptive multiple point blind equalization.

Hereafter, the following conditions are assumed for the development of the proposed equalization algorithm.

1. The locations of the target and noise sources may change slowly.
2. The microphones outnumber the sound sources.

The beamformers may also require a few additional conditions.

2.1. Multi-channel linear prediction

Let $y_1(t), \cdots, y_M(t)$ denote subband representations of the microphone signals, where M denotes the number of microphones. The subband index is not explicitly shown for the sake of conciseness. The adaptive equalizer is designed to produce M subband signals $x_1(t), \cdots, x_M(t)$ that are less reverberant than the input signals by using FIR filters. Below, superscripts T and H denote the non-conjugate and the conjugate transpose, respectively.

Figure 2 shows a block diagram of the proposed adaptive multiple point blind equalizer. This equalizer is based on multi-channel linear prediction, which is robust against filter order overestimation [15]. This approach first predicts the input signal vector from its finite past samples and then calculates the prediction residual. Thus, as shown in Fig. 2, the input signal vector, $\boldsymbol{y}(t) = [y_1(t), \cdots, y_M(t)]^T$, is delayed by Δ and fed into an M-input M-output adaptive filter of order K that predicts the input signal vector. Then, the predicted value is subtracted from the the original input to produce a vector of output signals, represented as $\boldsymbol{x}(t) = [x_1(t), \cdots, x_M(t)]^T$. Therefore, the equalizer output is calculated as

$$\boldsymbol{x}(t) = \boldsymbol{y}(t) - \sum_{k=0}^{K-1} \boldsymbol{G}_k^H(t)\boldsymbol{y}(t-k-\Delta). \qquad (1)$$

Each $\boldsymbol{G}_k(t)$ represents an M-by-M prediction filter coefficient matrix, which is updated anew at each discrete time point t according to the criterion described later.

In what follows, we represent the filter specified by $(\boldsymbol{G}_k(t))_{0 \le k \le K-1}$ as $\mathcal{G}(t)$. We also denote the (i, j)th element of the matrix filter $\mathcal{G}(t)$ as $\mathcal{G}_{ij}(t)$. We refer the reader to [16] for the theory of multi-channel linear prediction.

2.2. Optimization criterion

In this paper, we minimize the average log prediction error defined by (2) with respect to the prediction filter coefficient matrices at each discrete time point t in order to create anechoic spots at the source positions:

$$f(t) = \sum_{\tau=1}^{t} \frac{1}{t} \log \left| \det \left\langle \boldsymbol{x}(\tau)\boldsymbol{x}^H(\tau) \right\rangle_\tau \right|, \qquad (2)$$

where $\langle \cdot \rangle_t$ means an ensemble average at t. As shown in [8], this loss function relates to the degree of a temporal correlation of non-stationary multi-variate time series $(\boldsymbol{y}(t))_{t\in\mathbb{N}}$, where $\mathbb{N}$ denotes the set of all natural numbers. When we recall that, for many types of sound sources, signal samples taken from adjacent discrete time points are almost uncorrelated when the filter-bank analysis algorithm and the decimation factor are carefully chosen, we can intuitively see that minimizing this loss function would result in dereverberating the microphone signals. This applies particularly to speech signals. See [8] for a rigorous discussion.

It is noteworthy that the loss function given by (2) can also be understood from the viewpoint of sparse signal processing. The logarithmic function, $\log(x)$, decreases sharply as x approaches 0 whereas its increase is very slow for large x values. This means that mimimizing the loss function given by (2) is likely to yield sparse output signals that have many negligibly small samples. Keeping this in mind, let us recall that the smearing effect of reverberation means that the energy of reverberant speech signals tends to be broadly distributed over a time-frequency plane whereas that of anechoic signals is concentrated in limited time-frequency regions. Taking the above two facts into consideration, we may expect the minimization of (2) to make the signals less reverberant.

2.3. Batch processing-based optimization: Weighted prediction error minimization

Section 2.3 focuses on the batch processing-based optimization of the prediction filter as a preliminary to the derivation of the proposed adaptive optimization algorithm. The batch-processing algorithm proposed in [8] minimizes the loss function described above on condition that the prediction filter does not change over time. Specifically, this algorithm minimizes $f(T)$ with respect to $\mathcal{G} = (\boldsymbol{G}_k)_{1 \le k \le K}$, assuming that we observe T consecutive samples of the reverberant subband signals. Due to the time invariance assumption, the prediction filter, $\mathcal{G}$, is applied to all the discrete time points $t = 1, \cdots, T$.

The fundamental principle of this batch processing algorithm is to minimize the auxiliary loss function given below with respect alternately to $\mathcal{G}$ and $\mathcal{L} = (\boldsymbol{\Lambda}_t)_{1 \le t \le T}$, where each $\boldsymbol{\Lambda}_t$ is an M-dimensional auxiliary matrix:

$$F = \frac{1}{T} \sum_{t=1}^{T} \left\{ \mathrm{tr}\left(\boldsymbol{\Lambda}_t^{-1} \left\langle \boldsymbol{x}(t)\boldsymbol{x}(t)^H \right\rangle_t \right) - M + \log|\det \boldsymbol{\Lambda}_t| \right\}. \qquad (3)$$

The minimization of F with respect to $\mathcal{G}$ and $\mathcal{L}$ can be carried out using the following formulae:

$$\mathcal{G} = \underset{\mathcal{G}}{\text{argmin}} \frac{1}{T} \sum_{t=1}^{T} \text{tr}\left(\boldsymbol{\Lambda}_t^{-1} \left\langle \boldsymbol{x}(t)\boldsymbol{x}^H(t) \right\rangle_t\right)$$

$$\approx \underset{\mathcal{G}}{\text{argmin}} \frac{1}{T} \sum_{t=1}^{T} \text{tr}\left(\boldsymbol{\Lambda}_t^{-1} \boldsymbol{x}(t)\boldsymbol{x}^H(t)\right) \tag{4}$$

$$\boldsymbol{\Lambda}_t = \left\langle \boldsymbol{x}(t)\boldsymbol{x}(t)^H \right\rangle_t \quad \text{for } 1 \leq \forall t \leq T. \tag{5}$$

An advantage of this iterative algorithm is that optimization problem (4) can be easily solved since the loss function is quadratic. It was shown that iterating these two optimization processes monotonically decreases the original loss function given by $f(T)$.

As proposed in [8], a further simplification could be realized to reduce the computational cost. Let us assume that each $\boldsymbol{\Lambda}_t$ is a scalar matrix with a time-varying scaling factor λ_t. Then, denoting l^2-norm and an identity matrix as $||\cdot||$ and $\boldsymbol{I}$, respectively, $\boldsymbol{\Lambda}_t$ is simplified as

$$\boldsymbol{\Lambda}_t = \lambda_t \boldsymbol{I}, \tag{6}$$

with λ_t being calculated by

$$\lambda_t = \frac{||\boldsymbol{x}_t||^2}{M}. \tag{7}$$

As a result of this modification, the prediction filter update formula given by (4) is also simplified as

$$\mathcal{G} = \underset{\mathcal{G}}{\text{argmin}} \frac{1}{T} \sum_{t=1}^{T} \frac{||\boldsymbol{x}(t)||^2}{\lambda_t}. \tag{8}$$

Iterating (7) and (8) means that we solve the weighted least squares problem given by (8) at each iteration step, where the weights change from iteration to iteration according to (7). Thus, this algorithm is called weighted prediction error minimization. (The algorithm iterating (4) and (5) is correspondingly called generalized weighted prediction error minimization.)

It is important to note the similarity between the above algorithm and the maximum likelihood dereverberation method presented in [11]. This maximum likelihood method was developed on the assumption that only one sound source is present in the room. Despite this single source assumption, the resultant algorithm consists of two processing steps similar to (7) and (8). It is noteworthy that the counterpart of λ_t in this maximum likelihood method represents the magnitude of the anechoic sound at discrete time point t.

2.4. Adaptive optimization

Now, we turn our attention to the adaptive optimization of the prediction filter. The goal is to derive an efficient adaptive prediction filter optimization algorithm from (7) and (8). We achieve this by using subopitimal estimates of λ_t and the recursive least squares (RLS) method.

For the time being, we assume that the optimal λ_t value is known for each t. Then, at each discrete time point t, we simply need to minimize the weighted prediction errors averaged up to the tth time point. Specifically, we solve the following problem for $t = 1, 2, \cdots$:

$$\mathcal{G}_t = \underset{\mathcal{G}_t}{\text{argmin}} \frac{1}{t} \sum_{\tau=1}^{t} \gamma^{t-\tau} \frac{||\boldsymbol{x}(\tau)||^2}{\lambda_\tau}, \tag{9}$$

where γ is a forgetting factor and satisfies $0 < \gamma < 1$. If we use the following notations:

$$\boldsymbol{G}(t) = \begin{bmatrix} \boldsymbol{G}_0(t) \\ \vdots \\ \boldsymbol{G}_{K-1}(t) \end{bmatrix}, \quad \boldsymbol{Y}(t) = \begin{bmatrix} \boldsymbol{y}(t) \\ \vdots \\ \boldsymbol{y}(t-K+1) \end{bmatrix}, \tag{10}$$

then the algorithm can be described as shown in Table 1. Since the algorithm can be derived according to the known recipe for RLS algorithm derivation [17], we show the details in Appendix.

Table 1. Adaptive optimization algorithm for prediction filter.

1. *Initialization*
   Initialize the prediction filter and an associate covariance matrix, which we denote by $\boldsymbol{\Phi}(t)$, as $\boldsymbol{G}(0) = \boldsymbol{O}$ and $\boldsymbol{\Phi}(0) = \epsilon\boldsymbol{I}$, respectively, where ϵ is a small constant.
2. *Adaptation*
   Update the prediction filter and the covariance matrix whenever we observe a new sample of reverberant speech signal $\boldsymbol{y}(t)$ as

$$\boldsymbol{G}(t) = \boldsymbol{G}(t-1) + \boldsymbol{k}(t)\boldsymbol{x}_{\text{pred}}^H(t) \tag{11}$$

$$\boldsymbol{\Phi}(t) = \frac{1}{\gamma}\left(\boldsymbol{\Phi}(t-1) - \boldsymbol{k}(t)\boldsymbol{Y}^H(t-\Delta)\boldsymbol{\Phi}(t-1)\right), \tag{12}$$

   respectively. The auxiliary variables, $\boldsymbol{x}_{\text{pred}}(t)$ and $\boldsymbol{k}(t)$, are defined as

$$\boldsymbol{x}_{\text{pred}}(t) = \boldsymbol{y}(t) - \boldsymbol{G}^H(t-1)\boldsymbol{Y}(t-\Delta) \tag{13}$$

$$\boldsymbol{k}(t) = \frac{\boldsymbol{\Phi}(t-1)\boldsymbol{Y}(t-\Delta)}{\gamma\lambda_t + \boldsymbol{Y}^H(t-\Delta)\boldsymbol{\Phi}(t-1)\boldsymbol{Y}(t-\Delta)}, \tag{14}$$

   respectively.

Now that we have obtained a basic framework for prediction filter adaptation, we finally consider how to estimate the optimal λ_t value. Let us recall that the optimal λ_t value is the magnitude of the anechoic sound at the tth time point. Fortunately, this value can be efficiently estimated to a certain degree by using the spectral subtractive dereverberation methods [5, 7]. For example, we may use the method reported in [5], which results in the following simple estimator:

$$\lambda_t = \frac{1}{M}\left(||\boldsymbol{y}(t)||^2 - \exp^{-2\sigma\theta} ||\boldsymbol{y}(t-\Theta)||^2\right), \tag{15}$$

where σ represents a decay rate and is calculated based on reverberation time T_{60} as

$$\sigma = \frac{3}{T_{60}\log_{10}(\exp(1))}. \tag{16}$$

Constant θ is a time span outside of which source signals do not have significant auto-correlation coefficients. With speech signals, 50 ms is often chosen for this value. Θ is the number of sample intervals corresponding to θ.

In summary, the present algorithm adaptively optimizes the prediction filter by using λ_t estimated as above and the algorithm shown in Table 1. Then, for each discrete time point t, we calculate the equalizer output signal vector, $\boldsymbol{x}(t)$, as

$$\boldsymbol{x}(t) = \boldsymbol{y}(t) - \boldsymbol{G}^H(t+\alpha)\boldsymbol{y}(t-\Delta). \tag{17}$$

Fig. 3. Left: picture of four people in conversation; right: picture of microphone array. Since the conversation was recorded as part of a meeting transcription project at NTT Communication Science Laboratories, the microphone array was installed on top of an omnidirectional camera as seen in the right picture.

We employ (17), which uses a future prediction filter, instead of (1) in order to prevent the equalizer from acting unstably when the speakers suddenly change their positions.

3. EXPERIMENTAL RESULTS

We conducted an experiment to investigate how the proposed adaptive multiple point blind equalizer improves the noise reduction performance of beamforming microphone arrays. We recorded an English conversation engaged in by four people in a meeting room. The room had concrete walls and a carpeted floor and a reverberation time of approximately 350 ms. There was a small amount of background noise, which was generated by personal computers, air conditioners, and some other noise sources. The four participants were sitting at a round table, on which an eight-channel small microphone array was placed to record the conversation. The microphone array consisted of two layers of microphones: the first layer had five microphones that were placed on the edges of a 7-cm equilateral pentagon; the second layer had three microphones that were placed on the edges of a 4-cm equilateral triangle. The signals were sampled at 16 kHz. Figure 3 shows pictures of the room and the microphones.

In this experiment, we regarded one of the four participants as a target sound source and the other three as localized noise sources. The microphone system was configured as follows. Filter-bank analysis was carried out by using a short-time Fourier transform with a 512-point hamming window shifted by 128 points. The order, K, of the prediction filter and the amount of delay, Δ, in prediction were set at 6 and 3, respectively. Parameter α, which specifies the amount of delay in output signal generation (see (17)), was set at an equivalence of 100 ms. The forgetting factor was calculated as $\gamma = 1 - 0.5^{\Gamma}$, where integer Γ was varied from 2 to 10.

The proposed microphone system, comprising the above-described adaptive multiple point blind equalizer and an adaptive beamformer, was compared with the conventional beamforming microphone array described in [2]. The beamforming component of the proposed system also employed the method described in [2]. This beamformer requires knowledge of noise only periods. To keep the evaluation results independent of the voice activity detection accuracy, the noise-only periods were explicitly given based on a manual transcription of the conversation. The effectiveness of the microphone systems were assessed based on the noise reduction rate (NRR) and perceptual evaluation of speech quality (PESQ). The NRR represents the difference between input and output signal-to-noise ratios. Because the conversation was really recorded, no genuine source signals were available. Therefore, the NRR was estimated based on the differences between the magnitudes of the

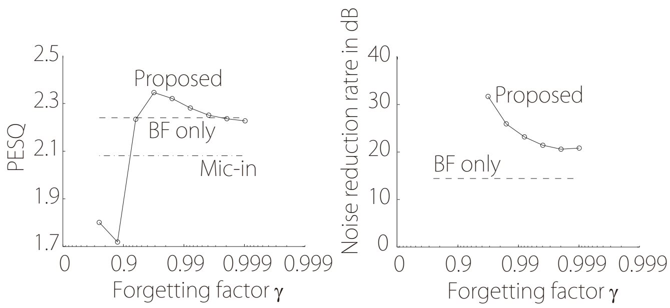

Fig. 4. PESQ and NRR as functions of forgetting factor.

normalized input and output signals during noise-only periods. This approach worked well for this experiment because the utterances of different speakers did not overlap very frequently and thus the input and output signals could be accurately normalized. The PESQ was calculated by using the headset microphone signal of the target speaker as a reference. The headset recording was made in synchronization with the microphone array signals. Specifically, we calculated the PESQ score for each target speech period and then averaged these local scores. The rare utterance overlaps mean that the PESQ scores mainly represent the degree of distortion imposed on the target speaker's voice.

The left plot in Fig. 4 shows the average PESQ scores as a function of forgetting factor γ. It can be observed that the proposed system with $\gamma > 0.95$ slightly outperformed the conventional system in terms of PESQ. The behavior of the equalizer was unstable when $\gamma < 0.95$, resulting in significant distortion of the target voice. Therefore, we excluded these parameter settings from the subsequent NRR-based evaluation. The right plot in Fig. 4 shows the NRR scores. It is clear that the proposed microphone system removed a larger amount of noise than the conventional one. This indicates that the noise reduction performance can be improved in reverberant rooms simply by pre-processing the microphone signals with the proposed equalizer. One interesting observation is that the effectiveness of the equalizer deteriorated when we set the forgetting factor at approximately 1.0. This indicates the importance of adaptiveness in dealing with fluctuations in room impulse responses resulting, for example, from changes in head postures.

4. CONCLUSION

This paper described a novel adaptive multiple point blind equalization algorithm and showed that this algorithm improved the noise reduction performance of conventional beamforming microphone arrays when it is used prior to beamforming. This algorithm attempts to shorten all the room impulse responses between the sound sources and the microphones, thereby creating anechoic spots at all the source positions. Therefore, it can prevent conventional microphone array beamformers from deteriorating in reverberant rooms.

Although the experimental results reveal the potential the present approach, the experiments are limited and do not cover the wide range of real acoustic environments. A thorough evaluation should be undertaken in the future to evaluate how effective the present algorithm is in a range of acoustic conditions represented by different rooms, source-to-microphone distances, noise source types, and signal-to-noise ratios.

5. APPENDIX

This appendix derives the adaptive prediction filter optimization algorithm shown in Table 1, starting from the problem given by (9). Since $\boldsymbol{x}(t)$, defined by (1), can be expressed as

$$\boldsymbol{x}(t) = \boldsymbol{y}(t) - \boldsymbol{G}^H(t)\boldsymbol{Y}(t-\Delta), \tag{18}$$

the solution to (9) is given by

$$\boldsymbol{G}(t) = \left(\sum_{\tau=1}^{t} \frac{\gamma^{t-\tau}}{\lambda_\tau} \boldsymbol{Y}(\tau-\Delta)\boldsymbol{Y}^H(\tau-\Delta)\right)^{-1} \left(\sum_{\tau=1}^{t} \frac{\gamma^{t-\tau}}{\lambda_\tau} \boldsymbol{Y}(\tau-\Delta)\boldsymbol{y}^H(\tau)\right). \tag{19}$$

The first term on the right hand side (including the matrix inversion operation), which appears in the first line of (19), represents the error covariance matrix associated with this estimator and is hereafter represented as $\boldsymbol{\Phi}(t)$. $\boldsymbol{G}(t)$ and $\boldsymbol{\Phi}(t)$ can be described using recursive expressions as follows:

$$\boldsymbol{G}(t) = \boldsymbol{\Phi}(t)\left(\frac{1}{\lambda_t}\boldsymbol{Y}(t-\Delta)\boldsymbol{y}^H(t) + \gamma\boldsymbol{\Phi}^{-1}(t-1)\boldsymbol{G}(t-1)\right) \tag{20}$$

$$\boldsymbol{\Phi}(t) = \left(\frac{1}{\lambda_t}\boldsymbol{Y}(t-\Delta)\boldsymbol{Y}^H(t-\Delta) + \gamma\boldsymbol{\Phi}^{-1}(t-1)\right)^{-1}. \tag{21}$$

Unfortunately, these equations include matrix inversion operations and thus incur a large computational cost. Therefore, our goal is to find alternative recursive expressions that can be calculated efficiently.

To avoid matrix inversion, we apply the Woodbury identity to (21), whereby we obtain

$$\boldsymbol{\Phi}(t) = \frac{1}{\gamma}\left(\boldsymbol{\Phi}(t-1) - \boldsymbol{\Phi}(t-1)\frac{\boldsymbol{Y}(t-\Delta)\boldsymbol{Y}^H(t-\Delta)}{\gamma\lambda_t + \boldsymbol{Y}^H(t-\Delta)\boldsymbol{\Phi}(t-1)\boldsymbol{Y}(t-\Delta)}\boldsymbol{\Phi}(t-1)\right). \tag{22}$$

This can be further rewritten by using vector $\boldsymbol{k}(t)$, which is defined by (14), as

$$\boldsymbol{\Phi}(t) = \frac{1}{\gamma}\left(\boldsymbol{\Phi}(t-1) - \boldsymbol{k}(t)\boldsymbol{Y}^H(t-\Delta)\boldsymbol{\Phi}(t-1)\right). \tag{23}$$

Combining (14) and (23) allows us to express $\boldsymbol{k}(t)$ compactly as

$$\boldsymbol{k}(t) = \frac{\boldsymbol{\Phi}(t)\boldsymbol{Y}(t-\Delta)}{\lambda_t}. \tag{24}$$

With the above expressions, we can obtain a new representation of $\boldsymbol{G}(t)$ as

$$\begin{aligned}\boldsymbol{G}(t) &= \frac{1}{\lambda_t}\boldsymbol{\Phi}(t)\boldsymbol{Y}(t-\Delta)\boldsymbol{y}^H(t) + \boldsymbol{G}(t-1) - \boldsymbol{k}(t)\boldsymbol{Y}^H(t-\Delta)\boldsymbol{G}(t-1)\\ &= \boldsymbol{k}(t)\boldsymbol{y}^H(t) + \boldsymbol{G}(t-1) - \boldsymbol{k}(t)\boldsymbol{Y}^H(t-\Delta)\boldsymbol{G}(t-1)\\ &= \boldsymbol{G}(t-1) + \boldsymbol{k}(t)\boldsymbol{x}^H_{\text{pred}}(t).\end{aligned} \tag{25}$$

Therefore, the optimization algorithm can be summarized as shown in Table 1, which does not involve matrix inversion operations.

6. REFERENCES

[1] O. Hoshuyama, A. Sugiyama, and A. Hirano, "A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters," *IEEE Trans. Signal Process.*, vol. 47, no. 10, pp. 2677–2684, 1999.

[2] M. Souden, J. Benesty, and S. Affes, "On optimal frequency-domain multichannel linear filtering for noise reduction," *IEEE Trans. Audio, Speech, Language Process.*, vol. 18, no. 2, pp. 260–276, 2010.

[3] S. Gannot and I. Cohen, "Adaptive beamforming and postfiltering," in *Springer Handbook of Speech Processing*, J. Benesty, M. M. Sondhi, and Y. Huang, Eds., pp. 945–978. Springer, 2008.

[4] T. Yoshioka, A. Sehr, M. Delcroix, K. Kinoshita, R. Maas, T. Nakatani, and W. Kellermann, "Making machines understand us in reverberant rooms: robustness against reverberation for automatic speech recognition," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 114–126, 2012.

[5] K. Lebart, J. M. Boucher, and P. N. Denbigh, "A new method based on spectral subtraction for speech dereverberation," *Acta Acustica United with Acustica*, vol. 87, pp. 359–366, 2001.

[6] J. S. Erkelens and R. Heusdens, "Correlation-based and model-based blind single-channel late-reverberation suppression in noisy time-varying acoustical environments," *IEEE Trans. Audio, Speech, Language Processing*, vol. 18, no. 7, pp. 1746–1765, 2010.

[7] M. Yu and F. K. Soong, "Constrained multichannel speech dereverberation," in *Proc. Interspeech*, 2012.

[8] T. Yoshioka and T. Nakatani, "Generalization of multi-channel linear prediction methods for blind MIMO impulse response shortening," *IEEE Trans. Audio, Speech, Language Process.*, vol. 20, no. 10, pp. 2707–2720, 2012.

[9] M. I. Gurelli and C. L. Nikias, "EVAM: An eigenvector-based algorithm for multichannel blind deconvolution of input colored signals," *IEEE Trans. Signal Process.*, vol. 43, no. 1, pp. 134–149, 1995.

[10] B. W. Gillespie, H. S. Malvar, and D. A. F. Florêncio, "Speech dereverberation via maximum-kurtosis subband adaptive filtering," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, 2001, vol. VI, pp. 3701–3704.

[11] T. Nakatani, T. Yoshioka, K. Kinoshita, M. Miyoshi, and B.-H. Juang, "Speech dereverberation based on variance-normalized delayed linear predictor," *IEEE Trans. Audio, Speech, Language Process.*, vol. 18, no. 7, pp. 1717–1731, 2010.

[12] S. Mosayyebpour, H. Sheikhzadeh, T. A. Gulliver, and M. Esmaeili, "Single-microphone LP residual skewness-based inverse filtering of the room impulse response," *IEEE Trans. Audio, Speech, Language Process.*, vol. 20, no. 5, pp. 1617–1632, 2012.

[13] T. Yoshioka, T. Nakatani, M. Miyoshi, and H. G. Okuno, "Blind separation and dereverberation of speech mixtures by joint optimization," *IEEE Trans. Audio, Speech, Language Process.*, vol. 19, no. 1, pp. 69–74, 2011.

[14] M. Togami, Y. Kawaguchi, R. Takeda, Y. Obuchi, and N. Nukaga, "Optimized speech dereverberation from probabilistic perspective for time varying acoustic transfer function," *IEEE Trans. Audio, Speech, Language Process.*, vol. 21, no. 7, pp. 1369–1380, 2013.

[15] K. Abed-Meraim, E. Moulines, and P. Loubaton, "Prediction error method for second-order blind identification," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 694–705, 1997.

[16] Z. Ding, "Linear predictive algorithms for blind multichannel identification," in *Signal Processing Advances in Wireless and Mobile Communications*, G. B. Giannakis, Y. Hua, P. Stoica, and L. Tong, Eds., vol. 1, pp. 179–209. Prentice Hall, 2001.

[17] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, fourth edition, 2001.