Developing a Context-Dependent Tuning Framework of Multi-channel Biometrics that Combine Audio-Visual Characteristics for Secure Access of an eHealth Platform

Marios Spanakis[1], Georgios C. Manikis[1], Sakshi Porwal[2], and Emmanouil G. Spanakis[1](✉)

[1] Computational BioMedicine Laboratory, Institute of Computer Science, Foundation for Research and Technology - Hellas, 71110 Heraklion, Crete, Greece
{marspan,gmanikis,spanakis}@ics.forth.gr
[2] Tech Inspire Ltd., Salford, UK
porwalsakshi@gmail.com

Abstract. The efficiency of a biometric system is identified by the detection error tradeoff (DET) curve, which is a visual characterization of the trade-off between the False Acceptance Rate (FAR) and the False Rejection Rate (FRR). A DET curve is a plot of FAR against FRR for various threshold values, ***t***. FRR refers to the expected probability that two mate samples (samples of the same biometric trait obtained from the same user) will be falsely declared as a non-match whereas FAR is the expected probability that two non-mate samples will be incorrectly recognized as a match. The threshold ***t*** defines how much the biometric characteristics must be similar, in order to make a positive comparison, so it measures the correspondence between characteristic to check and template stored in the database. By elevating the threshold, the risk that not authorized users can fool the system diminishes, but, on the other hand, it is more probable that some authorized users can sometimes be refused. In this work, we present the results for SpeechXRays multi-modal biometric system that uses audio-visual characteristics for user authentication in an eHealth platform for osteoarthritis management. Using the privacy and security mechanism provided by SpeechXrays based on audio and video biometrics medical personnel is able to be verified and subsequently identified to the eHealth application for osteoarthritis.

Keywords: Biometrics · Decision threshold · Equal Error Rate (EER) Detection error tradeoff (DET) · eHealth · Osteoarthritis

1 Introduction

Personal health systems and E-Health platforms aim on improving the interaction among health care professionals with their patients [1–6] and provide the means for secure access to sensitive medical information to people regarding their health status and disease management [7, 8]. To this respect, there is a tremendous need for secure and stratified access to health information [9] by adopting the use of modern ICT technology.

P. Perego et al. (Eds.): MobiHealth 2017, LNICST 247, pp. 190–198, 2018.
https://doi.org/10.1007/978-3-319-98551-0_22

Today, biometric authentication, is gaining as the leading technological achievement for the verification of a person's identity using a physical trait or behavioral characteristic in order to accept the identity of the person and verify him/her as an authorized user [10, 11]. These systems mainly rely on models derived from pattern recognition, where several characteristics from a person (e.g. voice, facial expression, etc.) are first transformed into one (unimodal) or many features (multimodal) and then are processed to accept or reject the verification and identification of a user [12]. A major prerequisite in this process is the so called training phase of the model composed of a pipeline process in which (a) captured biometric characteristics from specific users are stored in a database, and (b) used for training the model on the basis of that known content. Once training has been performed accurately, the biometric system can be applied for verification and identification.

SpeechXRays[1] is aiming to develop and test in real-life environments (i.e. medical units) a user recognition platform based on voice acoustics analysis and audio-visual identity verification. SpeechXRays provides a state of the art, accurate and user-friendly solution allowing storage and analysis of biometric data for authentication. The e-Health pilot of SpeechXrays will involve more than 400 medical personnel who through SpeechXrays will gain access to hospital's medical image/radiology archiving system as well as a personal health application designed for the management of osteoarthritis (OA) [6]. Recently an evaluation survey and preliminary results regarding acceptability of the approach its functionality, efficiency and user friendly environment were presented along with the acceptance of using the biometric system proposed from SpeechXRays for user authentication [9]. The aim of this work is to present and evaluate the verification rates from various decision thresholds that must be defined and adapted based on the level of security required by the SpeechXRays platform and the confidentiality of the data that the user is attempting to access.

1.1 Biometric System Errors

A biometric verification system usually makes two types of errors: (i) mistaking biometric measurements from two different persons to be from the same person (false match), and (ii) mistaking two biometric measurements from the same person to be from two different persons (false non-match). These two types of errors are often termed as false accept and false reject, respectively [13] and most commonly are described by *FMR (false match rate)/FAR (false acceptance rate), FNMR (false non-match rate)/FRR (false rejection rate, Failure to capture (FTC) and Failure to enroll (FTE)*. It is possible to reduce the errors by trying to record more biometric characteristics for every user so that, in case of variations on a template, the other can be used. On the other hand, there are natural variations to biometric characteristics which may not be erased but that could be minimized through the appropriate equipment. Another possibility is to act on the threshold of the system. This threshold defines how much the biometric characteristics must be similar, in order to make a positive comparison, so it measures the correspondence between the characteristic to check and the template stored in database. By

[1] http://www.speechxrays.com/.

elevating the threshold, the risk that not authorized users can fool the system drops (FAR is reduced), but, on the other hand, it is more probable that some authorized users can sometimes be refused (FRR increases). Biometric system errors can occur due to various reasons such as: Sampling (imperfect imaging conditions); Changes in characteristics (i.e. bruises or voice changes due to illness); Ambient conditions (temperature humidity); User interaction with the sensor (i.e. distance) and Sensors (different smartphones).

1.2 Efficiency of Biometric System and Application in SpeechXRays Data

The efficiency of any biometric system can be described by a visual characterization of the trade-off between the FAR and the FRR. The most basic and robust method is the calculation of the receiver operating characteristic (ROC) curve. To assess a biometric system performance, detection error tradeoff (DET) curve which is similar to the ROC curve analysis is followed to discriminate between two states that usually overlap such as genuine and impostor users. A DET curve is a plotted as FMR against FNMR for various **thresholds**, ***t***. Similarly, EER can be also estimated by the Receiver Operator Characteristic (ROC) curve. A specific threshold ***t*** can be calculated automatically based on the Error Rate curve of the FAR and FRR. This threshold is the value where FAR and FRR are equal (i.e., where FAR = FRR), and is called Equal Error Rate (EER) (intersection point of FAR & FRR) [10, 13, 14]. Additionally, the more the EER is near to 0% better is the performance of the target system.. In this light, EER is a performance metric with FAR and FRR used as performance criteria simultaneously, since EER is defined by both metrics with the constraint that they are equal.

2 Methodology

The performance of the SpeechXRays biometric system in its implementation will be assessed using experimental protocols based on both unimodal and multimodal data. Data will be randomly separated into enrolment data and data used at the verification level in order to simulate a real case scenario. However, the estimation of the system performance can be influenced by the selection of data used for enrolment and verification, affecting a good generalization in performance. SpeechXRays biometric system will be applied to health domain applications [18, 19], thus such model needs to demonstrate at first adequate verification capability on the data used for designing the system.

In SpeechXRays audio and voice data will be first partitioned into k equally (or nearly equally) sized folds following a k-fold cross-validation method for assessing the generalization performance of the system. Subsequently k iterations of enrolment and verification will be performed such that, within iterations, a different fold of the data is held-out for verification purposes while the remaining k-1 folds will be used as data for the enrolment phase. Finally, k-fold cross-validation will run several times, increasing the number of estimates, where data from the experimental protocol will be reshuffled and re-stratified before each run.

Security requirements such as confidentiality, integrity, authenticity, non-repudiation and availability are essential for computer and network based systems. Following steps were performed by the SpeechXRays biometric system to enrol the person by acquiring and storing the appropriate data, verify by comparing the captured data against the database, and authenticating or revoking access based on the comparison/classification of the biometric trait.

In the context of SpeechXRays, the following phases are studied in a pipeline process for enrolment, verification, authentication, and revocation through the biometric context dependent detection system. **Enrolment Phase:** Medical personnel will provide speech and facial imaging data under different times, environment conditions (i.e. noisy background, low light, etc.), and facial expressions. This multimodal information will be extracted using feature extraction techniques for voice and face data, and stored individually as a biometric template to a database. This template will be linked to a specific token (i.e. ID, name, etc.) related to each medical personnel. Once the information gathering is finished, two separate unimodal biometric systems will be applied to the data in order to construct SpeechXRays biometric system. Through enrolment of the system, appropriate thresholds will be estimated and assigned specifically to the medical data with different levels of security/sensitivity. **Verification Phase:** At the verification level, a medical specialist requires access to medical data classified with a specific security level degree. The user presents a token, facial and speech data, and the biometric feature template associated with the user is retrieved from the database. The system processes the given data, extracts the facial and speech features and compares them with the features stored at the database at the enrolment phase. **Authentication Phase:** if matching, secure session is opened between two parties and if not, the access is denied and **Revocation Phases** – revoke access based on security risk such as template leakage, spoofing attempts, etc. The user can then proceed again to the verification process and depending on the security levels assigned to the medical data a number of attempts can be made.

2.1 Demonstration of the Context Dependent Tuning Framework of SpeechXRays

SpeechXRays unimodal system* - *Experimental Protocol

According to the SpeechXRays developing procedure, individual unimodal biometric systems will be first applied to both speech and face data. The derived matching scores will be afterwards fused to conclude to the final decision at the verification phase. For that reason, a unimodal biometric system was implemented first using publicly available data from the ORL database[2], composed of 400 facial images of size 112x92. Ten different images of each of 40 distinct subjects-persons were captured in different times, lightning, facial expressions (i.e. open/closed eyes, smiling/not smiling) and details (i.e. glasses/no glasses). All images were acquired against a dark homogeneous background with faces in an upright position in frontal view, with a tolerance for some side movement. Facial features were exported using the methodology described in [15]. Following

[2] www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html.

an iterative process to assess the verification accuracy of the SpeechXRays biometric system, the entire data of each of the 40 distinct subjects was randomly partitioned into subsets act as enrolment and verification data respectively. Particularly, 80% of the images of each subject contributed to the enrolment phase of the biometric system, while the remaining 20% served as the set for verification. A 5-fold cross-validation was applied to the subset simulating the enrolment phase, to estimate the generalization performance of the system. Linear Discriminant Analysis (LDA) was applied during the enrolment phase to reduce feature dimensionality and construct linear combinations of the available features. Matching score calculations were performed using a nearest neighbour classifier and impostor and genuine distributions were calculated using the Euclidean distance measure.

SpeechXRays bimodal system* - *Experimental Protocol
Face and speech information in SpeechXRays were integrated according to the matching score fusion process from the post-classification techniques. To follow this approach, matching scores from face and speech were first calculated from individual unimodal systems and then normalized using min-max normalization to produce scores varying from 0 to 1. The data used for building the bimodal biometric system relied on the MOBIO database composed of 152 people with speech and facial data [16]. At last, fusion was achieved using linear regression techniques [17]. Genuine and impostor distributions were calculated and presented in the results.

3 Results

3.1 SpeechXRays Unimodal System

The resulted distributions at the training phase and for each fold are presented in Fig. 1. At the cross-validation phase, verification accept rates were plotted against the associated false accept rates (Fig. 2A). DET curve analysis based on the relationship between the FAR and FRR measurements is shown in Fig. 2B.

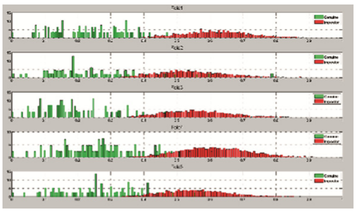

Fig. 1. Genuine and impostor distributions for each fold

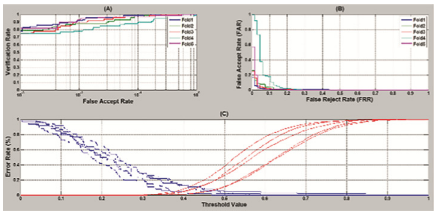

Fig. 2. Curves based on: (A) verification accept rate vs FAR and (B) relationship between FAR and FRR. (C) Error rates of FAR and FRR for different folds used

The error graphs of FAR and FRR are shown in Fig. 2C. In any given subset through the cross-validation procedure, the intersection point of these two graphs resulted to the

EER. The calculated value for EER was used to give automatically the threshold of the biometric system. Lower EER results to better system's performance, as the total error rate (sum of the FAR and the FRR at the point of the EER) decreases. A quantitative representation of the identification accuracy of the biometric system is given in the following table for indicative points in the DET curves (Table 1). All folds at the cross-validation process contributed equivalently to the estimation of the performance and an average value was calculated. The measures presented in Table 1 correspond to verification rate of 1%, 0.1, and 0.01 FAR to 82.60%, 72.27%, and 0.25% respectively.

Table 1. Verification rates based on different EER representing different levels of security for ORL dataset in SpeechXRays

Threshold	0% success	50% success	100% success
EER	13/40 subjects	14/40 subjects	13/40 subjects
EER – 10% EER (more strict)	18/40 subjects	11/40 subjects	11/40 subjects
EER – 20% EER (more strict)	19/40 subjects	13/40 subjects	8/40 subjects
EER + 10% EER (less strict)	11/40 subjects	13/40 subjects	16/40 subjects
EER + 20% EER (less strict)	9/40 subjects	11/40 subjects	20/40 subjects
Threshold	0% success	50% success	100% success

3.2 SpeechXRays Bimodal System

The data used for building the bimodal biometric system relied on the MOBIO database [16] and fusion was achieved using linear regression techniques and genuine and impostor distributions were calculated and presented in the Fig. 3. The evaluation of the bimodal SpeechXRays biometric system is also displayed in a quantitative way using specific points at the DET curve according to the Table 2.

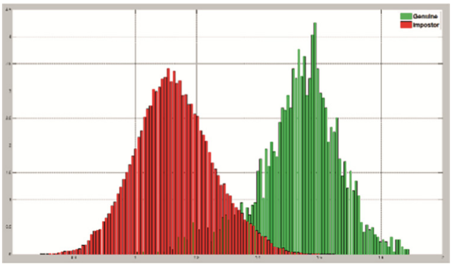

Fig. 3. Genuine and impostor distributions for MOBIO dataset

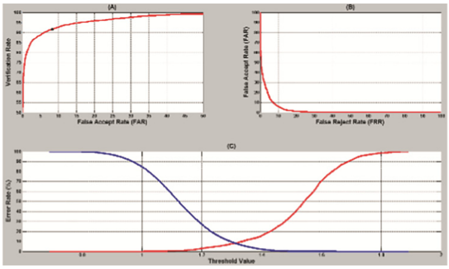

Fig. 4. Curves based on: (A) verification Rate vs FAR and (B) relationship between FAR and FRR. (C) Error rates of FAR and FRR using MOBIO bimodal data

Table 2. FAR and FRR rates based on different thresholds for bimodal operation in SpeechXRays for MOBIO dataset

Threshold	FAR	FRR
EER	8.33%	8.33%
EER – 10% EER	32.56%	1.98%
EER – 20% EER	73.12%	0.07%
EER + 10% EER	0.87%	22.18%
EER + 20% EER	0.03%	59.68%
EER	8.33%	8.33%

DET curve analysis were performed at the enrolment phase in which: (a) verification accept rates were plotted against the associated FAR, and (b) FAR against FRR are given in Fig. 4. The error graphs of FAR and FRR, (Fig. 4C) were also defined as the probability that an unauthorized user is accepted as authorized, and that an authorized user is rejected as unauthorized. According to the normalized integrated data, the calculated EER was measured with a threshold of **1.3109** and provided as an indicative threshold in the system. Access to the medical data will be given related to the sensitivity of the data in terms of their security levels. For high security the EER will be increased thus making a more secure environment for the user but inconvenient at the same time. On the contrary, medical data that are assigned, as "less secure" information can be accessed using SpeechXRays system with a threshold equal to the EER or less (Table 2). The measures presented in Table 2 correspond to verification rate of 1%, 0.1, and 0.01 FAR to 78.73%, 55.63%, and 29.44% respectively.

4 Discussion

Detection error trade off (DET) curves are used to assess the performance of a biometric system as a trade-off between the False Acceptance Rate (FAR) and the False Rejection Rate (FRR). A DET curve is a plot of FAR against FRR for various threshold values, $\boldsymbol{t}$. In this work the evaluation on verifying medical personnel's authentication through DET curve analysis for the multimodal biometric system of SpeechXRays was presented. Based on unimodal and multimodal datasets a demonstration of the context dependent tuning framework through DET curve analysis was described in order to test the verification rates based on different thresholds. Through this approach it was evaluated the user acceptance and the various matching thresholds that must be defined and adapted based on the level of security required by the service and the sensitivity of the medical data that the user is attempting to access. SpeexhRays' scope is to bring superior anti-spoofing capabilities and integrate them into an existing healthcare service. Upcoming updates of the SpeechXRays biometric system will include more advanced pattern recognition models for calculating the matching scores (i.e. Support Vector Machines).

Acknowledgement. This work is supported by the research project "SpeechXRays" which receives funding from the European Commission (EC) through Horizon 2020 Grant agreement No 653586.

References

1. Chiarugi, F., Spanakis, M., Lees, P.J., Chronaki, C.E., Tsiknakis, M., Orphanoudakis, S.C.: ECG in your hands: a multi-vendor ECG viewer for personal digital assistants. In: Computers in Cardiology 2003, 21–24 September 2003, pp. 359–362 (2003). https://doi.org/10.1109/cic.2003.1291166
2. Chiarugi, F., et al.: Real-time cardiac monitoring over a regional health network: preliminary results from initial field testing. In: Computers in Cardiology, 22–25 September 2002, pp. 347–350 (2002). https://doi.org/10.1109/cic.2002.1166780
3. Chronaki, C.E., et al.: An eHealth platform for instant interaction among health professionals. In: Computers in Cardiology 2003, 21–24 September 2003, pp. 101–104 (2003). https://doi.org/10.1109/cic.2003.1291100
4. Spat, S., et al.: A mobile android-based application for in-hospital glucose management in compliance with the medical device directive for software. In: Nikita, K.S., Lin, J.C., Fotiadis, D.I., Arredondo Waldmeyer, M.-T. (eds.) MobiHealth 2011. LNICST, vol. 83, pp. 211–216. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-29734-2_29
5. Tsiknakis, M., Spanakis, M.: Adoption of innovative eHealth services in prehospital emergency management: a case study. In: Proceedings of the 10th IEEE International Conference on Information Technology and Applications in Biomedicine, 3–5 November 2010, pp. 1–5 (2010). https://doi.org/10.1109/itab.2010.5687752
6. Maniadi, E., Spanakis, E.G., Karantanas, A., Marias, K.: A supportive environment for the long term management of knee osteoarthritis condition. In: Proceedings of the 5th EAI International Conference on Wireless Mobile Communication and Healthcare, London, Great Britain (2015)
7. Kondylakis, H., et al.: Digital patient: personalized and translational data management through the MyHealthAvatar EU project. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Annual Conference 2015, pp. 1397–1400 (2015). https://doi.org/10.1109/embc.2015.7318630
8. Maniadi, E., et al.: Designing a digital patient avatar in the context of the MyHealthAvatar project initiative. In: 13th IEEE International Conference on BioInformatics and BioEngineering, 10–13 November 2013 pp. 1–4 (2013). https://doi.org/10.1109/bibe.2013.6701560
9. Spanakis, E.G., Spanakis, M., Karantanas, A., Marias, K.: Secure access to patient's health records using SpeechXRays a mutli-channel biometrics platform for user authentication. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Annual Conference 2016, pp. 2541-2544 (2016). https://doi.org/10.1109/embc.2016.7591248
10. Jain, A.K., Flynn, P., Ross, A.A.: Handbook of Biometrics. Springer, New York (2007). https://doi.org/10.1007/978-0-387-71041-9
11. Li, S.Z., Jain, A.K.: Encyclopedia of Biometrics: I-Z, vol. 1. Springer, US (2009)
12. Ross, A.A., Nandakumar, K., Jain, A.: Handbook of Multibiometrics. Springer, US (2006)
13. Jain, A.K., Ross, A., Prabhakar, S.: An introduction to biometric recognition. IEEE Trans. Cir. Sys. Video Technol. **14**(1), 4–20 (2004). https://doi.org/10.1109/tcsvt.2003.818349
14. Jain, A.K., Ross, A.A., Nandakumar, K.: Introduction to Biometrics. Springer, US (2011)
15. Shen, L., Bai, L., Fairhurst, M.: Gabor wavelets and general discriminant analysis for face identification and verification. Image Vis. Comput. **25**(5), 553–563 (2007). https://doi.org/10.1016/j.imavis.2006.05.002

16. McCool, C., et al.: Bi-modal person recognition on a mobile phone: using mobile phone data. In: 2012 IEEE International Conference on Multimedia and Expo Workshops, 9–13 July 2012, pp. 635–640 (2012). https://doi.org/10.1109/icmew.2012.116
17. Ghayoumi, M.: A review of multimodal biometric systems: fusion methods and their applications. In: 2015 IEEE/ACIS 14th International Conference on Computer and Information Science (ICIS), 28 June–1 July 2015, pp. 131–136 (2015). https://doi.org/10.1109/icis.2015.7166582
18. Chiarugi, F., Trypakis, D., Spanakis, E.G.: Problems and solutions for storing and sharing data from medical devices in eHealth applications. In: 2nd OpenECG Workshop, 1–3 April 2004, Berlin (2004)
19. Spanakis, E.G., et al.: Technology-based innovations to foster personalized healthy lifestyles and well-being: a targeted review. J. Med. Internet Res. **18**(6), e128 (2016). https://doi.org/10.2196/jmir.4863, PMID: 27342137