5Q-02

Evaluation of 2D bird localization algorithm using microphone arrays

Daniel Gabriel[1], Ryosuke Kojima[2], Kotaro Hoshiba[1], Kazuhiro Nakadai[1,3]
1 Tokyo Institute of Technology 2 Kyoto University 3 Honda Research Institute Japan Co., Ltd.

1 Introduction

In bird song analysis, knowing the positions of the singing birds is crucial. It can provide information about the birds' behaviour and territorial displacement. The recent technological advancements in the acoustic engineering field, especially in microphone array development, enable ornithologists to record and analyze bird songs. Several studies have reported using microphone arrays in different acoustic scene analysis problems [1] [2] [3] and also in the bird song analysis problems [4] [5] [6]. By obtaining the Time Difference of Arrival (TDoA) of the sound to each microphone in the array, it is possible to obtain the direction of the sound source. However, by using multiple microphone arrays, it is possible to search for the intersecting points of the sound source directions to obtain 2D sound source positions. We proposed such a triangulation-based algorithm by taking noise-robustness into consideration with preliminary results [7]. This paper assesses the algorithm, and discusses its effectiveness and limitations.

2 System outline

To understand the results and discussion on limitation of the system, its outline will be presented in this section. Flow chart of the system is presented in Fig.1 and each paragraph will describe three main parts of the algorithm.

First is the initial sound processing. To obtain the sound data of singing birds, stationary microphone arrays are used. In the proposed system we assume having three or more microphone arrays distributed in a birds' natural environment. The sound is then recorded by the microphones and analysed using the beamformer-based MUltiple SIgnal Classification (MUSIC) [8] algorithm to obtain sound source directions. Also, the Geometric High-order Dicorrelation-based Source Separation (GHDSS) [9] was used to separate the recorded sound into sound sources, even if these sound sources exist at the same time. These two algorithm are implemented in HARK [10] an open source software for robot audition. We can design the number and position of microphone arrays depending on

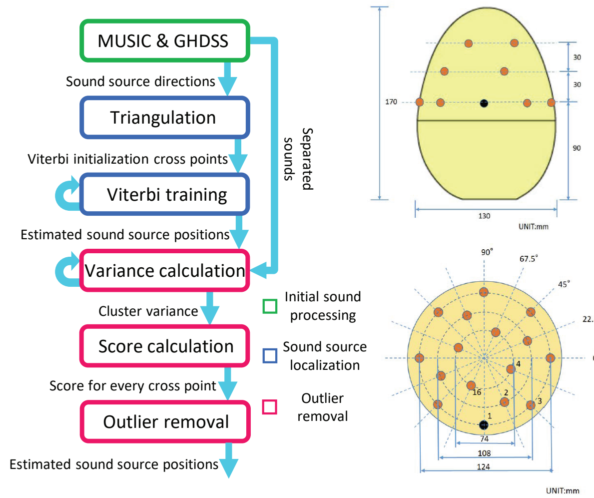
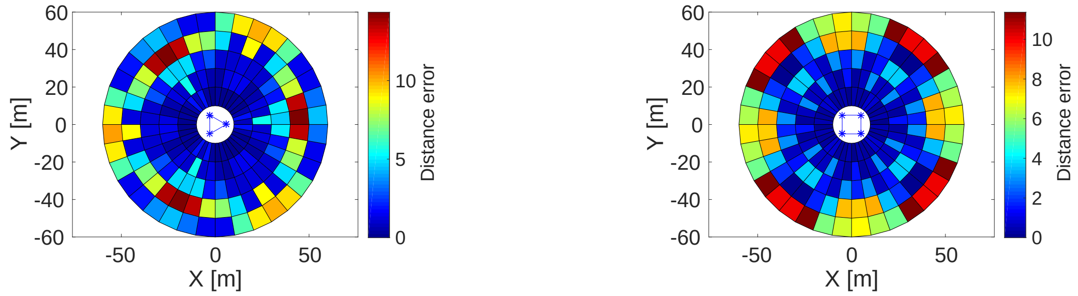


Fig. 1 On the left: Flow graph of the bird localization system, On the right: Egg-like shaped array design

a target application. Microphones distributed linearly or in a circle can be applied to effectively localize sounds on the same 2D plane as the geometry of the microphone array. By adding microphones with different elevations, 3D sound source localization can mprove. We used an egg-shaped microphone array, such as the one shown in Fig.1.

To obtain a 2D position on a plane from multiple 1D sound source directions, a triangulation algorithm is applied. A sound source position can be calculated as intersecting points of sound source directions estimated by different microphone arrays at each time window. An issue with using three or more microphone arrays is that, in a real, outdoor scenario, these directions will almost never intersect in one unique point and thus multiple intersecting points are generated for each sound source. To solve this problem, we expanded the algorithm to use the center of gravity (CoG) of the intersectig points [5].

In the case of a single sound source, the extended triangulation algorithm described above can estimate a sound source position properly. However, when multiple sound sources exist simultaneously, there is a difficulty in deciding a set of intersecting points for each sound source. To solve this problem, an outlier removal algorithm was introduced by analyzing the separated sounds using GHDSS.



a) three microphone arrays b) four microphone arrays

Fig. 2 Averaged error in sound source position estimation for the three types of bird songs.

3 Evaluation with Numerical Simulation

We conduct numerical simulation to assess the performance of 2D localization by changing the number and position of the microphone arrays in respect to the position and type of the sound source. A 16-ch egg-shaped microphone array is used with microphones distributed on the surface in three dimensions. For the evaluation, two types of microphone array layouts were tested: three and four microphone arrays distributed in the vertices of their respective regular polygons with the edge length of $10\,m$. Three different types of bird song recordings were used as sound sources: the Eastern crowned warbler, the Narcissus flycatcher and the Japanese bush warbler. The sound source positions changed in each simulation in distance and angle in respect to the array configuration. Each sound source was placed in $10\,m$, $20\,m$, $30\,m$, $40\,m$ and $50\,m$ away from the center of the microphone arrays, with the azimuth changing every $10°$.

4 Discussion

As can be noticed in Fig.2, the microphone array configuration has a big influence on the localization error. At the singular azimuths, where the directions overlap or are parallel to each other, the error has the highest value. Furthermore, these positions correspond to the most diverse distances between the sound sources and each of the microphone arrays. This is depicted by the red colors on the result figures. The best localization results will be obtained when the sound source exists near the microphone arrays, in direction of the edges of the configuration polygon, where the distances between each microphone array and the sound source are most balanced.

Another factor, which causes a large localization error is the resolution of the transfer function of the microphone arrays. Each array can localize sound sources with a $5°$ azimuth accuracy. The effect of a poorly positioned sound source can be observed on the left side of Fig.2, where the red spaces correspond to the estimated positions with the largest accuracy error, exactly in between the $5°$ azimuth estimation. In future work we plan on further develop the system to make it useful in real-time, real-environment applications. To achieve this goal, further evaluation of the various parameters in the algorithm, as well as evaluation of the outlier removal method is needed.

Acknowledgement

This work was supported by KAKENHI 16H02884, 16K00294, 17K00365 and ImPACT Tough Robotics Challenge.

References

[1] K. Nakadai et al., Active audition for humanoid, *AAAI-2000*, pp. 832-839, 2000

[2] Y. Sasaki et al., "Map-generation and identification of multiple sound sources from robot in motion", *IROS 2010*, pp. 437-443, 2010

[3] K. Sekiguchi et al., Online Simultaneous Localization and Mapping of Multiple Sound Sources and Asynchronous Microphone Arrays, *IROS 2016*, pp. 1973-1979, 2016.

[4] R. Kojima et al., "Semi-Automatic Bird Song Analysis by Spatial-Cue-Based Integration of Sound Source Detection, Localization, Separation, and Identification", *IROS 2016*, pp. 1287-1292, 2016.

[5] S. Matsubayashi et al., "Acoustic Monitoring of the Great Reed Warbler Using Multiple Microphone Arrays and Robot Audition", *JRM*, 29(1), pp. 224-235, 2017.

[6] R. Suzuki et al., "Localizing bird songs using an open source robot audition system with a microphone array", *INTERSPEECH 2016*, pp. 2626-2630, 2016

[7] D. Gabriel et al., "Iterative Outlier Removal Method Using In-Cluster Variance Changes in Multi-Microphone Array Sound Source Localization", *National Convention of IPSJ*, 6M-06, 2017

[8] R.O. Schmidt, "Multiple emitter location and signal parameter estimation", *IEEE Trans. Antennas and Propagation*, AP-34(3), pp. 276-280, 1986

[9] H. Nakajima et al., "Blind source separation with parameter-free adaptive step-size method for robot audition", *IEEE Trans. ASLP*, 18(6), pp. 1476-1485, 2010

[10] K. Nakadai et al., Design and Implementation of Robot Audition System "HARK", *Advanced Robotics*, 24(5–6), pp.739-761, 2010.