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.
© 2015 Elsevier Ltd. All rights reserved.

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
0885-2308/© 2015 Elsevier Ltd. All rights reserved.

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,

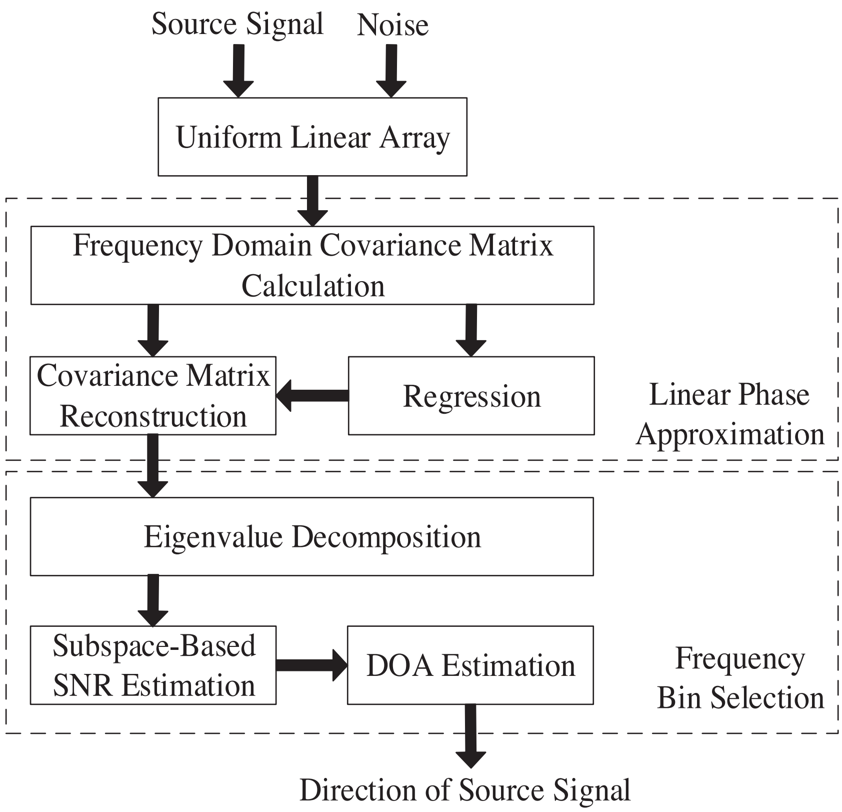


Fig. 1. Overview of DOA estimation system with the proposed DOA preprocessing techniques—linear phase approximation and frequency bin selection.

lower-dimensional matrix operation, flexible DOA method for source signals, and practical evaluation for noisy environments. In the DOA method, two DOA preprocessing techniques, called linear phase approximation and frequency bin selection, are applied. The linear phase approximation can rectify the distorted phase by iterative regression. To increase the accuracy of DOA estimation, a subspace-based signal-to-noise ratio (SNR) calculation is utilized to identify and filter out the noisy frequency bin in frequency bin selection. After the proposed preprocessing techniques are applied, the compensated covariance matrix of the desired signals can be utilized in the estimation of DOA. The performance of subspace-based DOA estimation is experimentally compared with two conventional DOA estimations and the proposed method is found to be suited to estimating the direction of signal in noisy environments. This work provides another feasible method for DOA estimation in real noisy environments.

The rest of this paper is organized as follows. Section 2 presents an overview of the proposed system for estimation DOA. Section 3 then describes in detail the proposed DOA system and the preprocessing method. Section 4 considers the experimental results. Section 5 briefly draws conclusions.

2. System overview

Fig. 1 displays the overview of the proposed system. Firstly, a uniform linear array of a pair of microphones is used to collect a sound, which is indicated as source signal in Fig. 1 and the following description. In the linear phase approximation preprocessing, the covariance matrix of the observed signals is firstly estimated to obtain noise-contaminated phase data. Subsequently, the regression procedure is developed to rectify the distorted phase data of the observed signals. In covariance matrix reconstruction, the covariance matrix of the signals is reconstructed using the improved compensative phase data.

Following the linear phase approximation, an eigenvalue decomposition (EVD) is performed to decompose the reconstructed covariance matrix at the beginning of frequency bin selection. In subspace-based DOA estimation, the

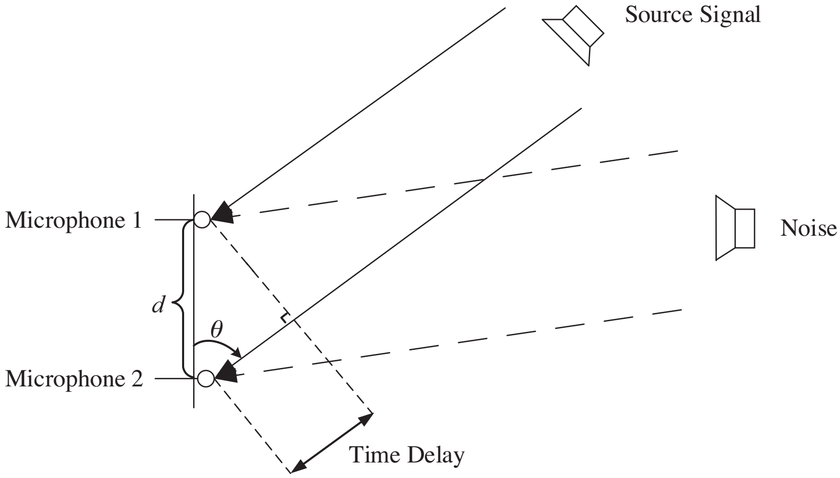


Fig. 2. Uniform linear microphone array for collecting a source signal with ambient noise. The direction to source signal from the microphones is θ, and the spacing between pair of microphones is d.

EVD is adopted to decompose the covariance matrix into the source subspace and the noise subspace. The signal power can then be calculated from the EVD results. Before the DOA estimation is performed, a proposed subspace-based SNR estimation is conducted to determine the amount of noise in each frequency bin. The frequency bins that contain noise can be identified and filtered out using the threshold value that was calculated in the subspace-based SNR estimation.

Compared with the system by Yang et al. (2008), the main difference is that the order between phase approximation and frequency selection. According to the method by Yang et al., the frequency bins which have local maximum power are firstly picked, and then the selected frequency bins are adopted in phase regression. However, the method of frequency selection is not much objective because the information of noise is unknown. When the power magnitude of the noise is larger than that of the source signal in some frequency bins, the phase regression would generate false phases to DOA estimation and cause biased DOA result. To avoid this situation, the proposed method selects all the frequency bins in linear phase approximation to obtain the preliminary compensative phases. Subsequently, these compensative phases are examined again in the proposed frequency bin selection. Finally, the significant frequency bins can be reserved for DOA estimation.

In DOA estimation, a subspace-based DOA algorithm is utilized because it supports high-resolution DOA performance using the proposed DOA method. The performance of covariance-based DOA estimation without the EVD can be improved using linear phase approximation. The following sections describe the proposed method and experimental results.

3. Proposed linear phase approximation and frequency bin selection

3.1. Uniform linear array

Fig. 2 presents the microphone signals $x_1(t)$ and $x_2(t)$ from the source signal are received by the uniform linear microphone array in a noisy environment. A pair of microphones, with a spacing of d, are placed in the environment with the source signal, $y(t)$, and the noise, $n(t)$. The direction of the source signal from the microphones is θ, and the first microphone signal $x_1(t)$ is set as the reference. The observed signals in the time domain can be modeled using the matrices and vectors that satisfy Eqs. (1) and (2), where t and τ are the time index and the time delay.

$$\mathbf{X}(t) = \mathbf{Y}(t) + \mathbf{N}(t) \tag{1}$$

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} y_1(t) \\ y_1(t-\tau) \end{bmatrix} + \begin{bmatrix} n_1(t) \\ n_2(t) \end{bmatrix} \tag{2}$$

The relationship between the time delay and the direction of the source signal can be formulated as Eq. (3), where c is the velocity of sound in air.

$$\tau = d\cos(\theta)c^{-1} \tag{3}$$

In the frequency domain, Eqs. (1) and (2) can be expressed as Eqs. (4) and (5), where f is the frequency.

$$\mathbf{X}(f) = \mathbf{Y}(f) + \mathbf{N}(f) \tag{4}$$

$$\begin{bmatrix} X_1(f) \\ X_2(f) \end{bmatrix} = \begin{bmatrix} Y_1(f) \\ Y_1(f)e^{-j2\pi f\tau} \end{bmatrix} + \begin{bmatrix} N_1(f) \\ N_2(f) \end{bmatrix} \tag{5}$$

3.2. Linear phase approximation

To simplify the representation of the observed signals and to determine the description of phase in noiseless and noisy environments, the noise that is described in Eq. (5) is ignored. In addition, the unwrapping factor for phase information is also omitted for simplicity in following description. In a noiseless environment, the observed signals can be regarded as a source signal. Equation (6) gives the relationship between the observed signal and the steering vector $\mathbf{A}(f,\theta)$.

$$\begin{aligned} \mathbf{X}(f) &= \begin{bmatrix} Y_1(f) & Y_1(f)^{-j2\pi f\tau} \end{bmatrix}^T \\ &= \begin{bmatrix} 1 & e^{-j2\pi f\tau} \end{bmatrix}^T Y_1(f) \\ &= \begin{bmatrix} 1 & e^{-j2\pi f d\cos\theta c^{-1}} \end{bmatrix}^T Y_1(f) \\ &= \mathbf{A}(f,\theta)Y_1(f) \end{aligned} \tag{6}$$

According to Eq. (6), the covariance matrix of the observed signals can be defined as Eq. (7), where $\cdot^*(f)$ is the conjugate transpose of the complex-valued signal, and $E[\cdot]$ represents the expectation value of a matrix. The exponential term in Eq. (7) contains information about the time delay. In subsequent discussion of the time delay, a part of the exponential term in Eq. (7) is replaced by the symbol $\zeta(f, \tau)$, which is defined in Eq. (8). The symbol $\zeta(f, \tau)$ denotes phase data or just "phase" in the following description.

$$\begin{aligned} \mathbf{R_{XX}} &= E[\mathbf{X}(f)\mathbf{X}^*(f)] \\ &= E\begin{bmatrix} Y_1(f)Y_1^*(f) & Y_1(f)Y_2^*(f) \\ Y_2(f)Y_1^*(f) & Y_2(f)Y_2^*(f) \end{bmatrix} \\ &= E\begin{bmatrix} Y_1(f)^2 & Y_1(f)^2 e^{j2\pi f\tau} \\ Y_1(f)^2 e^{-j2\pi f\tau} & Y_1(f)^2 \end{bmatrix} \end{aligned} \tag{7}$$

$$\zeta(f,\tau) = 2\pi f\tau \tag{8}$$

Closely examining Eq. (8) reveals that, in a noiseless environment, the phase $\zeta(f, \tau)$ is directly proportional to the frequency when the time delay τ is the same for all frequency bins. In Fig. 3, three different directions of a source signal (speech is used herein) are individually tested in a noiseless environment. The phase data of observed signals in different directions are depicted as straight lines. Such a line will be called a phase line hereafter. Fig. 3 also indicates that the slope of a phase line depends on the time delay τ, which is determined by the direction of observed signals.

However, in a noisy environment, noise makes the time delay τ vary. Fig. 4 shows the coordinates of the phase data associated with various frequency bins in a noisy environment. Comparing Fig. 3 with Fig. 4 reveals that noise destroys the linearity of the phase line. The noise-contaminated phase data generate phase distortion and lead to biased DOA estimation.

To reduce the phase distortion, preprocessing by the proposed linear phase approximation is utilized to estimate and rectify the phase data for each frequency bin. The linear phase approximation has following three stages, which are covariance matrix calculation, regression, and covariance matrix reconstruction. Regression is the principal process in linear phase approximation, which is presented in detail in Fig. 5.

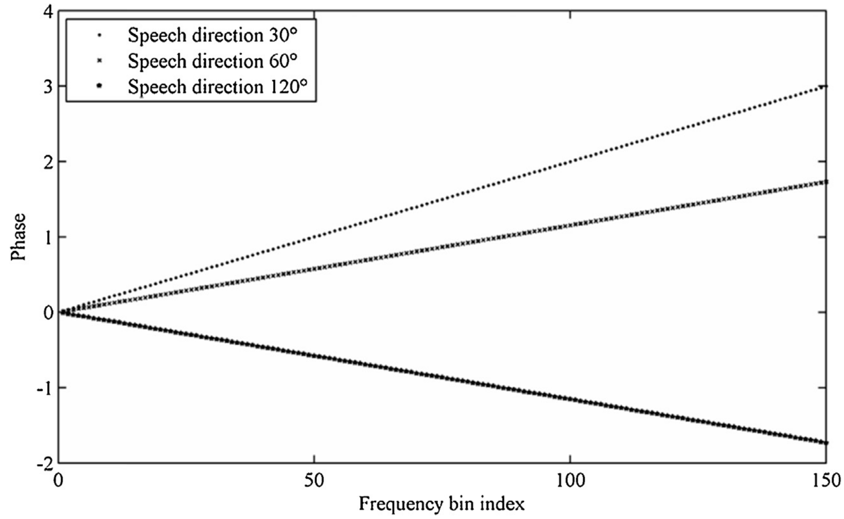


Fig. 3. Theoretical phase line (first 150 indices) for various directions of speech in noiseless environments. The directions of speech are 30°, 60°, and 120°; the spacing between pair of microphones is 0.08 m. Original phase data are linearly related to the frequency.

The purpose of linear phase approximation is to rectify the noise-contaminated phase using a flexible linear regression approach. Following the proposed processing, the compensative phase is obtained to rectify the observed phase. In Fig. 5, the observed phase $\zeta(f, \tau)$ is estimated using the covariance matrix that is calculated in the first stage. However, this phase is distorted by noise. To eliminate the distortion, a first-order regression is adopted in the phase line regression to evaluate the coefficients of a linear equation that fits the phase line in a least squares sense. The linear equation is Eq. (9), where p_1 and p_2 are the coefficients of the linear equation, and x is the index of frequency bin. The straight line that satisfies Eq. (9) can be regarded as the predicted ideal phase line in the absence of noise.

$$p(x) = p_1 x + p_2 \tag{9}$$

In phase update, the objective is iteratively to adapt the phase line to the ideal one. For the observed data, each frequency from the phase data $\zeta(f, \tau)$ can match a corresponding index of frequency bin x. After substituting the

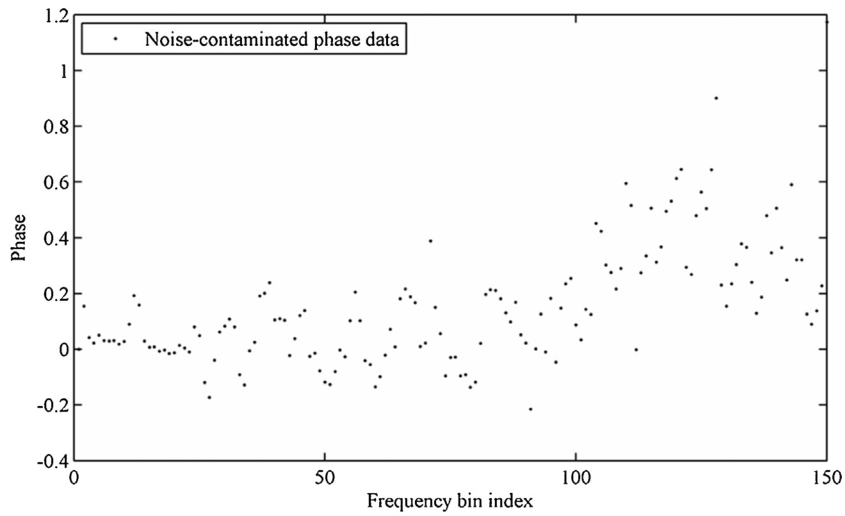


Fig. 4. Distorted phase line in noisy environment. Relation between phase and frequency bin is irregular because time delay τ is contaminated by noise.

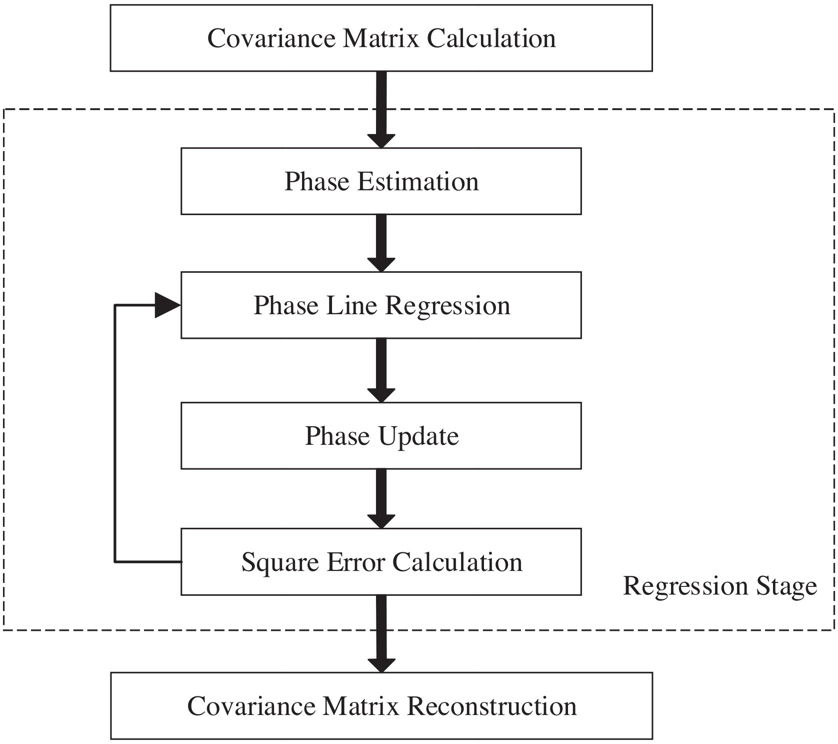


Fig. 5. Linear phase approximation, which has three stages – covariance matrix calculation, regression, and covariance matrix reconstruction.

index of frequency bin for the linear equation indicated in Eq. (9), the output $p(x)$ can be obtained according to the coefficients (i.e., p_1 and p_2) of the linear equation. The output can be taken as the ideal value of phase data, and all the outputs with different indices of frequency bins form a predicted phase line without the influence of noise.

In the comparative procedure, each $\zeta(f, \tau)$ is compared with the corresponding output $p(x)$. At the beginning of the linear phase approximation, the first index of the frequency bin means that the frequency is zero. Therefore, in the first comparison, the value of the phase data is directly set as zero without comparing with the corresponding value $p(x)$. Subsequently, the following phase data are individually compared with the corresponding value during the second comparison to the final comparison. The phase data that are lower than the corresponding value $p(x)$ are updated to the corresponding value $p(x)$. The updated phase data can be rewritten as $\zeta(f, \hat{\tau})$. Equation (10) yields the representation of updated phase data, where $\hat{\tau}$ is the compensative time delay in the frequency bin.

$$\zeta(f, \hat{\tau}) = 2\pi f \hat{\tau} \tag{10}$$

In the iteration after an update, a refinement that is based on $\zeta(f, \hat{\tau})$ is performed in the next iteration. To find an optimal phase line, a criterion that involves a square error calculation is applied. The phase line is iteratively updated if the square error exceeds a predefined threshold. Otherwise, the system outputs the phase data from the previous iteration to the next stage. Fig. 6 plots the rectified phase line that is obtained after the regression stage.

In the regression stage, the noise-contaminated phase data are enhanced by compensating for the time delay. The observed signals that satisfy Eqs. (4) and (5) can be written as Eqs. (11) and (12) in terms of $\hat{\mathbf{X}}(f)$, which includes the hypothetical modified source signal $\hat{\mathbf{Y}}(f)$ and the modified noise $\hat{\mathbf{N}}(f)$.

$$\hat{\mathbf{X}}(f) = \hat{\mathbf{Y}}(f) + \hat{\mathbf{N}}(f) \tag{11}$$

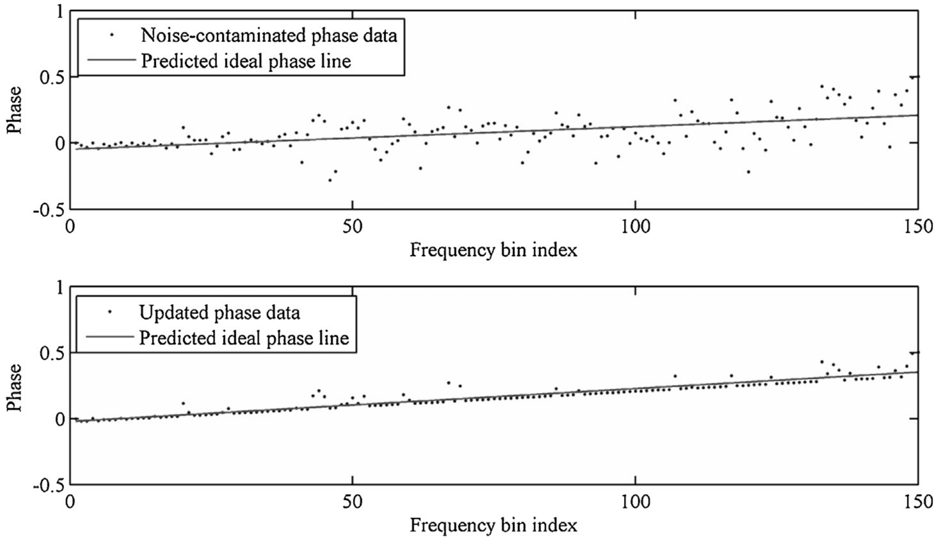


Fig. 6. Noise-contaminated phase data for frequency bins (first 150 indices) and phase data updated using linear phase approximation.

$$\begin{aligned}\begin{bmatrix}\hat{X}_1(f)\\ \hat{X}_2(f)\end{bmatrix} &= \begin{bmatrix}\hat{Y}_1(f)\\ \hat{Y}_1(f)e^{-j2\pi f\hat{\tau}}\end{bmatrix} + \hat{\mathbf{N}}(f)\\ &= \begin{bmatrix}X_1(f)\\ X_1(f)e^{-j2\pi f\hat{\tau}}\end{bmatrix}\\ &= \begin{bmatrix}X_1(f)\\ X_1(f)e^{-j\zeta(f,\hat{\tau})}\end{bmatrix}\end{aligned} \tag{12}$$

In covariance matrix reconstruction, the system reconstructs a new covariance matrix in each frequency bin based on the regressed phase data $\zeta(f, \hat{\tau})$. Equation (13) represents the reconstructed covariance matrix.

$$\begin{aligned}\mathbf{R}_{\hat{\mathbf{X}}\hat{\mathbf{X}}} &= E[\hat{\mathbf{X}}(f)\hat{\mathbf{X}}^*(f)]\\ &= E\begin{bmatrix}\hat{X}_1(f)\hat{X}_1^*(f) & \hat{X}_1(f)\hat{X}_2^*(f)\\ \hat{X}_2(f)\hat{X}_1^*(f) & \hat{X}_2(f)\hat{X}_2^*(f)\end{bmatrix}\\ &= E\begin{bmatrix}X_1(f)^2 & X_1(f)^2e^{j2\pi f\hat{\tau}}\\ X_1(f)^2e^{-j2\pi f\hat{\tau}} & X_1(f)^2\end{bmatrix}\\ &= E\begin{bmatrix}X_1(f)^2 & X_1(f)^2e^{j\zeta(f,\hat{\tau})}\\ X_1(f)^2e^{-j\zeta(f,\hat{\tau})} & X_1(f)^2\end{bmatrix}\end{aligned} \tag{13}$$

3.3. Frequency bin selection

In the above stages, the system reconstructs covariance matrices for each frequency bin. However, the slope of the predicted ideal phase line still differs from that of the theoretical one. Fig. 7 illustrates this difference.

To improve the performance of the DOA estimation, another proposed DOA preprocessing called frequency bin selection is used to identify the frequency bins, which contain the intense noise-contaminated phases. Fig. 8 presents the procedure of the proposed frequency bin selection.

At the beginning of frequency bin selection, the covariance matrix can be obtained by applying Eq. (13). In the following eigenvalue decomposition stage, the EVD of $\mathbf{R}_{\hat{\mathbf{X}}\hat{\mathbf{X}}}$, which is given by Eq. (13) yields Eq. (14), where $\mathbf{Q}$ is a

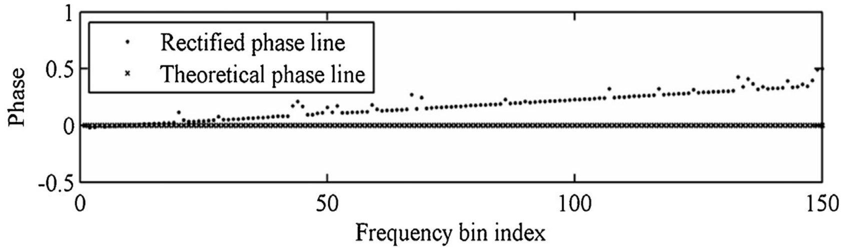


Fig. 7. Difference between rectified phase line after linear phase approximation (dotted line) and theoretical phase line.

square matrix whose ith column is the eigenvector $\mathbf{q}_i$, and $\mathbf{\Lambda}$ is a diagonal matrix. The diagonal elements $\lambda_{\hat{\mathbf{Y}}}$ and $\lambda_{\hat{\mathbf{N}}}$ are the eigenvalues of the source and the noise.

$$\begin{aligned} \mathbf{R}_{\hat{\mathbf{X}}\hat{\mathbf{X}}} &= \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1} \\ \mathbf{\Lambda} &= \begin{bmatrix} \lambda_{\hat{\mathbf{Y}}} & 0 \\ 0 & \lambda_{\hat{\mathbf{N}}} \end{bmatrix} \\ \mathbf{Q} &= \begin{bmatrix} \mathbf{q}_{\hat{\mathbf{Y}}} & \mathbf{q}_{\hat{\mathbf{N}}} \end{bmatrix} \end{aligned} \tag{14}$$

In subspace-based SNR estimation, the system quantifies the noise level at each frequency bin by parameter ε. The parameter ε, defined by Eq. (15), specifies the logarithm of the ratio of $\lambda_{\hat{\mathbf{Y}}}$ to $\lambda_{\hat{\mathbf{N}}}$.

$$\varepsilon = \log \frac{\lambda_{\hat{\mathbf{Y}}}}{\lambda_{\hat{\mathbf{N}}}} \tag{15}$$

In noise detection, the parameter ε is compared with a designated threshold χ. If ε exceeds the threshold χ, then the observed signal contains more of the wanted signal than noise. Therefore, the corresponding covariance matrix is preserved in DOA estimation. Otherwise, the system simply neglects the frequency bin.

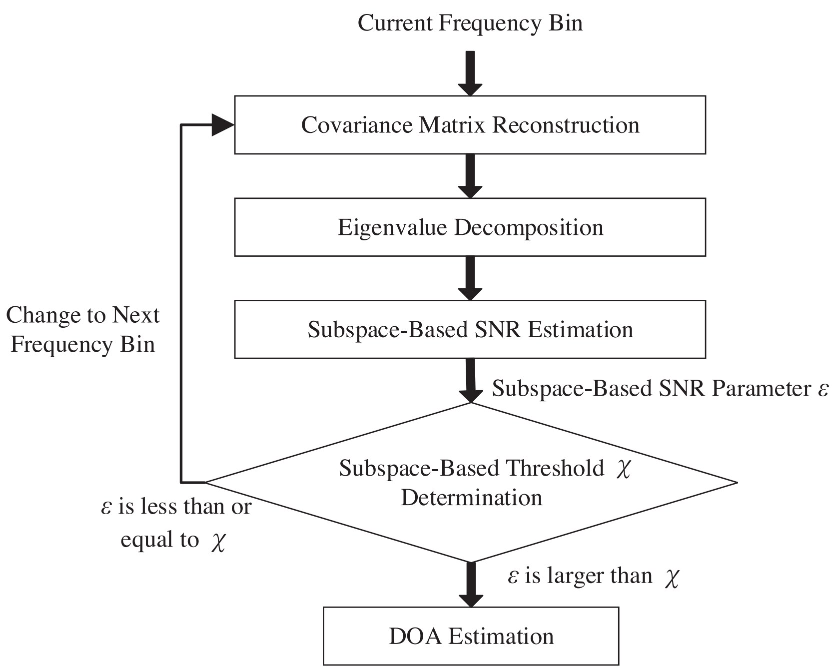


Fig. 8. Procedure of proposed frequency bin selection. Following the calculation of the covariance matrix and EVD, the subspace-based SNR can be estimated. Using the subspace-based threshold, the system determines whether the covariance matrix of a frequency bin should be used for DOA estimation.

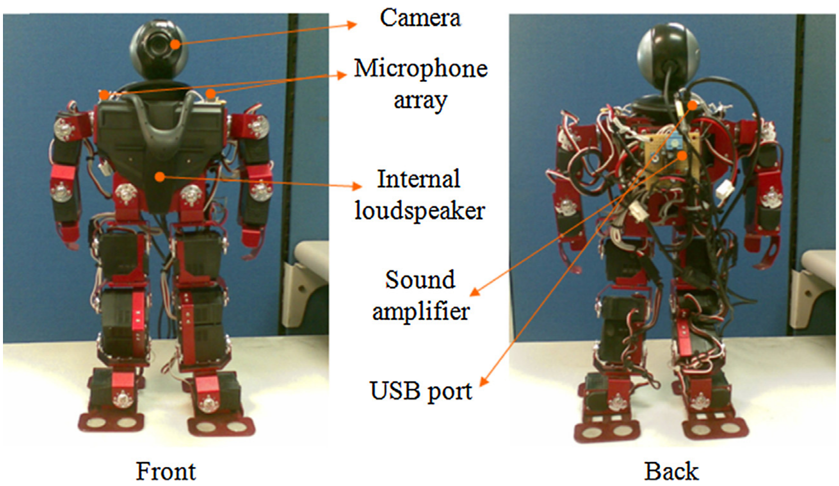


Fig. 9. Humanoid robot 16-DOF Robotinno$^{\text{TM}}$.

In the final stage, a subspace-based method called MUSIC is used in DOA estimation. In the MUSIC procedure, the signal can be separated into the source subspace and the noise subspace by using EVD to decompose its covariance matrix. From Eqs. (11) and (12), the observed signal following the linear phase approximation can be written as Eq. (16) in terms of the steering vector $\hat{\mathbf{A}}(f, \theta)$.

$$\begin{aligned}\hat{\mathbf{X}}(f) &= \hat{\mathbf{Y}}(f) + \hat{\mathbf{N}}(f)\\ &= \begin{bmatrix}1 & e^{-j2\pi f\hat{\tau}}\end{bmatrix}^T \hat{Y}_1(f) + \hat{\mathbf{N}}(f)\\ &= \hat{\mathbf{A}}(f, \theta)\hat{Y}_1(f) + \hat{\mathbf{N}}(f)\end{aligned} \tag{16}$$

According to Eqs. (14) and (16), the direction of the signal can be estimated from the eigenvector of the covariance matrix and the steering vector. The maximum power at the test angle θ can thus be obtained. The direction is calculated using Eq. (17).

$$P_{\text{MUSIC}}(\theta) = \frac{1}{\hat{\mathbf{A}}^H(f, \theta)\mathbf{q}_{\hat{\mathbf{N}}}\mathbf{q}_{\hat{\mathbf{N}}}{}^H\hat{\mathbf{A}}(f, \theta)} \tag{17}$$

During the calculation of signal power for a frequency bin whose subspace-based SNR parameter is larger than the designated threshold in the frequency bin selection, the signal power at different trial angles (0–180°) can be obtained by using Eq. (17). Then the values of signal power about this frequency bin at each trial angle are collected. When the power values from all the selected frequency bins are calculated, an accumulative signal power at each trial angle can be acquired according to the pre-calculated signal power from the selected frequency bins. Finally, the DOA result can be gained by finding the angle that has the maximum accumulated signal power.

4. Experimental results

4.1. Experimental setup

The proposed system is embedded in a humanoid robot called 16-DOF Robotinno$^{\text{TM}}$, which was manufactured by the Innovati Company. Fig. 9 displays in detail this humanoid robot. The linear microphone array that is placed on the shoulder of the robot comprises two omni-directional microphones; the microphone spacing is 0.08 m. About the selection of test source signals, a human speech signal is used as the source signal in following experiments.

With respect to the experimental environment, Fig. 10 presents the layout of the proposed speech location system in the meeting chamber (7.2 m long, and 6.1 m wide). The linear microphone array on the robot records a single segment of speech source and the background white noise at a sampling rate of 8 kHz. The distance between the robot and the speaker is 3.0 m. When the robot receives the speech, it turns its body toward the source of the speech, consistent with

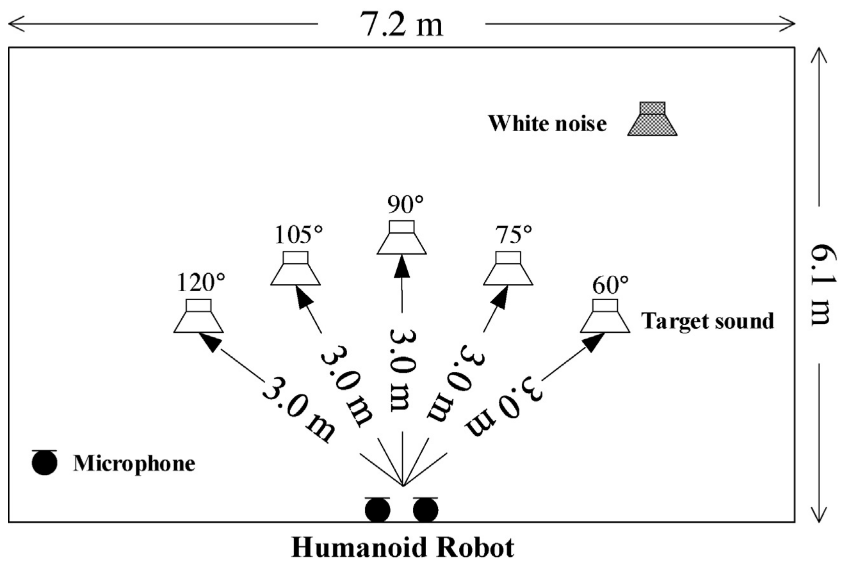


Fig. 10. Layout of chamber for the developed speech location system.

the DOA result. The predefined threshold of the square error calculation in the linear phase approximation is set to 25 in the experiments. In the DOA algorithms, both MUSIC and MVDR are evaluated using the proposed preprocessing techniques.

About the detailed setting of the proposed method, a 512-point fast Fourier transform (FFT) with Hamming window is adopted in the experiments. The window length is 512, and the shift size is zero (without overlap). The reverberation time (RT60) of the test environment is 1.237 s. For recorded test sentences, the length of test sentences is less than 3 s; the number of time frames is rounded 45 for Eq. (13) to calculate the covariance matrix. In linear phase approximation, the first 150 phase data points are selected for the phase regression.

In the test, an estimated error η is adopted to evaluate the performance of the proposed DOA method. Equation (18) yields the estimated error, where N is the total number of test speech segments, θ_i and $\hat{\theta}_i$ are the actual angle and the estimated angle to the ith test speech segment.

$$\eta = \frac{\sum_{i=1}^{N} |\theta_i - \hat{\theta}_i|}{N} \tag{18}$$

To demonstrate the feasibility of the proposed method, experimental results concerning four scenarios are obtained. In the first scenario, the performance of the proposed method with various subspace-based thresholds χ is evaluated. In the second, the MUSIC and MVDR methods are compared. In the third experiment, the proposed method is compared with the related method (Yang et al., 2008). In the final scenario, the method is evaluated using test speech segments with various SNRs.

4.2. Various subspace-based thresholds in frequency bin selection

Fig. 11 presents the results of DOA estimation that were obtained using the proposed linear phase approximation and frequency bin selection with various subspace-based thresholds. The angle from speaker to the linear microphone array is 75°, and the DOA estimation method is based on MUSIC. To evaluate the effectiveness of the frequency bin selection, four subspace-based thresholds are used to confirm the DOA estimation performance. According to Fig. 8 and Eq. (15), the function of frequency bin selection is almost disable when the subspace-based threshold χ is set to zero. Since the range of the ratio of $\lambda_{\hat{\mathbf{Y}}_1}$ to $\lambda_{\hat{\mathbf{N}}}$ is 1, each frequency bin can easily pass the condition of frequency bin selection and estimate the signal power in DOA estimation. Fig. 11 displays two source peaks at 66° and 72° when χ is set to zero, only the second source peak is a correct angle in DOA estimation. Increasing the threshold χ, the source peak exists at 72° and 74° when χ is set to one and two, respectively. In the extreme case (when χ is set to three), all of the frequency bins are ignored in the DOA estimation.

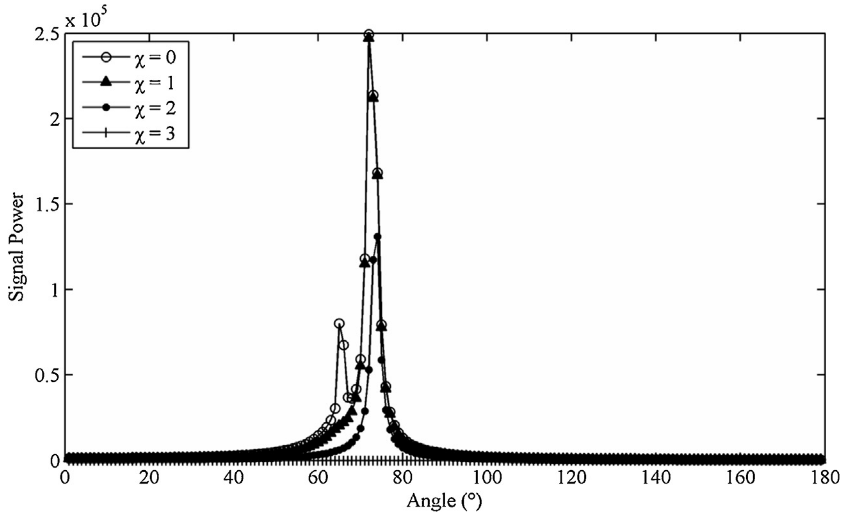


Fig. 11. Results of DOA estimation using linear phase approximation and frequency bin selection with various subspace-based thresholds. The angle from the speaker to the linear microphone array is 75°.

According to the DOA estimation results, the DOA estimation performs best when the subspace-based threshold is set to two. Table 1 presents the detailed estimated error for each test angle and threshold setting; for each test angle, 20 test speech segments were recorded and the estimated error evaluated with four threshold settings in frequency bin selection.

4.3. Comparison results using MUSIC and MVDR methods

Experiments on subspace-based and covariance-based DOA estimations yield the second set of experimental results. In the experiments, each of 11 speakers utters a sentence that takes less than 3 s 30 times in each of the specified directions. (Therefore, a total of 330 test sentences are spoken with different SNRs in each direction.) Five test directions (60°, 75°, 90°, 105°, and 120°) were used.

Table 2 compares the results obtained using conventional MUSIC and MUSIC with the proposed method in the subspace-based DOA experiments. The results indicate that the estimated error of the conventional MUSIC increases as the direction approaches 60° or 120°, because the time delay increases as the direction approaches 60° or 120°. The phases become large and are easily distorted by ambient noise. MUSIC with the proposed method outperforms the conventional MUSIC method in each direction. The mean of the five estimated errors is thus reduced by 7.61°.

A general MVDR estimation is carried out in the experiments to evaluate the covariance-based DOA performance. Since MVDR does not include the EVD procedure, the experimental results are concerned only the linear phase approximation method. Table 3 compares MVDR and MVDR with linear phase approximation. The results demonstrate that the proposed method outperforms the conventional MVDR approach. The mean of the estimated errors is thus reduced by 4.98°.

Table 1
Estimated error for each test angle and subspace-based threshold setting in frequency bin selection.

Threshold	Angle				
	60°	75°	90°	105°	120°
$\chi = 0$	4.80°	5.40°	2.50°	4.00°	7.15°
$\chi = 1$	4.55°	5.00°	2.50°	3.85°	7.15°
$\chi = 2$	1.85°	2.35°	1.65°	2.40°	5.00°
$\chi = 3$	22.40°	21.35°	62.85°	8.50°	91.50°

Table 2
Estimated error of conventional MUSIC and MUSIC with the proposed linear phase approximation and frequency bin selection.

Angle	Method	
	MUSIC	Proposed
60°	25.22°	13.96°
75°	17.48°	14.89°
90°	17.20°	10.80°
105°	16.50°	13.61°
120°	25.85°	10.94°

Table 3
Estimated error of conventional MVDR and MVDR with proposed linear phase approximation.

Angle	Method	
	MVDR	Proposed
60°	19.08°	12.44°
75°	10.54°	5.36°
90°	10.13°	7.89°
105°	10.62°	7.63°
120°	22.93°	15.08°

4.4. Experimental results using related and proposed methods

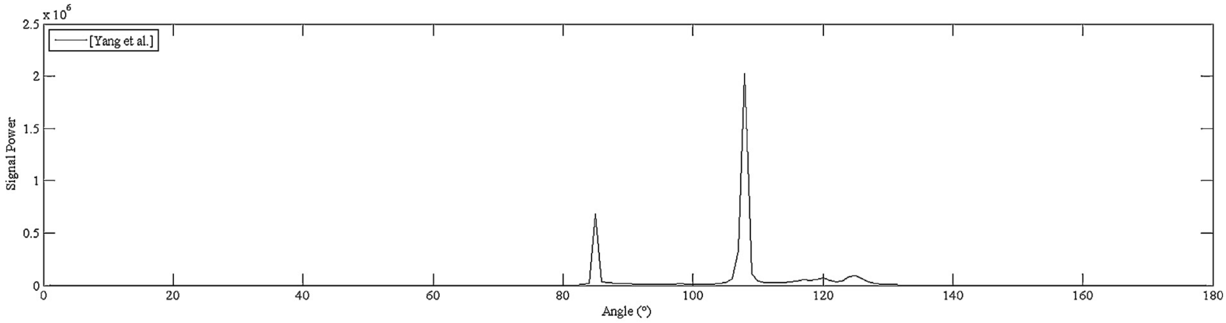
In the third set of experimental results, the proposed method is compared with a preprocessing-based DOA method that is presented by Yang et al. (2008). Comparing the proposed method with the method by Yang et al., the main difference between the method by Yang et al. and the proposed method is the manner of frequency selection. Examining the frequency selection in the method by Yang et al., each of frequency bins is taken to calculate its power magnitude with a logarithmic scale. Then the frequency bins which have local maximum power magnitude are selected as the desired frequency bins for least square-based phase regression. To compare the resolution of DOA, both related work and proposed work are experimented in following experiments.

In the experiments, five test directions (60°, 75°, 90°, 105°, and 120°) are used to record the test sentences. Totally 250 sentences are recorded for the experiments. (50 sentences are recorded in each of particular directions.) About the DOA estimation, MUSIC algorithm is utilized in both the related work and the proposed work. Table 4 shows the estimated error of the method by Yang et al. and the proposed method. The results demonstrate that the proposed method is superior to the related work; the average estimated error of the proposed method is lower than the related work about 1.96°.

Fig. 12 presents the DOA resolution by using the method by Yang et al. and the proposed method. The direction from the speaker to the microphone array is 105°. The results show that two peaks (86° and 108°) exist in the method by Yang et al., and one peak exists at 103° in the proposed method. Fig. 12 also illustrates that the proposed method can provide a suitable preprocessing method for subspace-based DOA estimation.

Table 4
Estimated error of the related method (Yang et al., 2008) and the proposed method.

Angle	Method	
	Yang et al. (2008)	Proposed
60°	8.90°	6.12°
75°	7.14°	5.26°
90°	9.48°	7.70°
105°	8.42°	7.04°
120°	11.16°	9.20°



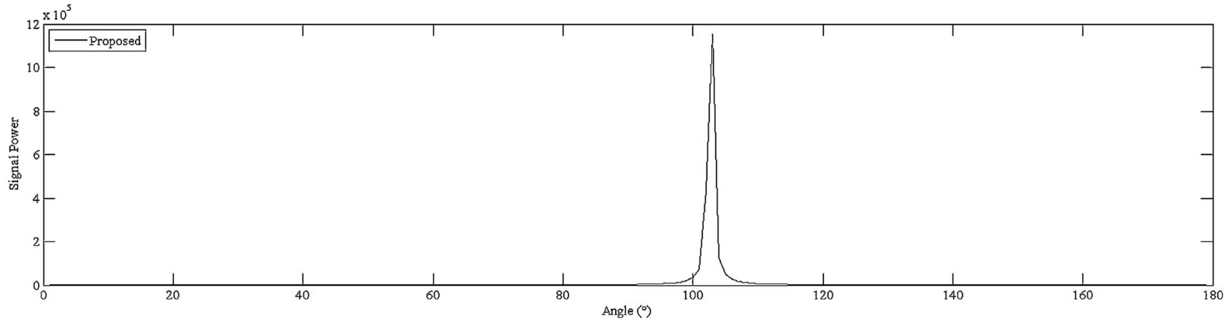


Fig. 12. DOA resolution by using the related method (Yang et al., 2008) and the proposed method. The angle from the speaker to the linear microphone array is 105°.

4.5. Evaluation of estimated error for speech with various SNRs

In the final experiments, 120 sentences with various SNRs (10, 20, and 30 dB) and angles (45° and 90°) are used to evaluate MUSIC and MUSIC with the proposed method. Fig. 13 presents the performance of the MUSIC approach and of the MUSIC approach with proposed linear phase approximation and frequency bin selection. Low SNR speech yields a high estimated error because of the intense effect of the noise. The results demonstrate that the estimated error can be reduced by using the proposed method.

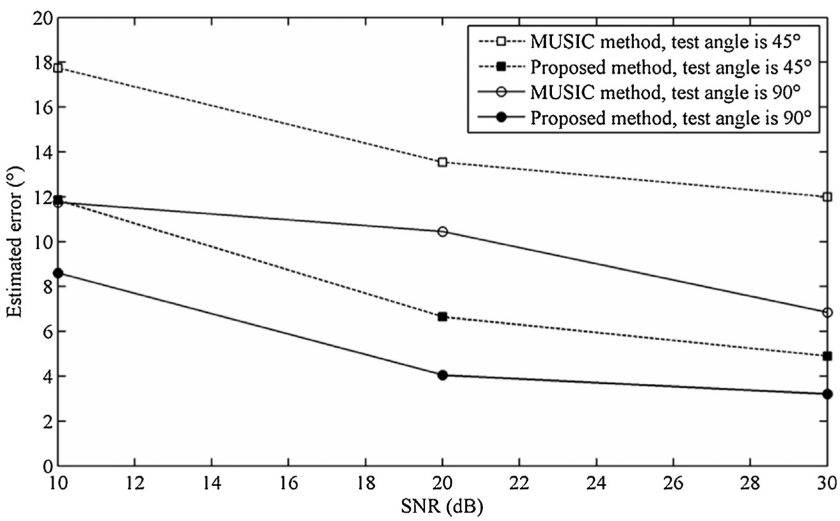


Fig. 13. DOA performance for test speeches with various SNRs at 45° and 90°. MUSIC with the presented method outperforms conventional MUSIC.

In conclusion, noise can affect the phase of a received signal from a microphone array. Phase rectification can be used to reduce the effect of noise. The experimental results demonstrate that the proposed method improves DOA estimation because it rectifies the distorted phase using the linear phase approximation. To improve DOA accuracy further, another method, called frequency bin selection, is proposed to identify and filter out the frequency bin that contains too much noise before the DOA is estimated. The experimental results also reveal that the proposed method is feasible for DOA estimation.

5. Conclusion

This work develops an effective method of DOA estimation that can be used in an interactive robot to estimate the direction from source signal in a noisy environment. To reduce the effect of noise on DOA estimation, two DOA preprocessing techniques called linear phase approximation and frequency bin selection, are proposed. The linear phase approximation rectifies the distorted phase by regression, and the frequency bin selection improves DOA performance by identifying noisy frequency bins. The experimental results herein indicate that the proposed method yields smaller DOA estimation error for various test speech angles and SNRs than conventional DOA methods. The DOA estimation error is also lower than the related method in experiments. In future work, the proposed DOA method can be integrated with other methods such as surveillance, speech enhancement, noise reduction, and speech recognition, to facilitate interactions between human and robots.

References

Caylar, S., 2009. A new neural network DOA estimation technique based on subarray beamforming. In: Proc. 2009 IEEE Int. Conf. Electromagnetics in Advanced Applications, Torino, Italy, September 14–18, pp. 732–734.

Chen, H.-H., Chan, S.-C., Zhang, Z.-G., Ho, K.-L., 2008. Adaptive beamforming and recursive DOA estimation using frequency-invariant uniform concentric spherical arrays. IEEE Trans. Circuit Syst. I: Regular Papers 55 (April (10)), 3077–3089.

Cho, K., Okumura, H., Nishiura, T., Yamashita, Y., 2008. Localization of multiple sound sources based on inter-channel correlation using a distributed microphone system. In: Proc. 9th Annual Conf. International Speech Communication Association, Brisbane, Australia, September 22–26, pp. 443–446.

Cho, K., Nishiura, T., Tamashita, Y., 2010. Robust speaker localization in a disturbance noise environment using a distributed microphone system. In: Proc. 2010 IEEE 7th Int. Symp. Chinese Spoken Language, Tainan, Taiwan, November 29–December 3, pp. 209–213.

Hioka, Y., Koizumi, Y., Hamada, N., 2003. Improvement of DOA estimation using virtually generated multichannel data from two-channel microphone array. J. Signal Process. 7 (February (1)), 105–109.

Hwang, S., Sarkar, T.K., 2005. Interpolation technique for direction of arrival (DOA) estimation using a transformation matrix through singular value decomposition. In: Proc. 2005 IEEE Int. Symp. Antennas and Propagation Society, Washington, USA, July 3–8, pp. 130–133.

Innovati Company [Online]. Available: http://www.innovati.com.tw

Krishnaveni, V., Kesavamurthy, T., Aparna, B., 2013. Beamforming for direction-of-arrival (DOA) estimation – a survey. Int. J. Comput. Appl. 61 (January (11)), 4–11.

Li, P., 2005. An efficient DOA estimation technique for multi-user multi-path signals in TD-SCDMA systems. In: Proc. 2005 IEEE Int. Symp. Microwave, Antenna, Propagation and EMC Technologies for Wireless Communication, Beijing, China, August 8–12, pp. 1382–1385.

Li, J., Halder, B., Stoica, P., Viberg, M., 1995. Computationally efficient angle estimation for signals with known waveforms. IEEE Trans. Signal Process. 43 (September (9)), 2154–2163.

Liu, J.-W., Chen, H., 2010. A novel DOA estimation technique for coherent real-valued sources. In: Proc. 2010 IEEE 2nd Int. Conf. Industrial Mechatronics and Automation, Wuhan, China, May 30–31, pp. 135–138.

Mestre, X., Lagunas, M.A., 2006. Finite sample size effect on minimum variance beamformers: optimum diagonal loading factor for large array. IEEE Trans. Signal Process. 54 (January (1)), 69–82.

Mestre, X., Lagunas, M.A., 2008. Modified subspace algorithms for DOA estimation with large arrays. IEEE Trans. Signal Process. 56 (January (2)), 598–614.

Pirinen, T., Pertila, P., Visa, A., 2003. Toward intelligent sensors reliability for time delay based direction of arrival estimates. In: Proc. 2003 IEEE Int. Conf. Acoustics, Speech, and Signal Processing, Hong Kong, Hong Kong, April 6–10, pp. 197–200.

Swindlehurst, A.L., 1998. Time delay and spatial signature estimation using known asynchronous signals. IEEE Trans. Signal Process. 46 (February (2)), 449–461.

Tanigawa, S., Hamada, N., 2003. Direction-of-arrival estimation of speech using virtually generated multichannel data from two-channel microphone array. Electron. Commun. Jpn. Part 3 86 (February (2)), 33–42.

Togami, M., Amano, A., Sumiyoshi, T., Obuchi, Y., 2009. DOA estimation method based on sparseness of speech sources for human symbiotic robots. In: Proc. 2009 IEEE Int. Conf. Acoustics, Speech, and Signal Processing, Taipei, Taiwan, April 19–24, pp. 3693–3696.

Wang, Q., You, H., 2008. MVDR beam-space pre-processing for wideband sources DOA estimation. In: Proc. 2008 IEEE Int. Symp. Intelligent Information Technology Application, Shanghai, China, December 21–22, pp. 1091–1094.

Wang, N., Agathoklis, P., Antoniou, A., 2006. A new DOA estimation technique based on subarray beamforming. IEEE Trans. Signal Process. 54 (August (9)), 3279–3290.

Wang, J., Du, R., Liu, F., 2010. A new method based on the spatial differencing technique for DOA estimation. In: Proc. 2010 IEEE Int. Conf. Networking, Sensing, and Control, Chicago, USA, April 10–12, pp. 44–48.

Xiao, Y., Ma, L., Khorasani, K., 2004. A novel wideband DOA estimation technique based on harmonic source model for a uniform linear array. In: Proc. 2004 IEEE Int. Symp. Circuits and Systems, Vancouver, Canada, May 23–26, pp. 445–448.

Yang, J.-M., Choi, M.-S., Kang, H.-G., 2008. Two-channel DOA estimation using frequency selective music algorithm with a phase compensation in reverberant room. In: Proc. 2008 IEEE 5th Int. Workshop on Sensor Array and Multichannel Signal Processing, Darmstadt, Germany, July 21–23, pp. 365–368.