Global Information Assurance Certification Paper

Shedding some light on Voice Authentication

Dualta Currie

GSEC- V1.4b

Abstract

Biometric authentication technology and development has grown over the last 6 years from being something we have seen on Science fiction television shows into a reality where we can now purchase the systems and implement them both in our business and private lives.

In this paper I will attempt to explain, in non-technical language, the technologies behind one particular type of biometric authentication, voice authentication. I will look at the human voice, how this is captured by technology, and how this can then be used to verify that the person is who they claim to be.

Introduction:

People have always kept secrets and protected their possessions. It has evolved from the physical to the technological, where now technology is used to restrict access to our resources and user authentication is the 'key to the door'. Authentication techniques have developed with it, and now who can access our resources is controlled using three main methods: [1]

- Knowledge-based authentication is based on information authorised individuals will know, and unauthorised individuals will not. E.g. A PIN or a password. information.
- Object-based authentication is based on possessing a token or tool that permits the person access to the controlled resource. E.g. Keys, pass cards or a SecureId.
- Biometric-based authentication measures individuals' unique physical or behavioural characteristics. It exists today in various forms such as fingerprint verification, retinal scans, facial analysis, analysis of vein structures and voice authentication.

Reasons for using Biometric Authentication:

Biometric authentication has some key advantages over knowledge and token-based authentication techniques. Biometric characteristics are not easily forgotten, like a password, or lost like a key. One can hardly lend someone your finger nor can someone easily steal your eye. That makes them fairly secure, and convenient. Unfortunately, they've had to wait for technology to catch up to the level where it can support their effective use. Only recently has technology provided the statistical, analytical and data processing techniques to support it properly.

Why choose voice authentication?

For the majority of biometric authentication techniques, sophisticated equipment and the physical presence of the person being authenticated is required. For example, fingerprint scanning, pen signatures and retinal scans – not so with voice authentication, where authentication may be given remotely via a device commonly known as the telephone. Given the use of the correct analytical techniques, a person's voiceprint can be as unique as any other biometric characteristic, but yet can be used for authentication remotely and has the added benefit of being less personally intrusive than say, subjecting the person to a retinal or fingerprint scan.

The concept :

Voice authentication is a fairly simple process. To register, a user records sample(s) of their voice which are stored in the authenticating system and become known as their 'voiceprint'. Then, to access this resource subsequently, they supply a sample of their voice to the system, and it decides if it matches their voiceprint before allowing them access.

The risks of it's application:

When deciding whether or not to employ a voice authentication system it is important to consider the application. If it is to be used to authenticate a user to administer their bank accounts for example, this is a completely different risk than say accessing their voicemail on their mobile phone. Should a false acceptance result in the banking application, the consequences would be considered much more severe.

Remember : Risk = Threat x Vulnerability

The elements of any voice authentication system need to be analysed and it must be ensured that individually and collectively, the probability of a vulnerability arising is low, and the potential for an individual or group to exploit the vulnerability unlikely.

Voice biometrics explained:

Biometric characteristics fall into two broad categories :[2]

- Physiological Biometrics are concerned with the unique physical traits of the individual, for example retinal scans, fingerprints and face geometries.
- Behavioural Biometrics are concerned with the unique way individuals perform certain actions, for example conventional pen signatures and key stroke detection.

In the case of voice authentication, there is both a Physiological biometric component (for example, voice tone and pitch) and a behavioural component (for example, accent). This makes it very useful for biometric authentication.

Voice authentication, identification and speech recognition.[3]

Voice authentication, also referred to as 'verification', is just one of the voice-based technologies. Others include voice identification, interactive voice response (IVR), and degrees of speech recognition. These technologies share base technologies and methodologies, but differ considerably in terms of the extent to which reliance is placed on certain sub-technologies.

To allay any confusion between the various technologies, I will briefly describe the differences between them, before focusing particularly on the issue of voice authentication:

Voice (or speech) authentication attempts to verify that the individual speaking is, in fact, who they claim to be.[14] This is normally accomplished by comparing an individual's voice with a previously recorded "voiceprint" sample of their speech.

Voice identification attempts to identify the individual's voice. This normally involves comparing an individual's voice with a number of previously recorded samples of speech, in an attempt to ascertain which, if any, it closely resembles.

Speech recognition does not attempt to give any information as to the identity of the speaker, but instead attempts to determine what they are saying.

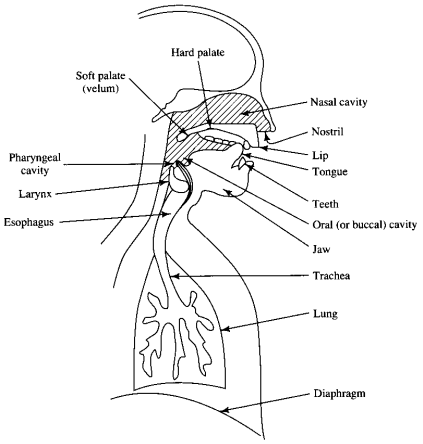
These technologies converge depending on the application, where often speech recognition is employed in conjunction with identification and authentication, hence the confusion.

Speech [4,5,6,15]:

Now that we are aware of the technologies voice authentication employs, we need to look at how the voice is produced, the characteristics that allow us to extract meaning from it, and the method by which it can be converted into a form that can be handled by computer systems. In doing this particular attention will be paid to those characteristics of the voice that render it unique for each individual, therefore allowing their identification. To do this, we can examine the physiological component of human speech, which is produced by the human voice tract.

In simple terms, the voice is created by air passing over the larynx or other parts of the vocal tract. The larynx vibrates creating an acoustic wave, essentially a hum, which is modified by the motion of the palate, tongue and lips. Sounds produced by the larynx are called *phonated* or *voiced* sounds. Examples of voiced sounds would be the *m* in "mud" or the *r* in "ram". Simultaneously, other sounds are produced by other parts of the vocal tract, for instance whispered sounds are created by air rushing over the arytenoids cartilage at the back of the throat. Sounds not originating in the larynx are called *unvoiced* sounds. Examples of these would be the *f* in "fish" or the *s* in "sea".[7]

Figure 1: The vocal tract[8].



All sounds produced are, at the same time, fundamentally influenced by the actual shape of the vocal tract. This shape is brought about both as a consequence of hereditary and developmental factors, and of environmental factors.

In parallel to these physiological characteristics, speech contains a behavioural component. This manifests itself as the accent of the voice, and affects how quickly words are spoken, how sounds are pronounced and emphasized, and what other mannerisms are applied to speech.

Together these physiological and behavioural factors combine to produce voice patterns that are essentially unique for every individual, and are difficult or impossible to duplicate short of recording the voice.

When analysed using modern technology, human speech appears to be rather inefficient, in terms of time and energy expended to transfer information. Speech is constructed out of various sounds, termed *phonemes*. Common English usage utilises around 40 phonemes, analogous to the characters of the phonetic alphabet seen in most dictionaries. For instance the word "mud" uses three phonemes denoted /m/ (the mmm sound), /u/ (the uh), and /d/.

Almost all the information in each phoneme can, in fact, be deduced from only a small fraction of the entire phoneme sound. For example, the *n* sound in the word "man" may take one or two tenths of a second to say. Yet, for analytical purposes, only the first 20 or 30 milliseconds and the last 10 or 20 milliseconds of the sound is

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.