CEPSTRUM ANALYSIS AND BLIND SYSTEM IDENTIFICATION FOR LOCALIZATION OF PSEUDO-PERIODIC SOUNDS

Daniele Salvati
AVIRES Lab.
Dep. of Mathematics and Computer Science
University of Udine, Italy
daniele.salvati@uniud.it

Sergio Canazza
Sound and Music Computing Group
Dep. of Information Engineering
University of Padova, Italy
canazza@dei.unipd.it

ABSTRACT

Blind System Identification (BSI) focuses on the impulse responses between the source and the microphones to estimate the Time Difference Of Arrival (TDOA) of an acoustic source in reverberant environments. Considering the Adaptive Eigenvalue Decomposition (AED) method based on Normalized MultiChannel Frequency Least Mean Square (NMCFLMS) with sparse priors imposition, this paper shows that the use of cepstrum analysis of microphone pair signals allows to locate pseudo-periodic sounds by selecting the appropriate frequency bins, reducing the effect of reverberant ambient noise that covers most of the spectral range. Experimental results of a thirteen sinusoidal components wave in different reverberant conditions are reported.

1. INTRODUCTION

The aim of an acoustic source localization system is to estimate the position of sound sources in space by analyzing the sound field with a microphone array, a set of microphones arranged to capture the spatial information of sound. The AED [1] is a time delay estimation method based on the BSI, which focuses on the impulse responses between the source and the microphones. The AED takes the reverberation fully into account with the advantages to obtain a better performance under highly reverberant conditions. It can be efficiently implemented with the NMCFLMS with sparse priors imposition [2]. However, when the source of interest is harmonic, or generally pseudo-periodic, the NMCFLMS converges to an incorrect value because reverberant ambient noise covers most of the spectral range.

This paper proposes the use of cepstrum spectrum to select the frequency bins of pseudo-periodic signals, and to consider only these spectrum components in NMCFLMS filter to successfully obtain the TDOA estimation of a microphone pair. The interest in locating pseudo-periodic sounds using microphone arrays may be attractive for human-computer interaction in musical applications. It has been shown in [3, 4] that there is some potential in using the sound source localization to directly control an audio processing by moving the sound produced by its own musical instrument.

2. BSI CEPSTRUM ALGORITHM

AED [1] is a BSI estimation method for time delay between a microphone pair based on a reverberant model using eigenvalue decomposition. Introducing the channel impulse response $g_m$ from the source to microphone $m$, we can express the reverberant model as

$$x_m(k) = g_m * s(k - k_s - \tau_m) + v_m(k)$$  \hspace{1cm} (1)

where * denotes convolution, $s(k)$ is the unknown uncorrelated source signal, $k$ is the sample time index, $k_s$ is the propagation time from the unknown source to the reference sensor, $\tau_m$ is the TDOA of the signal between the $m_{th}$ microphone and the reference array, and $v_m(k)$ is the additive noise signal at the $m_{th}$ sensor, assumed to be uncorrelated with not only all of the source signals but also with the noise observed at the other sensors. AED assumes that the room is linear and time invariant; therefore, neglecting the influence of noise we can write

$$x_1 * g_2 = g_1 * s * g_2 = x_2 * g_1.$$  \hspace{1cm} (2)

The vectors of the signal samples at the microphone outputs and the impulse response vectors of length $L$ can be expressed as

$$x_i(k) = [x_i(k), x_i(k - 1), \ldots, x_i(k - L + 1)]^T$$  \hspace{1cm} (3)

and

$$g_i = [g_i(0), g_i(1), \ldots, g_i(L - 1)]^T.$$  \hspace{1cm} (4)

Substituting the vectors, equation (2) becomes

$$x_1^T g_2 - x_2^T g_1 = 0.$$  \hspace{1cm} (5)

Then, introducing the correlation matrix, we can define the following equation in matrix notation from (5)

$$Ru = 0.$$  \hspace{1cm} (6)

where $R$ is the correlation matrix and $u$ is a $2M \times 1$ vector formed by the juxtaposition of the two impulse responses

$$R = \begin{bmatrix} R_{x_1 x_1} & R_{x_1 x_2} \\ R_{x_2 x_1} & R_{x_2 x_2} \end{bmatrix}$$

and

$$u = [g_2 - g_1]^T.$$
and \( R_{x_i x_j} = E[x_i(k)x_j^*(k)] \). The equation (6) implies that \( \mathbf{u} \) is the eigenvector corresponding to the zero-valued eigenvalue of \( \mathbf{R} \). If the two impulse responses have no common zeros and the autocorrelation matrix of \( s(n) \) is full rank, there is only a single zero-valued eigenvalue. In practice, only estimate of the sample correlation matrix is available, and, therefore, instead of the zero-valued eigenvalue, we search for the minimum eigenvalue of \( \mathbf{R} \) using the iterative method. To speed up the convergence and to achieve efficient estimation, a NMCFLMS can be used [2]. Finally, the time delay estimation is

\[
\hat{\tau}_{12}^{AED} = \arg \max_l \hat{g}_{1,l} - \arg \max_l \hat{g}_{2,l}. \tag{7}
\]

An improved AED method [2] imposes sparse priors on the responses to reduce the temporal whitening and provide a more accurate and robust time delay estimation.

The performance of the AED technique is dramatically reduced in the case of pseudo-periodic sounds because reverberant ambient noise covers most of the spectral range, and consequently the NMCFLMS converges to an incorrect value of TDOA.

Cepstrum [5] analysis provides to select the spectral components to be used in the NMCFLMS filter. The cepstrum is calculated from the FFT by taking the real logarithm and performing an IFFT, which leads to the real cepstrum

\[
e(f) = \frac{1}{L} \sum_{j=0}^{L-1} \log \sum_{k=0}^{L-1} x(k) e^{-\frac{2\pi j f k}{L}} e^{\frac{2\pi j k}{L}}. \tag{8}
\]

The local maxima of cepstrum spectrum are the selected frequency bins of signal to be used in the NMCFLMS filter, and considering that the other values, i.e., those not selected by the cepstrum analysis, of FFT are assumed to be zero.

### 3. EXPERIMENTAL RESULTS

A thirteen sinusoidal components wave is used in experiments to test the proposed BSI cepstrum algorithm in different reverberation conditions. AED method based on NMCFLMS with sparse priors imposition and an Hann window analysis of 2048 samples at 44.1 kHz are used.

The first six cepstrum components are selected to evaluate the NMCFLMS performance. The Matlab code that implements the image-source method developed by E. A. Lehmann is used to simulate reverberant audio data in room acoustics [6]. A room of \((3.5 \times 4.5 \times 3)\) meters size is considered. The position of microphones are \((1.4, 1, 1.3)\), \((1.6, 1, 1.3)\), and the source is \((2.5, 3, 1.3)\) m. In this case the acoustic source reaches the microphones with a TDOA of -11 samples. Table 1 summarizes the TDOAs estimated by NMCFLMS with sparse priors imposition, comparing the results without and with cepstrum analysis, considering different number of cepstrum components. In all reverberant conditions the NMCFLMS without cepstrum analysis converges to the an incorrect value of TDOA, while the NMCFLMS with cepstrum properly works to a minimum of five frequency components.

<table>
<thead>
<tr>
<th>RT60 (s)</th>
<th>NMCFLMS without cepstrum</th>
<th>NMCFLMS with cepstrum</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>2</td>
<td>-12</td>
</tr>
<tr>
<td>0.5</td>
<td>0</td>
<td>-10</td>
</tr>
<tr>
<td>1.0</td>
<td>0</td>
<td>-10</td>
</tr>
</tbody>
</table>

Table 1. The comparison of TDOA estimation (values are in samples). The cepstrum NMCFLMS is evaluated with different numbers of frequency components.

### 4. CONCLUSIONS

The use of cepstrum analysis in the AED method based on NMCFLMS with sparse priors imposition allows to estimate TDOA of harmonics, or generally pseudo-periodic, sounds in reverberant environments. Some experimental results with a thirteen sinusoidal components wave in different reverberant conditions show that the selection of frequency component bins, obtained by means of cepstrum spectrum and picking the local maxima, reduces the influence of reverberant ambient noise that covers most of spectral range. The proposed BSI cepstrum algorithm converges to the correct TDOA value with a minimum of five frequency components up to RT60 of 1 s.

### 5. REFERENCES


