JAIST Repository
https://dspace.jaist.ac.jp/

Title	Method for estimating monaural DOA and distance using modulation spectrum analysis
Author(s)	王, 鋭
Citation	
Issue Date	2021-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/17084
Rights	
Description	Supervisor:鵜木 祐史, 先端科学技術研究科, 修士（情報科学）

JAPAN ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

Method for estimating monaural DOA and distance using modulation spectrum analysis

1910037 Wang Rui

Human has an amazing ability to locate target sound source. Sound localization is one of the most important senses of human hearing include direction of arrival (DOA) and distance. It happens every minute when we hear some voices. For example, people can spot a car from its sound to avoid danger. In general, people with two normal ears hearing can perceive the sound location easily. Due to the different position of the two ears relative to the sound source, the time and level of the sound reaching the two ears will also be different. This is called binaural cues include interaural time differences (ITD) and interaural level differences (ILD). But for people that have only one ear hearing, the binaural cues can not be used in sound localization. Yet some studies found that people with monaural hearing (single ear) can perceive sound location include DOA and distance by using monaural cues. However, it is not yet clear how human can use these cues to perceive the direction of arrival (DOA) of a sound source of 3D space. Therefore, how can human perceives the sound location become an important research issues in the related study of hearing mechanism.

On the other hand, in the application of engineering, sound localization is a common used technology in many acoustic fields such as noise reduction, speech recognition, robot acoustic system, target tracking and so on. In general, typical DOA and distance estimation method is based on microphone arrays by using binural cues like Time Difference of Arrival (TDOA) method. Schmidt also proposed Multiple signal classification algorithm (MUSIC) in 1979, which has became a classical method in sound localization.

However, no matter TDOA and MUSIC method, they are all multi-channel signal processing methods. Such method always request a certain number of microphones. And these microphones are required to be located at certain intervals in application, which will bring spatial and mechanical constraint in many environments. Some researchers start to investigate the possibility of single microphone localization. Takashima proposed a monaural sound localization method based on GMM (Gaussian Mixture Model). However, this method needs huge training data and the related features in signal still are still unknown. In addition, most of current monaural sound localization method are only useful in 2D plane. In this case, if we can apply human ability for monaural sound localization to the engineering issues, a method of estimating DOA of the target sound in 3D space using monaural cues can be realized as applications of single-channel auditory signal processing.

Recently, it has been widely acknowledged that human can use the monaural cues for sound localization. It has been reported that the possible monaural cues are often referred to spectral cues in the head-related transfer function (HRTF). Moreover, it has also been suggested that modulation cues play an important role in monaural DOA estimates. Based on these researches, some studies have indicated interest in using the monaural modulation spectrum (MMS) in monaural DOA estimation. Ando *et al.* reported the feasibility of estimating monaural DOA based on MMS by using the modulation transfer function (MTF) but it is only useful for 2D plane. Bui *et al.* then proposed a 3D monaural DOA estimation method using MMS. However, this method has a large root-mean-square-error (RMSE) in monaural DOA estimation and the performance varies greatly in different positions. In addition, in all the MMS based monaural localization researches, only monaural DOA was studied. The possibility of monaural sound soure distance estimation by using MMS has not been investigated.

This paper aims to investigate the monaural cues related with DOA and distance of sound source in MMS and propose a method for estimating monaural DOA and distance in 3D space using modulation spectrum analysis. It is believed that MMS can be used to estimate monaural DOA as well as distance of sound source. Our previous proposed method has proved that MMS can be used in monaural DOA estimation in 3D space. But, unfortunately, there is a large RMSE in that method. So, to achieve the research goal, there are two steps in this study. First, improve the performance of MMS based monaural DOA estimation method in 3D space. Secondly, investigate the related features in MMS that could be used in monaural distance estimation and proposed a method for monaural distance estimation using MMS features.

In step 1, this study made a investigation of the possible reasons that cause the large RMSE in our previously proposed method. As a result, it was found that unreasonable boundary conditions in the estimation algorithm and the under-fitting problem in the polynomial regression model give rise to the large RMSE in monaural DOA estimation. Then, the unreasonable boundary conditions were removed so that the RMSE of DOA estimation of some points in the edge of 3D space decreases. After that, this paper tried to increase the polynomial regression order to reduce the RMSE in DOA estimation. In this process, the influence of the polynomial regression order to the RMSE of regression model was investigated. The optimal order of regression model was determined to be 8. All the processes are described in Chapter 3.

Then in step 2, a monaural distance estimation method based on MMS features is proposed. In our investigation of monaural distance perception, it was found that, under a certain reverberation, long distance transmission will reduce the modulation depth of the signal. Different distances will cause different change in modulation spectrum. Sounds arriving at a long distance are harder to hear than at a short distance. The received signal will be affected by reverberation to different degrees depending on the distance of sound source. In our investigation, the slope feature of MMS was found to be related with absolute monaural distance estimation since it can reflect the change of MMS shape caused by distance in transmission. These contents are contained in Chapter 4.

Simulations with several signal types and multiple subjects were carried to simultaneously estimate the DOA and distance of an incoming sound source in 3D space. HRTF database and image source method (ISM) were used to simulate sound sources with different directions and distances. The results of separate estimation of DOA and distance are reflected in Chapter 3 and Chapter 4 respectively. After that, a joint estimation include DOA and distance in 3D space was described in Chapter 5. Finally, there is a general discussion about the overall estimation method. The results show the effectiveness of our proposed method for estimating monaural DOA and distance using modulation spectrum analysis.