Sound Source Tracking Using Multiple Microphone Arrays Mounted to an Unmanned Aerial Vehicle

Taiki Yamada[1], Katsutoshi Itoyama[1], Kenji Nishida[1], Kazuhiro Nakadai[1,2]

***Abstract*— In this paper, we propose a source tracking method using multiple microphone arrays. While a single microphone array is generally used for sound source direction estimation, we use multiple microphone arrays in order to estimate the sound source position by triangulation. This method has the merit that the estimation error in the distance direction is suppressed and there is no need to use an auxiliary data. Experimental results through numerical simulation showed that the proposed method estimates the sound source trajectory with distance error of 4.3% on average.**

I. INTRODUCTION

Sound source tracking is an essential function for microphone array processing. However, most studies focus on tracking only sound source directions. When it comes to position estimation, there are researches that estimate the sound source direction by applying stochastic filters to the estimated direction [1] and researches that calculate the sound source position by combining the estimated direction and auxiliary data [2]. These approaches have problems such that the estimation errors in the distance direction become large, or auxiliary data such as terrain, camera, etc. is required. To solve these problems, we propose a sound source tracking method using multiple microphone arrays. Each microphone array estimates the direction of a sound source and then estimates the 3D position of the sound source by triangulation. This triangulation can suppress the error in the distance direction. In addition, the method can be applied to a moving sound source and its trajectory can be obtained. This paper overviews the situation to be considered, explain the proposed method with the evaluation results through numerical simulation.

II. PROBLEM STATEMENT

We consider a three dimensional field where the UAV can move freely in three dimensions and rotate freely in the yaw direction. The state of the UAV is $\boldsymbol{r}(t) = [x_r(t), y_r(t), z_r(t), \phi_r(t)]^T$ where t indicates time, $x_r(t), y_r(t), z_r(t)$ are the three-dimensional coordinates of the UAV in the absolute coordinate system, and $\phi_r(t)$ is the yaw angle of the UAV. Multiple microphone arrays are rigidly mounted to the UAV, which means the position of the microphone array is also known. For simplicity, the microphone array is assumed to be mounted on a circle of radius d whose center is the center of the UAV. The sound source is assumed to be single and able to move freely on three dimensions. The emitted sound is constant. The three-dimensional coordinates of the sound source in the absolute coordinate system is $\boldsymbol{e}(t) = [x_e(t), y_e(t), z_e(t)]^T$. Let $N \geq 2$ be the amount of microphone arrays. The main problem of this paper is to estimate the sound source trajectory given the acoustic signal recorded by the N microphone arrays and the state of the UAV $\boldsymbol{r}$.

[1]Department of Systems and Control Engineering, School of Engineering, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo 152-8552, Japan yamada@ra.sc.e.titech.ac.jp

[2]Honda Research Institute Japan Co., Ltd., 8-1 Honcho, Wako, Saitama 351-0188, Japan

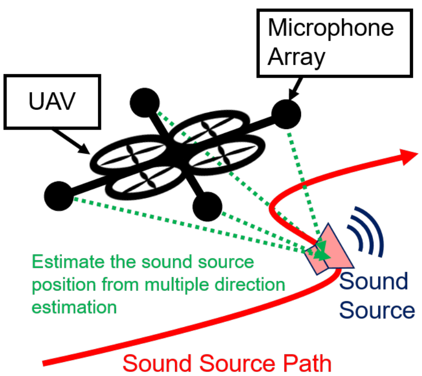

Fig. 1. Image of sound source tracking using multiple microphone arrays

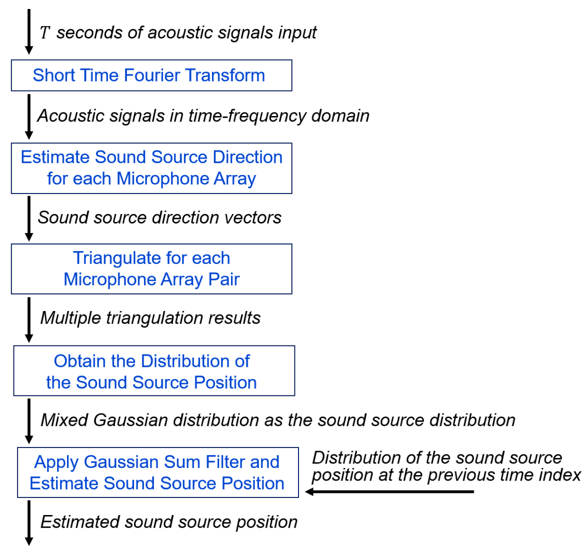

Fig. 2. Procedure in a single iteration of the proposed method

III. METHODS

The sound source trajectory is obtained by repeating sound source position estimation and filtering. The whole procedure of the iteration is illustrated in Fig. 2. After recording the acoustic signal of T seconds and applying STFT to the

TABLE I

INITIAL STATES AND DYNAMICS OF THE UAV AND THE SOUND SOURCE

	Initial State	Dynamics
Sound Source($\boldsymbol{e}$)	$(0.0, 1.0, 0.0)$ $[\mathrm{m}, \mathrm{m}, \mathrm{m}]$	$(0.1, 0.2, 0)$ $[\mathrm{m/s}, \mathrm{m/s}, \mathrm{m/s}]$
UAV($\boldsymbol{r}$)	$(0.0, 0.0, 2.0, 0.0)$ $[\mathrm{m}, \mathrm{m}, \mathrm{m}, \mathrm{rad}]$	$(0.2, 0.2, 0.0, -0.12)$ $[\mathrm{m/s}, \mathrm{m/s}, \mathrm{m/s}, \mathrm{rad/s}]$

recorded acoustic signal, we obtain the acoustic signal in the time-frequency domain. Next, the MUSIC method [3] is applied to estimate the sound source direction in each microphone array. The MUSIC method is a method of sound source direction estimation by analyzing the eigenspace of the spatial correlation matrix. After estimating the sound source direction by each microphone array, the sound source position is estimated by triangulation. In general, estimated direction vectors obtained from each microphone array do not intersect, so triangulation is performed by calculating the midpoint between two points constituting a line segment having a minimum distance between both vectors. By performing triangulation on all the microphone array pairs, $J = {}_N C_2$ triangulation results are obtained, which can be regarded as candidate points of the sound source position in a sense. In order to treat these candidates as one, the sound source position distribution is regarded as a mixed Gaussian distribution represented by the sum of Gaussian distributions in which each mean is the coordinates of the corresponding intersection point as following.

$$\frac{1}{J}\sum_{j=1}^{J}\mathcal{N}\left(\mu^j, \sigma^2 I_{3\times3}\right) \tag{1}$$

Then, a Gaussian sum filter is applied to equation (1) to correct the sound source distribution from its previous distribution. Assume that the distribution of sound sources at time index k can also be expressed as a mixed Gaussian distribution as follows.

$$\sum_{i=1}^{I_k} w_k^i \mathcal{N}\left(\boldsymbol{e}_{k|k}^i, \boldsymbol{P}_{k|k}^i\right) \tag{2}$$

I_k is the number of Gaussian distributions consisting the corrected sound source distribution at time index k. The equation equation (2) is obtained by multiplying the unscented transformation of the equation equation (2) at time index $k-1$ by the equation (1) and normalizing the weights of the product. Finally, in order to treat equation (2) as a sound source trajectory, we calculate the weighted average of each Gaussian distribution composed of a mixed Gaussian distribution.

IV. EVALUATION

Numerical simulations is done by MATLAB to evaluate the proposed method. We consider a single sound source moving while a single UAV is flying around. $N = 4$ microphone arrays are mounted to the UAV with equal intervals. The sound source is assumed to make continuous white noise. Table I illustrates the initial state and the dynamics of the sound source and the UAV. Note that the dynamics of the UAVs are defined in its own body coordinate system. The acoustic signal is recorded at 16 bits, 44.1[kHz] while white noise is added to each microphone input with signal-noise ratio -6[dB]. Fig. 3 shows the estimation results in the bird's eye view. We can see that the proposed method well estimated the sound source trajectory. Since the proposed method explicitly estimates the 3D position of the sound source, there is little error in the direction distance from the start of the estimation. In the simulation for 10 seconds, the estimation was performed within 10% of distance error and the average distance error was about 4.3%. However in the final stage, the estimation error increased as the distance between the UAV and the source increased. This is because the accuracy of the triangulation deteriorates as the distance between the UAV and the source increase.

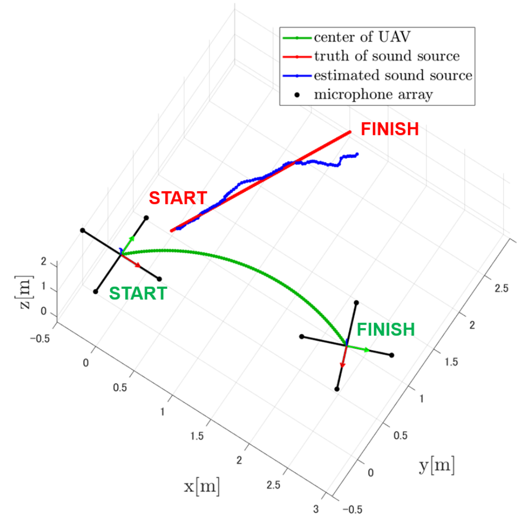

Fig. 3. Bird's eye view of the simulation result

V. CONCLUSION

In this paper, we proposed a source tracking method using multiple microphone arrays mounted to an UAV and evaluated the method by numerical simulation. The proposed method estimated the sound source trajectory with an average distance error of 4.3%. Using multiple microphone arrays allows us to perform explicit 3D sound source estimation without using auxiliary data. Sound source tracking under the situation when the UAV and the sound source having a long distance, and tracking multiple sound sources are the future tasks for this research.

ACKNOWLEDGMENT

This work was supported by KAKENHI 16H02884, 17K00365, and 19K12017.

REFERENCES

[1] A. Portello, *et al.*, Active localization of an intermittent sound source from a moving binaural sensor, Forum Acusticum, 2014.

[2] K. Hoshiba, *et al.*, Design of UAV-embedded microphone array system for sound source localization in outdoor environments, Sensors 2017, 17(11), 2535, 2017.

[3] R. Schmidt, Multiple emitter location and signal parameter estimation, IEEE Trans. on Antennas and Propagation, 34(3), pp.276–28, 1986.