Acoustic Ear Recognition

Ton H.M. Akkermans, Tom A.M. Kevenaar, and Daniel W.E. Schobben

Philips Research, Prof. Holstlaan 4, 5656 AA Eindhoven, The Netherlands
{ton.h.akkermans, tom.kevenaar, daniel.schobben}@philips.com

Abstract. We investigate how the acoustic properties of the pinna – i.e. the outer flap of the ear- and the ear canal can be used as a biometric. The acoustic properties can be measured relatively easy with an inexpensive sensor and feature vectors can be derived with little effort. Classification results for three platforms are given (headphone, earphone, mobile phone) using noise as an input signal. Furthermore, preliminary results are given for the mobile phone platform where we use music as an input signal. We achieve equal error rates in the order of 1%-5%, depending on the platform that is used to do the measurement.

1 Introduction

Well-known biometric methods for identity verification are based on modalities such as fingerprints, irises, faces, or speech to distinguish individuals. In some situations, however, these well-known modalities cannot be used due to the price and/or form factor of the required sensor or the required effort to derive feature vectors from measurements. Therefore we investigated if the acoustic properties of the pinna - i.e., the outer flap of the ear - and the ear canal can be used as a biometric. The acoustic properties can be measured relatively simple and economically and we found that the acoustic properties differ substantially between individuals. Therefore ear recognition is a possible candidate to replace pin codes in devices such as mobile phones or to automatically personalize headphones or other audio equipment. An additional advantage of ear recognition is that, unlike real fingerprints that are left behind on glasses or desks, "ear-fingerprints" are not left behind and can also not be captured as easily as an image of a face. In this respect acoustic ear recognition may lead to a more secure biometric.

The shape of the outer ear, such as the folds of the pinna, the length and shape of the ear canal are very different between humans as can be observed when visually comparing the ears of two individuals. These differences are even more pronounced for acoustic measurements of the transfer function of the pinna and ear canal using a loudspeaker close to the ear and a microphone close to, or in, the ear canal as shown in Figure 1. Such transfer functions can be seen as a kind of "fingerprint" of the ear canal and/or pinna. The spectrum of an acoustic transfer function can be used almost directly as the feature vector for a given individual. Using the acoustic properties of the ear as a biometric has first been published in [1] but there has been no public data on performance and application of the technology.

2 Acoustic Properties of the Ear Canal

It is well known that the optical properties of human ears can be used in biometric identification [2,3,4]. In [5], the authors investigate the relationship between optical and acoustic properties of the ear. In [6], acoustic ear biometrics have been used to develop and evaluate a recently developed biometric template protection system [7,8,9]. In [1], the Sandia Corporation US claimed the first US patent on acoustic ear recognition. In the current paper we focus mainly on the acoustic properties of the ear and its potential to be used as biometric modality.

The ear canal is a resonant system, which together with the pinna provides rich features. In a coarse approximation it is a one-dimensional system that resonates at one quarter of the acoustic wavelength. The resonance will typically be around 2500Hz but it will vary from person to person. Typical resonance frequencies are corresponding to typical lengths and shapes of pinna and ear canal.

The length of the ear canal and the curvatures of the pinna have dimensions that vary from millimeters to a few centimeters. To be able to detect these shapes and curvatures the acoustic probing waves should have proper wavelengths. Restricting ourselves to low cost loudspeakers and microphones we can easily generate and measure sound waves from 100Hz up to 15 kHz. Assuming that we can resolve structures in the order of 1/10 of the wavelength, the minimum resolving power becomes about 2mm, which seems appropriate to capture most distinguishing features.

3 Set-Up

The principle of the measurement set-up is shown in Figure 1. A loudspeaker close to the ear canal generates an excitation signal while a microphone measures the reflected echo responses. In general the excitation can be any acoustic signal like noise or music that has a fairly flat frequency spectrum. Alternatively the excitation signal may be preprocessed in such a way that those frequencies are emphasized that allow for a good discrimination between individuals.

In our current set-up we measure the transfer function of the ear by sending a noise signal into the pinna and outer ear. Figure 2 shows a possible method for determining
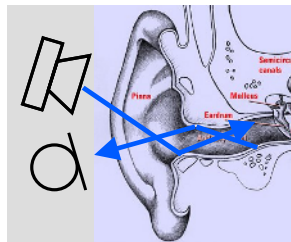

Fig. 1. An acoustic probe wave is send into the ear canal while a microphone receives the response
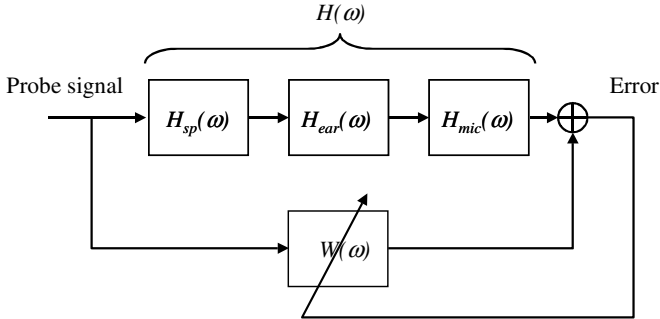
Fig. 2. Measuring the transfer function

this transfer function. The excitation signal is fed into the transfer function $H(\omega)$ that should be identified. The finite impulse response filter $W(\omega)$ is adaptively optimized using a steepest descent adaptive filter that minimizes the error signal which is the difference between the microphone signal and the output of the adaptive filter $W(\omega)$. Both the system $H(\omega)$ to be identified and its estimate $W(\omega)$ consist of the cascade of the transfer functions of the loudspeaker, pinna and ear-canal, and microphone. An alternative approach for determining the transfer function is to directly divide, in the frequency domain, the response signal coming from the microphone by the input signal [10]. Although both approaches gives similar results for noisy signals, the approach depicted in Figure 2 is more flexible when non-stationary input signals such as music are used as a probe signal (see also section 6.4).

4 Acquiring a Feature Vector

The estimate of the transfer function is a complex entity. Although it is expected that delays and phase shifts still contain significant discriminating information about an individual, they may also lead to larger intra-class variations, i.e. variations amongst various measured transfer functions for the same subject due to unwanted phase shift introduced by the measurement system. In order to eliminate these phase shifts we extracted the amplitude of the ear transfer function frequency spectrum as the biometric feature vector. As an example, Figure 3 shows transfer functions for three individuals.
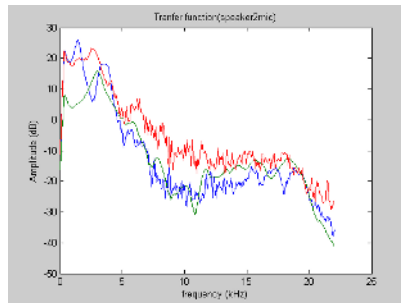

Fig. 3. Amplitude of the frequency response of the ear transfer function

Obviously, information of the biometric modality is lost by choosing the feature extraction method mentioned above. Therefore in Section 6.4 we give some results for the mobile phone platform where the response signal in the time domain is used.

5   Test Platforms

Often the performance of recognition systems relies strongly on the way the feature extraction method is implemented in the specific application. Therefore we investigated the robustness of the acoustic ear recognition system based on different platforms. The pictures show these platforms and the position of their microphones marked by the arrow.

Fig. 4. The three platforms: headphones, earphones and mobile phone all with extra microphones indicated by arrows

The headphones in Figure 4 (Philips SBC HP 890) have 1 microphone per side that is mounted underneath the cloth that covers the loudspeaker. A tube is mounted onto each microphone that allows for measuring the sound pressure at the entrance of the ear canal. The earphones in Figure 4 have 1 microphone per ear-piece which is mounted underneath the original factory fit rubber cover.

The mobile phone of Figure 4 has 1 microphone next to the speaker whereas the other two platforms of Figure 4 each have 2 sensing microphones (1 microphone per ear) resulting in feature vector lengths of 256 and 512 components, respectively.

6   Results

In order to derive results, we collected the following measurements. For both the headphone and earphones based platform, 8 ear transfer functions were measured for each of 31 subjects and collected in two separate databases. For the mobile phone platform we enrolled 17 persons with 8 measurements per person that were stored in a third database. In the remainder of this section we will show some results obtained using these databases.

6.1   Correlation Between Ears

In order to determine the similarity between the two ears of an individual we determined the average correlation between the measurements of the two ears. We define the correlation as

$$C = \frac{x^T y}{\|x\| \|y\|} \qquad (1)$$

where x and y are two feature vectors taken relative to the mean of the whole population. The average correlation C_j between the left and right ear of an individual j is taken as the average over the correlations between every possible combination of a measurement in the headphone database of the left and the right ear of this individual. The overall correlation between the left and right ear of the whole population is then defined as the average over the C_j's of all individuals. The reason for using the headphone database was that it shows the lowest intra-class variability per ear and is therefore most suitable to determine the biometric difference between left and right ear. In order to minimize the loss of information, the time responses rather than frequency responses were used and they were manually compensated for undesirable time delays. It turns out that the correlation between measurements of one ear of one individual is on average 90%. Comparing left and right ear gives an average correlation of roughly 80%. In conclusion we can say that using both ears only gives marginally better discrimination capabilities since the acoustic left and right ear responses are quite similar and differs 10% in terms of correlation.

6.2   Recognition Performance

To test the performance of the acoustic ear recognition system the FAR (False Acceptance Rate) and FRR (False Rejection Rate) have been investigated using the impostor and genuine distributions using the correlation measure (1). The probing noise signal contained frequencies in the range 1.5kHz-22kHz. Figure 5 shows the Receiver Operating Characteristics (ROC) of the unprocessed frequency response data. We observe that the headphones and earphones give roughly the same performance resulting in an equal error rate of respectively 7% and 6 %.

As a second experiment, Fischer Linear Discriminant Analysis (LDA) was applied to the three ear databases to selects the most discriminating components among the subjects. In order to determine the eigenvalues and eigenvectors, the generalized eigenvalue problem

$$S_b q = \lambda S_w q \qquad (2)$$

was solved for q and λ where S_b and S_w are the estimated between-class and within-class covariance matrices, respectively. We used a regulation parameter to avoid singularity problems in S_w.

Figure 6 again shows the ROC performance but now a Fisher LDA transformation is applied to the frequency impulse responses. It can be seen from Figure 6 that the performance improves significantly, especially for the headphones and earphones platform. Furthermore a slight increase in FRR will significantly reduce the FAR leading to a high security level. The mobile phone performance is worse due to two reasons. Firstly, the between-class variation of mobile phones is much larger due to uncontrolled position and pressing of the mobile phone against the pinna. This is also observed when we consider the 'signal-to-noise ratio' of the feature vector
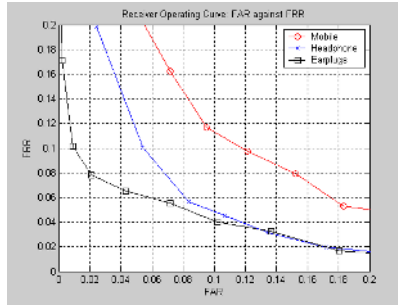
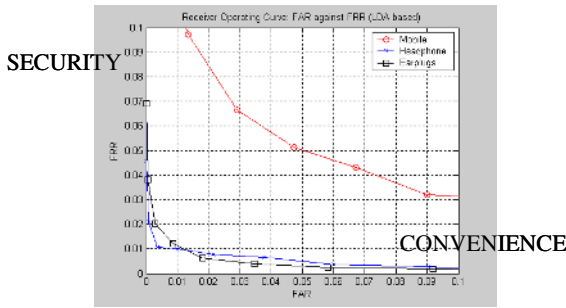**Fig. 5.** Receiver operating curves without Fisher LDA transformation

Fig. 6. Receiver Operating Curves using Fisher LDA a transformation

components after LDA. The average over all users and all components for the headphone and earphone database is in the order of 40 while for the mobile phone is it in the order of 16. A second reason is that, although the correlation between the two ears of one individual is very high, measuring two ears rather than one still gives slightly better discrimination between individuals.

6.3 Relevant Frequency Ranges

We also investigated how the applied frequency range used in the excitation signal influences the classification performance. Table 1 gives an overview of the Equal Error probability as function of the applied frequency range of the acoustic probe signal using the Fisher LDA transformation.

Table 1. Ear recognition performance (EER) as a function of the frequency range of the excitation signal

Freq.Range (Hz)	Headphones	Earphones	Mobile Phone
1.5k-22k	0.8	1	5.5
1.5k-10k	0.8	1.4	6.5
10k-22k	2.5	2.5	10
16k-22k	8	6	18

Although these figures depend quite heavily on the individual loudspeaker and microphone performances (especially in mobile phones the loudspeaker transfer at frequencies above 10 kHz deteriorate significantly), it can be seen that a wider frequency range gives better classification results. It is further interesting to notice that the frequency range 16kHz-22kHz still leads to reasonable classification results indicating that ultrasonic characterisation might be an option.

6.4   Experiments with Music and Time Domain Signals

In order to enhance user convenience we performed experiments where the excitation signal is a music signal rather than a noise signal. In our case we used a music signal in MP3 format which has the advantage that it has inaudible noise components in its spectrum due to the underlying Human Auditory System model used to compress music signals. These noise components improve the estimate of the transfer function. The initial experiments used a database of 12 persons with 10 measurements per person. The output signal from the microphone in the frequency domain rather than the transfer function $H(\omega)$ was used directly as a feature vector. Consequently, a user should always be probed with the same piece of music.

In Figure 7 two ROCs are given, one for a noise input and one for a music input where the curve referring to a noise input signal is copied from Figure 6. It can be seen that both systems give similar classification results.
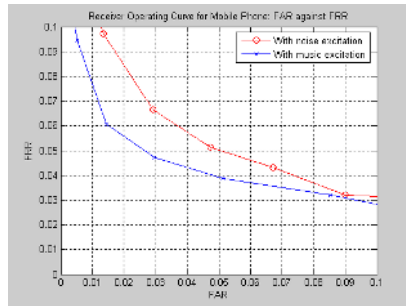

Fig. 7. The Receiver Operating Curves for a noise and music input signal for a mobile phone
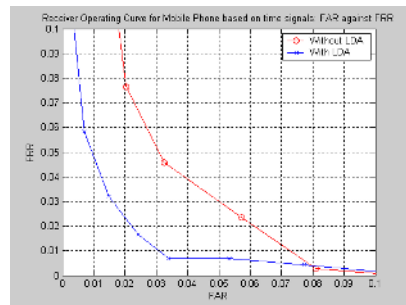

Fig. 8. The Receiver Operating Curves for a mobile phone based on time signals

As mentioned above, discarding the phase information in the feature vectors might deteriorate classification results but is practically necessary to handle random phase shifts in the measurement system. In order to estimate the influence of discarding the phase, we used the time-domain signal coming from the microphone as a feature vector where we manually compensated for the system delay. The results are given in Figure 8 where, compared to Figure 6, we see an improvement in classification results. In practical systems a pilot tone can be inserted to handle random system delays.

7   Conclusions

This paper describes a novel biometric system based on the acoustic properties of the human ear. Three practical platforms were developed including a mobile phone, headphones and earphones where using noise as a probing signal. The amplitude of the frequency spectrum of the ear transfer function has been found to provide stable and rich features. False acceptance and rejection rates have been derived from measurements taken from various subjects. Applying a Fisher LDA transform greatly improve the performance. In order to enhance user convenience we also used music as a probing signal which resulted in comparable ROCs. Finally we used a time signal rather than the amplitude of the transfer function as a feature vector resulting in improved classification results. Further research consists of deriving the transfer function for an arbitrary piece of music and retaining the phase information in the measurement signal.

References

1. Sandia Corporation, patent US 5,787,187, "Systems and methods for biometric identification using the acoustic properties of the ear canal."
2. B. Moreno, A. Sanchez, J.F. Velez,On the use of outer ear images for personal identification in security applications, Proceedings. IEEE 33rd Annual 1999 International Carnahan Conference on Security Technology,5-7, pp469 – 476, Oct. 1999.
3. M. Burge, W. Burger, Ear biometrics in computer vision, Proc. 15th International Conference on Pattern Recognition, Vol. 2, 3-7, pp822 – 826, Sept 2000.
4. K.H. Pun and Y.S. Moon, Recent advances in ear biometrics, Proc. Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 17-19, 164 – 169, May 2004.
5. Y. Tao, A.I. Tew, and S.J. Porter, The Differential Pressure Synthesis Method for Estimating Acoustic Pressures on Human Heads, 112th Audio Engineering Society Convention, 10–13 Munich, Germany, May 2002.
6. P. Tuyls, E. Verbitskiy, T. Ignatenko, D. Schobben and T.Akkermans. Privacy Protected Biometric Templates:Acoustic Ear Identification. Proc.SPIE,Vol.5404, pp176-182, April, 2004.
7. J-P. Linnartz and P. Tuyls, New shielding functions to enhance privacy and prevent misuse of biometric templates, Proc.4th Int. Conf. on Audio- and Video-Based Biometric Person Authentication (AVBPA 2003), Springer LNCS 2688, pp 393-402, 2003.

8. P. Tuyls and J. Goseling, Capacity and Examples of Template Protection in Biometric Authentication systems, Biometric Authentication Workshop (BioAW 2004), LNCS 30087, pp158-170, Prague, 2004.
9. P. Tuyls, A. Akkermans, T. Kevenaar, G-J Schrijen, A. Bazen and R. Veldhuis. Practical Biometric Authentication with Template Protection. Proc. 5th Int. Conf. on Audio- and Video-Based Biometric Person Authentication (AVBPA2005), Springer LNCS 3546, pp436-446, 2005.
10. A.H.M. Akkermans, T.A.M. Kevenaar, D.W.E. Schobben. Acoustic Ear Recognition for Person Identification. Accepted for the IEEE AutoID Workshop, October 17-18, 2005 Buffalo, New York, USA.