TDDC03 Projects, Spring 2006

Biometric Authentication using Voice

Manuela Marzotti
Cristina Nardini

Supervisor: Niclas Wadstromer

Biometric Authentication using Voice

Manuela Marzotti Cristina Nardini
Linköpings universitetet, Sweden
Email: {manma325,crina723}@student.liu.se

Abstract

The biometric systems are increasingly used in our society. In this paper, we will address on one of these system: the biometric authentication using voice. Starting from the general background of the existing biometric systems we will proceed to analyze each step of a voice authentication system. We will describe the principles of application and at the same time the main problems related to this system, like security and reliability of this kind of system. Like every other authentication system, this one can be target of threats and attacks to its security and we will try to explain the main ones. Therefore we will give an overview about how the system is used in practice.

1. History

The term biometrics is a composite word and it is formed by "bio", which means living creature and "metrics" which means to measure an object, generally in a quantitative way, but in some occasions it refers to a qualitative criteria.

The first scientific method of biometric identification was developed in the laboratories of the prison of Paris by Alphonse Bertillon, who created an identification system called anthropometry, in the seventeenth century. [10] When Bertillon was 26, he began to annotate all the physical characteristics of the prisoners, until he had create a system which identified the criminals called **Bertillonage**, which was founded on a combination of physical measures. It was based on:

- The human skeleton does not change from the twentieth year
- Every skeleton is different for every individual

The Bertillon system was composed from two part: in one he revealed the physical descriptions of the human body and in the other one he revealed the physical measures of determined parts of the human body.

Bertillon introduced in the card the physical measures of the trunk, like height, arms' range of the arms, measure of the bust, measure of the right ear.

An other section of the card regards he introduced the physical measures of the head, which are of remarkable importance, because they don't change. They are width and length of the head. Subsequently the measures of the limbs are taken, which are left medium finger, left foot, left forearm. Then he introduced the descriptions of the face, the hand and the body of the prisoner like ear, forehead, nose, eye. At the end the hand of the prisoner is checked describing the palm, the face and the fingers of the hand and the operator looked for some distinguishing mark or tattoo on the body.

We reproduce a card for example below:

Figure 1. The first criminal identification card filed by the New York State Bertillon Bureau [9]

It was quite difficult to take accurate measures because the procedure was very complicated and there were a lot of number of cards to check. But the inconvenience that involved the failure of the Bertillon system as an identification device was that the physical measures of the Bertillonage were not unique, like in the event happened in 1903, in the federal prison of Leavenworth, when the operators noticed that the characteristics of a prisoner and his picture were similar to someone else physical measures and pictures. That failure may have been caused either from the fact that physical measures are not unique or from lacking accuracy in the measurements. Due to the inconsistency of physical measurements, it was necessary to move towards another kind of identification system which was not affected by the lack of reliability related to skeleton size and these other problems just described. Fingerprints were immune from these issues and thus they eventually

became the most common and widespread identification method.

From the 1960's till today, many biometric systems have been develops in order to change the authentication systems.

2. Introduction

It is possible to verify a user through three different approaches: something that he knows, like a password or a PIN, something that he has, like a key, or something that he is, biometric characteristics. The biometric systems are more simple because since the user does not have to remember the password or to be afraid of losing it, "but they are not secret. You leave your fingerprints on everything you touch, and your iris patterns can be observed anywhere you look." [7]. The development of the biometric systems is tightly coupled to the IT technologies, and this is the reason why today they are very used.

There is a large number of biometric methods: fingerprint, iris, signature, gait, hand geometry, voice, retinal pattern, etc..

It is possible to distinguish these characteristics into main fields:

- **physiological**: fingerprint, iris, hand, face;
- **behavioural**: voice, signature, gait.

According to their dissimilar quality they can be used in different environments.

In this paper, we will examine the biometric authentication using voice, specifying first of all how it works, the problems connected to its usage (like legal and privacy issues) and ultimately, the attack risks that such a system may suffer. After we will give an overview of how a system is used in practice work.

3. Background

"Biometric authentication is the verification of a user's identity by means of a physical trait or behavioural characteristic that can't easily be changed, such as a fingerprint." [2]. This is the strength that the biometric systems have. In fact, the value of most part of the biometric properties of a person is assumed not to change for a long period, so after their storage, it is possible recognize someone through corporal and behavioural characteristics.

The key idea of the biometric systems is to accept the identity of the person who is demanding a service, and to verify him or her as an authorized user. This procedure is implemented through measurements and comparisons of the biological characteristics of the individual with others one previously checked and recorded in a database by a mechanism, called Enrollment.

At this point, is important make a distinction between two kind of systems built to make a confrontation:

- **verification** : "*Am I whom I claim I am*?". In this case, the user claims to have a certain identity, and the system compares the person characteristic with only one template in the database, giving a result that could be true, if they match, or false, on the contrary.
- **identification**: "*Who am I*?". In this case, the user doesn't claim to have an identity, but the system compares the person characteristic with every template pre-stored in the database. It may match with one of the identities, or it may be a "No match" on the contrary.

The difference between a biometric system and an traditional access system is the kind of answer that is supplied as a result of one query. During the acquisition of a biometric characteristic, due to external and technological factors, some differences between the recording and the consecutive surveys will find and this makes necessary a probabilistic answer. For instance, fingerprints are not always stable because there may be some variations due to a wrong finger pressure on the device. A biometric system says therefore how much a current characteristic is similar to that one saved in the database, therefore "matching is always approximate, never exact" [13]. The immediately following problem is the identification of the threshold on which the system will have to operate the choices of acceptance or refusal of the biometric characteristic to check.

3.1 Biometric System Errors

A Biometric system can make two different kind of mistakes, that are well-known as:

- **FMR** (false match rate): "mistaking biometric measurements from two different fingers to be from the same finger" [3]. This is also called FAR (false acceptance rate).
- **FNMR** (false non-match rate): "mistaking two biometric measurements from the same finger to be from two different fingers" [3]. This is also called FRR (false rejection rate).

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. However, on the other hand, there are natural variations to biometric characteristics which may not be erased but that could be minimized through the appropriate equipments. 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 characteristic to check and template stored in the database. By elevating the threshold, the risk that not authorized users can fool the system diminishes (FAR is reduced), but, on the other hand, it is more probable that some authorized users can sometimes be refused (FRR increases).

The relation between this two kind of mistakes is represented below:

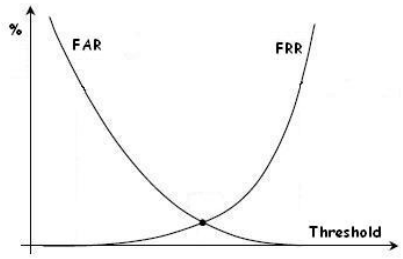

Figure 2. Relation between FRR and FAR

3.2 Requirements on a biometric identifier

The requirements of a biometric feature used for authentication purposes are seven, and they have to take into consideration the following property:

1. *Universality*: each person has a biometric identifier all his own
2. *Distinctiveness*: the biometric identifiers of two different person should be dissimilar in order to make a distinction of them.
3. *Permanence*: the biometric identifier don't change over a period time.
4. *Collectability*: the biometric identifier can be measured in a quantitative way.
5. *Performance*: shows the resources, the procedures and the environments that define a very careful identification.
6. *Acceptability*: shows the level on which each person accepts a particular biometric identifier in his lives.
7. *Circumvention*: which reflects that is possible to fool the system with the greatest of ease.

"A practical biometric system should have acceptable recognition accuracy and speed with reasonable resource requirements, harmless to the users, accepted by the intended population, and sufficiently robust to various fraudulent methods." [3]

4. Voice Biometric Authentication

4.1 Voice biometric system

In general, a voice biometric system works in the following way. It digitizes a person's speech converting each spoken word into segments consisting of "formants", which are numerous dominant frequencies. The tones in the segment identifies the person's voice prints, that are stored in the database. The wave formed by a user's words produces an electrical signal output from the microphone, called analog waveform or signal. "The objective of digitizing is to derive from an analog waveform a digital waveform that can be used in digital machines as a faithful representation." [4] This analog waveform is first sampled and then quantized in order to obtain a digital signal that will be store in the database ready for the comparison.

In order to distinguish different speakers, the voice verification system combines various characteristics that make a voice one of a kind and that identifies univocally the person who is speaking. The physiological aspects are mainly related to "the physical shape of the vocal tract" [11]. This depends by the conformation of pharynx, larynx, mouth, jaw, tongue and lungs. The different structure of all those elements in any human being makes the sound produced by the movement of the air quite different. The behavioural aspects are related to all different aspects like movements and manners that may influence the speech. It is thanks to the analysis of all those aspects that a voice verification systems can identify a voice by processing a spoken sentence, dividing it in segments and converting it to a digital format. Each segment is analyzed by all its distinctive vocal characteristics, like tone, pitch, and cadence.

To ensure that a voice sample has enough quality in order to be compared in the future, the user has to repeat several times a phrase or a sequence number so that the sample will be accepted and then memorized in the database. When the person pronounces the authorization sentence some parts of it are analyzed and compared with the very same parts of the sample previously stored by that specific user.

There are different systems that use different methods to recognize a person's voice. Some of them are text-dependent, while others are called text-independent. The first ones ask the speaker the insertion of a predetermined sentence for his identification. This phrase is also known as "pass phrase" [11].

On the other hand, the second kind of systems have a different approach. They do not rely on a pass phrase, they require the user to pronounce a speech quite longer. There isn't any bond to the text of the speaker for his identification. This gives the system the possibility to evaluate the voice of the person and compare the many different aspects that make it unique.

4.2 Phases

Like in all biometric system, the voice one is divided in different phases. First of all it is necessary to capture the signal and to store it in the database.

This process is called *Enrollment*, and executes the following steps [3]:

- *Biometric acquisition*: the first acquisition has to be with high quality so that the successive identifications could be consistent

– *Quality check*: "it is generally performed to ensure that the acquired sample can be reliably processed by successive stages" [3]

– *Features extractor*: the measures found are processed and combined in order to obtain a valid representation, called template.

Depending on the application, it is possible to store the obtained template in a magnetic card, in a central database, in a local workstation or directly in the acquisition device.

When someone will interrogate the system, we can talk about an authentication's phase, and as we previously saw, it could probably occur a verification or an identification request. The first application is concerned to determine if "a speaker is the person they claim to be" [5], and the second one "is the process of determining which speaker in a group of known speakers most closely matches the unknown speaker." [5]. The answer, in both cases, will be an acceptance or a refusal that means the user can have access or not to the system.

4.3 How it works

In a biometric system the signal enters in the system and it is manipulated to extract some mathematical information. The first step is signal detection (*segmentation*) in order to understand whether the incoming signal is voice. To do this "zero crossing rate" and height of the pitch peak in the autocorrelation function are used [5]. The first parameter is the number of cross between the signal and the x-axis, in a given time. It was calculated that if this number is greater than 70 the signal is not a voice signal[5]. An autocorrelation function is applied to the incoming signal to heighten the peak, this way we can determined if it is periodic or not. Probably a periodic signal is voice.

After the phase, the voice signal is ready for the feature extraction, while the others signals are discarded. There are two main techniques: MFCC (Mel Frequency Cepstrals Coefficients) and RASTA-PLP. With this MFCC, applying a series of mathematical transformations, it extracts a series of coefficients (cepstrals coefficients) represent the vocal signal. The signal is sampled at a frequency of 8 kHz, the samples are quantized and filtered, after they are grouped in a 10 ms frame. For each frame, the power spectre is obtained by the Fast Fourier Transform (FFT) and it is passed in a series of rectangular or triangular "pass band" filters. In each frequency band the energy is log-compressed, discrete cosine transform (DCT) is applied, in order to obtain the cepstral parameters. With this technique it can be represented a vocal signal through a vector of cepstral coefficients, that they not only contain information on the frequency and the intensity of the voice, but also on the power and the associated energy, on which the particular tone of voice depends.

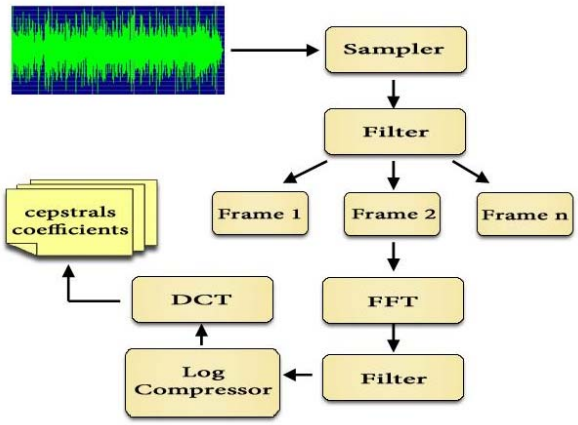

Figure 3. MFCC schema

The RASTA-PLP technique is an improvement of the PLP (Perceptual Linear Prediction), developed from H. Hermansky, it uses techniques of linear prediction in order to extract the parameters with which representing vocal signal and it has a series of transformations in order to simulate the human ear. The RASTA-PLP for the firsts phases works like MFCC, i.e. the signal is sampled and grouped and the FFT is applied. The associated energy to every band is weighed by the EqualLoudness function, it approximates the unequal sensibility of the human ear to different frequencies. The energy calculate in this way is log-compressed and the signal is filtered in a RASTA filter(RelAtive SpecTrA). It considers, for each band, the energy trend in the time, and filters this function in a "pass band" filter; this way the very fast or very slow variations of the energy are "smooth". After an inverse log transformation is applied at the energy's values, the energy spectrum is compressed and the cepstral parameters are calculated.

The two methods previously exposed concur to represent the voice with a vector of coefficients, stored in the database during the enrollment phase. For the authentication, the user has to emit vocal signal, that will be opportunely elaborate, in order to extract the cepstral coefficients. For the identification, the search inside the database will be looking for the signal more close to the input signal by executing of a nearest-neighbour query, using, for example, the Euclidean distance (since we use vectors). For the verification, the vector relates to the user's claimed identity, in the database, is used for the comparison with the incoming one. The authentication phase accepts or rejects the user, the threshold of rejection or acceptance depends on the particular system.

Whether the user is rejected the access to the system is denied.

4.4 Problems

A big problem in the biometric authentication using voice is that "the biometric community has not yet established upper limits for face and voice biometrics" [17]. It is impossible to know how many types of voice exist or if there are persons who speak in the same way. It is also impossible to measure how much a person's voice can change, and this means that one could be mistaken to be someone else and someone can imitate voice of someone else. This is one of the greatest open questions in the field of the biometric authentication using voice. [18]

"When a person speaks, compressed air form the lips as a *sound pressure wave*", [4] so the speech produced by each single person depends on the movement of the mouth, on the lung pressure and on the vocal tract. Sometimes also the very same individual may produce a different voice sample compared to the reference sample due to some alterations he or she might be experience such as a physical illness (like a cold) or unusual excitement or depression. As a consequence, the voice emitted by the person may be different from the template and it will not match, even if the identity is correct. This is a only the first problem a developer of one of these systems may encounter. There are some other issues that influence the effectiveness of voice recognition in general (and voice verification specifically). In first place, the quality of the input device has a big role concerning the quality of the voice sample recording. Many low-cost microphones are not able to capture some minimal differences of the characteristics of each segment of a voice sample. For this reason sometimes the biometric system is integrated with the speech recognition technology, so it could be able to authenticate the user whether on "something you are" or on "something you know" [14]. To this problem, we have also to add some problems due to background noise and, in the case of remote communication, due to line interference. In this case the user could spent more then one hour in front of the terminal in order to register his voice at the moment of the initialization.

The biometric system using voice is the only one that does not demand the presence of the user at the moment of the identification and also of the recording, therefore it can be used for telephony-based applications where it is possible to execute one remote verification. But this is also a disadvantage because it is necessary that the user does his authentication through the same device used for the recording. "The use of different channels affects the transmission of the voice features, thereby affecting the matching results." [14]

The authentication technologies are not flawless because it is managed from computer. A really big problem is that the computer can't know if the person which provides data to the machine is effectively the owner of the data. For a computer it is important that the given data during the authentication process is present in its database and the hackers play on this weak point of the system, that they try to "identify himself" as an authorized user.

4.5 Attacks

The problems related to security are connected to the devices used, like the database, the network link and the acquisition devices. [13] The threats to security can be different. The attacks can be headed to the device, when someone provides the replacement of the devices in order to steal the information, and also for the link among devices, hosts and workstation during the communication between extractor and matcher, or from enrollment or matcher to database [21].

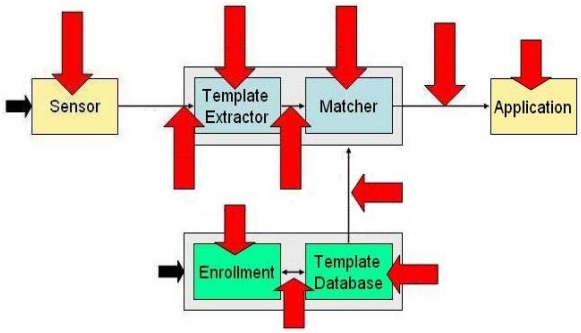

Figure 4. Pattern Recognition Model [21]

The biometric system using voice can be easily breached by replay attack [11] that can be resolved using digital encryption and time-stamping [11] in order to code the data flow and to sign it digitally.

The most clear attack is that a voice biometric system could be fooled using a simulation of a person's speech or recording it, so that it can be easily imitated with recording procedures. Therefore, if sometimes we misrecognize the voice of someone on the telephone, for the computers it is even harder the identification.

It is also possible to insert a Trojan which will give false data to the system, so that the calculated biometric print corresponds an authorized person and it could not recognize actually the honest user. This could happen whether in order to replace the characteristic in the database or in order to influence the match result, he could modify the information how he wants, "e.g. an hacker could replace the biometric data on a server with data that always declares a true match for a particular person." [21].

Naturally, in order to avoid a good part of these attacks, a first line of defense is represented by encryption, but for the rest ones, most of the times there is no solution yet.

4.6 Privacy issues

The biometric characteristics provide very personal information about its owner and this causes privacy problems. The privacy issue is a vexed question in many cases it slows down the spreading of biometrics into society. The people are afraid of use of the information found from the biometric systems, especially they have fear to be traced and to be watched in every moment. It cannot be avoided because biometric characteristics are inseparable and indissoluble from its owner. Of course, the way biometric invades the privacy depends from how and what kind of technology is used. We think that the biometric authentication using voice is less invasive than other types of biometric, for example the DNA. In fact, with the DNA other information on the person (hereditary disease) can be extracted, while with the voice this is not possible. The voice has an high level of "acceptability". However, there are privacy problems for the biometric voice in some situations, for example in surveillance applications: the persons must have the control on their biometric information, but these could secretly be collected and used to monitor the owner. The voice has this problem. The speech recognition can be used also to monitor the conversations, for example between the employees of a company and increasing the surveillance of them. "In the future, technologies like voice mimic and concatenated synthesis could be used in real time to generate utterances in the voice of a specific individual"[8]. Therefore, the introduction of the biometric systems in the society must be planned with some regulation, so that personal rights are not violated.

5. Areas of application

The application of this technology is mainly for security use, in many environments where it is necessary a certain protection for the access, for example:

- *eCommerce and credit card buying*, for the security of the user that buy;
- *Banking and trading*, for the secure management of the bank account and financial transactions;
- *Legal use*, to confirm the identity of a remote access and for block the access at the document;
- *Insurance use*, for avoid frauds;
- *Health use*, secure access to medical records and responsibility for doctor's orders;
- *Network security*, for the access to LAN and WLAN.

A real system, developed by IBM [22] and Sentry-com [16] and called VoiceShield, uses the voice recognition to secure access to the web during the business transaction process. The voice recognition is the third layer (last one) in the authentication process. The other two layers are related to the recognition of the hardware platform and to the introducing of a 4-digit PIN. First the system matches the user's computer hardware identify and the PIN code inserted by the user. If the match is positive the access to the system is allowed. The voice recognition is done afterwards, it "has to be used for information access only"[16]. The biometric technology is used to add security at the system and minimize the possibility of stolen identities. The developers assert that "the chance of two randomly chosen individual to have similar VoicePrints has been shown in Field-Test trial to be in the order of 1%"[16]. The kind of tests made to achieve this result are not explained, thus it is difficult for us to determine how does the system work in a realistic environment. VoiceShield is a text-depended system and allows only the user's verification. The system server receives the encrypted voice print and checks the identity using a decision algorithm. The comparison is between the input voice and the record in the database related to the PIN code; if the identity is verified the access is permitted.

In the cellular field there are researches in order to find system for the speaker verification, like the IBM's NIST, that uses a combination of multiple, data-perturbed system designed to improve the performances [19]. Mitsubishi Electric Telecom uses the voice recognition for some of its mobiles. The voice assures the secure access to the cellular, when it is turn on a sentence is pronounced and it will be compared with the model in the SIM card [20].

6. Conclusions

Currently the biometric seems to be the natural evolution of the traditional systems resulting from technologies improvement. Certainly the voice is one of the more studied technologies. The traditional access techniques have a series of problems, they can be lost, stolen or lent in an unauthorized way. Moreover they do not control the effective identity of the customer and need that the user remembers codes or password. The biometric key is generated from the personal characteristics of the individual, therefore is not subject to the problems listed before. However there are others problems. For example the decision is probabilistic, it does not indicates an absolute certainty, it expresses only likeness: a malfunctioning microphone or a cold could distort the result.

Another difficulty is that a voice can be imitated. We believe that a reliable system should be able to

understand which part of the voice signal is not possible to imitate in order to use it as a key point during the authentication phase, but we don't know how it is possible in realistic applications.

The voice recognition is a field in which numerous researches and studies are possible, and it has all presuppositions needed to be an important component in the future security systems.

References

[1] Miller B., Vital signs of identity. February 1994 *IEEE Spectrum*, 31, 22-30.

[2] Russell Kay, QuickStudy: Biometric authentication, April 2005 http://www.computerworld.com/securitytopics/security/story/0,10801,100772,00.html

[3] Maltoni, D et.al., Handbook of Fingerprint Recognition, Springer Verlag

[4] Gordon E. Peltron, Voice Processing, Uyless Black, Consulting Editor

[5] Richard L.Klevans, Robert D.Rodman, Voice Recognition, Artech House, Boston – London

[6] The trouble with biometrics, http://www.sagecertification.org/publications/login/1999-8/features/biometrics.html

[7] Biometrics: Uses and Abuses, http://www.schneier.com/essay-019.html

[8] Speaker verification and Privacy, http://www.jmarkowitz.com

[9] Division of Criminal Justice Services, http://criminaljustice.state.ny.us/ojis/history/ph_bert1.htm

[10] Anthropometry, http://en.wikipedia.org/wiki/Bertillonage

[11] Voice verification, http://www.globalsecurity.org/security/systems/voice.htm

[12] James L. Wayman, Digital Signal Processing in Biometric Identification: a Review, College of Engineering and Office of Research and Graduate Studies

[13] The Biometric Dilemma, Rick Smith, Ph.D., CISSP - 28 October 2001

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

[15] Zdenek Ríha & Václav Matyáš, Biometric Authentication Systems, Copyright 2001

[16] The authentication voice of security, http://www.sentry-com.co.il/index.html

[17] An Introduction to Evaluating Biometric Systems, http://www.frvt.org/DLs/FERET7.pdf

[18] Sadaoki Furui, Recent advances in Speaker Recognition

[19] G.N. Ramaswamy, J.Navratil, U.V. Chaudhari, R.D. Zilca, The IBM system for the NIST 2002 cellular speaker verification evaluation, April 2003, ICASSP-2003, Hong Kong,

[20] Mitsubishi electric, http://www.mitsubishielectric.com/

[21] Leandro A. Loss, Thwarting Attacks, www.cse.unr.edu/~bebis/CS790Q/Lect/Chapter_12.ppt

[22] IBM, http://www.ibm.com/us/