AN IMPROVED METHOD FOR TDOA-BASED SPEECH SOURCE LOCALIZATION

Hamid Reza Abutalebi\(^1\), and Hossein Momenzadeh\(^2\)

\(^{1,2}\) Signal Processing Research Lab, Electrical Engineering Department, Yazd University
Yazd, Iran
habutalebi@yazduni.ac.ir

Abstract

TDOA (Time Difference Of Arrival)-based algorithms are common methods for speech source localization. Generalized Cross Correlation (GCC) method is the most important approach for estimating TDOA between microphone pairs. The performance of this method significantly degrades in the presence of noise and reverberation. In this paper, we firstly propose a modification to make GCC-PHAT method robust against environment noise. Then, we use an iterative technique that employs the location estimation to improve TDOAs accuracy. Extensive experiments show the capability of the proposed methods in significant increment of TDOA accuracy, and consequently, more accurate estimations of source location.

1. INTRODUCTION

Speech source localization using microphone arrays is one of the most important research topics in speech signal processing. A common way for localization is the use of TDOA (Time Difference Of Arrival)-based algorithms [1]. In these algorithms, we firstly determine the TDOA of signals between different microphone pairs (TDOA estimation stage); then, the source location is estimated based on these TDOAs (location estimation stage).

The accuracy of estimated TDOAs is very important, since any error in TDOAs leads to a high error in localization [2]. In real acoustic environments, the accuracy of TDOAs is degraded due to noise and/or reverberation [2, 3].

GCC method is the most common and fastest two-channel algorithm for TDOA estimation [3]. This algorithm encounters several problems in real acoustic environment. In this paper, we have firstly explained the GCC basics and its variants. Then, noting the defects of these techniques in real (practical) applications, we have proposed a novel method for TDOA estimation and evaluated its performance in noisy and reverberant environments.

Furthermore, we have also proposed a hybrid localization method to improve the accuracy. In this algorithm, TDOA estimation is iteratively combined with source localization estimation to improve the accuracy of TDOA estimation. This, in turn, makes the source localization more accurate. In the proposed method, TDOA estimation is modified according to the primary estimated location of source (that is estimated by a closed form method such as Spherical Interpolation (SI) or Spherical Intersection (SX) [4]). Moreover, we have added an outlier removal technique in the system that improves the localization accuracy.

By implementing the proposed modifications and evaluating the whole system on simulated and real (practical) data, we have demonstrated the superiority of the proposed methods in
accurate speech source localization. The rest of this paper is organized as follows. In section 2, GCC method is described. The modified GCC-PHAT method is presented in section 3. In section 4 hybrid localization method and outlier removal are presented. Section 5 and 6 include the results of simulations and experiments in simulated and real room situations. Finally, some concluding remarks are given in section 7.

2. GENERALIZED CROSS CORRELATION METHOD

GCC algorithm uses time delay information from only one pair of microphones [5]. Due to the use of FFT, computational complexity of GCC is low; therefore, it is a common choice for real-time applications.

In this method, delay estimation is obtained via [3, 5]:

\[
\hat{r}_{GCC} = \arg \max_m \Psi_{GCC}[m],
\]

where

\[
\Psi_{GCC}[m] = \sum_{k=0}^{K-1} \Phi[k] S_{xx}[k] e^{i2\pi mk/K},
\]

is so-called GCC Function (GCCF). \( S_{xx}[k] \) is the cross spectrum, and is approximately equal to \( S_{xx}[k] \approx X_n[k] X^*_n[k] \), where \( X_n[k] \) is the DFT of \( x_n[m] \). Also, \( \Phi[k] \) is a weighting function. Several weighting functions have been proposed in the literature, two most important of them will be described in the following.

2.1 GCC-PHAT Algorithm

In this method, the weighting function is applied by PHAse Transform (PHAT) function defined as [5]:

\[
\Phi_{PHAT}[k] = \frac{1}{|S_{xx}[k]|},
\]

Neglecting noise effect in (2), we can deduce that the weighted cross correlation spectrum is free from the source signal and depends only on the channel response. This way, we can justify good performance of the method in reverberant situations.

2.2 GCC-ML Algorithm

In this case, the weighting function is a Maximum Likelihood (ML) filter defined as [5]:

\[
\Phi_{ML}[k] = \frac{|X_n[k]|^2|X[k]|^2}{|X_n[k]|^2|X[k]|^2 + |N[k]|^2|X[k]|^2},
\]

where \( N_n[k] \) is the noise power spectrum in \( n^{th} \) microphone [3]. In ML filter, signal and noise are assumed independent and stationary. So, in reverberant environments where these