Available online at www.sciencedirect.com

ScienceDirect

Computer Speech and Language 34 (2015) 113–128

COMPUTER SPEECH AND LANGUAGE

www.elsevier.com/locate/csl

Subspace-based DOA with linear phase approximation and frequency bin selection preprocessing for interactive robots in noisy environments

Sheng-Chieh Lee*, Bo-Wei Chen, Jhing-Fa Wang, Min-Jian Liao, Wen Ji

Department of Electrical Engineering, National Cheng Kung University, Taiwan

Received 24 May 2014; received in revised form 12 February 2015; accepted 10 March 2015
Available online 20 March 2015

Abstract

This work develops a method of estimating subspace-based direction of arrival (DOA) that uses two proposed preprocesses. The method can be used in applications that involve interactive robots to calculate the direction to a noise-contaminated signal in noisy environments. The proposed method can be divided into two parts, which are linear phase approximation and frequency bin selection. Linear phase approximation rectifies the phases of the two-channel signals that are affected by noise, and reconstructs the covariance matrix of the received signals according to the compensative phases using phase line regression. To increase the accuracy of DOA result, a method of frequency bin selection that is based on eigenvalue decomposition (EVD) is utilized to detect and filter out the noisy frequency bins of the microphone signals. The proposed techniques are adopted in a method of subspace-based DOA estimation that is called multiple signal classification (MUSIC). Experimental results reveal that the mean estimation error obtained using proposed method can be reduced by 7.61° from the conventional MUSIC method. The proposed method is compared with the covariance-based DOA method that is called the minimum variance distortionless response (MVDR). The DOA improves the mean estimation accuracy by 4.98° relative to the conventional MVDR method. The experimental results demonstrate that both subspace-based and covariance-based DOA algorithms with the proposed preprocessing method outperform the DOA estimation in detecting the direction of signal in a noisy environment.

Keyword: Sound source location; Direction of arrival (DOA); Angular correction; Linear phase approximation; Covariance matrix reconstruction; Frequency bin selection; Interactive robot; Human–robot interaction

1. Introduction

The objective of a sound source location system is to determine the direction of a sound source using the DOA algorithm. The DOA algorithm finds the coherence of received signals using a correlation matrix, and calculates the angle of the sound source from a restrictive space. Much great deal of research has been done on sound source location, and especially speech source location in several scenarios and applications, such as human–robot interaction,

* Corresponding author. Tel.: +886 932 854 913; fax: +886 6 2381249.
E-mail addresses: u9554806@nkfust.edu.tw, appleway1984@gmail.com (S.-C. Lee).

http://dx.doi.org/10.1016/j.csl.2015.03.002

human–computer interaction, surveillance systems, audio–visual conferencing, and speech recognition (Chen et al., 2008; Mestre and Lagunas, 2008; Togami et al., 2009; Pirinen et al., 2003; Xiao et al., 2004; Caylar, 2009; Liu and Chen, 2010; Li, 2005; Wang et al., 2010; Hwang and Sarkar, 2005). For DOA methods, a conventional beamformer (Krishnaveni et al., 2013; Mestre and Lagunas, 2006) utilizes a linear array of equally spaced sensors to receive sound, and coherently sum the outputs of the distributed sensors using weight vectors. The weight vectors contain information about the time delay that is associated with path length variation. Maximum likelihood (ML) methods (Li et al., 1995; Swindlehurst, 1998) have been developed to exploit DOA estimation when the narrow-band plane waves of the received signals are already known.

Much of the literature reveals that a noisy ambient environment worsens the performance of a sound source location system. Ambient noise is a signal that can influence the components of the covariance matrix of collected signals, resulting in phase distortion in the frequency domain. Phase distortion can generate an estimation error in the analysis of the direction of a sound. Over the past few decades many methods that involve beamscan algorithms and subspace algorithms have been developed to mitigate the effect of noise in DOA (Cho et al., 2008, 2010; Wang et al., 2006; Yang et al., 2008; Wang and You, 2008; Hioka et al., 2003; Tanigawa and Hamada, 2003).

Cho et al. (2008, 2010) proposed a distributed microphone system that was composed of 16 microphones to locate a sound source position using the accumulated correlation coefficient between the multiple channel pairs. The experimental results indicate that the position of the sound can be calculated by cross-power spectrum phase (CSP) analysis. Wang et al. (2006) presented a new beamforming DOA using two virtual subarrays. The presented method yields a phase-shifted reference signal whose phase relative to the reference signal is a function of the target DOA. Yang et al. (2008) proposed the least square-based (LS-based) phase estimation method for preprocessing in the multiple signal classification (MUSIC) algorithm to find information about the direction to the speech source. The experimental results, based on computer simulation, reveal the accuracy of the estimated DOA of test speech. Wang and You (2008) presented a wideband beam-space preprocessing method, in which a rotational signal subspace (RSS) algorithm is utilized to obtain the effective data on reference frequency bins. In the computer simulations, a linear array with 20 sensors is used to receive the wideband signal (such as Gaussian white noise), and then to evaluate the beam-space covariance matrix for use in the DOA algorithm. Hamada et al. (Hioka et al., 2003; Tanigawa and Hamada, 2003) introduced a two-channel method for estimating DOA using virtually generated multichannel data; their method was based on the harmonic structure of vowels. Their simulation results demonstrated that the proposed method was suitable for sound localization in noisy environments.

To increase the interaction between human beings and robots, DOA methods also have been adopted in the application of interactive robots in recent years. The interactive robot using DOA technique can compute the position of user and then give feedbacks such as communication, message, and personal information to the user. Therefore, how to reduce the noise interference on DOA estimation for robot interaction becomes more noteworthy. Reviewing the previous DOA methods (Cho et al., 2008, 2010; Wang et al., 2006; Yang et al., 2008; Wang and You, 2008; Hioka et al., 2003; Tanigawa and Hamada, 2003), the main contribution of the methods is to reduce the influences of ambient noises or room reverberations. However, there are several additional aspects should be considered for the application of interactive robots. The first one is the geometry of microphone array for interactive robots. Cho et al. (2008, 2010) installed 16 microphones in a 4×4 lattice condition, and the distance between the microphones was 135 cm. This size of microphone array does not suitable for interactive robots. The second one is the number of microphones that are used in a microphone array. The microphone array using more microphones can achieve better performance on estimating the direction of a target sound source (Cho et al., 2008, 2010; Wang and You, 2008). Nevertheless, for the application of interactive robots, it is another consideration that how to select a suitable microphone array and avoid the high dimensional matrix operation from the microphone signals. The third one is the restricted condition for test voices such as harmonic structure of vowel signals (Hioka et al., 2003; Tanigawa and Hamada, 2003). Hamada et al. only used the vowel voice "a" to illustrate the DOA performance in experimental results; it is not enough to demonstrate the flexibility of DOA method in noisy environments. The final one is experimental evaluation; the type of evaluation can be classified as computer simulation (Wang et al., 2006; Yang et al., 2008; Wang and You, 2008) and practical evaluation (Cho et al., 2008, 2010; Hioka et al., 2003; Tanigawa and Hamada, 2003). To confirm the feasibility of the DOA method, this work tends to actualize a DOA estimation that can be really adopted in practical environments.

Integrating the above-mentioned DOA methods and the aspects for applications of interactive robots, this work presents a DOA method for interactive robots to estimate the direction of a source signal in noisy environments. The proposed method has following four properties: Appropriate size of a microphone array for interactive robots,