Ear Acoustic Biometrics Using Inaudible Signals and Its Application to Continuous User Authentication

Shivangi Mahto, Takayuki Arakawa, Takafumi Koshinaka
Data Science Research Laboratories
NEC Corporation, Japan
Email: s-mahto@cp.jp.nec.com, t-arakawa@cp.jp.nec.com, koshinak@ap.jp.nec.com

***Abstract*—This paper presents an improved version of a previously-proposed ear-acoustic biometric system for personal authentication. Even though the previous system provided a fast, accurate, and easy means of authentication, it employed noticeably audible probe signals to extract ear acoustic-features, signals which might interrupt user activities. To overcome this problem, this paper presents silent user authentication by employing inaudible signals in the place of audible signals for capturing ear acoustic-features. A comparative study using a number of audible and inaudible signals demonstrates that inaudible signals provide accurate authentication under the condition that the relative position of the earphone device against the ear canal is constant, which is a requirement for continuous user authentication. On the other hand, audible signals offer better accuracy when the earphone position changes, which is often the case in initial user authentication. This suggests the idea of a hybrid system that employs both audible and inaudible signals for, respectively, accurate initial authentication and user-friendly continuous authentication.**

I. Introduction

Biometric authentication minimizes the risk of information being lost, forgotten, stolen, or leaked. Various kinds of biometric authentication have been studied over the past several decades, including fingerprint, facial, iris, retina, and voice recognition [1], [2], [3]. These approaches normally require users to perform some kind of action, such as putting a finger on a scanner or facing a camera. With the proliferation of in-ear personal-assistant devices, more commonly referred to as hearable devices, such as 'The Dash' [4] and 'Xperia Ear' [5], there is a greater need for in-ear biometric authentication. Biometric authentication based on human ears has been mostly discussed in the context of image recognition [6]. Like the ridge patterns in fingerprints, outer ear (pinna) patterns carry personal identity information and are more stable than those of faces, which change with changes in facial expressions. While considerable progress has recently been made in ear biometrics, there still remain several technical difficulties arising from such factors as illumination variation and occlusion, which are often encountered in outdoor conditions [7].

Another promising approach to in-ear authentication is based on the unique acoustic characteristics of individual ear canals. Evidence for personal identity in ear acoustics can be found in research done on virtual reality systems in which users sense 3D sounds via binaural earphones [8]. Methods for modeling and estimating external ear acoustics has been presented in [9], [10]. These studies, however, do not focus on authentication of a person by means of ear canal acoustics. An initial investigation on using the acoustic properties of the pinna and ear canal for recognition has been reported in [11], [12]. These works refer to this kind of recognition as *acoustic ear recognition*. The term *acoustic* was used to differentiate it from ear recognition [6] that focuses on recognizing pinna pattern shapes on the basis of image processing. Since our work hinges on the acoustic characterization of the ear canal for personal authentication, we refer to this as ear acoustic biometrics. Recently, there has been a revival in ear acoustic biometrics research. Advances in MEMS technology has allowed transducers (microphones and loud-speakers) to be built in far more miniature sizes and with relatively flat responses.

In earlier studies, [11] and [12], audio signals with audible frequency ranges up to 15 kHz were used to capture ear acoustics of individuals, and they achieved error rates ranging from 0.8% to 18.0%, depending on the type of capturing device. Another work in this line has reported holistic development of a biometric system based on acoustic ear recognition, but its best performance showed an error rate of 14.9% [13]. The high error rates of such methods restrict their applicability to real-world applications.

In [14], the authors have proposed a unique biometric authentication method that exploits the acoustic characteristics of human ears and achieves an EER of less than 1%. It transmits a probe sound signal from an earphone device to the ear canal of an individual and records an echo signal. Then, using the probe and echo signals, it extracts ear acoustics for the individual. The system is fast, accurate, and does not require the user to perform any kind of action. However, the audible sound signals used in this method are noticeable and may often interrupt user activities during authentication.

In this paper, we attempt to overcome this problem by employing inaudible sound signals to achieve silent authentication, so that repetitive and continuous authentication [15] will not irritate or interrupt the user. To the best of our knowledge, this is the first attempt reported in the literature to use inaudible signals for ear biometric authentication. In our study, we analyze and demonstrate its efficacy in comparison to conventional audible-signal-based authentication.

This paper is organized as follows: Section II describes our ear acoustic authentication system; Section III presents application of inaudible signals for silent authentication and proposes a hybrid ear authentication method; Section IV

ISBN 978-90-827970-1-5 © EURASIP 2018

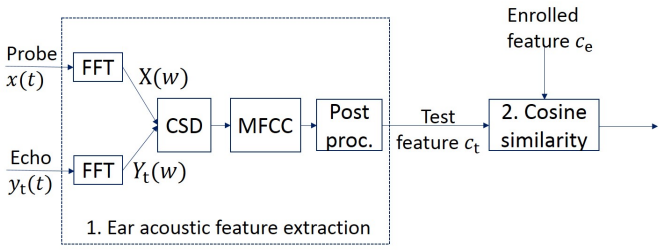


Fig. 1: Overview of ear authentication system. FFT refers to Fast Fourier Transformation, CSD to Cross Spectral Density extraction, MFCC to MFCC extraction, and Post proc. to further processing steps, such as normalization and dimensionality reduction.

describes data collection and presents results of performance evaluations with respect to inaudible and audible probe signals. In Section V, we summarize our work and discuss issues to be covered in future work.

II. Ear acoustic authentication system

Fig. 1 shows the system architecture for ear acoustic authentication [14]. For each authentication trial, a probe signal $x(t)$ is transmitted through the ear canal, and the echo signal $y(t)$ is recorded. Given a pair of such echo signals, $(y_e(t), y_t(t))$, recorded over enrollment and test stages, we attempt to determine whether or not they belong to the same individual, with probe signal $x(t)$ assumed to be known prior to the authentication.

A. Ear acoustic feature extraction

For each of the two echo signals, ear acoustic features are extracted using the cross spectrum between $y_i(t), i \in \{e, t\}$ and $x(t)$. The steps are as follows:

- **Fast Fourier Transformation (FFT)** is applied to $x(t)$ and $y(t)$ to obtain their Discrete Fourier Transforms $X(w)$ and $Y(w)$, respectively. Here, $w = \frac{2\pi k}{N}$, where $k \in \{0, 1, ..., N-1\}$ and N is the length of input sequences $x(t)$ and $y(t)$.
- **Cross spectral density (CSD)** between $X(w)$ and $Y(w)$ is calculated as

$$H(w) = \frac{X(w)^* Y(w)}{X(w)^* X(w)} \tag{1}$$

, where $X(w)^*$ is the complex conjugate of $X(w)$. Notice that $H(w)$ represents the ear canal transfer function.
- **Mel-frequency Cepstral Coefficients (MFCCs)** [16] are extracted from the magnitude spectrum $|H(w)|$ of the ear canal transfer function. To this end, Mel-filter banks are applied, and a logarithmic compression and discrete cosine transform (DCT) are then applied. After denoting the output of an M-channel filter bank as $S(m), m = 1, 2, ..., M$, the MFCCs can then be computed as

$$c_n = \sum_{m=1}^{M} \log S(m) \cos\left[\frac{\pi n}{M}(m - \frac{1}{2})\right] \tag{2}$$

, where $n = 1, ..., N_f < N_d$ is the index of cepstral coefficients. In comparison to other features (e.g., linear frequency cepstral coefficients), we have found that mel-filters help more in reducing intra-class variance in ear acoustic features for higher frequencies, particularly for inaudible signals. Also, the application of MFCCs is well-known for its great success in various fields of speech information processing, including automatic speech recognition (ASR), speaker recognition, and emotion recognition. We consider it to potentially have more than a little advantage in acquiring the characteristics of ear acoustics at quite low dimensions (e.g., 20 in our experiments).
- Further, we post-process the MFCCs by removing the mean and scaling each dimension to unit variance. The normalized MFCCs are then transformed to a low-dimensional space using linear discriminant analysis (LDA) [11].

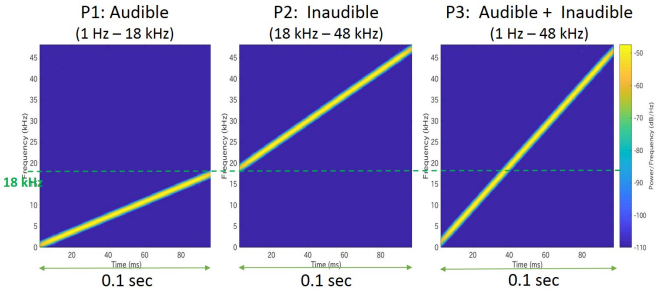


Fig. 2: Spectrograms of the three linear chirps used for experimental purposes. P1 sweeps across the entire range of audible frequencies (1 Hz - 18 kHz). The P2 signal covers inaudible frequencies (18 kHz - 48 kHz) and P3 sweeps across the entire available frequency range (1 Hz - 48 kHz) covering both audible and inaudible ranges.

B. Similarity Measure

For computing the similarity between extracted ear acoustic features (c_e, c_t) corresponding to those extracted using the two echo signals $(y_e(t), y_t(t))$, cosine similarity is defined as follows:

$$s(c_e, c_t) = \frac{{c_e}^{\mathrm{T}} c_t}{\sqrt{\|c_e\| \, \|c_t\|}} \tag{3}$$

where $c = [c_1, c_2, ..., c_{N_f}]^{\mathrm{T}}$. If $s(c_e, c_t) > \theta$, then the two features are considered to be from the same person (ear). Here, θ is a predetermined threshold.

III. Silent authentication using inaudible signals

To achieve silent authentication, we attempted to use inaudible signals for capturing ear acoustics. We believe that inaudible signals are as capable as audible signals in capturing unique individual ear-canal features, as will be shown later in our experimental results. Since the audible frequency range of an adult usually lies in the range of 20 Hz to 16 kHz, we regard audio signals with frequencies of 18 kHz or above as inaudible signals.

While the ear canal acts as a resonant system with a typical resonance frequency at around 2.5 kHz, it may vary from person to person [11]. Hence, the audible frequency range can be used to capture the dominant formants representing person-dependent traits. However, the length of an ear canal and its curvatures have dimensions that vary from millimeters

ISBN 978-90-827970-1-5 © EURASIP 2018

to centimeters, which can only be captured using smaller wavelengths, and, for this purpose, frequencies in an inaudible range can be helpful in modeling minuscule variations.

In earlier work, due to hardware limitations, only frequencies up to 16 kHz could be used to collect the ear acoustic data. With recent advancements in microphone-integrated earphones, frequencies as high as 48 kHz can be transmitted through them, and they have become increasingly available on the market. For microphone-integrated earphones, we used a developmental version of EarPods that has a maximum operating frequency of 48 kHz, on a desktop computer with an audio interface having a 96 kHz sampling frequency and a 24-bit depth. Hence, we were able to transmit inaudible signals with a maximum frequency as high as 48 kHz.

For our experimental purposes, we collected ear acoustic data for individuals using three kinds of sinusoidal linear chirps, each 0.1-second-long, as probe signals (P1, P2 and P3 as shown in Fig. 2). A linear chirp is a signal in which the frequency varies exactly linearly with time. Chirp is used because it covers all the frequencies and is easy to manage. Also, inaudible signals can offer sufficient robustness against background noise, such as speech and music, as the frequency bands of the inaudible signal and the noise do not overlap.

A. Analysis of the effects on captured ear acoustics of different positioning of an earphone against the ear canal

To investigate the effects on captured ear acoustics of changes in the relative position of an earphone against the ear canal, we analyzed the mel-spectrum features of the ear acoustics of an individual, captured by the means of the probe signals sweeping over the entire frequency range of 1 Hz to 48 kHz (P3 as shown in Fig. 2) for 2 scenarios. The first scenario compared captured ear acoustic features when the relative position of earphone against the ear canal was fixed across all the recordings of ear acoustics (Fig. 3a), while the second scenario analyzed captured ear acoustic features when the positioning of the earphone was varied across all the recordings (Fig. 3b). To introduce variability in the positioning of earphones, we asked the individuals to remove and replace their earphones after each recording. Note that a single recording involved transmitting a probe signal one time and receiving the corresponding echo signal. Features captured by means of the P3 probe signals were carefully analyzed because they contained both audible and inaudible parts, and, hence, we were able to analyze the earphone's positioning effect on the two frequency ranges as well.

In the first scenario, we observed that the captured ear acoustic features had negligible intra-class variance for both the audible and inaudible parts. In contrast to this, with different positioning of earphones, we observed a large variation across the captured ear acoustic features. Also, the inaudible parts had higher variability across different recordings than the audible parts. Similar observations were obtained for other individuals as well. This indicates that the placement of an earphone against the ear canal contributes significantly to the captured ear acoustic features. Also, the smaller wavelengths of inaudible signals are more sensitive than those of audible signals to the relative position of the earphone device against the ear canal. That is why we propose a hybrid system that employs both audible and inaudible signals at, respectively, initial and continuous authentication stages. Such a hybrid system is able to maintain both high authentication accuracy and good usability.

TABLE I: Authentication steps, relative position of earphone device against ear canal across recordings during those authentication steps, and the probe signals used in the corresponding authentication steps

Authentication step	**Relative position of earphone**	**Probe signal**
Initial	Different position	Audible
Continuous	Fixed position	Inaudible

B. Proposed hybrid system

We here propose an authentication system employing the 2-step authentication (Table I), as explained below:

- Initial authentication: First time authentication after wearing the earphone. Here, the relative position of the earphone device will always be different between enrollment and test authentication sessions. According to our above-described analysis, this condition will lead to higher variability in the ear acoustic features captured by inaudible signals as compared to those of audible ones. Thus, for higher accuracy, audible probe signals will be used for this authentication.
- Continuous authentication: Authentication when the earphone device has not been removed after the initial authentication. The relative position of the earphone device against the ear canal does not change across authentication sessions and, hence, we expect inaudible signals to perform as well as audible, along with no interruption in user activities. Therefore, for this step inaudible probe signals are used for silent, continuous authentication.

Both of the authentication steps follow the authentication procedure described in Section II.

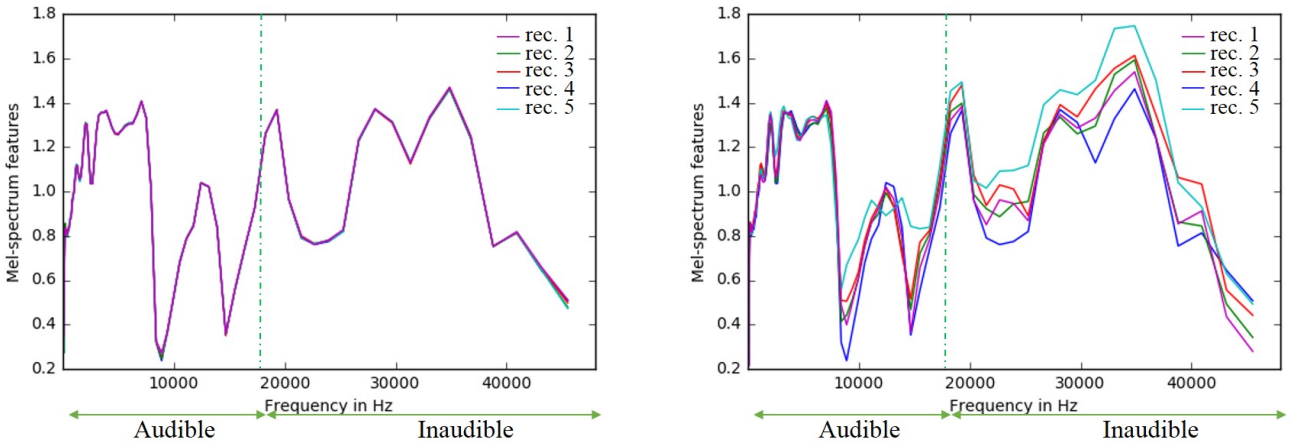
IV. Evaluation

A. Data collection

To evaluate the performance of ear acoustic authentication using inaudible as well as audible signals, we collected ear acoustic data for 25 individuals, using the audible (P1) and inaudible (P2) probe signals described in Section III. For each individual, ear acoustics were collected over five recording sessions in which the earphone was removed and then replaced into the ear canal after each session so as to introduce variability in wearing conditions. During each session, the relative position of the earphone against the ear canal was assumed to be fixed, and each of the two probe signals was transmitted five times to capture ear acoustics under the condition of fixed earphone position.

For each individual for each of the two probe signals, then, we were able to capture five sets of ear acoustics, all featuring different earphone wearing conditions, to be used for the evaluation of the initial authentication step. Also, each

ISBN 978-90-827970-1-5 © EURASIP 2018



(a) Fixed earphone position across all recordings.

(b) Different earphone positions across all recordings.

Fig. 3: Mel-spectrum features of ear acoustics belonging to the same person, captured during various recordings (rec.). Fig. 3a shows the ear acoustic features captured when the relative position of the earphone against the ear canal was fixed for all five recordings, while Fig. 3b shows the ear acoustic features captured when the relative position of the earphone against the ear canal was different for all five recordings because of the repositioning of the earphone after each recording.

TABLE II: Systems used for evaluation

System	Feature	Post proc.	Similarity measure
System 1	MFCCs	Normalization	Cosine Similarity
System 2	MFCCs	Norm.+ LDA	Cosine Similarity

of the five sets contained five ear acoustics, all captured with the same earphone positioning against the ear canal, to be used for evaluation of the continuous authentication step.

B. Experimental setup

We compared the performance of audible and inaudible probe signals for both of the initial and continuous authentication steps using two evaluation systems. Each of the two systems basically conformed to the architecture shown in Fig. 1, with two different types of post-processing to MFCCs (Table II). For authentication, 9600-point FFT was applied on both the probe and echo signals to extract 20-dimensional MFCCs representing the ear acoustics of an individual. Training data was prepared for training the post-processing steps on the MFCCs. A set of 25 subjects was randomly divided, 5 times, into two subsets of 15 and 10 subjects. For each of the 5 times, ear acoustics features for 15 subjects were used for calculation of means and variances for normalization, as well as for training for LDA transformations. LDA transformation matrices were learned in order to transform 20 dimensional MFCCs into 12 dimensional features. The above-mentioned parameters were chosen on the basis of earlier evaluations that are explained in [14]. Ear acoustic features corresponding to the other 10 subjects were used for creating evaluation data.

For evaluating the performance of the proposed system, we created genuine and impostor pairs for both the initial and continuous authentication steps. A "genuine pair" is one in which enrollment and test data belong to the same individual, whereas, for an impostor pair, the enrollment and test data belong to different persons. Equal error rates (ERR) were observed for each of the 5 times of evaluation, on the basis of which the overall performance of the proposed systems was evaluated by averaging EERs over the 5 evaluations. An EER is one having an operating point (threshold θ) for which the false acceptance rate and false rejection rate are equal.

For the continuous authentication step, genuine pairs for an individual were created by picking ear acoustic features captured during the same recording session. Such features would include the same earphone wearing conditions that are required for continuous authentication. For the initial authentication step, genuine pairs were created for an individual by picking ear acoustic features captured during 2 different recording sessions. Such features would have different earphone positions against the ear canal, which meets the initial-authentication requirement. For impostor pairs, in both the initial and continuous authentication steps, a pair of ear acoustic features corresponding to two different individuals was chosen. It should be noted that, for impostor pairs, the position of the earphone against the ear canal for two individuals would always be different, and, hence, along with between-person variability, variability due to different earphone positions was also implicitly included in the impostor pairs.

For the initial authentication step, to handle intra-class variations in ear acoustic features arising from different earphone wearing conditions, we performed triple-enrollment evaluations in addition to the usual single enrollment evaluations. Under triple-enrollment evaluations, a given test feature was compared against 3 different enrollment features, and the maximum of 3 resulting scores was chosen as the final score. This kind of evaluation helps in handling large intra-class variability [17].

ISBN 978-90-827970-1-5 © EURASIP 2018

TABLE III: Performance (EER %) of the audible and inaudible probe signals in the continuous authentication step for the 2 evaluation systems.

	System 1 (Norm)	System 2 (Norm + LDA)
P1:Audible	0.95	0.28
P2:**Inaudible**	**<0.01**	**<0.01**

TABLE IV: Performance (EER %) of audible and inaudible probe signals in initial authentication step for the 2 evaluation systems. Evaluations were done under both single and triple-enrollment conditions.

	System 1 (Norm)		System 2 (Norm + LDA)	
	1-enr	3-enr	1-enr	3-enr
P1:**Audible**	6.76	2.27	4.47	**1.44**
P2:Inaudible	11.96	3.95	11.00	4.72

C. Experimental results

1) For continuous authentication, inaudible signals performed best: Table III shows the performance of the audible and inaudible probe signals in the continuous authentication step for the two evaluation systems. Both of the probe-signal types performed well, with inaudible signals having the lowest EER. This shows that use of inaudible signals offers potential for silent continuous authentication along with high accuracy as compared to the conventional use of audible signals alone.

2) For initial authentication, audible signals performed best: Table IV shows the performance of the audible and inaudible probe signals in the initial authentication step for the two evaluation systems. Evaluations were done under both single and triple-enrollment conditions. It can be seen that triple-enrollment evaluation resulted in better performance for both the probe-signal types than did single-enrollment evaluation. Also, audible signals (P1) with System 2 settings offered the best performance, with an EER of 1.44%. Hence, they should be used for initial authentication in order to obtain high accuracy.

3) Different earphone-wearing conditions led to higher variability in the captured ear acoustic space of an individual: Tables III and IV show that for each of the two probe-signal types, performance in the initial authentication step was worse than that in the continuous authentication step. Also, for inaudible probe signals, degradation was much higher than that for audible signals. These results agree with our earlier observations that different earphone-wearing conditions result in greater within-person variability in captured ear acoustic features. Also, the variability is much larger for inaudible frequencies and, hence, authentication performance worsens, irrespective of the type of probe signals.

V. Summary and Future Work

We have described a biometric authentication system using inaudible signals for capturing the acoustic characteristics of ears and offering the key feature of silent continuous authentication. Such authentication is not affected by background audible noise, such as speech or music. Hence, we can authenticate a person while he/she is listening to music or talking to someone. However, initial authentication is challenging because of the variability introduced by repositioning of earphones against ear canals. Also, inaudible signals are more sensitive than audible signals to such repositioning. In response to this, we have proposed a hybrid system using audible probe signals for initial authentication alone, followed by silent continuous authentication using inaudible probe signals.

We intend in future work to focus on improving the performance of initial authentication and to include experiments with a larger amount of training data, detailed analyses of the nature of ear acoustic features captured by audible and inaudible probe signals in different recording sessions, and extraction of robust ear acoustic features by removing factors that depend on earphone-wearing conditions.

References

[1] M. Mizoguchi and M. Hara, "Fingerprint/palmprint matching identification technology," *NEC Technical Journal*, vol. 5, pp. 18–22, 2010.

[2] H. Imaoka, "Face recognition research: Beyond the limit of accuracy," *The IAPR 2014 Biometrics Lecture, The 2014 International Joint Conference on Biometrics (IJCB 2014)*, 2014.

[3] T. Koshinaka, O. Hoshuyama, Y. Onishi, R. Isotani, and M. Tani, "Speech/acoustic analysis technology-its application in support of public solutions," *NEC Technical Journal*, vol. 9, no. 1, pp. 82–85, 2015.

[4] "The Dash from Bragi," https://www.bragi.com/thedash/.

[5] "Xperia Ear from SONY," https://www.sonymobile.com/global-en/products/smart-products/xperia-ear/.

[6] Ž. Emeršič, V. Štruc, and P. Peer, "Ear recognition: More than a survey," *Neurocomputing*, vol. 255, pp. 26–39, 2017.

[7] A. Abaza, A. Ross, C. Hebert, M. A. F. Harrison, and M. S. Nixon, "A survey on ear biometrics," *ACM computing surveys (CSUR)*, vol. 45, no. 2, p. 22, 2013.

[8] S. Yano, H. Hokari, and S. Shimada, "A study on personal difference in the transfer functions of sound localization using stereo earphones," *IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 83, no. 5, pp. 877–887, 2000.

[9] M. Hiipakka, M. Tikander, and M. Karjalainen, "Modeling the external ear acoustics for insert headphone usage," *Journal of the Audio Engineering Society*, vol. 58, no. 4, pp. 269–281, 2010.

[10] H. Deng and J. Yang, "Modeling and estimating acoustic transfer functions of external ears with or without headphones," *The Journal of the Acoustical Society of America*, vol. 138, no. 2, pp. 694–707, 2015.

[11] A. H. Akkermans, T. A. Kevenaar, and D. W. Schobben, "Acoustic ear recognition for person identification," in *Automatic Identification Advanced Technologies, 2005. Fourth IEEE Workshop on*. IEEE, 2005, pp. 219–223.

[12] T. H. Akkermans, T. A. Kevenaar, and D. W. Schobben, "Acoustic ear recognition," in *International Conference on Biometrics*. Springer, 2006, pp. 697–705.

[13] M. Derawi, P. Bours, and R. Chen, "Biometric acoustic ear recognition," in *Biometric Security and Privacy*. Springer, 2017, pp. 71–120.

[14] T. Arakawa, T. Koshinaka, S. Yano, H. Irisawa, R. Miyahara, and H. Imaoka, "Fast and accurate personal authentication using ear acoustics," in *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2016 Asia-Pacific*. IEEE, 2016, pp. 1–4.

[15] V. M. Patel, R. Chellappa, D. Chandra, and B. Barbello, "Continuous user authentication on mobile devices: Recent progress and remaining challenges," *IEEE Signal Processing Magazine*, vol. 33, no. 4, pp. 49–61, 2016.

[16] P. Mermelstein, "Distance measures for speech recognition, psychological and instrumental," *Pattern recognition and artificial intelligence*, vol. 116, pp. 374–388, 1976.

[17] P. Rajan, A. Afanasyev, V. Hautamäki, and T. Kinnunen, "From single to multiple enrollment i-vectors: Practical plda scoring variants for speaker verification," *Digital Signal Processing*, vol. 31, pp. 93–101, 2014.

ISBN 978-90-827970-1-5 © EURASIP 2018