Detection of user's body movement for binaural hearing aids to control of directivity

Yoshifumi CHISAKI*, Shogo TANAKA† and Tsuyoshi USAGAWA‡
*Graduate School of Science and Technology, Kumamoto University, Kumamoto, JAPAN
E-mail: chisaki@cs.kumamoto-u.ac.jp Tel: +81-96-342-3621
†Graduate School of Science and Technology, Kumamoto University, Kumamoto, JAPAN
E-mail: shogo@hicc.cs.kumamoto-u.ac.jp Tel: +81-96-342-3621
‡Graduate School of Science and Technology, Kumamoto University, Kumamoto, JAPAN
E-mail: tuie@cs.kumamoto-u.ac.jp Tel: +81-96-342-3622

***Abstract*—Estimation of sound source directions and separation of the sound sources are implemented on many products widely, and one of the applications is binaural hearing aids. In conversation using binaural hearing aids, continuous tracking of sound sources with acoustics signals are sometimes complicated because sound sources move dynamically. In order to make the tracking of sound sources simple, it is considered to be helpful to use non-verbal information in communication. Since user's body movement, including a head, corresponds to speakers' positions, it is possible to estimate communication zone where sound sources locate by the head direction. In this paper, a head movement in conversation, as non-verbal information in communication, is focused, and two zone detection methods are discussed. A rotational angle of head movement is estimated by both acceleration by an accelerometer and angular velocity by angular velocity sensor which is attached to left ear position. The classification of spatial communication zone is performed by two methods, the threshold method and the support vector machine (SVM). As the results, performance on estimation of the target direction by the threshold-based method was slightly better than that by the SVM-based method.**

I. Introduction

A binaural signal processing has highly potential to be adopted to many applications, such as a binaural hearing aids. An expected function by introducing a binaural model into hearing aids is signal enhancement. The frequency domain binaural model (FDBM) as one of binaural models can estimate a sound source direction and enhance sound source signal based on the estimated direction of sound source [1]. The frequency domain binaural model can also estimate multiple sound sources directions simultaneously [2]. However, the method sometimes estimates wrong direction when multiple sound sources exist. This error causes discontinuity of enhanced speech signal in time because the process is performed frame by frame. Spatial area for enhancement can be expected to estimate using temporal information based on the estimated sound sources directions, however, a speaker's intention in non-verbal communication is not taken into account, and it is expected to require complex method for estimating spatial area for enhancement when a head moves in a conversation with other speakers.

A spatial area for listening in a conversation and a head movement have a crucial relationship in communication. The head movement is also concerned with a signal processing based on a head related transfer function, e.g. head tracking plays important role in virtual audio display [3]. Since accelerometer and angular velocity sensor become compact with highly-developed technology, a motion sensor can be buried in an ear canal hearing aid itself. Thus, taking into account of head movement estimation has potential to assist to estimate spatial area for signal enhancement.

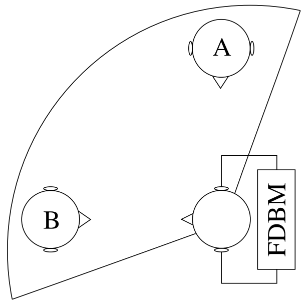

Fig. 1. Spatial area in a conversation with sound source A and B.

In this paper, head movement is detected by a combination of an angular velocity sensor and an accelerometer, and the sensors are attached to one side of ear position. Two rotation angle estimation methods are compared by experiments. One is based on threshold and the other one is based on support vector machine (SVM). Section II describes geometric mapping of a head movement onto a plane where sound sources locate. Classification methods and experiments are mentioned in section III and IV, respectively.

II. Conversation of gyro data into rotational angular

The directions of arrival can be estimated simultaneously by Frequency Domain Binaural Model (FDBM) [1][2]. In order to improve the performance on the directions estimation, the estimation of spatial area including A and B is useful as shown in Fig. 1. This section describes a detection method of head movement.

In this paper, an inertial sensor is attached to left ear position. The inertial sensor includes 3-axis accelerometer and

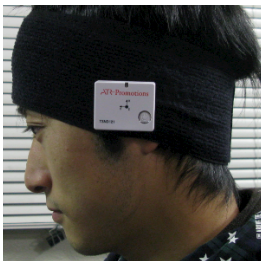

Fig. 2. A subject who wears inertial sensor TSND121 manufactured by ATR-Promotions.

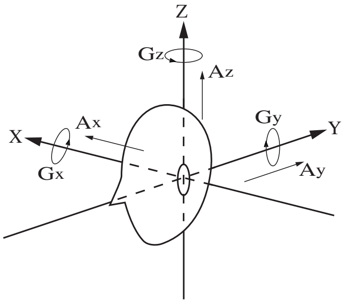

Fig. 3. Axes of inertial sensor.

3-axis angular velocity sensor. In order to make estimation process of spatial area simple, head movement and a sound source are mapped onto a horizontal plane. The conversion of the sensor output into a rotational angle on a plane is mentioned here. The relationship between axes for sensors and a rotational angle on a plane is mentioned in section II-A. In section II-B, mapping sensors data onto a horizontal plane where sound sources locate is described.

A. Sensor

Inertial sensor TSND121 manufactured by ATR-Promotions is attached to left ear position as shown in Fig. 2. The values Ax, Ay, Az by 3-axis accelerometer and Gx, Gy, Gz by 3-axis angular velocity sensor is obtained in a coordinate system as shown in Fig. 3.

B. Mapping on a rotational angle plane

Fig. 4 shows a relationship of coordinate systems. The upper figure is a head-based coordinate system, and the lower one is a gravity-based coordinate system. Fig. 5 and 6 also show a head-based coordinate system and a gravity-based coordinate system from a view along Z axis, respectively. θ in Fig. 5 and θ' in Fig. 6 are rotational angles on XY plane and X′Y′ plane, respectively. In order to θ on XY plane maps θ' onto X′Y′ plane, let us assume that coordinate (x_0,y_0,z_0) is a start point and (x_1,y_1,z_1) in an end point, respectively.

When XYZ axes are rotated with $\alpha°$, $\beta°$, $\gamma°$ against X′Y′Z′ axes, position (x_0',y_0',z_0') can be obtained by

$$\begin{bmatrix} x_0' \\ y_0' \\ z_0' \\ 1 \end{bmatrix} = R_z^{-1} R_y^{-1} R_x^{-1} \begin{bmatrix} x_0 \\ y_0 \\ z_0 \\ 1 \end{bmatrix}, \tag{1}$$

where the rotation matrices are as follows:

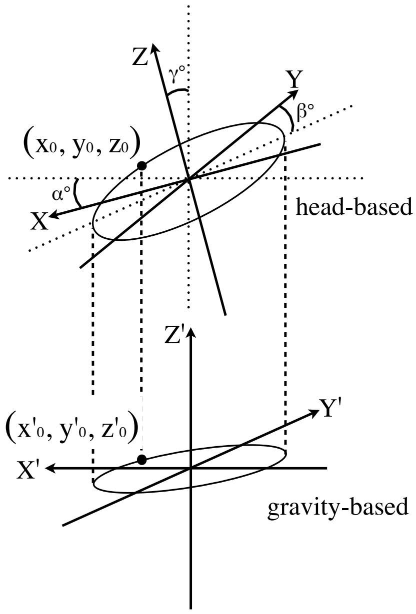

Fig. 4. Mapping onto a gravity-based horizontal plane.

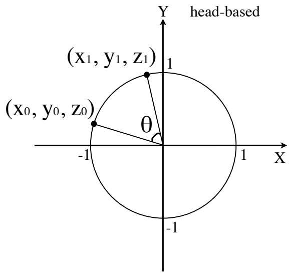

Fig. 5. θ is a rotational angle on XY plane for a head-based coordinate system. (x_0, y_0, z_0) and (x_1, y_1, z_1) are start and end points, namely, sampling points by an inertial sensor.

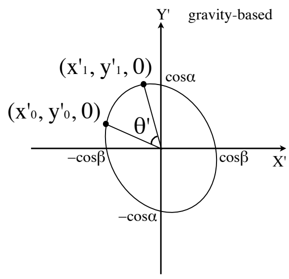

Fig. 6. θ' is a rotational angle on X′Y′ plane for a gravity-based coordinate system. (x_0, y_0, z_0) and (x_1, y_1, z_1) are mapped positions from (x_0, y_0, z_0) and (x_1, y_1, z_1).

$$R_x = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha & 0 \\ 0 & \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, R_y = \begin{bmatrix} \cos\beta & 0 & \sin\beta & 0 \\ 0 & 1 & 0 & 0 \\ -\sin\beta & 0 & \cos\beta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$R_z = \begin{bmatrix} \cos\gamma & -\sin\gamma & 0 & 0 \\ \sin\gamma & \cos\gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Since θ can be obtained by an inertial sensor, end point (x_1,y_1,z_1) is calculated by a rotational matrix. A relationship between start point (x_0,y_0,z_0) and end point (x_1,y_1,z_1) on XY plane as shown in Fig. 5 can be expressed as

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} \tag{2}$$

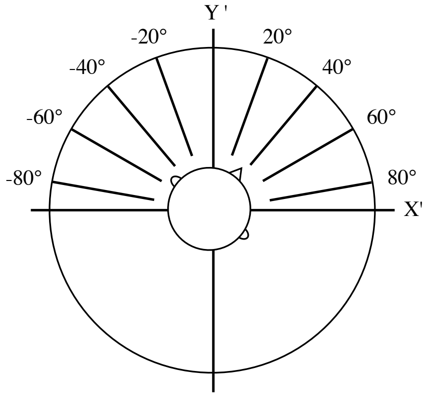

Fig. 7. A subject and target degrees.

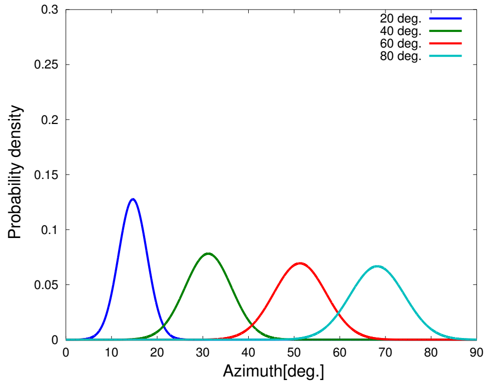

Fig. 8. Normal distribution of θ' in each target angle.

TABLE I
THRESHOLD T_{θ_i} OF BETWEEN EACH ANGLE.

Threshold					
-80°,-60°	-60°,-40°	-40°,-20°	20°,40°	40°,60°	60°,80°
-59.7	-40.8	-21.4	21.0	39.2	57.6

because z_0=z_1=0 when a head rotates on a head-based coordinate system. End point (x_1,y_1,z_1) also can be mapped onto X′Y′ plane as (x'_1,y'_1,z'_1).

Finally, a rotational angle θ' on X′Y′ plane can be obtained by (3)

$$\theta' = \arctan\left(\frac{x'_0y'_1 - x'_1y'_0}{x'_0x'_1 + y'_0y'_1}\right) \tag{3}$$

III. ESTIMATION METHODS

A rotational angle θ' can be estimated as mentioned in section II. When a conversation is performed, a center of human's head is not always fixed at the center, i.e. the origin of plane X′Y′. Furthermore, a main lobe in directivity of signal enhancement by a binaural model is not so sharp. Thus, zoning is considered to be pertinent, and a plane X′Y′ is divided into multiple zones.

In this paper, two mapping methods for rotational angle θ' onto a zone on a plane X′Y′ are discussed. One is based on a threshold and the other one is based on support vector machine (SVM). In the following sections, a data set D_θ for θ' is measured in advance for a threshold and machine learning. A data set D_θ consists of a measured rotational angle θ' by indicating a target direction with a visual marker to a subject.

A. Threshold method

Let us assume that two normal distributions N_{θ_i} and $N_{\theta_{i+1}}$ correspond to data sets of two directions D_{θ_i} and $D_{\theta_{i+1}}$ respectively. When those distributions cross over each other, the intersection of the distributions can be useful threshold. The threshold is used to decide whether a rotational angle θ' belongs to distribution N_{θ_i} or $N_{\theta_{i+1}}$. Finally, $\hat{\theta}$ is set to θ_i.

B. Support Vector Machine

Support vector machine (SVM) originated from the Optimal Separating Hyperplane developed by Vapnik et al. in the 1960s was extended to a nonlinear classifier combined with learning with kernels in the 1990s [4][5]. The SVM uses margin maximization criterion which separates positive examples and negative examples by a maximum margin hyperplane.

Since the SVM is a classifier of two classes basically, combinations of multiple SVMs is introduced to select a spatial zone labeled as $\hat{\theta}$. Classification plane H_{θ_i} between data sets D_{θ_i} and $D_{\theta_{i+1}}$ can be defined as

$$H_{\theta_i} := \mathrm{SVM}_{learn}(D_{\theta_i}, D_{\theta_{i+1}}), \tag{4}$$

where i is index of divided spatial zone. This learning process is performed in advance.

Finally, $\hat{\theta}$ can be obtained by

$$\hat{\theta} := \{\theta' | \forall i, \mathrm{SVM}_{classify}(H_{\theta_i}, \theta')\}. \tag{5}$$

The function $\mathrm{SVM}_{classify}$ gives θ_i for data set D_{θ_i}.

IV. EXPERIMENTS

Performance on spatial zone estimation is examined in this section. Conditions for data collection of head movement is mentioned in section IV-A, Section IV-B presents conditions and a classification result of a head rotational angle classification by the threshold method, and section IV-C mentions those by the SVM-based method.

A. Data collection

A data set for head movement is collected by the following conditions. Inertial sensor TSND121/ATR-Promotions is attached to left ear position as shown in Fig. 2. Ranges for accelerometer and angular velocity sensor are set to ±5 G and ±500 dps, respectively. These ranges are decided by a preliminary experiment. Sampling frequency for both sensors is 24 kHz. A center angle in azimuth for a zone is illustrated in Fig. 7, and the center angle is every 20° from -80° to -80°. A

TABLE II
CLASSIFICATION RESULTS BY THE THRESHOLD METHOD (%).

		Classification result			
		20°	40°	60°	80°
Target angle	20°	96.0	4.0		
	40°	3.0	94.7	2.3	
	60°		3.7	91.0	5.3
	80°			7.3	92.7

(a) Right side by the threshold method.

		Classification result			
		-20°	-40°	-60°	-80°
Target angle	-20°	94.0	6.0		
	-40°	2.0	96.3	1.7	
	-60°		2.3	91.7	6.0
	-80°			7.3	92.7

(b) Left side by the threshold method.

TABLE III
CLASSIFICATION RESULTS BY SVM (%).

		Classification result			
		20°	40°	60°	80°
Target angle	20°	93.0	7.0		
	40°	4.3	93.7	2.0	
	60°		5.7	91.3	3.0
	80°			9.0	91.0

(a) Right side by SVM.

		Classification result			
		-20°	-40°	-60°	-80°
Target angle	-20°	90.7	9.3		
	-40°	4.7	93.0	2.3	
	-60°		3.0	87.3	9.7
	-80°			8.3	91.7

(b) Left side by SVM.

subject was seated on a chair in front of table. The number of subjects is five and those have no disorder of motor function. The subject was instructed to turn a tip of nose towards a target angle from a front side and return to a front side. Each subject swinger a head 60 times per a session and five sessions for each angle are performed. It is also directed to keep a head at a center on $X'Y'$ plane and not to move lower half of body as possible. SInce movable area of a neck is around $\pm 60°$ in general, the movement of upper half of body is included in estimation of a head rotation.

B. Threshold method

1) Threshold: Thresholds between data sets are calculated by the mean and the variance of measured data for each data set. Fig. 8 shows distributions centered on the means of each data set. The thresholds at intersections of these distributions are shown in table I. According to the results, all the average angles in azimuth is shifted to the front side θ=0°, and the difference between the target angle and the threshold becomes larger as the azimuth increases. The results also suggest the necessity of zoning or calibration of an estimated rotational angle θ'.

2) Results of classification: Results of classification are shown in table II. The results in a table are a percentage of classification, row and column indices show the target angle θ_i and the estimated angle $\hat{\theta}$, respectively. According to the results, the threshold method can estimate the correct zone over 91%. However, the errors are slightly large in case that a target angle is 60° or 80°. The reason is considered that movable area of a neck is around $\pm 60°$.

C. Classification of head rotational angle by SVM

In SVM learning, K-fold cross-validation is used for inspection [6]. In this experiment, K is six. Thus, a data set D_{θ_i} are divided into six parts for learning data. Evaluation is performed against each learning data. The results by the SVM showed the same tendency as those by the threshold method.

V. CONCLUSIONS

In this paper, spatial area estimation with a single inertial sensor is introduced to improve performance on direction of arrival for a target sound source. Two estimation methods, threshold method and support vector machine (SVM) based method are evaluated. The accuracy on correct zone selection by both methods was over 90%, and the performance by the threshold method is slightly better than that by the SVM-based method.

According to the results, it is expected to estimate a spatial area by the single inertial sensor and it could be helpful to reduce the discontinuity of direction of arrival in time. Furthermore, the threshold method is quite simple and expected to implement on a communication devices, such as Google Glass.

In future, the zone detection method is combined with a binaural hearing aids. Therefore, it will be discussed how to integrate sound source directions estimated by both the binaural model and estimated spatial areas.

ACKNOWLEDGMENT

Part of this work was supported by JSPS KAKENHI Grant Number 24500147.

REFERENCES

[1] H. Nakashima, Y. Chisaki, T. Usagawa and M. Ebata Frequency domain binaural model based on interaural phase and level differences, *Acoustical Science & Technology*, Vol.24, No.4, pp.172-178, 2003.7.

[2] Y. Chisaki, S. Kawano, K. Nagata, K. Matsuo, H. Nakashima and T. Usagawa, Azimuthal and elevation localization of two sound sources using interaural phase and level differences, *Acoustical Science & Technology*, Vol.29, No.2, pp.139-148, 2008.3.

[3] M. Otani, T. Hirahara, Auditory artifacts due to switching head-related transfer functions of a dynamic virtual auditory display, *IEICE Trnas on Fundamentals of Electronics, Communications and Computer Sciences*, E91-A(6), pp.1320-1328 (2008).

[4] C. Cortes and V. Vapnik, Support Vector Networks, *Machine Learning*, vol.20, pp.273297, 1995.

[5] T. Kurita, "Support Vector Machine and Generalization," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, Vol.8, No.2, 2004.

[6] K. Fukunaga, *K. Introduction to Statistical Pattern Recognition, 2nd Ed*, Academic Press: San Diego, CA, USA, 1990.