TDDC03 Projects, Spring 2004

An overview of Biometric Authentication using Voice

Xiaobo Wang, Bo Sun

Supervisor: Fredrik Claesson, Kristin Anderson

An overview of Biometric Authentication using Voice

Bo Sun
820330-P216
bosu417@student.liu.se

Xiaobo Wang
760902-P279
xiawa640@student.liu.se

Abstract
Biometrics authentication is a hot topic in information society. There are some biometrics technologies that are mature and widely used today, and voice biometric is one of them with significant strength and obvious weakness. In this paper, an overview of biometric authentication system using voice is provided.

Key words
Voice, Biometric, Identifier, Authentication, Verification, Identification, Unimodal biometric, Multimodal biometric

1. Introduction

With the 21st century leading us into information age, computer systems, which provides a set of services and resource for the services, are playing a more and more important role in our life. However, access control to the service should also be provided by a secure system for information security. Since the traditional access control measures like passwords and PIN code have many disadvantages, one of which is that it has to be always remembered by people; And in order to enhance access authorization efficiently, it has become more and more important to recognize and apply human's biometric characteristics. Many biometric techniques have been evaluated and applied for access authorization applications, such as voice, fingerprints, face, iris, etc. In this paper, we will give an overview of biometric authentication using voice from two dimensions:

1. The current status of Voice Biometric Authentication.
2. The operation models existing in Voice Biometric Authentication.

2. Background

As an access control method, biometric authentication is implemented in a biometric system. What kind of system could we call a biometric system? "A *biometric system* is essentially a pattern recognition system that recognizes a person by determining the authenticity of a specific physiological and/or behavioral characteristic possessed by that person"[17]. A basic point to design a practical biometric system is to determine how a person can be recognized. According to the application context, a biometric system may be either a *verification* system or an *identification* system.

1. Verification means *"Am I whom I claim I am?"* A verification system is such a system that when someone claims to be a certain identity, the biometric characteristic captured from this person will be compared with his/her own biometric template(s) pre-stored in the database of the system. In a verification system, one-to-one comparison is made, and the result, true or false, is used to authenticate the person.
2. Identification means *"Who am I?"* In an identification system, the biometric characteristic captured from a person is compared with every template in the database for a match to recognize the person's identity. It is a one-to-many comparison process. If the comparison result is "No Match", the system rejects the person's access request, otherwise, the person's identity is identified and the request is accepted. An identification system confirms a person's identity without he/she claiming his/her identity (*Who am I?*).

Besides the two different concepts above, there are three important terms that are often mentions in biometric field:

1. **FRR** (False Rejection Rate): "a probability that a biometric technology denies access to an authorized user"[8].
2. **FAR** (False Acceptance Rate): "a probability that a biometric technology grants access to a unauthorized user"[8].
3. **EER** (Equal Error Rate): the point at which the **FRR** and **FAR** cross [8], which means a system will deny as many authorized users as will grant unauthorized users. The lower the equal error rate the more accurate any particular device is. A tight threshold setting will reduce the potential for false acceptance errors but it would increase the false rejection errors.

The three terms can be demonstrated in Figure 1.

As we mentioned at the beginning of background, a biometric system authenticates a person's identity through his/her biometric characteristic. There are 7 requirements for choosing a characteristic as a biometric identifier for a biometric system:

1. *Universality*, which means that the biometric identifier should be owned by each person.
2. *Distinctiveness*, which means that the difference of the biometric identifier between two persons should be distinctive enough to distinguish them.

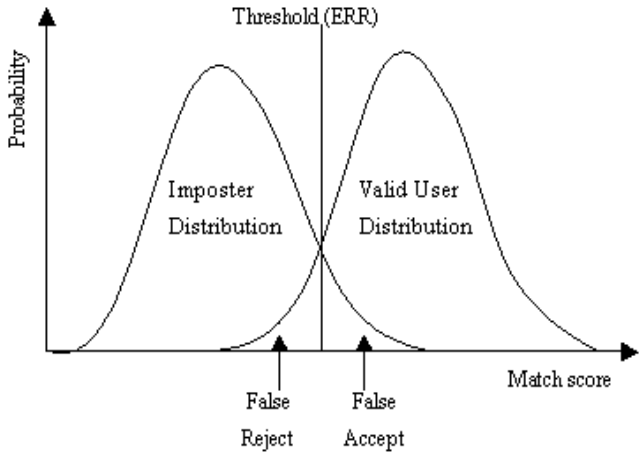

Figure 1, Relation of FRR, FAR and ERR [8]

3. *Permanence*, which means that the biometric

identifier should be stable enough within a certain long period.

4. *Collectability*, which means that the biometric identifier could be captured and analyzed in quantitative method.
5. *Performance*, which represents how fast, accurate and robust the biometric system can achieve, and to what extent the biometric system should be able to perform independently environment condition. It also represents how flexible the biometric system is in adjusting threshold settings depending on the security level of the application.
6. *Acceptability*, which represents to what extent people are willing to accept a particular physiological characteristic as a biometric identifier.
7. *Circumvention*, which means to what extent a biometric system can be cheated.

Fingerprints, voice, face, iris, retinal**,** etc are major biometric identifiers so far. Among them, the technique for voice biometric authentication has been very matured.

3. The current status of Voice Biometric Authentication

3.1 Voice biometric system

"*Voice biometrics* (also referred to as speaker verification or voice verification) involves verification of a speaker based on the unique geometry of the speaker's vocal tract such as the vocal tract length, ratio of larynx [1] to sinuses, and the resulting harmonics, pitch, and range."[8] It should be distinguished with speech recognition that does not aim to authenticate a person's identity but understand the meaning of words spoken by a person. Usually, the voice biometric system consists of enrollment process and authentication process. The enrollment process is prior to the authentication process. In enrollment process, the user must be enrolled into the database by creating a reference template of the user's voice-print. In the authentication process, the system authenticates a person's identity by matching a live voice sample with voice-print stored in the system database.

3.2 Work process of the voice biometric system

For further understanding of the voice biometric system, we should know how this system works and why a person can be authenticated by using this system. The basic work process of a voice biometric system can be divided into two stages and illustrated with two figures below: [22][23]

During the first stage (Enrollment, Figure 2), a person's speech is firstly captured and sent into feature extraction area, where each spoken words can be reduced into segments which compose of several dominant frequencies after passing through Fourier Transform function; Then after Magnitude [23], Sampling & Quantifying, these frequencies can be captured in digital form – feature vectors (binary code), which will be sent into Modeling Creation area. In Modeling Creation area, these feature vectors can be transformed into Gausssian Mixture Model (GMM) representation – voice-print by using GMM [24][p12, 22] method, which will be stored with a defined threshold in the database of the system for comparison in the future.

The second stage (Authentication, Figure 3) could be described like this [p4, 22]: When a user attempts to get access to enter the system, his/her speech utterance will be compared with his/her voice-print pre-stored in the database. In Similarity Score Calculation area, the similarity score extracted from user's utterance would be compared with his/her pre-set threshold in the database. If the similarity score is higher than threshold, this person will be accepted by the system; otherwise, he/she will be denied.

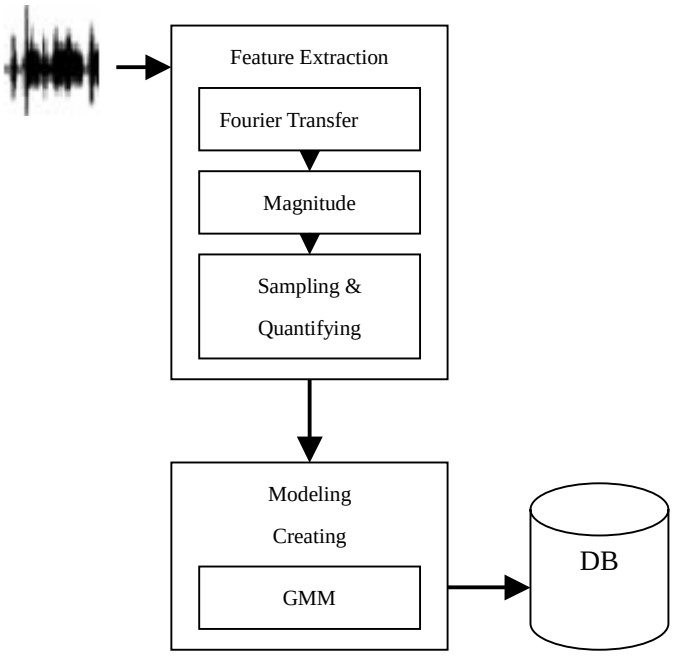

Figure 2. Enrollment [22][23]

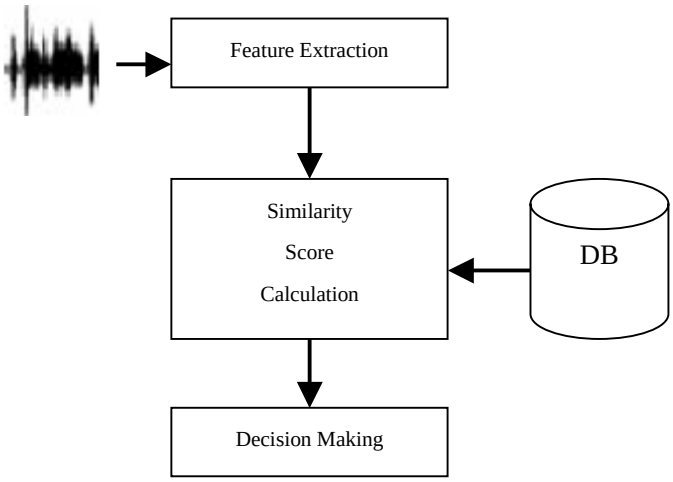

Figure 3. Authentication [22]

However, in order to decrease the FAR and FRR, purely voice-print based biometrics systems can be classified into two modes: text-dependent or text-independent. In both modes, a variation includes the possibility of using a challenge-response dialog that prompts users to repeat certain phrases [8]. In text-dependent mode, the phrases need to have been

[1] Larynx: part of throat

spoken during authentication as well as enrollment. And in text-independent mode, a much longer speech sample must be collected from the caller during enrollment instead of during authentication.

3.3 Advantage and disadvantage of voice biometric system

In the following section, we list the advantages and disadvantages of the voice biometric system.

Advantages:

"*Intuitive technology:* Voice biometric is an intuitive and natural technology since it uses the human voice"[8]. Speech is an obviously outside character of people so that the voice authentication is a natural process that is familiar with and widely accepted by people. Therefore it satisfies the requirement "acceptability" on a big level.

Cost effective: There is no necessity to equip any special hardware, but standard phones or microphones, to extract the voice features and authenticate the identity. So the voice biometric costs lower than other biometrics do. It can be instantly deployed, and mostly support the anytime, anywhere concept in practice compared to other biometrics.

"*Remote authentication:* Voice biometric is the only biometrics that can provide remote authentication without the need for the physical presence of the user."[8] For instance, if the voice biometric system works in text-dependent mode, and the text is generated randomly or time-variant, it will be hard to cheat the voice biometric system at remote terminal by using some tricks, such as using the tape that records the speech of the user. So the voice biometric authentication system in text-dependent mode can be used to achieve the remote authentication and satisfies the requirement "circumvention", while the other biometrics authentication systems cannot prevent artificial extracting devices from forging when authenticating the remote identities.

Disadvantages:

Variability: Speech, the synthetic of vocal tracts, mouth, nasal cavities, and lips, as a predominant biometric identifier, can be used to distinguish people uniquely. However background noises, temporary voice alterations, or poor quality phone interference makes the voice variable. Due to voice's variability, it is hard to select the threshold to balance the FAR and the FRR. Therefore the voice biometric system may not offer large-scale recognition. The scale of FRR and FAR associated with voice biometric system are 10-20% and 2-5% respectively [20].

Poor accuracy and low security: Because the voice-based features are sensitive to a number of factors, the typical voice biometric authentication system performs poor accuracy and voiced-based authentication is currently restricted to low-security applications.

"*Cross channel scenario:* The biggest challenge faced by voice biometric is that of cross channel use"[8]. A cross channel condition can arise if a user enrolls through one type of phone and tries to authenticate on another type of phone. The distortions, resulting from different types of phone, are different because of the different power and different gauss noise distribution in the phone. And finally it will result in the mismatch of the authentication system.

3.4 The privacy problems of voice biometric system

The speech is familiar with and widely accepted by most people, except those who have the obstacle with speaking, so voice biometric almost has no privacy problems. In addition, compared with other biometrics, voice biometric satisfies the requirement "acceptability" on a high level, while the fingerprint biometric on a medium level and iris biometric on a low level [1].

4. The modes of operation existing in Voice Biometric Authentication

4.1 Modes of operation

There are three operation models of a biometric authentication system using voice: the *unimodal biometric system* [13], which relies on the evidence of a single biometric trait, the system combining biometric with other identifier, and the multimodal biometric system.

A unimodal biometric system by using voice can be text-dependent or text-independent. With text-independent, because of the fatal weakness of voice identifier - voice's variability, it is hard to decrease the FRR and FAR. And with the text-dependent, the system will be simpler and can achieve higher accuracies than the system by using text-independent technique, but it is obtrusive and time-consuming. In addition, the unimodal biometric system is an identification system, and the variability of voice will increase the mismatch rate of the system. So it is hard to build a good authentication system only based on purely voice-prints.

In the system combining biometric with other identifier, people often use some identifiers, such as smartcard (what I have) or PIN (what I know), to make up the bad influence resulted from voice's variability, thereby to decrease the FRR and FAR. In addition, combining with what I know or what I have, the system is not an identification system (one-to-many), but a verification system (one-to-one), and the accuracy and robustness of the authentication system will be improved. Both the accuracy and complexity of this operation model are in the middle level in these three models, and is widely used today [4][21].

The third operation model using voice biometric identifier is the multimodal biometric system, in which some limitations imposed by unimodal biometric systems can be overcome. Such as multimodal biometric systems can ensure sufficient population coverage by using multiple traits. And multimodal biometric systems provide anti-spoofing measures by making it difficult for an intruder to simultaneously spoof the multiple biometric traits of a legitimate user. Thereby the third operation model is the most accurate and complex model among these three operation models.

4.2 Practical applications

The IBM's speech biometrics system [19] is a unimodal biometric system by using text-independent mode. It works by combining text-independent acoustic voiceprint match with conversational knowledge match. First step is the conversational knowledge match, in which the system matches the user's login account by recognizing what the user speaks with the user's password that the user input. If the conversation knowledge match results ok, the system goes to the second step, text-independent acoustic voiceprint match, which is a verification process by using text-independent mode. The

conversational knowledge match improves the accuracy and robustness of a purely voice-print based authentication system by decreasing the FRR and FAR to 3.2% and 10^{-5}% in about 20sec speech respectively.

There are some researches on multimodal biometric systems by using voice biometric. For example a hybrid person authentication prototype integrating multiple biometric devices is presented in [7]. In this authentication prototype, frontal face and text-dependent voice biometric are chosen to authenticate a user. Another example is an application of existing face and speaker identification techniques to a person identification task on a handheld device [10]. And both of the two multimodal biometric systems achieve reduction in equal error rate over the better of the two independent systems.

4.3 Potential attacks

The potential attacks to voice biometric system exist since the vulnerabilities of voice biometric are obvious. For example, the variability of voice biometric makes it hard to configure the threshold of a voice biometric system. The following scenarios specify potential attacks that threaten the poor threshold configuration.

First we assume there is a purely voice-print based authentication system, and there is a large-scale of voice-print in the system. We also assume that there are two voice-prints are similar with each other. Then the false acceptance will arise.

Second we assume there is a unimodal biometric system by using voice combined with PIN verification, and there is a small-scale of voice-print in the system. We also assume that there are two users' voice-prints are similar with each other, and one of the two users knows this situation and the other's PIN. Then the one can login system with the other's authorities.

To avoid such attacks, the solution should mainly be in two dimensions based on the realistic situation: one is to deploy a suitable operation model, and the other is to build a reasonable authentication policy.

There is a conflict between the accuracy and complexity when choosing an operation model for an application biometric authentication system. The high accuracy will be achieved through complex authentication processes. And the simpler the authentication process is, the less accurate the biometric system will be. It is the relationship (trade-off) between the accuracy and complexity of a voice biometric system.

5. Conclusion

Biometric authentication is a hot point in our present information society. More and more resource (human, fund, etc) has been invested in this field for research and application, which biometric identifier (fingerprints, voice, face, iris, retinal) is the best way to authenticate a person will never be a definite conclusion. However, since voice biometric authentication has some unique merit such as intuitive technology, remote authentication, etc, it has become one of methods that have been applied extensively.

Certainly, we should also notice there are many weakness which makes biometric authentication using voice hard to implement independently (the purely voice-print based authentication system), such as variability, cross channel scenario, etc.

We realize there is still much space for improvements in voice biometric authentication, for example, establishing a better policy to set up a threshold to make a better balance between FRR and FAR. Further study on user's requirements and continuously developing information technology is also helpful for improving a biometric system using voice.

6. Reference

[1] Anil Jain, Lin Hong, Sharath Pankanti, "Biometric identification", February 2000 *Communications of the ACM*, Volume 43 Issue 2

[2] Stéphane H. Maes, Jirí Navrátil and Upendra V. Chaudhari, "Conversational Speech Biometrics", Chapter in *E-Commerce Agents Marketplace Solutions*, Security Issues, and Supply and Demand

[3] J.-L. Dugelay, J.-C. Junqua, C. Kotropoulos, R. Kuhn, F. Perronnin1 I. Pitas, "Recent Advances in Biometric Person Authenticattion", IEEE Int.Conf.on *Acoustics Speech and Signal Processing (ICASSP), special session on biometric*, May 2002, Orlando, Florida.

[4] Judith A. Markowitz, "Voice biometrics", September 2000 *Communications of the ACM*, Volume 43 Issue 9

[5] Ganesh N. Ramaswamy Ran D. Zilca Oleg Alecksandrovich, "A Programmable Policy Manager For Conversational Biometrics", IBM Thomas J. Watson Research Center Yorktown Heights, New York

[6] Jirí Navrátil, Jan Kleindienst and Stéphane H. Maes, "An Instantiable Speech Biometrics Module with Natural Language Interface: Implementation in the Telephony Environment", Istanbul, Turkey, June 2000. *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*

[7] Norman Poh, Jerzy Korczak, "Hybrid Biometric Person Authentication Using Face and Voice Features", Paper presented in the Third International Conference, *Audio- and Video-Based Biometric Person, Authentication AVBPA 2001*, Halmstad, Sweden, proceedings pages 348-353, June 2001.

[8] Frost & Sullivan, "Secure Automation Solutions Using Voice Authentication", Copyright 2003

[9] Andrew R. Mark, "The development of destination-specific biometric authentication", April 2000, *Proceedings of the tenth conference on Computers, freedom and privacy: challenging the assumptions*

[10] Timothy J. Hazen, Eugene Weinstein, Alex Park, "Posters: Towards robust person recognition on handheld devices using face and speaker identification technologies", November 2003, *Proceedings of the 5th international conference on Multimodal interfaces*

[11] Gerrit Bleumer, "Biometric Authentication and Multilateral Security", AT&T Labs-Research, Shannon Laboratory, Florham Park, NJ

[12] Irma van der Ploeg, "Written on the body: biometrics and identity", March 1999, *ACM SIGCAS Computers and Society*, Volume 29 Issue 1

[13] Anil K. Jain, Arun Ross, "Multimodal interfaces that flex, adapt, and persist: Multibiometric systems", January 2004, *Communications of the ACM*, Volume 47 Issue 1

[14] Gerik v.Graevenitz, "Biometrics in Access

Control", Bergdata Biometrics GmbH, Bonn, Germany
[15] Ari Juels, Martin Wattenberg, "A fuzzy commitment scheme", November 1999, *Proceedings of the 6th ACM conference on Computer and communications security*
[16] Andrew S. Patrick, A. Chris Long, Scott Flinn, "Workshops: HCI and security systems", April 2003, CHI '03 *extended abstracts on Human factors in computing systems*
[17] Maltoni, D et.al. "Handbook of Fingerprint Recognition", Springer Verlag
[18] J. Navratil, J. Kleindienst, S.H. Maes, "An instantiable speech biometrics module with natural language interface: Implementation in the telephony environment", *ICASSP* 2000, Istanbul, Turkey, June 2000.
[19] J. Navratil, U.V. Chaudhari, S.H. Maes, "A Speech Biometrics System With Multi-Grained Speaker Modeling", *KONVENS 2000*, Ilmenau, Germany, October 2000
[20] National Institute of Standards and Technology, www.nist.gov/speech/tests/spk/2000
[21] Voice Identification And Authentication Roundup March 1, 2003
[22] Suphi Umut Naci, "Self Score Normalization and Frame Pruning Techniques for Speaker Verification Systems", B.S. in E.E., Bogaziçi University, 2001
[23] Dr. Hüseyin Abut, "Digitized And Digital Signatures for Personal Identification", *Distinguished Lecture Class 2002 IEEE Signal Processing Society*
[24] The Gausssian Mixture Model, http://www-math.univ-fcomte.fr/MIXMOD/statdoc/node27.html