HUMAN ACOUSTIC FINGERPRINTS: A NOVEL BIOMETRIC MODALITY FOR MOBILE SECURITY

Yuxi Liu, Dimitrios Hatzinakos

The Edward S. Roger Sr. Department of Electrical and Computer Engineering,
University of Toronto,
10 King's College Road, Toronto, ON, Canada, M5S 3G4

ABSTRACT

Recently, the demand for more robust protection against unauthorized use of mobile devices has been rapidly growing. This paper presents a novel biometric modality Transient Evoked Otoacoustic Emission (TEOAE) for mobile security. Prior works have investigated TEOAE for biometrics in a setting where an individual is to be identified among a pre-enrolled identity gallery. However, this limits the applicability to mobile environment, where attacks in most cases are from imposters unknown to the system before. Therefore, we employ an unsupervised learning approach based on Autoencoder Neural Network to tackle such blind recognition problem. The learning model is trained upon a generic dataset and used to verify an individual in a random population. We also introduce the framework of mobile biometric system considering practical application. Experiments show the merits of the proposed method and system performance is further evaluated by cross-validation with an average EER 2.41% achieved.

Index Terms— Mobile Security, Biometric Verification, Otoacoustic Emission, Time-frequency Analysis, Autoencoder Neural Network

1. INTRODUCTION

Nowadays, mobile devices have outgrown their initial use for communication and have added many powerful functionalities, such as data storage, web access and remote commerce, etc. The increasing need for greater mobile security has been opening up new areas in biometrics.

However, the reliability of conventional biometric modalities for mobile security has been questioned due to the threat of advanced spoofing techniques. For example, fingerprints can be easily taken from the surface of phone screen or anywhere touched, and an artificial clone can be created by plastic mold and gelatin [1]. Transient Evoked Otoacoustic Emission (TEOAE) for biometrics that we investigate in this work, however, is naturally immune to falsification or replay attacks as a physiological signal.

TEOAE is an approximate 20 ms acoustic response generated by an active process occurring inside the cochlea after a low level short click stimulus [2]. TEOAE echoes back to the middle ear and ear canal, and thus can be easily collected by an earphone with built-in microphones. Importantly, its uniqueness among individuals which was primarily discovered in biological and clinical studies [3, 4] has been utilized for identity recognition [5, 6, 7]. Moreover, it is almost impossible to fabricate someone's auditory system and extremely difficult to steal his/her TEOAE, as an outcome of physiological activity of the auditory perception in response to a specific acoustic stimulation. Overall, its compatibility with mobile devices and robustness make it an advantageous modality for mobile security.

Rest of the paper is organized as follows. Section 2 is a review of previous work and elaboration on the relation to this work here. Section 3 provides detailed methodology followed by the mobile biometric system proposed in Section 4. After that, Section 5 presents experimental results and analysis. Lastly Section 6 concludes our work.

2. RELATED WORKS

In [6], Grabham et al. conducted a quantitative study of TEOAE for biometrics by approximating the distribution of the intra-subject and inter-subject distance through maximum likelihood estimation to the time series data (raw signal). Later, our previous work [7] argued that analyzing TEOAE in only time domain is suboptimal due to the fact that TEOAE is a cumulative response consisting of different frequency components evoked by the broadband stimulus. Wavelet analysis was therefore employed to derive the time-frequency representation of the signal prior to recognition procedure.

Previous work [7] presented high accuracy in identification by applying linear discriminant analysis (LDA) to maximize between-class scatter. As it assumed a closed-set gallery among which an individual is to be identified as a pre-enrolled identity, it is however not applicable to mobile setting. In

This work has been supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

such an open-ended scenario, attacks in most cases are from imposters unregistered or unknown to the system in advance (for instance a thief attempts to be authenticated on a smart phone).

The method proposed here translates previous work to mobile security setting by adopting an unsupervised learning approach based on Autoencoder Neural Network [8, 9]. The learning model is trained upon a generic dataset and used to verify an individual in a random population, which tackles the blind recognition problem.

3. VERIFICATION ALGORITHM

3.1. Time-frequency Representation

First, the time-frequency representation algorithm in [7] as a pre-processing procedure is herein revisited.

According to [10], TEOAE is generated from the vibration pattern of basilar membrane in cochlea depending on the spectral energy of the stimulus. When a broadband stimulus is used, TEOAE becomes a cumulative response consisting of several dominant frequency components. As a time-frequency analysis tool, continuous wavelet transform (CWT) is therefore used to separate the mixed frequency components. Let $\psi(t)$ be the mother wavelet, a narrow band-pass function whose Fourier transform is centered around frequency f_0. The wavelet transform of a signal $x(t)$ regarding $\psi(t)$ is defined as:

$$WT_x(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t)\psi^*(\frac{t-b}{a})dt \qquad (1)$$

where $\psi(\frac{t-b}{a})$ is the adjustable window function with scale factor a and time translation factor b. The Fourier version of $\psi(\frac{t-b}{a})$ is the same band-pass function centered around f_0/a. Thus, with different scale factor, corresponding frequency component can be extracted from the original signal.

Previous work [7] showed that with Daubechies 5 as mother wavelet, frequency component extracted at scale = 8 serves as good representation of the original signal, i.e., it can well differentiate ("identify") itself from those of other sources. Take two subjects from a TEOAE dataset [11] for example (Figure 1), representations reveal high similarity within the same subject (A) and low similarity between different subjects (A and B), while such similarity is difficult to detect in the original signals.

3.2. Feature Extraction Based on Autoencoder

In pattern verification, for high-dimensional features, dimension reduction is usually implemented prior to similarity matching to avoid possible redundance and irrelevance of the features. An unsupervised learning model Autoencoder Neural Network [8, 9] is employed in this work to further extract effective features.

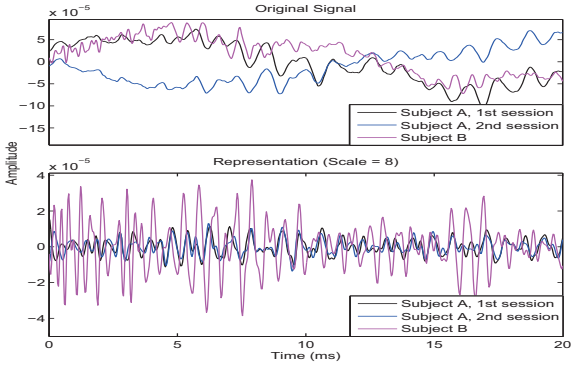

Fig. 1. Original signal from subject A of two different sessions, and from subject B respectively, along with each corresponding representation.

Autoencoder is an artificial neural network with an input layer, an output layer which has the same meaning as the input one, and one or more hidden layer(s). In this work, the autoencoder (Figure 2) trains a signal hidden layer $\boldsymbol{h}$ such that the output layer $\widehat{\boldsymbol{z}}$ is forced to reproduce the input layer $\boldsymbol{z}$. In particular, provided that an input sample is of d dimension, the hidden layer consisting of a smaller subset n ($n < d$) nodes is parameterized by a weight $\boldsymbol{W}_1 \in \Re^{d\times n}$ and bias $\boldsymbol{b}_1 \in \Re^n$, and is computed by:

$$\boldsymbol{h} = sig(\boldsymbol{W}_1^T \boldsymbol{z} + \boldsymbol{b}_1) \qquad (2)$$

where $sig\{*\}$ is the activation function with sigmoid kernel. Similarly, the approximation output is computed by:

$$\widehat{\boldsymbol{z}} = sig(\boldsymbol{W}_2^T \boldsymbol{h} + \boldsymbol{b}_2) \qquad (3)$$

where weight $\boldsymbol{W}_2 \in \Re^{n\times d}$ and bias $\boldsymbol{b}_2 \in \Re^d$.

Given a training set of M CWT resulting vectors $\{\boldsymbol{z}^{(1)}, \boldsymbol{z}^{(2)}, ..., \boldsymbol{z}^{(M)}\}$, back-propagation algorithm is applied to learn the autoencoder model $\boldsymbol{\Theta} = \{\boldsymbol{W}_1, \boldsymbol{b}_1, \boldsymbol{W}_2, \boldsymbol{b}_2\}$ by minimizing the following cost function:

$$\begin{aligned} J(\boldsymbol{\Theta}) &= J_{MSE} + J_{weight} \\ &= \frac{1}{M}\sum_{i=1}^{M}(\frac{1}{2}\|\widehat{\boldsymbol{z}}^{(i)} - \boldsymbol{z}^{(i)}\|^2) + \frac{\lambda}{2}(\|\boldsymbol{W}_1\|^2 + \|\boldsymbol{W}_2\|^2) \end{aligned} \qquad (4)$$

where the first term J_{MSE} is the mean squared error of reconstruction, qualifying the difference between $\boldsymbol{z}^{(i)}$ and the corresponding reconstructed version $\widehat{\boldsymbol{z}}^{(i)}$; J_{weight} is the weight decay term, with a regularization factor λ, tending to prevent overfitting. The minimization of (4) is typically accomplished by an iterative optimization method such as gradient descent or Newton's method (e.g., L-BFGS as the one we employ).

The well-trained model $\boldsymbol{\Theta}^*$ is then used to extract the features $\boldsymbol{h}^{(i)}$ of the training samples $\boldsymbol{z}^{(i)}$ via (2) with optimal $\boldsymbol{W}_1^*$ and $\boldsymbol{b}_1^*$. Moreover, given a new input $\boldsymbol{z}'$, its feature $\boldsymbol{h}'$ can be extracted in the same way, i.e., $\boldsymbol{h}' = sig(\boldsymbol{W}_1^{*T}\boldsymbol{z}' + \boldsymbol{b}_1^*)$.

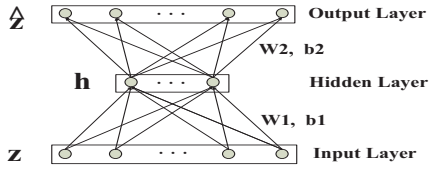

Fig. 2. Illustration of the applied autoencoder model.

3.3. Matching

Given two vectors $\boldsymbol{z}^{(a)}$ and $\boldsymbol{z}^{(b)}$ representing TEOAE signals $\boldsymbol{x}^{(a)}$ and $\boldsymbol{x}^{(b)}$ after CWT respectively, verification is conducted on the corresponding features $\boldsymbol{h}^{(a)}$ and $\boldsymbol{h}^{(b)}$. In this work, Euclidean distance is chosen as a metric, where similarity score S between $\boldsymbol{x}^{(a)}$ and $\boldsymbol{x}^{(b)}$ is calculated by:

$$S(\boldsymbol{x}^{(a)}, \boldsymbol{x}^{(b)}) = \sqrt{(\boldsymbol{h}^{(a)} - \boldsymbol{h}^{(b)})^T (\boldsymbol{h}^{(a)} - \boldsymbol{h}^{(b)})} \quad (5)$$

A smaller value means a higher similarity between two signals, i.e., they are more likely to represent the same person. An ACCEPT decision is made only if $S(\boldsymbol{x}^{(a)}, \boldsymbol{x}^{(b)}) \leq t$, where t is a distance threshold.

4. MOBILE BIOMETRIC SYSTEM

4.1. System Framework

In this section, a complete framework of the mobile biometric system is proposed, which involves a generic dataset setup stage, an enrollment stage and a verification stage. A block diagram of the proposed system is depicted in Figure 3.

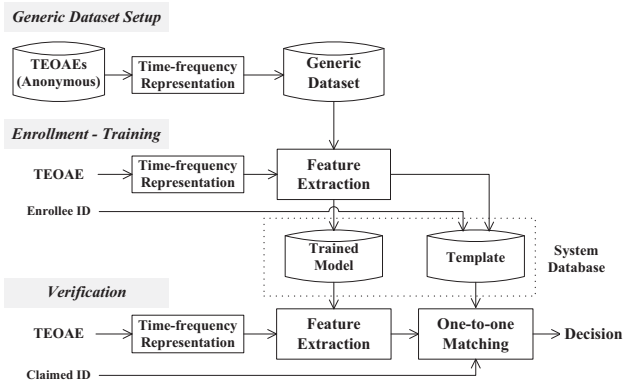

Fig. 3. Framework of the proposed mobile biometric system.

Since in mobile biometric environment, imposters may be unknown to the system, the generic pool is intended for future training in order to create a paradigm of TEOAE morphologies. It is an anonymous collection of representation vectors of TEOAE samples from a large population after CWT.

During the enrollment session, TEOAE is acquired from an enrollee and subjected to the same time-frequency analysis. Resulting representation vector $\boldsymbol{z}_{en}^{(k)}$ is then combined with the generic pool and used to drive the autoencoder-based learning algorithm. Such an enrollee-oriented (personalized) training approach that takes the enrollee's sample against the generic pool helps handle false accept. The well-trained model with parameters $\boldsymbol{W}_1^*$ and $\boldsymbol{b}_1^*$, and the learned template consisting of the feature set $\boldsymbol{h}_{en}^{(k)}$ extracted and a user specific decision threshold $t^{(k)}$, establish the system database (see the dash line in Figure 3).

As for verification operation, an individual gets recorded and claims an identity k. Following CWT, feature $\boldsymbol{h}'$ is extracted according to the trained model. The biometric system will retrieve the corresponding template $\boldsymbol{h}_{en}^{(k)}$ from the gallery and conduct a one-to-one matching between $\boldsymbol{h}'$ and $\boldsymbol{h}_{en}^{(k)}$. Decision of either accept or reject the claim is finally made.

4.2. Application Scenario

As an example, Figure 4 exhibits how the proposed system is applied to mobile commerce (e.g., remote banking, phone purchase). To fulfill registration for the secure system, a client's phone is responsible for TEOAE data acquisition and transmission to the remote server in the administrator's side, where data processing algorithms are afterwards implemented. When a user (who is either the authorized client or an illegitimate one) is signing in, the server receives data sent from user's phone and performs matching after computation. It finally responds to both the administrator and the user whether follow-up transaction is allowed or refused. It is expected that security questions whose answers can be easily discovered (e.g., "What is your mother's maiden name?", "What is your date of birth?") will no longer be asked.

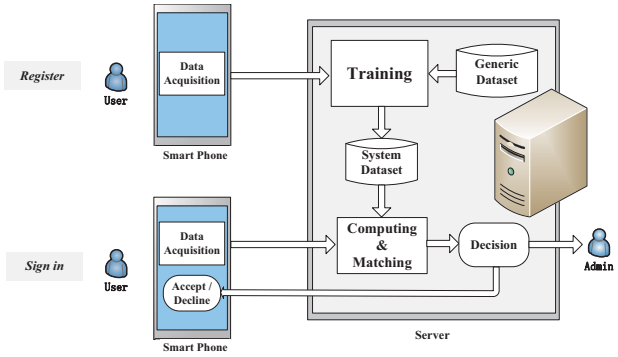

Fig. 4. Architecture of the secure biometric system in mobile commerce.

Further, local application (e.g., user verification, private data access) of the proposed biometric system share the same architecture, except that datasets are stored and computation is conducted locally or on a client-server with which the device needs to communicate.

5. EXPERIMENTS AND RESULTS

Experiments in this work are conducted on a biometric setting dataset [11] with 54 subjects. For each subject, more than 20 TEOAE recordings were collected from both ears in each of two separate sessions (with time interval of at least one week for biometric evaluation purpose). For simplicity, we initially investigate data collected from the left ear. Last 10 recordings from 30 subjects in the first session are used to simulate the generic dataset; last 10 recordings from the remaining 24 subjects in the first and second session are adopted to constitute the enrollment set and the testing (verification) set respectively. Verification performance is measured in false accept rate (FAR), false reject rate (FRR) and equal error rate (EER), i.e., the rate at which FAR and FRR are equal.

Firstly, to justify the merit of the enrollee-oriented training approach, comparison is conducted between training with and without subject's enrollment sample. Figure 5 shows the detection error tradeoff (DET) curve under each approach, where an EER 19.58% is obtained by appending enrollee's sample to the training set, while 24.58% without such involvement. It proves that personalized training can effectively handle false accept and hence reduce EER. Moreover, EER rises to 30.34% by using the LDA approach [7], which indicates the closed setting assumption is inapplicable to the open-ended scenario.

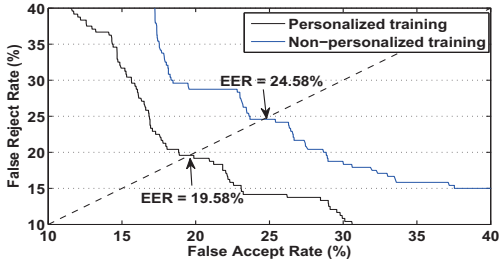

Fig. 5. DET curve under personalized training and non-personalized training.

On the other hand, it is observed that certain individuals have stronger TEOAE templates with smaller intra-subject variability than others, while some bear relatively larger fluctuation over time (e.g., subject 24 versus subject 12 as shown in Figure 6 individual ROC plots). This is generally known for all biometric modalities and it is therefore unadvisable to impose an universal decision threshold to all users. Instead, based on personal ROC plots, system can choose the desired user-specific threshold $t^{(k)}$ for an individual k depending on requirement of a particular application scenario, and store it for future use. If for example a number of intruders are expected, low FAR is preferred over FRR and a smaller threshold is therefore suggested. Or as shown in Table 1, individual threshold is selected at which FAR equates to FRR, guaranteeing a good trade-off between both risks. In this way, the particular intra-subject variability of different user is fully considered and system performance is greatly improved with an average EER **2.29%** achieved.

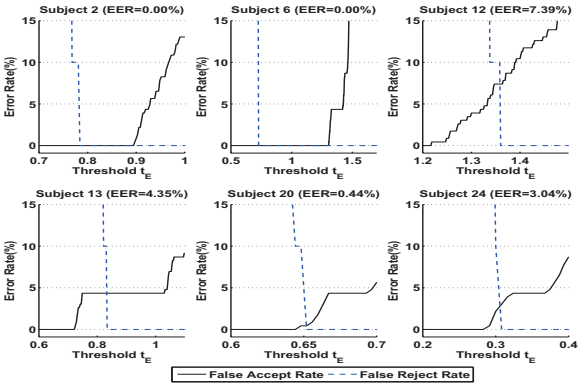

Fig. 6. Individual ROC plots for selected subjects.

Table 1. EER for individual subjects with user-specific threshold.

Subject ID	$t^{(k)}$	EER	Improved by
1	1.180	0.00%	19.58%
2	0.784	0.00%	19.58%
3	1.404	0.00%	19.58%
4	0.837	4.35%	15.23%
5	1.626	0.44%	19.14%
6	0.728	0.00%	19.58%
7	1.371	9.13%	10.45%
...	...	...	...
...	...	...	...
...	...	...	...
24	0.332	3.04%	16.54%
Average	–	**2.29%**	17.29%

Cross-validation is finally performed to estimate system performance. For each round, 30 subjects whose recordings constitute the generic dataset are randomly chosen; recordings from the remaining 24 subjects in the first and second session establish the enrollment and testing set respectively. Results (EERs) are then averaged over 200 iterations to produce a performance evaluation, **2.41%**$\pm$1.38% (mean$\pm$std) for the left ear and **2.81%**$\pm$1.35% for the right.

6. CONCLUSION

In this paper, we have presented securing model devices through personal acoustic "fingerprints" TEOAE, which is immune to fabrication or falsification. To tackle the blind recognition problem in mobile environment, i.e., authentication among an unknown population, we resort to Autoencoder Neural Network to extract effective features after CWT-based time-frequency analysis. In addition, we tailor verification of each enrollee via personalized training approach and user-specific decision thresholds. Experiment results have demonstrated the efficiency of the proposed methodology.

7. REFERENCES

[1] T. Matsumoto, H. Matsumoto, K. Yamada, and S. Hoshino, "Impact of artificial 'gummy' fingers on fingerprint systems.," *SPIE Proceedings*, vol. 4677, Optical Security and Counterfeit Deterrence Techniques IV, pp. 275–289, 2002.

[2] R. Probst, B. L. Lonsbury-Martin, and G. K. Martin, "A review of otoacoustic emissions," *The Journal of the Acoustical Society of America*, vol. 89, no. 5, pp. 2027–2067, 1991.

[3] D. McFadden and R. Mishra, "On the relation between hearing sensitivity and otoacoustic emissions," *Hearing Research*, vol. 71, pp. 208–213, 1993.

[4] D. McFadden and J. C. Loehlin, "On the herdibility of spontaneous otoacoustic emissions: A twins study," *Hearing Research*, vol. 85, pp. 181–198, 1995.

[5] J. Gao, F. Agrafioti, S. Wang, and D. Hatzinakos, "Transient otoacoustic emissions for biometric recognition," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2012, pp. 2249–2252.

[6] N. J. Grabham, M. A. Swabey, P. Chambers, M. E. Lutman, N. M. White, J. E. Chad, and S. P. Beeby, "An evaluation of otoacoustic emissions as a biometric.," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 1, pp. 174–183, 2013.

[7] Y. Liu and D. Hatzinakos, "Biometric identification based on transient evoked otoacoustic emission," in *Proceedings of the IEEE International Symposium on Signal Processing and Information Technology*, 2013, to appear.

[8] G E Hinton and R R Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, July 2006.

[9] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 2, pp. 210–227, 2009.

[10] H. P. Wit, J. C. Langevoort, and R. J. Ritsma, "Frequency spectra of cochlear acoustic emissions (kempechoes)," *The Journal of the Acoustical Society of America*, vol. 70, no. 2, 1981.

[11] "Biometrics Security Laboratory, University of Toronto," http://www.comm.utoronto.ca/ biometrics/databases.