Applied Acoustics 73 (2012) 50–55

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

Wide-band dereverberation method based on multichannel linear prediction using prewhitening filter

Takuma Okamoto [a,b,*], Yukio Iwaya [a,c], Yôiti Suzuki [a,c]

[a] Research Institute of Electrical Communications, Tohoku University, Katahira 2-1-1, Aoba-ku, Sendai 980-8577, Japan
[b] Graduate School of Engineering, Tohoku University, Katahira 2-1-1, Aoba-ku, Sendai 980-8577, Japan
[c] Graduate School of Information Sciences, Tohoku University, Katahira 2-1-1, Aoba-ku, Sendai 980-8577, Japan

ARTICLE INFO

Article history:
Received 28 February 2011
Received in revised form 4 July 2011
Accepted 7 July 2011
Available online 31 July 2011

Keywords:
Microphone array signal processing
Dereverberation
LIME algorithm
Colored source signal
Prewhitening filter

ABSTRACT

Several dereverberation algorithms have been studied. The sampling frequencies used in conventional studies are typically 8–16 kHz because their main purpose is preprocessing for improving the intelligibility of speech communication and articulation for automatic speech recognition. However, in next-generation communication systems, techniques to analyze and reproduce not only semantic information of sound but also more high-definition components such as spatial information and directivity will be increasingly necessary. To decompose these sound field characteristics with high definition, a dereverberation algorithm that is useful at high sampling frequencies is an important technique to process sound that includes high-frequency spectra such as musical sounds. The LInear-predictive Multichannel Equalization (LIME) algorithm is a promising dereverberation method. Using the LIME algorithm, however, a dereverberation signal cannot be solved at high sampling frequencies when the source signal is colored, such as in the case of speech and sound of musical signals. Because the rank of the correlation matrix calculated from such a colored signal is not full, the characteristic polynomial cannot be calculated precisely. To alleviate this problem, we propose preprocessing of all input signals with filters to whiten their spectra so that this algorithm can function for colored signals at high sampling frequencies.

1. Introduction

Conventionally, the main purpose of dereverberation algorithms is to improve speech communication intelligibility and the performance of automatic speech recognition systems [1]. Therefore, in these studies, sampling frequencies are typically 8–16 kHz. However, in next-generation high-definition communications systems, techniques to analyze and synthesize not only semantic information of sound but also more qualitative characteristics comparable to textures of sound must be very important. Therefore, it is necessary to extract room acoustic characteristics and original sound precisely from each sound source, along with the directivity of each sound source from the sound sensed in a room at a position that is distant from the sound sources [2–5]. Dereverberation is a key technique. The sampling frequency must be sufficiently high, such as 44.1 kHz. Recently, dereverberation of musical sounds has been attracting the interest of researchers [6]. Therefore, we conducted a study to develop an algorithm that can decompose reverberant components from received signals at high sampling frequencies.

To date, various dereverberation techniques have been proposed [1]. Among them, methods using inverse filtering of a room transfer function based on the Multi input–output INverse Theorem (MINT) [7] show good performances to extract original source signals from input signals, including reflected sound. In methods based on MINT, the impulse responses corresponding to room acoustics (room impulse responses, RIR) are necessary to calculate the inverse filters. However, it is difficult to estimate impulse responses solely from the input signals because of the great length of room impulse responses [8]. These techniques are known as blind multichannel identification based on second-order or higher-order statistical [9] or subspace methods [10]. Another type is blind dereverberation, where the inverse filter is estimated solely from the input signals [11–13]. Particularly, LInear-predictive Multi-input Equalization (LIME) [12] is widely used because this method shows good performance against whitening problems in blind dereverberation. In the LIME algorithm, the signal generating process is estimated from input signals based on multi-channel linear prediction; the original signal is extracted almost perfectly if the sampling frequency is as high as 16 kHz.

However, when the sampling frequency is higher and the source signal is colored as normally as typical music signals and speech signals, then the dereverberation performance is decreased

* Corresponding author at: Research Institute of Electrical Communications, Tohoku University, Katahira 2-1-1, Aoba-ku, Sendai 980-8577, Japan.
E-mail address: okamoto@ais.riec.tohoku.ac.jp (T. Okamoto).

0003-682X/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.apacoust.2011.07.004

even with the LIME algorithm. This performance degradation occurs because the matrix rank used in the calculation is decreased and becomes extremely ill-conditioned. To cope with this problem in this paper, we propose a new wide-band dereverberation algorithm by improving the LIME algorithm.

In Section 2, the LIME algorithm is briefly introduced. The problem of the LIME algorithm at high sampling frequency is discussed in Section 3. A new algorithm using a prewhitening filter is then proposed in Section 4. In Section 5, the performance of the proposed method is evaluated. Finally, our conclusions are presented in Section 6.

2. LIME algorithm

The LIME algorithm is based on the hypothesis that the original sound source signal $s(n)$ is modeled using linear prediction. The estimated source signal $\hat{s}(n)$ is recovered by calculating the estimated prediction residual $\hat{e}(n)$ and the estimated Auto-Regressive (AR) parameters $\hat{a}(z)$ from the matrix $\mathbf{Q}$ obtained from the input signals $x_i(n)$, as shown by the following equation:

$$\mathbf{Q} = (E\{\mathbf{x}(n-1)\mathbf{x}^T(n-1)\})^{+}E\{\mathbf{x}(n-1)\mathbf{x}^T(n)\}. \tag{1}$$

In that equation, $\mathbf{A}^+$ is the Moore–Penrose generalized inverse of matrix [15] $\mathbf{A}$, $\mathbf{x}_i(n) = [x_i(n),\ldots,x_i(n-(L-1))]^T$ and $\mathbf{x}(n) = [\mathbf{x}_1^T(n),\ldots,\mathbf{x}_P^T(n)]^T$.

The estimated AR parameters $\hat{a}(z)$, which correspond to the linear predictive coefficient $a(n)$ of the source signal $s(n)$, are obtained from the characteristic polynomial [14] of $\mathbf{Q}$. The LIME algorithm schema is depicted in Fig. 1. The order of $a(n)$ is defined as

$$N = M + L - 1, \tag{2}$$

where M is the order of the room impulse response $h_i(n)$ and where L signifies the order of the inverse filter based on MINT.

The order of $a(n)$ is obtainable as Eq. (2) by calculating $\mathbf{Q}$ using the higher order than the real order M of $h_i(n)$ despite unclear M. This constitutes a great advantage of LIME. The LIME algorithm has been introduced and discussed in detail in an earlier report [12].

3. Performance degradation of LIME at high sampling frequency

3.1. Occasion of performance degradation

In the LIME algorithm, the prediction residual $\hat{e}(n)$ and the estimated Auto-Regressive (AR) parameters $\hat{a}(z)$ of the source signal are calculated from the correlation matrix of received signals and its inverse given by Eq. (1). Therefore, the accuracy of LIME depends on the granularity of $\mathbf{Q}$ in Eq. (1). To calculate Eq. (1), the rank of the correlation matrix must be full because the precision of inverse in Eq. (1) depends on its rank. The rank of the correlation matrix is determined by its condition number, which is obtainable as a ratio between the maximum eigenvalue and minimum eigenvalue.

If the length of L in Eq. (2) is sufficiently large in LIME, then the eigenvalues of the correlation matrix $E\{\mathbf{x}(n-1)\mathbf{x}^T(n)\}$ in Eq. (1) are equivalent to the average power spectrum of $x(n)$ [16,17]. Therefore, if the average power spectrum of $x(n)$ is colored, then the difference between the maximum eigenvalue and the minimum eigenvalue is large and the rank of its correlation matrix becomes smaller than the full rank.

However, most actual sounds around us are frequently colored, but have energy bias. For example, in speech and musical instrumental signals, energy is concentrated in the low-frequency region. It decays concomitantly at higher frequencies. Such a tendency of colored characteristics increases when the sampling frequency is high.

At high sampling frequency, the correlation matrix $E\{\mathbf{x}(n-1)\mathbf{x}^T(n-1)\}$, defined as Eq. (1), is colored when the input signal is colored. Results show that the rank of the correlation matrix $E\{\mathbf{x}(n-1)\mathbf{x}^T(n-1)\}$ is reduced to less than N(= the order of AR polynomial $\hat{a}(z)$) and that its inverse and the characteristic polynomial of $\mathbf{Q}$ cannot be calculated accurately, which underscores an important problem: LIME does not work at high sampling frequencies if the input signals are colored.

3.2. Computer simulation

Using computer simulations, we demonstrate that the dereverberation performance is degraded when a source signal is a colored signal in LIME.

In the simulations, the room impulse responses $h_i(n)$ that were used were those measured using a time-stretched pulse (TSP) [18] in a soundproof room using a microphone array of 20 microphones (Fig. 2). The room size was 5.18 m × 3.38 m × 2.52 m. The soundproof room reverberation time was 0.15 s and sampling frequency was 44.1 kHz for deliberation at high sampling frequency. Therefore, the length of the impulse response $h_i(n)$ was around $44{,}100 \times 0.15 = 6150$ samples. Fig. 3 shows the temporal waveform and the averaged spectrum of a measured impulse response. The averaged spectra of $s(n) * h(n)$ at two sampling frequencies are shown in Fig. 4a–d, respectively. The original source signal in

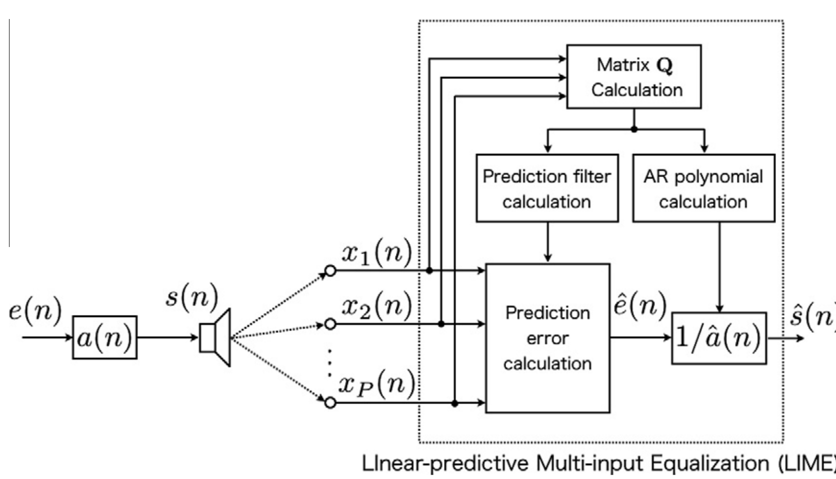

Fig. 1. Block diagram of the LIME Algorithm [12].