Multimodal Personal Ear Authentication Using Multiple Sensor Information

Shunji Itani†, Shunsuke KIta‡ and Yoshinobu Kajikawa†
† Kasnai University, Osaka, Japan
E-mail:k938025,kaji@kansai-u.ac.jp
‡ Osaka Research Institute of Industrial Science and Technology, Osaka, Japan
E-mail: kitas@tri-osaka.jp

***Abstract*—In this study, we examine personal authentication technology using individual differences of acoustic transfer function measured from the pinna ; the Pinna Related Transfer Function (PRTF). In this paper, we consider personal authentication using a smartphone, which is convenient and can be authenticated simply by holding it against the pinna. However, it is known that in the case of using a smartphone, PRTF changes according to the direction of signal emission to the pinna during measurement, so that the authentication rate is lower than that of using earphones and headphones. The position information of the smartphone is measured by the acceleration sensor and proximity sensor and used together with the PRTF for learning at the same time. We demonstrate that the proposed method can deal with the changes in the smartphone's position of each measurement and improve the robustness.**

I. INTRODUCTION

Recently, personal authentication technology has been introduced in Internet shopping, apartments, companies and so on. However, since authentication methods using ID cards and passwords are the mainstream at present, the number of damage such as spoofing caused by loss of ID cards and password theft is increasing. Therefore, biometric authentication is drawing attention. Biometric authentication is a personal authentication technology using human's biometric information with robustness and convenience.

In research on biometric authentication technology, authentication methods using moving images of the pinna have been studied [1]. Several studies investigated that the shape of the pinna are complicated and there are large individual differences [2-5]. On the other hand, the conventional research described the feasibility of an authentication system using acoustic transfer functions of the pinna and the ear canal [6].

In the conventional study, authentication experiments were conducted using the acoustic transfer function of the pinna; the Pinna Related Transfer Function (PRTF) measured by microphones attached to earphones, headphones, and mobile phones [6]. As a result, the authentication accuracy of the mobile phone was lower than those of the headphones and the earphones. This is because that headphones and earphones cover the pinna and the ear canal. Accurate measurement of PRTF improves authentication accuracy. This fact is also used for an authentication system using canal-type earphones [7]. The classification error is shown to be less than 1%. However, headphones and earphones need to be worn for each

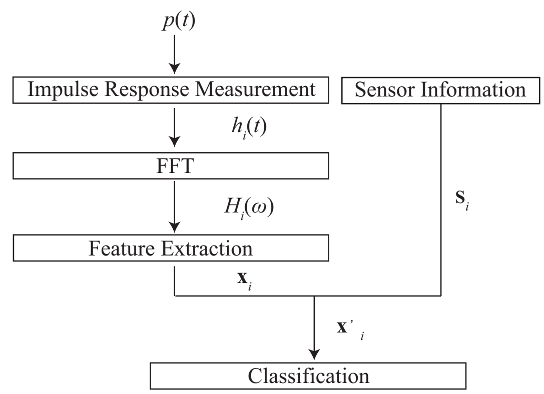

Fig. 1. Overview of personal authentication using PRTF and sensor data.

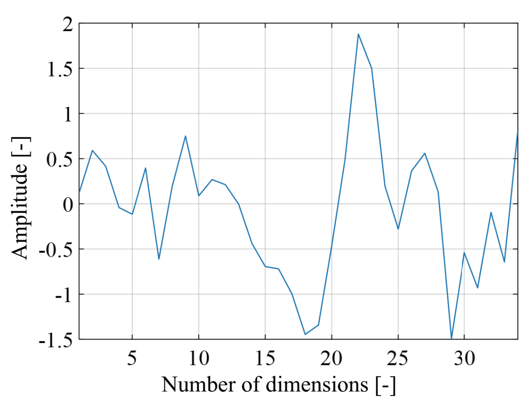

Fig. 2. Example of input data.

authentication. On the other hand, smart phone is considered more convenient because it just need to hold like calling. Therefore, we focus on an authentication system using the PRTF measured by smartphone that is as convenient as a mobile phone.

H. Takemoto et al. showed that the PRTF changes according to the direction in which the measurement signal is emitted to the pinna [8]. Furthermore, our previous study suggested that measuring PRTF from different directions and using them simultaneously for authentication improved the robustness of the authentication system. Also it is known that the error rate increases when authentication is performed with a PRTF measured at different position from the PRTF used for learning [9]. This is because the PRTF itself cannot identify the differences in position even for the same individual, and the PRTFs are judged to be other people's PRTFs due to changes in position.

From these, it is considered that authentication is possible if the measurement position is known. Therefore, by using sensor information to identify relative distance between pinna and smartphone at the time of measurement, the positional information can be interpolated through the sensor information for PRTFs that are measured at a different location from the time of learning. Thereby it is considered reducing the error rate. In this paper, we propose a multimodal method using positional information measured by a sensor. Three-dimensional acceleration and the distance of the pinna to the smartphone are measured by the acceleration sensor and proximity sensor mounted on the smartphone. The multimodal personal pinna authentication is conducted by using positional information and the PRTF together. Moreover, our previous study shows that it is effective to use support vector machine (SVM) as a classifier when the number of data is small in personal authentication using PRTF [10].

II. Personal Authentication System Using PRTFs

Fig. 1 shows the overview of the proposed personal authentication system using PRTF. First, the impulse responses (PRIR: Pinna Related Impulse Response) $h_i(t)$ and sensor information (three-dimensional acceleration and the distance between pinna and smartphone) $\mathbf{s}_i(t)$ are measured at the same time for registrants and non-registrants. Here, i indicates the number of data. In the measurement, the Time Stretched Pulse (TSP) signal $p(t)$ is used. The impulse responses are transformed to the frequency responses $H_i(\omega)$, $i \in N$ by FFT. Next, sub-banding is applied by dividing the PRTF into 30 subbands and taking the average of the sum of amplitude. The values of vector after sub-banding are extracted as a feature vector $\mathbf{x}_i$. The input data $\mathbf{x'}_i$ are made by combining the feature vector $\mathbf{x}_i$ and the sensor information $\mathbf{s}_i(t)$. Standardization is conducted so that the average is 0 and the variance is 1 for the entire training data, and label data which is $+1$ for registrants or -1 for others. An example of the multimodal data is shown in Fig. 2. The 1st to 30th dimensions represent PRTF, and the 31st to 34th dimensions represent sensor information. Finally SVM is used as two-class classification which divides users to registrants or others.

A. Support Vector Machine (SVM)

SVM is a classifier that constructs a discrimination plane between two classes as shown in Fig. 3. Classifier $g(\mathbf{x})$ is given by

$$g(\mathbf{x}) = \begin{cases} 1 & (f(\mathbf{x}) > 0) \\ -1 & (f(\mathbf{x}) < 0) \end{cases},$$

TABLE I
Measurement conditions

Number of Subject	9
Input signal	TSP signal
Number of averaging	8
Length of TSP signal	65536
Sampling rate	48000 Hz
FFT point	48000
Sensor Input Voltage	3.3V

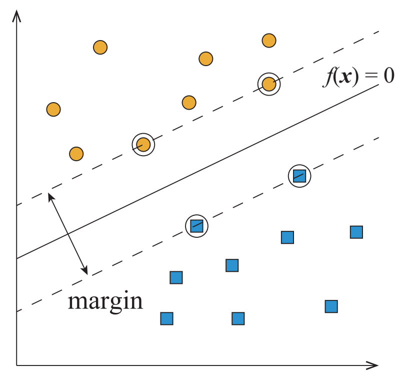

Fig. 3. Examples of classification boundary

TABLE II
Experimental conditions

Registrant training data	100
Imposter training data	100
Registrant test data	80
Imposter test data	80
Kernel function	Linear, RBF
Regularization coefficient C	0.001, 0.01, 0.1, 1, 10, 100
Kernel coefficients γ	0.001, 0.01, 0.1, 1, 10, 100
Estimation method	Cross validation (Number of divisions : 10)

$$f(\mathbf{x}) = \mathbf{w}^{\mathrm{T}}\mathbf{x} + b,$$

where $\mathbf{w} \in \mathbb{R}^M$ is the M-dimensional weight vector and $b \in \mathbb{R}$ is the scalar bias. Under the condition where all observed points are correctly classified, this optimization problem maximizing margin is shown as

$$\min_{\mathbf{w},b} ||\mathbf{w}||^2, \quad \text{s.t. } y_i(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b) \geq 1, i \in [n]. \tag{1}$$

However, there are often no discrimination planes that can completely separate the training data in reality. In such a case, the relaxation variable given by

$$\xi_i \geq 0, i \in [n], \tag{2}$$

to allow training data that is not correctly classified. Eq.(1) is replaced as

$$\min_{\mathbf{w},b,\xi} ||\mathbf{w}||^2 + C\sum_{i=1}^{N}\xi_i,$$
$$\text{s.t. } y_i(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b) \geq 1 - \xi, i \in [n], \tag{3}$$

where C is a positive constant called the regularization coefficient.

III. Experimental Method for Personal Authentication

Fig. 4. shows front and back of a handmade measurement device like a smart phone. We made our own measuring device because it is easy to control. The sizes of the measuring device is 140.0mm high, 70.0mm wide, and 1.0mm depth and created with a 3D printer. It is equipped with a micro speaker (Hosiden HDR9254), microphones (Hosiden KUB2823), proximity sensor (STMicroelectronics VL6180X) and acceleration sensor (Analog Devices ADXL354). Fig. 5. shows the block diagram of PRTF measurement. The PRTFs are measured by using the handmade measurement device shown in Fig. 4. The environment requires a personal computer with an external sound card (Fireface UC), microphone amplifier, and their respective connection cables shown in Fig.4. and Fig.5. The measurements were conducted in a sound proof room ($2.9 \times 3.1 \times 2.1$ m^3) that walls are treated with sound absorbing textile material to suppress reflection. Table I shows the measurement conditions. The subjects were 9 men and women in their 20's and the measurement were made in left ear, 20 times per person, and the measurement device was repositioned every time. However, to make the number of data the same for registrant and imposter, only the pinna used as a registrant was measured 200 times.

The experimental conditions for personal authentication are shown in Table II. The hyperparameters of SVM are determined by grid search.

IV. Results and Discussion

epstopdf test.eps In this section, we compare the results of the authentication accuracy between the conventional method and the proposed method. The conventional method uses only PRTFs [9], [10]. The proposed method utilizes multimodal. As the evaluation criteria, the false rejection rate (FRR), the false acceptance rate (FAR), and the half total error rate (HTER) are used.

The authentication result is shown in Table III. In the case of multimodal, FAR is improved by 2.9%, FRR is improved by 7.7%, and HTER is improved by 5.3%. The reason for the improvement in accuracy seems to be the positional information was interpolated by the sensor information used for learning together with the PRTFs. The authentication system became robust against changes in the PRTFs due to fluctuations in the measurement position that occur during authentication.

TABLE III
Classification Result

	FRR	FAR	HTER
PRTF only	3.9%	10.3%	7.1%
Multimodal	1.0%	2.6%	1.8%

TABLE IV
Classification Result (sensor information only)

FRR	FAR	HTER
13.9%	20.3%	17.1%

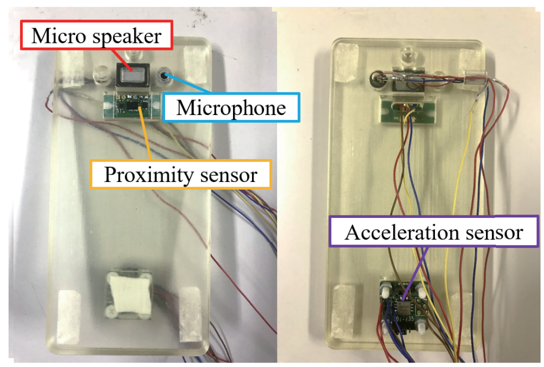

Fig. 4. A handmade measurement device

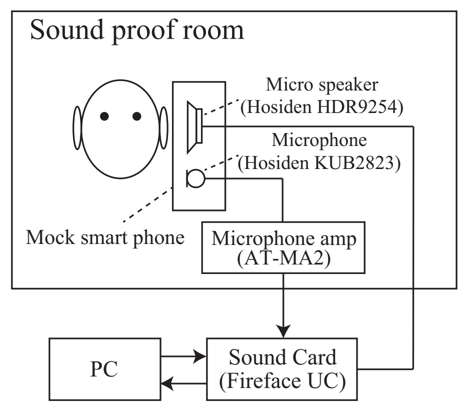

Fig. 5. Block diagram and configuration of a measurement system for PRTF.

Table IV shows the authentication result when the authentication is conducted using only the sensor information. The accuracy of authentication using acceleration and distance sensors is about 20%. Lee, W. H. et al. [11] shows personal authentication using sensors such as acceleration sensors, gyro sensors, and magnetic sensors mounted on smartphones. This research showed that it is possible to identify an individual only with the sensor information of the smartphone. It is also stated that the authentication accuracy is improved by increasing the types of features. From this, it is suggested that the relative distance between the smartphone and the pinna can be treated as a feature in this case as well. The number of effective features is increased, so that the authentication accuracy is improved.

V. Conclusion

In this paper, we proposed an authentication system that uses individual differences in the PRTF measured with a smart phone. We examined the effectiveness of a system using sensor information that represents positional information. As a result, it was possible to improve the authentication accuracy by acquiring the positional information from the sensor. The reason why the authentication accuracy was improved is that the positional information was interpolated by the sensor information.

That made the authentication system robust against changes in the PRTF due to fluctuations in the measurement position which occur during authentication. In addition, it was possible that the way of holding the smart phone obtained from the sensor information can be treated as a feature. Therefore the effective feature had increased, so that the authentication accuracy was improved.

As a future work, we are considering to use a lot of learning data of the pinna in order to implement into the practical use. A method of changing the shape of the model on the 3D data and obtaining the PRTFs using an acoustic numerical analysis method such as the finite element method can be considered as an effective method for increasing the number of pinna data. Also, considering a practical use, we need to conduct authentication experiments in an environment other than a sound proof room. The error rate will probably increase when the background noise increases, but we need to verify how much the background noise affects the authentication rate and consider how to deal with it.

Acknowledgement

This paper is financially supported by JSPS KAKENHI (18K19791).

References

[1] N. Tashiro, M. Abe, T. Okamura, K. Shinohara, “A study of individual identification using length of components on pinna using principal component analysis,” ITE Technical Report, vol. 24, no. 41, pp. 43–48, Jul. 2000.

[2] D. Watabe, T. Minamidani, H. Sai, T. Maeda, T. Yamazaki, and J. Cao, “Estimating a shooting direction in ear recognition,” 2015 International Conference on Biometrics and Kansei Engineering, pp. 559–568, Warsaw, Poland, Sep. 2015.

[3] D. Watabe, T. Minamidani, W. Zhao, H. Sai, and J. Cao, “Effect of barrel distortion and super-resolution for single-view-based ear biometrics rotated in depth,” 2013 International Conference on Biometrics and Kansei Engineering, pp. 183–188, Tokyo, Japan, Jul. 2013.

[4] D. Watabe, H. Sai, K. Sakai, and O. Nakamura, “Improving the robustness of single-view ear-based recognition under a rotated in depth perspective,” 2011 International Conference on Biometrics and Kansei Engineering, pp. 179–184, Kagawa, Japan, Sep. 2011.

[5] K. Shinohara, “Individual authentication with pinna : Analysis of individual with pinna,” ITE Technical Report, vol. 25, no. 22, pp.25–30. Mar. 2001.

[6] A. H. M. Akkermans, T. A. M. Kevenaar, D. W. E. Schobben, “Acoustic ear recognition for person identification,” Fourth IEEE Workshop on Automatic Identification Advanced Technologies, pp.219–223, Eindhoven, Netherlands, Oct. 2005.

[7] S. Yano, T. Arakawa, T. Koshinaka, H. Imaoka, H. Irisawa, “Improving acoustic ear recognition accuracy for personal identification by averaging biometrics data and spreading measurement errors over a wide frequency range,” IEICE Trans. Fundamentals (Japanese Edition), vol. J100-A, no. 4, pp. 161–168, Apr. 2017.

[8] H. Takemoto, P. Mokhtari, H. Kato, and R. Nishimura., “Mechanism for generating peaks and notches of head-related transfer functions in the median plane,” The Journal of the Acoustical Society of America, vol.132, no.6, pp.3832–3841, Dec. 2012.

[9] S. Itani, Y. Kajikawa, S. Kita, “Improving Robustness of Authentication System Based on Pinna Related Transfer Function,” 2019 IEEE 8th Global Conference on Consumer Electronics, Osaka, Japan, pp. 1079–1080, Oct. 2019.

[10] Y. Higashiguchi, Y. Kajikawa, S. Kita, “A personal authentication system based on pinna related transfer function,” International Conference on Biometrics and Kansei Engineering, Kyoto, Japan, pp. 123–126, Sep. 2017.

[11] Lee, W. H., & Lee, R. B., ”Multi-sensor authentication to improve smartphone security.“ 2015 International Conference on Information Systems Security and Privacy (ICISSP). IEEE, pp. 1-11, Feb. 2015.