Audio Voice Authentication (November 2004)

Jacqueline Chow, Janet Tse, and Jeff Tung, *Students, UBC ECE*

Abstract — In this present day of vast technological advances that demand the highest security possible to protect legitimate users and their data from impostors, current security technologies may not prove to be sufficient. The current, widely used authentication schemes are based upon the establishment of one's identity by "what he has" and "what he knows" – the most common example of such is the bank card and their respective PIN number. This form of authentication has worked well in the past; however, these systems are vulnerable to the ploy of an imposter. Biometrics, as a method to partly or entirely certify the user's claim of identity is becoming progressively more attractive since it establishes one's identity based on "who he is". Among the various forms of biometrics for use as a form of security is voice biometrics – the use of a person's voice to certify they are who they claim to be. In this paper, we investigate the feasibility, advantages and disadvantages of user authentication using voice biometrics.

Index Terms — Authentication, biometric, recognition, security, speaker identification, verification, voice.

I. Introduction

IN order to ensure that legitimate users and their information are protected from imposters, authentication techniques has been deemed necessary for many decades. Presently, with the advance of embedded CPU technology leading to more applications of embedded systems, the traditional ways of authentication have been seen as unreliable. Ideas of using biometrics to determine one's identity are growing within many organizations that see the uniqueness associated with this method. There are three types of security and authentication:

1. what you have – a card key, token (like a bank card);
2. what you know – a password, PIN, piece of personal information (like mother's maiden name);
3. who you are – a biometric.

Biometrics is the most secure and convenient of the three types since it does not rely upon information that the user has and knows to grant access. Instead, it is about who the user is by use of his/her physiological and/or behavioral characteristics. As [2] states, any human characteristic can be used as a biometric characteristic if it satisfies all of the following four requirements:

- Universality: each person should have the characteristic.
- Distinctiveness: any two persons should be sufficiently different in terms of the characteristic.
- Permanence: the characteristic should be sufficiently invariant (with respect to a matching criterion) over a period of time.
- Collectability: the characteristic can be measured quantitatively.

Biometrics using physiological characteristics would include ear, fingerprint, palm and hand geometry, iris, retina, facial characteristics. Common behavioral biometrics includes voice, keystroke pattern, signature, and gait. Among all the various biometrics, fingerprint, iris, retina, voice and signature are some of the more developed types for use as authentication techniques [1]. Figure 1 shows the process of using biometrics in security.

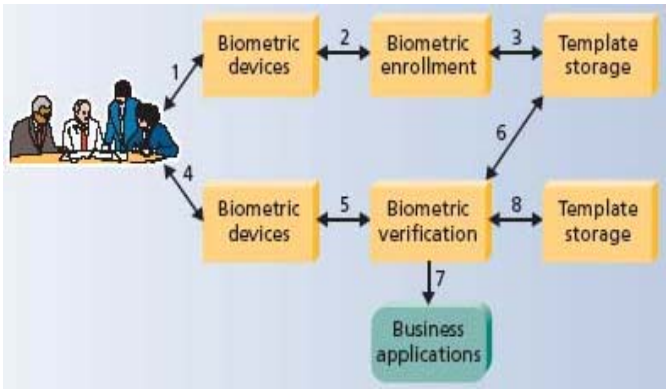

Figure 1. Process of Using Biometrics in Security.

1. Capture the chosen biometric by use of biometric devices;
2. Process the biometric information, extract and enroll biometric template;

J. Chow is currently an undergraduate in the Electrical and Computer Engineering Department at the University of British Columbia, student number: 83556019, e-mail: jachow@ece.ubc.ca.ca

J. Tse is currently an undergraduate in the Electrical and Computer Engineering Department at the University of British Columbia, student number: 83313015, e-mail: jtse@ece.ubc.ca.ca

J. Tung is currently an undergraduate in the Electrical and Computer Engineering Department at the University of British Columbia, student number: 50891985, e-mail: jtung@ece.ubc.ca.ca

3. Storage of the biometric template in repository – local, central, portable form (like smart cards);
4. Live-scan the chosen biometric;
5. Process the biometric information, extract the biometric template;
6. Compare the scanned template with the templates in repository for authentication;
7. Provide matching score to business applications;
8. Record a secure audit trail with respect to system use.

II. VOICE BIOMETIC AUTHENTICATION

Voice, usually considered as a form of behavioral biometric is in fact a combination of both physiological and behavioral biometrics. Because no actual personal characteristics are available in the voice, a voiceprint system must convert the speech signal into the physical characteristics that they represent. For example, the vocal tract shape is represented in the properties of the spectral peaks and the glottal source is tied in with the pitch striations. Figure 2 shows the physical characteristics that affect speech production and Figure 3 shows the corresponding speech production model. Hence, the design of voice authentication technology is not based upon voice recognition; but rather, it is based upon voice-to-print authentication [3].

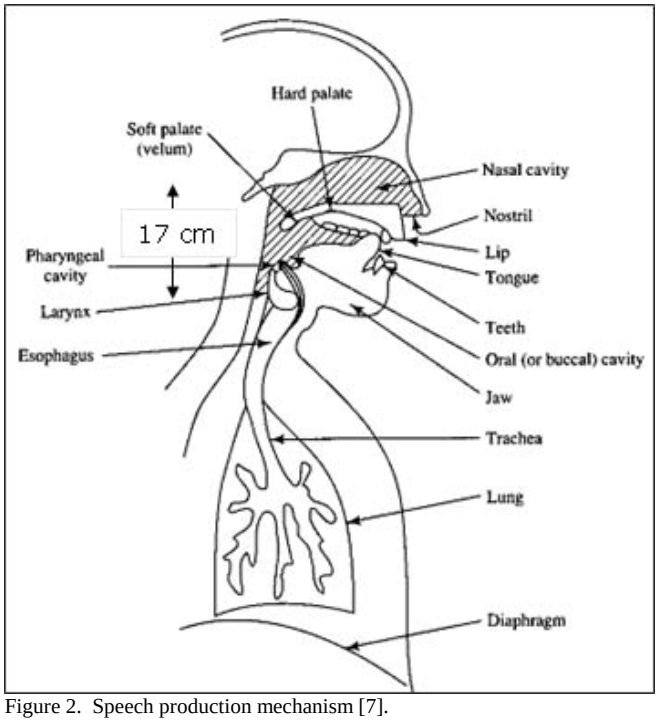

Figure 2. Speech production mechanism [7].

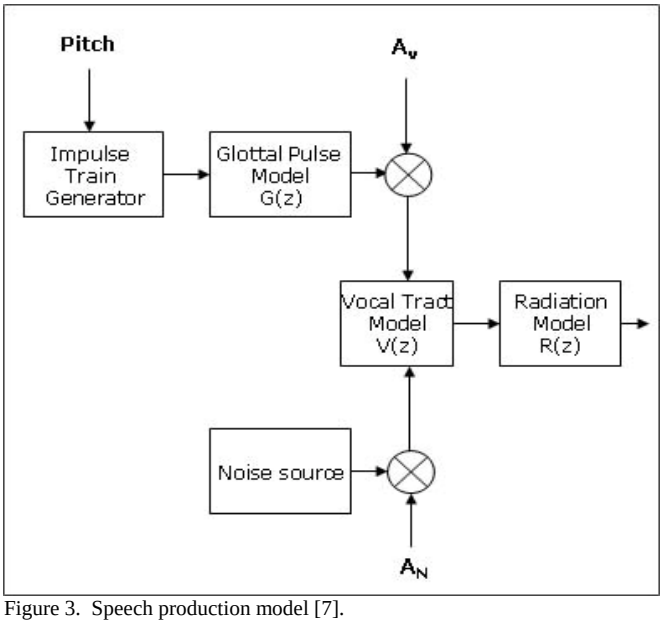

Figure 3. Speech production model [7].

There are two main types of systems used to identify users; text-dependant or text-independent speech input. Figure 4 shows a decision tree for speaker identification. A text-dependant speech input system would typically involve a randomly generated pass phrase to combat replay attacks and have the users recite the phrase. The system would then compare the properties of the input waveform to the saved property of the user and see if there is a match. Text-independent speech input approach, on the other hand, can verify the user's identity without need of text. This is more difficult but more flexible approach, such as background verification when the user is conducting other speech interactions [1], [3].

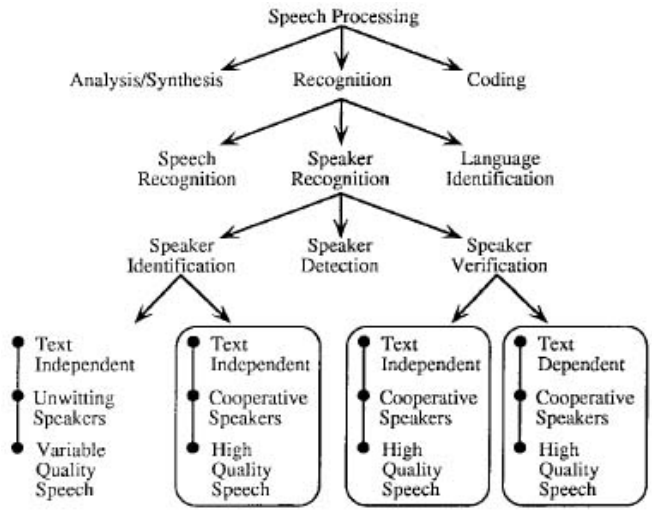

Figure 4. Speaker identification decision tree [6].

The advantage of the text-dependant system is that because the user is an active part of the authentication program it allows for a more controlled authentication procedure which will produce a more accurate measurement allowing for a smaller window of error. However, it also allows for replay attacks if an attacker can figure out how the phrases are

generated. A text-independent system in contrast can be used to verify users more naturally during normal conversation and can even verify people without them knowing about it. Although the hardware and software to do this is more complicated as it must deal with more background noise and uncontrolled inputs.

There are many different procedures for verifying users using voice. The first of which is a simple feature extraction/pattern matching algorithm. There are five main steps in the verification process shown below in figure 5. The first is obtaining the signal. Then the signal is filtered and converted to a digital format for processing. The signal then has the main features that are looked at by the system and the signal is split up into small vectors typically spanning around 20 ms. Each of these vectors is then compared to a stored vector to generate scores for each vector according to how closely they match. The scores are then looked at as a whole to generate an accept or reject response to the attempt at authentication.

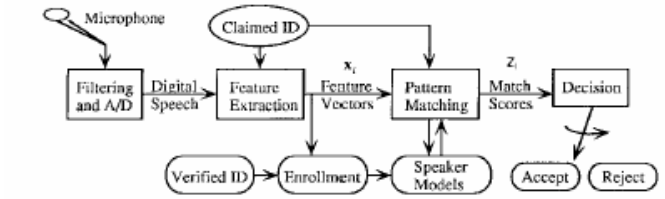

Figure 5. Generic speaker-verification system [6].

However, this technique requires a very controlled environment for deployment. If there is a large amount of background noise the readings taken by the microphone will not be as clear as one taken on a quiet day. Such a variance in readings on the system above would cause noticeable variations on the authentication strictness. On a loud day it may take a user more times to authenticate then on a quiet day. The system would also need to lower requirement so that it is possible for a user to authenticate in the worst case scenario [6].

Because of that a more sophisticated method is usually used in practice. In this method along with the users' speaker model, a generic model is created to help authenticate. If the generic model can be made sufficiently complex enough to take into account background noise and other current environmental aspects then it can simulate what a random user would sound like if it was trying to authenticate on the system. If the system has two models to compare it to then when a user tries to authenticate then it can use probability to test which of the two models the generic or the specific model it most likely matches with. The second generic model improves performance because it provides a neutral model accounting for background noise to normalize the system and provide the decision making mechanism with a simulation of what a false user would be.

Beyond the basic structure the system uses, a determining factor of how strong a voice authentication system is on the pattern matching algorithms it implements. The traditional method is to use mathematical models to measure the difference and use error tolerances to determine if users are accepted. However, neural networks are also emerging as an alternative to these methods allowing for more sophisticated probability analysis.

Dynamic Time Warping is commonly used in speech recognition as it is one of the simplest to implement. It does not model probability and only computes the difference between two samples. The sample is first split into frames for analysis. Each frame has its main features extracted such as large peaks. Then the difference between the stored file and the new file is taken and squared. This forms the basis of the scoring of a file. The squaring in the formula ensures that large differences are scaled up because they are more likely to suggest a false user [7].

Hidden Markov Modeling is a more advanced technique that directly deals with probability. It does this by creating a state machine with many different states corresponding to features within a voice signal. Then a transition matrix is stored which holds the probability of state transitions. A Markov model is then created based on a speech sample as can be compared to the stored model and a comparison of the two can be made. Although this technique is more computational intensive it takes into account transition between frames which makes it harder to fool than a Dynamic Time Warping.

Probability calculations can be made even more powerful with the advent of neural networks. Naturally, as humans we recognize voice very easily and it is one of the primary ways we can identify people. However, our ability to do this is hard to describe in a mathematical formula, perhaps because we analyze many parts of the voice at once rather than in pieces. Neural networks are able to simulate how animal or human brains function so they are able to identify speakers similar to the way we do. Because of this neural networks out test Markov Models in speaker identification rate anywhere from 12% to 24% [8].

III. ADVANTAGES OF VOICE BIOMETRIC

With the vast number of biometric authentication choices available, voice biometrics has its own distinctions that cannot be matched by others. These include the lack of new hardware needed, not intrusive, and can naturally thwart spoofing attacks by use of different authentication approaches.

Among all the various types of biometrics that can be used in authentication, voice biometrics has the most potential for growth. In addition, speech does not project to the users as threatening or intrusive to be provided since it is a natural signal that is produced. In many applications, speech may be the main and perhaps only modality – telephone system for example. Even for other applications, with non-telephone related signal delivery, such as computer-based

applications, no new hardware is required since most computers today are packaged with sound cards and some with built-in microphones. In addition to simple hardware voice biometrics have the most potential for growth due to the nature of the data it gathers. Voice signals tend to be smaller in size than a fingerprint scan or a retinal scan. Voice Security System's even advertises a voiceprint from them is less than 1Kb in size. This would allow for a voice print to be carried around easily on a keychain for some sort of large scale authentication system. There could be generic terminals for authentication, and then a user could walk up and put in a voice print linked to a user and then talk to match the voiceprint and thus be authenticated to the system.

When combined with utterance verification, voice verification and authentication is one of the few biometrics that supports a natural "challenge-response" to help thwart spoofing attacks. This can be achieved by the text-dependant system approach that presents a series of randomized phrases for the user to repeat. It is possible for the system to not only verify that the voice match, but also that the actual required phrase repeated match. In addition, the use of automatic knowledge verification can also be used to compare the content of the spoken utterance in response to a question to that information stored in his/her personal profile [3]. An example would be the spoken utterance response to a question such as "What is the name of your pet?"

In addition the ease of implementation, because voice is how most humans communicate with one another, a voice biometric system that uses a challenge-response system could easily be extended to provide customer service related services. After authentication of the user over the phone, a voice system could easily be modified to ask a customer what kind of problem they were having or who they wished to speak to. The voice system could then forward the call or even play an automated response to simpler requests, thus reducing the number of employees needed [9].

IV. DIFFICULTIES/DISADVANTAGES OF VOICE BIOMETRIC

Despite the benefits associated with the use of biometrics as forms of authentication, there are other factors that need to be taken into consideration. With all biometrics, there exist two primary sources of error in biometric data: time and environmental conditions. As stated by [1], biometrics many change as an individual ages. Environmental conditions may either alter the biometric directly (for example, if a finger is cut and scarred) or interfere with the data collection (for instance, background noise when using a voice biometric).

Voice biometric is one of the biometrics that has both physiological as well as behavioral characteristics. The physiological characteristics of human speech are invariant for the individual; however, the behavioral characteristic of the speech of a person changes over time due to emotional state, age, and medical conditions (such as a common cold) [5]. As a result of this, the voice signal may not be consistently reproduced by the speaker.

Varied microphones and channels used can also cause difficulties since most voice authentication systems rely on low-level spectrum features susceptible to transducer/channel effects [3]. Moreover, the ambient noises in the environment upon which authentication systems is used can lead to complications during the capture of the voice biometrics. Furthermore, there may exist and acoustic mismatch between the training and the testing environments since the enrollment and testing voice may come from different headphones and networks [4].

In addition, the enrollment procedure has often been more complicated than with other forms of biometrics since the accuracy of the collected training data is critical to the performance of the authentication system. Even a true speaker might make a mistaken when repeating the training utterances or pass-phrases for several times. It is also seen as an inconvenience to the user as well as the system developer, who often has to supervise and ensure the quality o the collected data [4]; thus, leading to the perception that voice authentication is not user friendly.

Lastly voice biometrics suffers an amplified disadvantage that is inherent in all types of biometrics. Because biometrics is the measuring of a physical property of an individual if it is somehow stolen such as a digital scan of someone's fingerprint, it is insecure forever. You cannot change a fingerprint like a password. However, other forms of biometrics can make up for this by increasing the security between the data input point and the data processing point. An example of this would be not allowing any type of digital input into the system besides the fingerprint scanner. This is a much harder task because if someone is capable of reproducing your voice, it will be able to enter an authentication system just as easily through the microphone as the valid user.

V. CONCLUSION

Voice Biometrics is not at all a new concept. However, as little as five years ago very few people possessed any kind of deeper understanding of the idea. In the last few years the increase of processing power in the way of development of advanced neural networks and the size of devices needed to implement voice authentication shrinking it has reemerged as a viable solution for security issues.

Conceptually, voice has always been an ideal biometric for security. When compared to fingerprints or retinal scans, most people would prefer voice because it seems less intrusive. It is possible to authenticate or identify people without them knowing they are even under scrutiny due to

the small size of microphone technology. As well as small microphones, voiceprint records of a user are also much more portable and smaller due to new storage technologies. However, despite all these advantages, voice authentication was rarely used in practice due to concerns that it was not accurate enough.

These concerns were once quite valid as older systems when implemented on a large scale, typical voice authentication systems would have 60-80% success when a user was trying to authenticate. In addition to this, users would need to be trained in order for the system to be able to work with the user. These difficulties typically overshadowed the advantages of the voice system. Who would care how easy it would be implement authentication hardware if it required 10 minutes for each of a ten thousand user database to create necessary profiles.

With success levels rising to nearly 90% per try due to neural networks being integrated into the pattern matching stages of authentication, the problem of lack of accuracy is being addressed. These new techniques also allow for smaller samples being needed by the system which will lower the training time needed to profile a user. Because of these improvements voice authentication now becomes a viable option for many different situations. A company could easily setup a voice authentication for the accounts of its customers over the phone; although human operators would still need to be employed for the minority who would be unable to authenticate successfully such a system would increase security and reduce personal costs. Voice prints could also easily be stored onto the cards used by customers of a bank and voice authentication could help to authenticate people to bank machines. In addition to these large scale implementations, voice prints can easily be implemented on a local level. They are already showing up on cell phones to lock the phone in case of theft. The possibilities are almost endless for the applications of voice biometrics and it is almost certain they will soon make up a larger piece of the security market in the next few years.

REFERENCES

[1] S. Liu, and M.Silverman, "A Practical Guide to Biometrics Security Technology," IT Professional, vol. 3, issue 1, pp.27-32, January-February 2001.

[2] A.K. Jain, A.Ross, and S. Prabhakar, "An Introduction to Biometric Recognition," IEEE Transactions Circuits and Systems for Video Technology, vol. 14, issue 1, pp. 4-20, January 2004.

[3] J. Ortega-Garcia, J. Bigun, D. Reynolds, and J. Gonzalez-Rodriguez, "Authentication Gets Personal With Biometrics," IEEE Signal Processing Magazine, vol. 21, issue 2, pp. 50-62, March 2004

[4] Qi Li, Biing-Hwang Juang, and Chin-Hui Lee, "Automatic Verbal Information Verification for User Authentication," IEEE Transactions Speech and Audio Processing, vol. 8, issue 5, pp. 585-596, September 2000

[5] A.K.Jain, Lin Hong and S.Pankanti, "Biometric Identification," Communications of the ACM, vol. 43, pp. 90-98, February 2000.

[6] J.P. Campbell Jr., "Speaker Recognition: A Tutorial," Proceeding of the IEEE, vol. 85, issue 9, pp. 1437-1462, September 1997.

[7] S.S. Chikkerur, "Speaker Recognition," presented at the University at Buffalo Center for Unified Biometrics and Sensors [Online]. Available: http://www.eng.buffalo.edu/~ssc5/resources/presentations/lectures/Lec4.ppt

[8] Erik J. Zeek, "Speaker Recognition by Hidden Markov Models and Neural Networks," Master's Thesis, Air Force Inst. of Tech. Wright-Patterson, AFB OH School of Engineering and Management, December 1996. Available: http://www.stormingmedia.us/69/6960/A696023.html

[9] S.J. Vaughan-Nichols, "Voice Authentication Speaks to the Marketplace," IEEE Computer Society, vol. 37, issue 3, pp. 13-15, March 2004.

[10] W.M. Campbell, and C. C. Broun, “Low Complexity Speaker Authentication Techniques Using Polynomial Classifiers,” Proceedings of SPIE - The International Society for Optical Engineering, vol. 3722, pp. 357-367, 1999.