EFFECTS OF EQUIPMENT VARIATIONS ON SPEAKER RECOGNITION ERROR RATES

By

CLARK D. SHAVER

Master of Science in Electrical Engineering

Oklahoma State University

Stillwater, Oklahoma

2009

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
MASTER OF SCIENCE
December, 2009

EFFECTS OF EQUIPMENT VARIATIONS ON SPEAKER RECOGNITION ERROR RATES

Dr. John M. Acken
Thesis Adviser

Dr. Stephen S. Bell

Dr. Nazanin Rahnavard

Dr. A. Gordon Emslie
Dean of the Graduate College

ACKNOWLEDGMENTS

I would like to thank Dr. John Acken for his assistance, guidance and encouragement. Dr. Acken's insight and genuine excitement for this project has been motivating and has made this experience enjoyable. He is both a great mentor and a great friend. Also I would like to thank Dr. Stephen Bell and Dr. Nazanin Rahnavard for agreeing to sit on my committee.

I would like to thank my parents, Ronnie and Carolyn, who throughout my youth repeated *ad nauseam* with words and action, the phrase "*You can do anything you want to in this life, if you put your mind to it.*" Their influence continues to be a great factor in my life today. I would like to thank my brother, Craig. Our conversations always aspire to lofty goals. I am also indebted to my children, Quenlyn, Audrie, Craig, and Rosabel who have given me great motivation for finishing this thesis in a timely fashion.

Most of all I would like to thank my patient and loving wife, Jaimie. She has been nothing but supportive throughout my education. For many semesters she has served as both mother and father to our small children. She has fallen asleep in an empty bed many nights. To her great credit she has done all of this without a single word of complaint. Her great character is revealed not only by an absence of complaints in conversation, but she has been my greatest source of encouragement and support throughout my education. To her I dedicate this work.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER I

INTRODUCTION

The security of a system is dependent upon the security of each of the components. Subsystems for a distributed system include point of access, communication link, authorizing agent, and delivery mechanism. These subsystems are shown in Figure 1.1. In a distributed system, such as the Internet, bank ATMs, or credit card gas pumps, there is limited control over the point of access. Independent of access, the communication link must be secured. Therefore, the most commonly recognized component of a secure system is the encryption algorithm. The most commonly recognized problem in a secure communication system is encryption key management.

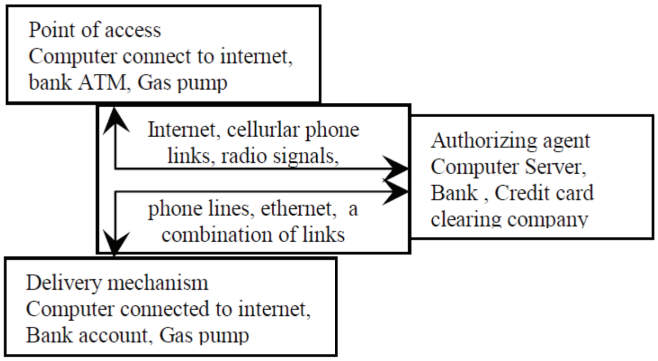

Point of access
Computer connect to internet, bank ATM, Gas pump
Internet, cellurlar phone links, radio signals,
phone lines, ethernet, a combination of links
Authorizing agent
Computer Server, Bank , Credit card clearing company
Delivery mechanism
Computer connected to internet, Bank account, Gas pump

Figure 1.1: Subsystems of a distributed, secure system

An alternative to encryption for distributed systems is to fragment the data so that interception of pieces is useless. Another alternative is steganography, which is hiding the message within other information [1]. Other components and issues for security systems are: hashing (for checking data integrity), identity authentication (for allowing access), electronic signatures (for preventing revocation of legitimate transactions), information labeling (for tracing location and times of transactions), and monitors (for identifying potential attacks on the system). Each of the components affects the overall security of the system. The weakest component limits the system's overall strength of security. There is a difficulty with automatic identity authentication in distributed systems. The difficulty is in part due to the human interaction with the system. Divulged passwords and stolen or lost credit cards present a human aspect that is difficult to manage. Identity authentication systems can decrease susceptibility to a security breach by adding extra dimensions or elements to the authentication of one's identity [2, 3]. One clear example of increasing strength by utilizing multiple elements of authentication is a bank ATM system. At the ATM, two forms of authentication is required, a debit card and a four digit pin number. By itself, a four digit pin number is a very susceptible security measure, but when used in a multimodal authentication system, it adds a significant level of security. Adding additional elements of identity authentication to a distributed system adds to the strength of security of the system.

In distributed environments, system developers often have little or no control over the point of access equipment utilized. The same information or system can be accessed by multiple terminals. For instance, one may gain access to the same privileged information on the internet via numerous terminals, i.e. cell phones, PDA, laptop, home, school and work PC's, etc. As distributed systems become more complex, protecting them also has also added layers of complexity. One general and widely used method of protecting distributed systems is by automatic identity authentication [4]. Identity authentication systems determine whether an individual has been properly authorized to access a system. There are three main elements in the methods of identity authentication; what you know (example: password or login), what you have (example: debit card or key) and what you are (biometrics) [3, 6-7]. What you know requires a user to input data into a system to confirm that the user provided data matches previously supplied data. The hardware used for authentication in a 'what you know' system is only required to gather the data, how that data is gathered is not a concern for the system. An example is in a login situation, where it does not matter if the data is transferred from user to system via keypad or a 'speech recognition / conversion' system. As long as the data is submitted correctly, the system can authorize access for that user. When utilizing the means of 'what you have' to authenticate one's identity, the system is completely hardware dependent. It is hardware dependent because the item 'you have' is hardware and the device that authenticates the item' you have' also is hardware. A 'what you have' system is a key-receptacle system. To be authenticated, the key and the receptacle must match. One cannot enter into their hotel room by swiping a credit card into the door key reader. One advantage of this system is that the system designer has control of both the key and the key reader. They are designed to work together.

Biometrics is different. Biometrics is the actual measure of what you are or what you do [5]. Instead of the user inputting data for comparison or furnishing an object for purposes of authentication, biometric authentication is the actual measure of a feature of the user. The measure of that feature is compared to an earlier measurement of the same feature of the user to authenticate one's identity. In the case of biometrics, the measurement devices may have an effect on the actual measurement. For instance, a simple biometric would be to measure one’s height. Assume that the height measurement is recorded correctly when the person was being enrolled as an authorized user. At a later date the user then attempts to gain access to the system. If the measurement device is one half inch higher than when at first, the user could be falsely rejected from gaining system access. In many distributed environment, the potential for measurement devices to vary is great. Measurement variation can have a significant detrimental effect on an identity authentication system.

This investigation is concerned with one particular biometric, the biometric of voice. Speaker recognition systems use one’s voice as a metric to detect a specific speaker [7]. For speaker recognition systems in a distributed environment, such as the internet, microphones are certain to vary. Frequency response to various microphones can vary widely. Two different microphones can produce two dissimilar signals for the exact same recording. In a speaker recognition system microphone dissimilarity may lead to, 1) a significant enough dissimilarity to cause the system to fail to recognize the speaker, or 2) a dissimilarity not significant enough to affect the system’s ability to recognize the speaker. The opposite is true for imposter speakers as well. Microphone effects may be significant / insignificant enough to alter/not alter the imposter rejection capability of the speaker recognition system. To discover whether or not the effects of varying microphones has a significant detrimental effect on the ability of a speaker recognition system perform identity authentication is the objective of this research. To accomplish this objective, voice samples from a group of people, spoken into a set of digital recording systems were submitted to a speaker recognition system and the error rates of each system were analyzed and compared.

CHAPTER II

BACKGROUND

Section 1 – Identity Authentication

With an ever growing networked-world, where a large amount of sensitive data is digitized and security is in high demand, identity authentication has come to play a vital role in security. Multimodal systems have been given more credence to increase a system's security [8-10]. A recent trip to a popular amusement park in south Texas affords a good example of a multimodal identity authentication system. Upon your first gate entrance to the park, your ticket is presented with your name on it, identity is verified via a driver's license or other accepted identification document and a thumbprint is scanned to enroll you into the amusement park database. Upon return trips, a thumbprint is scanned, and the ticket presented. By requiring both an item that you have and a verification of what you are, a significant increase in security is generated. The amusement park has utilized two of the three main elements of identity authentication. The elements of authentication are what you have, what you know and what you are [3, 6-7]. Many internet-based authentication systems only require one of the three elements, what you know. A typical web-based security application may require a username and password to gain access to certain information. Though the application may require two separate sets of information, it is still only requiring one of the three elements, what you know. Requiring two sets of information can be insufficient as the author's recent personal experience on an auction website has demonstrated. Others can, by various means, learn what you know. A more secure system is a typical bank ATM system. Here one is required to present a physical debit card, what you have, and a four digit numerical pin, what you know. Though a four digit pin number is a weak security measure, it adds significant strength to the overall authentication system when a physical card is required. The card increases security as one must learn the "what you know" and obtain the "what you have" in order to acquire access to the account. The addition of a third element would secure access to a system even further. By adding extra elements to an authentication system, one adds a significant degree of complexity to potential intruders. In order for a system to be considered level 3 according to NIST document 800-63 at least 2 of the three elements must be utilized in the authentication system [2]. One may improve password strength by increasing password lengths or by adding a secondary password [11]. By adding another authentication element to an authentication system, an even greater improvement in system strength can be realized (see Figure 2.1). The general increase in authentication system strength can be represented by the equation,

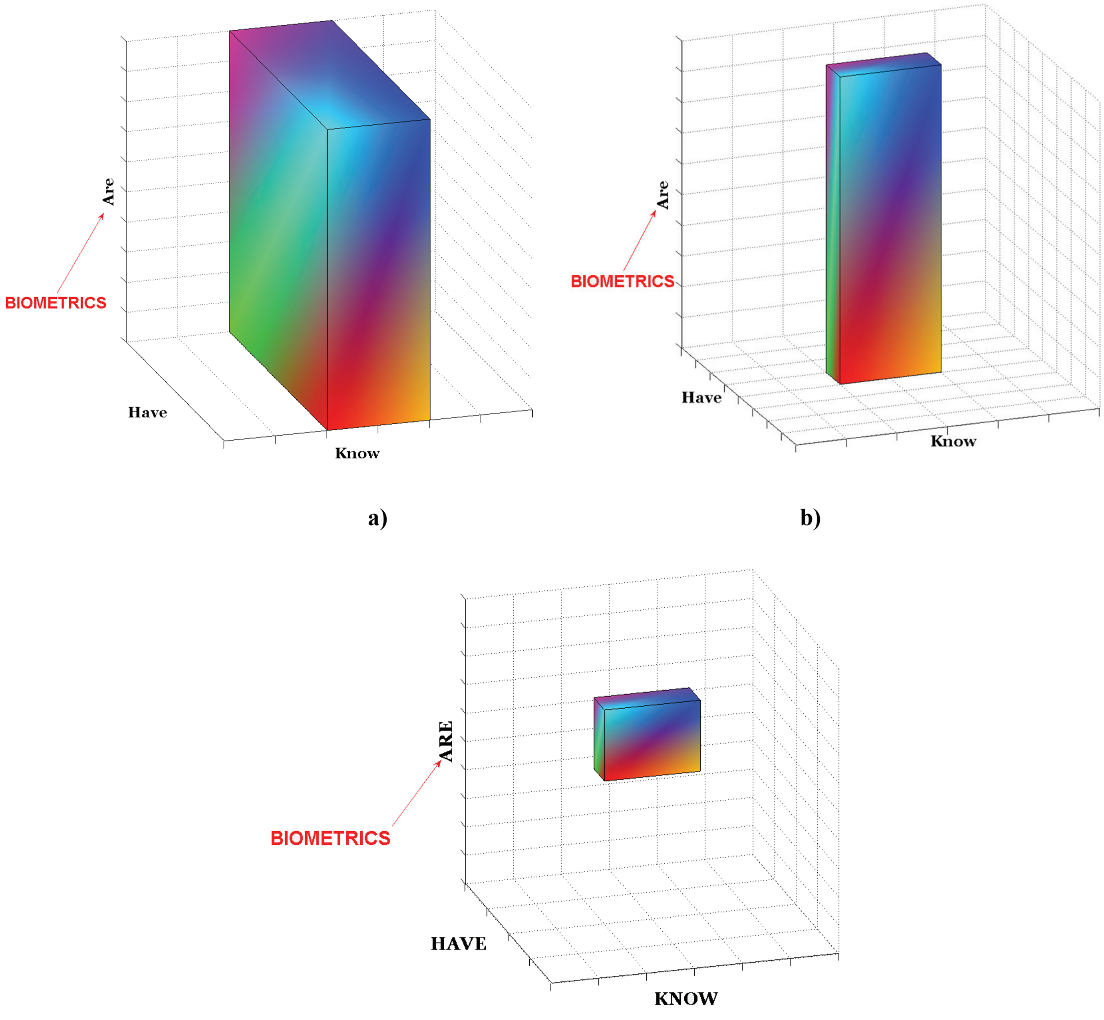
$$S_{total} = S_h \cdot S_a \cdot S_k \ , \tag{2.1}$$

And,

$$S_i < 1 \, , \tag{2.2}$$

Where,

S_h = Susceptibility to system security breach: what you have
S_k = Susceptibility to system security breach: what you know
S_a = Susceptibility to system security breach: what you are



a)

b)

c)

Figure 2.1: Illustration of susceptibility with 1-3 elements of authentication. a) represents a one element system, b) represents a two element system and c) represents a three element system.

Each axis in Figure 2.1 correlates to one of the three elements of identity authentication (KNOW, HAVE, and ARE). When only one element is utilized, the other two elements are 100% susceptible, because they are not utilized. By adding additional elements, the overall volume, which is equivalent to the system's susceptibility, is reduced. The maximum susceptibility, or maximum volume in Figure 2.1, is then $S_{max} = 1$. If one of the elements is impenetrable the susceptibility of the element and subsequently the system, is $S_i = S_{total} = 0$. If one of the three systems is non-existent then it is equivalent to a completely susceptible breach of security, or $S_i=1$. A system that requires only a password may increase security of the system by adding additional character requirements. In Figure 2.1a adding password characters is represented by a one dimensional reduction, specifically a reduction in the S_k dimension. By adding a physical token requirement, a second element is reduced, as is illustrated in Figure 2.1b. Even a relatively poor secondary element generates significantly less susceptibility. To equally reduce the susceptibility of the 'password only' system more and more characters are needed. A longer password is harder to guess or crack. "Cracking" a password can be done with software that repeatedly guesses at a password and keeps trying until access is granted. A long password takes a long time to guess or crack, reducing susceptibility. However, even valid users can forget or mistype long passwords. When a password exceed a person's ability to remember it, the person takes shortcuts. Consider that no amount of additional characters will increase the system security when a person writes their long password on a post-it note next to their terminal.

By adding a secondary or tertiary means of authentication, even a substandard means, system security is increased. For instance a banking system that requires a couple of items of knowledge can enhance their security by adding a required USB token that must be connected to your computer prior to account access. Such devices limit intruders to those with physical access to the token. What can be known is information. By comparing submitted information to expected and/or stored information, identity authentication can be accomplished. What one has is a physical device. Identity authentication is accomplished in a 'what you have' system by comparing a user possession, a physical device, to another physical device. Often a ‘what you have’ system is a key-receptacle type system. And “what one is” are their physical characteristics. An intrinsic property of one’s physical characteristics is the difficulty in transferring those characteristics to another. Information may be divulged, or a physical device may transfer hands, but as a rule it is much more difficult to transmit one’s attributes and/or features to another. Circumvention may be considered the cost to trick or falsify a system, as in the cost of guessing an x-character password. The difficulty of circumvention of a biometric is generally greater than that of the other two authentication elements. For that reason, adding the biometric element of authentication to a secure system generates a clear benefit.

Section 2 - Biometrics

Biometrics is a measure of what a person is or what a person does (produces). The nature of biometrics makes it generally the least vulnerable to intentional falsification of the three authentication elements. One may lose a credit card, or divulge a pin number, but it is significantly more difficult to give away what you are. Attributes can be mimicked. It is not impossible to lift a fingerprint, or replace your DNA sample with that of others. As a general assumption it would require a significant increase in effort to ‘fake’ what one is, as compared to the other two authentication elements.

There are two general types of biometric systems, static and dynamic. A static or physiological system measures purely what you are, such as a retina scan or a fingerprint. A dynamic, or behavioral, biometric measures your actions, such as facial expressions, signatures, behavioral patterns or voice generation [5, 7]. Because of the requirement of an action in dynamic biometrics, intrapersonal changes in an individual or changes in an environment play a role. A signature of an individual is never exactly the same and over time may evolve in its primary, measurable attributes. How to deal with the problem of intrapersonal variability is an issue and topic of research in dynamic biometrics, including speaker authentication [12-13]. Intrapersonal variability is not a significant issue in the relatively stable 'static' biometrics such as retina patterns or fingerprints, which for most people remain substantially constant throughout the majority of life. A user or set of users' acceptability of a method may limit certain static biometrics. For instance, in internet applications that require data to be digitized and sent over the net, a fingerprint or DNA data may not be a comfortable fit with some users. A third parameter may be access to technology. DNA analysis or fingerprint reading technology may not be wide spread. In an internet application a dynamic sample, such as a handwriting sample or voice sample, can be considered more acceptable to the user [14]. One advantage of voice as the biometric as opposed to signatures of thumbprints is the availability of the technology. In many applications, the sole mode of system access and/or identity authentication for remote users is speech and it is often not considered intrusive [15]. One example would be a telephone banking system. Speaker recognition is generally an acceptable, low cost, widely available technology.

There is another broad division in biometrics: authentication (verification) versus identification [15-19]. Identification asks "who is he?" where authentication asks, "Is he who he says he is?" The task of authentication is a much simpler matter as compared to identification. It is a closed-set versus an open-set problem. Generally any system used for identification could be utilized in an authentication application. The same cannot be said for authentication systems in identification applications. Authentication is utilized for secure access in distributed systems.

Section 3 – Basics of Speaker Recognition

The specific biometric of interest in this thesis, is voice as measured by a speaker recognition systems. Speaker recognition systems can be partitioned into one of two groups, text-dependent and text-independent [15, 19-20]. A text-dependent system is one in which the phrase or phrases that one speaks during enrollment are the same phrase or phrases as used when requesting authentication for system access. These systems have an advantage in accuracy due to common word usage, pronunciation, prosody (rhythm and emphasis) and phone usage singularities in ones speech [7, 21]. Simple pattern matching algorithms are used with some text-dependent systems to verify the proper person is saying the proper phrase. The basis of what is being measured in text-independent systems fundamentally differs from text-dependent systems. Text-dependent systems attempt recognition by identifying how a user says a specific phrase. Text-independent systems use fundamental voice data buried in voice signals to do speaker recognition. Because text-independent systems analyze basic voice information and not how a particular user says a particular phrase, the user is not required to speak any certain word or phrase.

There are several voice attributes that can be analyzed to verify identity. These attributes can be divided into two basic groups, low level and high level information. The low-level information, uses small time segments of the voice signals and analyzes the basic structure of one's voice, i.e. signal spectrum, tone, frequency, etc [22]. Recent research has also shown the viability of high level information used in conjunction with the more classic low level systems [22-27]. Some examples of high level information in speech include accent, pronunciation, often used words or phrases. High-level data is beginning to have a significant role in speaker recognition systems.

Figure 2.2 illustrates the classification of a low-level, text-independent speaker recognition system within the framework of identity authentication systems. Each box represents a possible classification at each step. The solid lines represent the decision path used to decide upon a low-level, text-independent, speaker authentication system. The bracketed items represent favorable attributes sought after in a biometric system

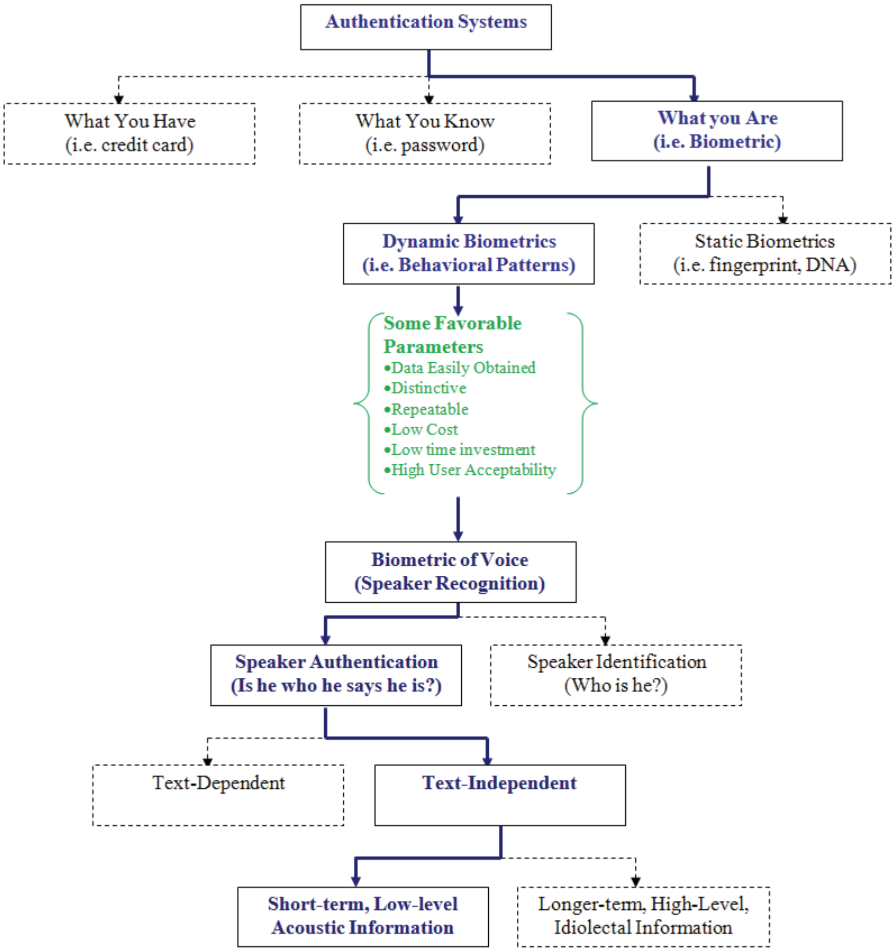


Figure 2.2: Classification of a short-term, text-independent, speaker authentication system

in a distributed environment. The first decision to be made is which combination of the three elements of identity authentication will be used in the system. If the combination of authentication elements includes biometrics, then one must decide which type of system to use, dynamic or static. The decision of which biometric to use is likely to be influenced by the partial list of parameters found in the brackets in Figure 2.2. A more complete list is found in Table 2.1. The characteristics used to evaluate and compare different biometrics include: cost, time, universality, distinctiveness, permanence, collectability, acceptability, circumvention, accuracy, repeatability, storage requirements, and availability of technology [14].

Table 2.1: Evaluation of Characteristics for Biometrics

Characteristic	DNA	fingerprints	Hand dimension	Height	Voice	Weight	Eye Iris Scan	Face dimensions
Cost	☹			++	++	++		
Time	☹			++	++	++		
Universality	++	++		++	++	++		
Distinctiveness	++	++				-		
Permanence	++	++				-		
Collectability	☹	++			++	++	☹	
Acceptability	☹	-	+	++	++	M	☹	M
Circumvention	++					☹		
Accuracy	++					☹		
Repeatability	++					☹		
Storage Requirements	☹	☹	++	++		++		++
Availability of Technology	☹	+	M	+	++	++	-	M

++ = Great (or cost is low, time is short; hard to circumvent) ; + = Good; M = Medium; - = Bad; ☹=- Terrible (cost is high); Blanks = no information

The cost parameter includes the money, time, equipment and expertise for the implementation of the system and the collection of the measurements. The time characteristic is specific to the measurement collection and analysis time. That is, the time from when an identity authentication request is made until the access is granted or denied. Universality is a measure of the portion of the sample population that are able to meet the requirements of the systems. For example, everyone has DNA, but not everyone has hair. So a DNA test is universally applicable, while hair color is not applicable to people without hair. Distinctiveness is a measure of how unique or different the measurements for an individual will be from other individuals. Finger prints are very distinct whereas weight is not. Permanence is a measure of intrapersonal variations, the change in the biometric with the passage of time. Collectability is the characteristic indicating how much effort is required to obtain samples for the biometric. Acceptability is a subjective measure of how willing a person is to submit to the biometric measurement. Most of us would not submit to a blood test just to enter a gas station. On the other hand, we readily submit to height measurements for carnival rides at the state fair. Circumvention is the ease or cost to trick or falsify the measurement. Measuring weight is easy to falsify by carrying lead in one's pockets. A falsified eye scan is a bit more difficult. The accuracy of a biometric is the probability that an individual will be properly authenticated. Specifically, it includes the probability of properly authenticating the identity or access for authorized individuals and properly rejecting the identity or access for unauthorized individuals. Repeatability is the variance of the biometric measurement over repeated trials. The data storage requirement is evaluated both for the individual measurement as well as the total database of each individual measurement. The availability of the technology is a make or break decision as well as a quantitative measure. A biometric is not an option for immediate deployment if it requires a technology that does not currently exists. However, even if the technology exists, the ready availability of the technology is a factor. For example, many computers and recording devices have the ability to capture a voice or a picture, but not many people have ready access to DNA or fingerprint collection devices. The selection of a biometric based upon these characteristics clearly involves many tradeoffs. How one weighs each of these decision factors, is a function of the application. For many applications, as indicated in Table 2.1, speaker recognition is the best candidate.

As stated previously, speaker recognition systems can also be divided by their specific objective: identification or authentication. In speaker identification the system identifies who a person is out of some set which may include all human beings. The system asks ‘who is he/she?’ In speaker authentication (or verification) a person’s identity is checked against a claimed identity. In authentication, a system asks “Is he who he says he is?” Speaker recognition systems, as well as all identity authentication systems, have two basic phases, enrollment and testing [20, 28]. In the enrollment phase, users train a system by providing an initial voice sample. The ‘training’ or enrollment sample is compared to the later samples submitted for authentication. The purpose of the enrollment phase of an identity authentication system is to generate a standard for the individual, which he/she will be measured against in the testing phase. In the speaker recognition system, a standard is generated by modeling a person’s voice. That model will later be used to check that an individual’s voice is the voice of an authorized individual. Models of a system provide an efficient method for comparison. During the authentication or testing phase, a basic speaker recognition system collects the analog voice signal, converts it to an analog electrical signal and then digitizes the signal. From the digital signal, some feature(s) of the voice signal is (are) extracted and measured. When the enrollment utterance is provided, a statistical model is generated. The model is later compared to features from the 'test' samples. After the two are compared a decision must be made if the person requesting authentication matched or not. The main steps of the described system are outlined in Figure 2.3. Additional processing enhancement steps can often be found in speaker recognition systems, such as filtering and score normalization.

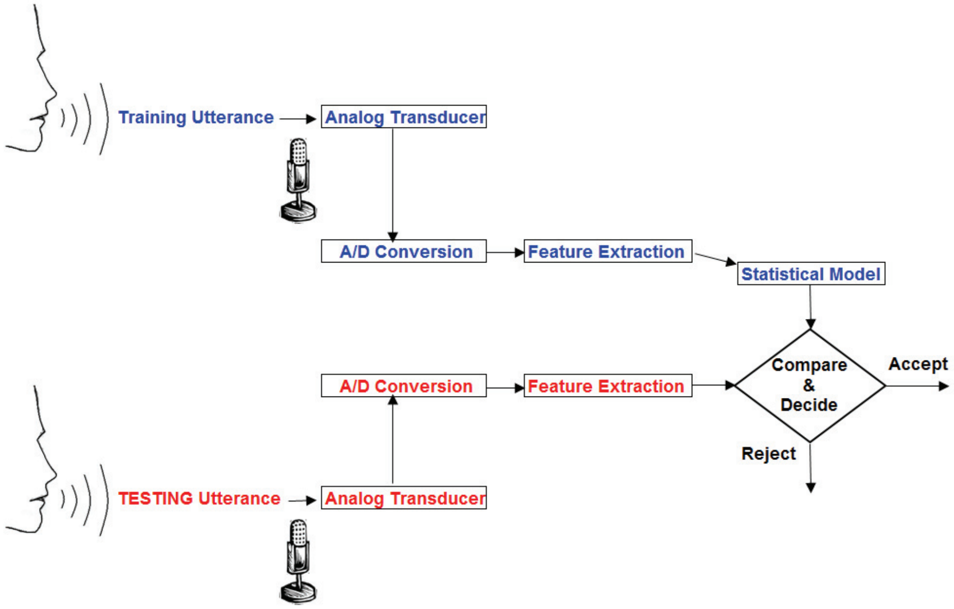


Figure 2.3: Overview of main components in a speaker recognition system

Voice is produced by air being pushed up from the lungs through the glottal folds (vocal folds) and then through the vocal tract and eventually out of the speakers mouth. The vocal folds produce a base sound that is manipulated into specific phonetic events by the vocal tract [29, 30]. Lip radiation, mouth geometry and other biological functions also play minor roles in voice production. For simplicity, these will be lumped together with the vocal tract in the following discussion. In voiced speech, vocal folds contract and relax creating a source of sound [30- 33]. The vocal fold sound is modified by the vocal tract to create specific noises such as vowels, consonants, etc. Mathematically speech can be modeled as a source-filter system [30, 31]. The air from the lungs being pushed through the vocal folds would be the source. The vocal tract would act as a filter. The source (vocal fold) and filter (vocal tract) would be convoluted together to generate the final voice signal. The speech signal as a convolution of the two signals is illustrated in Figure 2.4. The sound from the vocal folds is one of several features of a voice that humans use to identify an individual just by hearing one's speech. Other means, such as the high level features spoken of, are used for 'identification by ear' as well. One method of performing automatic, text-independent, speaker recognition is to take advantage of the identifying properties of the vocal fold signal. How can the vocal fold sound be analyzed independent of the vocal tract? One answer is, by deconvolution. Deconvolution can be used because speech is a convolution of the vocal folds and vocal tract. One method of deconvolution is by cepstral analysis. The cepstrum fundamentally is the spectrum of the log of a spectrum, or alternatively, the cepstrum [34]. A cepstrum is a technique used for deconvolution of a signal. The cepstrum is the inverse Fourier

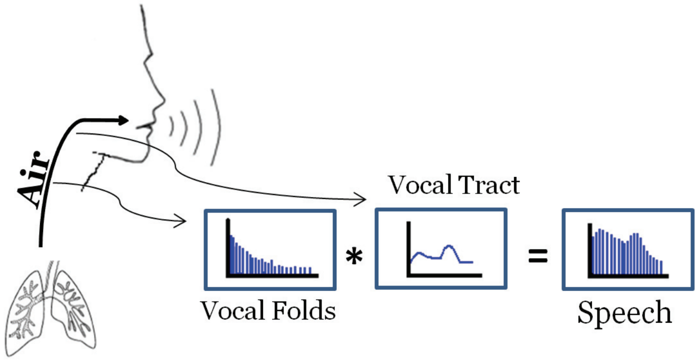


Figure 2.4: Speech model as a source-filter convolution

transform of the log-magnitude Fourier transform of the signal (see Equation (2.3)) [35]. The product property of logarithmic functions allows the spectrum of the voice spectrum to be mathematically separated into the log magnitudes of the vocal fold and vocal tract signals. Because of the relative difference in quefrencies (frequencies in the cepstrum domain) of the vocal folds and vocal tracts, separating these signals can be accomplished with a simple lifter [36]. A lifter is a filter in the cepstral domain [34]. One common method to apply the lifter is by passing the log power spectrum of the signal through a filterbank [20]. Common speaker recognition systems space filters in the filter bank on a mel-spaced frequency scale, which closely resembles the auditory scale of the human ear.

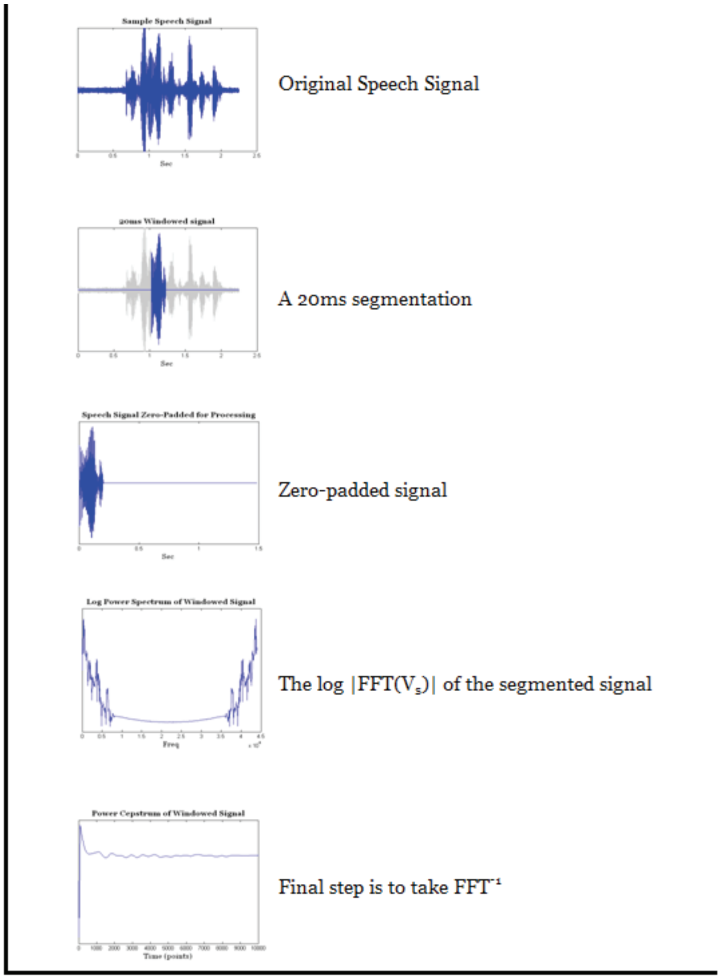
$$C = \frac{1}{2\pi}\int_{-\pi}^{\pi}\log\left|\int_{-\pi}^{\pi} x\cdot e^{-j\omega\cdot t}dt\right|\cdot e^{j\omega n}d\omega = \Im^{-1}\left(\log\left|\Im(x)\right|\right) \tag{2.3}$$

The mel-spaced scale emphasizes the lower frequencies while attenuating some of the upper frequencies [16]. Passing a speech sample through the mel-scale filterbank allows for the general isolation of the excitation signal of the vocal folds. The pseudo-code in Figure 2.5 illustrates the process for deconvolution of the two voice signals via cepstral analysis.

```
Vocal_Tract*Vocal_Fold = $V_f * V_t$ = Voice Signal          :Start with Voice Sample
$\log|\mathcal{F}(V_f * V_t)|$                                  :Take Log of FourierTransform(FT)
$\log|\mathcal{F}(V_f)\cdot\mathcal{F}(V_t)|$                    :FT of each signal portion
$\log|\mathcal{F}(V_f)| + \log|\mathcal{F}(V_t)|$                :Additive properties of logrithm
Filterbank($\log|\mathcal{F}(V_f)| + \log|\mathcal{F}(V_t)|$)    :filterbank isolates vocal fold signal
Mel-Cepstrum = $\mathcal{F}^{-1}$ [MelFilter ($\log|\mathcal{F}(V_f)| + \log|\mathcal{F}(V_t)|$)]    :Mel-Cepstrum of Voice
Mel-Cepstrum $\approx \mathcal{F}^{-1}$ [MelFilter ($\log|\mathcal{F}(V_f)|$ )]   : Approximates Mel-Cepst. of vocal folds
```

Figure 2.5: Pseudo-code demonstrating the cepstrum deconvolution process

In the implemented algorithm, the Inverse Discrete Fourier Transform (IDFT) of the filtered log-spectrum is taken [20, 37-39]. The IDFT gives N number of cepstral coefficients on the Mel-scale, called Mel-Frequency Cepstral Coefficients (MFCC) [35]. The MFCC’s of the entire utterance broken up into 20ms segments of speech are obtained. In the enrollment phase, the MFCC’s are the features that are obtained that will later act as a measure of one’s voice, and thus one’s ident*i*ty.



Obtaining a cepstrum from a voice sample

Figure 2.6: Visual step-by-step of the short-term cepstrum of a voice signal

The purpose of the enrollment phase of an identity authentication system is to generate a standard for the individual, which he/she will be measured against in the testing phase. In the speaker recognition system, standard generation is done by modeling a person's voice. That model will later be used to measure the identity of an individual's voice. What are being modeled in the speaker recognition system are the MFCC vectors $(\vec{x}_i)$. The feature vectors are modeled using a tool called a Gaussian Mixture Model (GMM) [40]. A GMM is the combination of D-Variate Gaussians added piece-wise. The component probability densities are given by Equation (2.4). The GMM itself is a sum of the weighted densities, show in Equation (2.5) [40, 41].

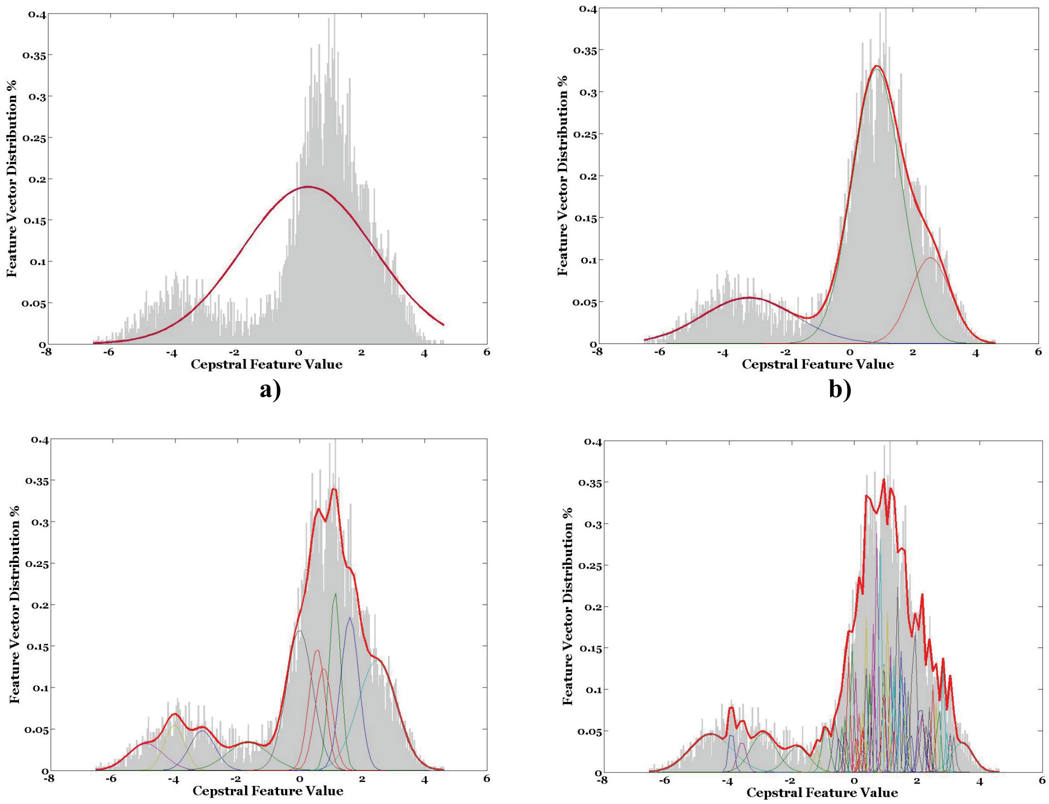
$$f_{x_i}(\vec{x}) = \frac{1}{(2\pi)^{D/2}|K|^{1/2}} e^{\left\{-\frac{1}{2}(\vec{x}-\vec{\mu})^T K^{-1}(\vec{x}-\vec{\mu})\right\}} \tag{2.4}$$

$$GMM = \sum_{i=1}^{D}\left\{c_i \cdot f_{x_i}(\vec{x})\right\} \tag{2.5}$$

N-number of MFCC's are taken every 20 milliseconds. For just a few seconds of sample speech, the amount of data for just one vector can occupy several thousand words of memory. A specific system will have a number (N) of feature vectors, containing L_X/.002s length data and modeled by D number of Gaussians, where L_X is the length of the voice signal. Other methods of metric creation were utilized prior to the application of GMM's to speaker recognition. Earlier methods include Hidden Markov Models

(HMM) and vector quantization (VQ) [15, 42]. Both HMM and VQ have proven to be more computational intensive with no, or only modest error rate improvements [21].

The selection of the number of Gaussians (D) has an effect on performance [40]. A uni- or bi-variate Gaussian mixture is not likely to describe a feature's distribution very well. On the other end of the spectrum, as the number of Gaussians increase, the amount of information about the signal that each adds will decrease. In fact, too detailed a model which contains information about background noise, or environmental acoustics can be detrimental to error rates [40]. Figure 2.7 shows the resulting GMM of the same speech signal feature distribution for various values of D. As the number of Gaussians is increased, the model matches ever more closely to the actual feature distribution. Figure 2.7a uses a single variate Gaussian, giving a very loose approximation. In Figure 2.7b the 3-variate GMM models the voice feature distribution's basic contour well. The 3-variate GMM represents a significant increase in accuracy over the 1-variate system. Figure 2.7c, a 10-variate GMM also models this feature contour well. Additionally the 10-variate system picks up some singularities that could potentially distinguish an authenticated user from an imposter. Figure 2.7d adds even more detail. Figure 2.7d shows the voice feature as modeled by a 64-variate GMM. The 64-variate model picks up some of the same singularities as the 10-variate system. It also models minor idiosyncrasies in the feature distribution, which are most likely singular to the particular environment where the sample was taken or to the particular phrase that was spoken.



a)

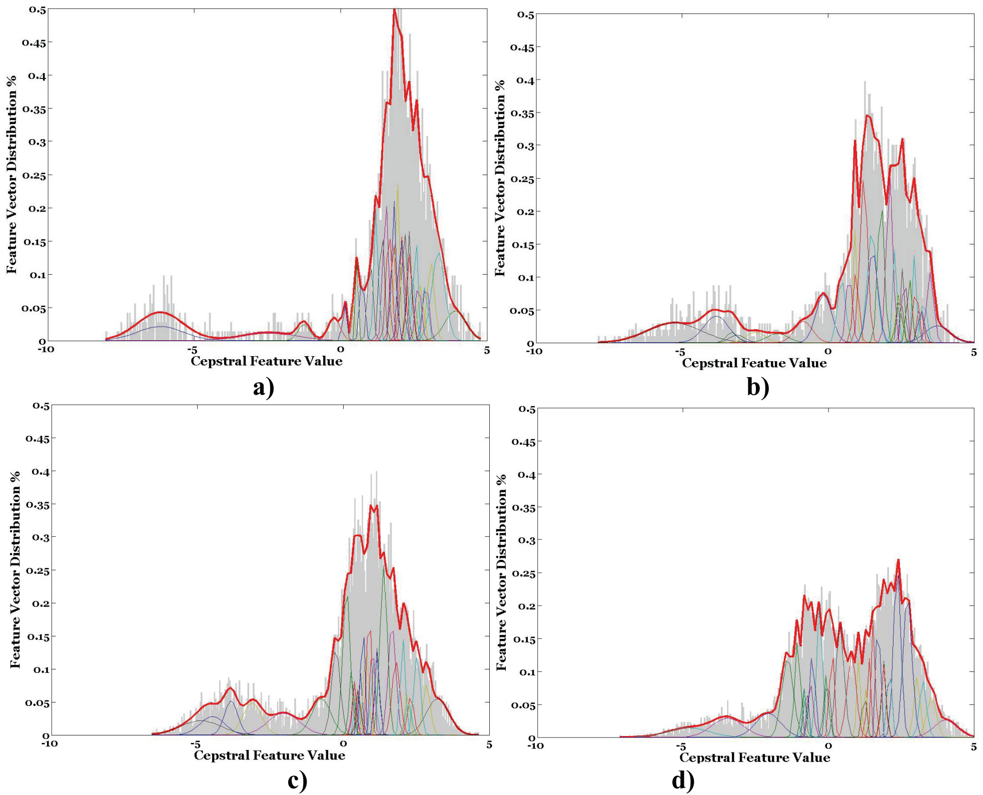
b)

c)

d)

Figure 2.7: a) 1-variate GMM, b) two-variate GMM, c) 10-variate GMM, d) 64-variate GMM

Figure 2.8 illustrates differences in voice feature distributions. Each of the four graphs represents the 2nd MFCC vector of sample utterances. Figures 2.8a, 2.8b and 2.8c are three different voice samples from the same speaker. Figure 2.8d is from a different speaker. Figures 2.8a and 2.8b came from the same microphone. The distributions in Figure 2.8a – 2.8c are similar in shape, but contain significant differences in detail. Because of the differences in detail, there is a limit to the efficacy of adding Gaussians to a GMM model. Research has shown that minimal error rate improvement is realized by adding more than about 32 component Gaussians to a GMM [40]. In fact, adding too many component Gaussians can have a detrimental effect on error rates [40].



a) b) c) d)

Figure 2.8: 2nd cepstral coefficients GMM. a)-c) are the same speaker saying different phrases. d) is a different speaker on system #2 saying same phrase as c).

With the enrollment model in the system, a voice sample being tested for authentication can then be compared and scored against the model of the enrollment speech signal. The task of authentication is to determine if the speaker is who he/she claims to be. Basically, the task is a hypothesis test. The hypothesis is 'the speech sample Y_0 is from the modeled speaker Y_M' [43]. The hypothesis test can produce one of 4 results [3, 14, 44]. A true accept (TA) occurs when the system correctly authenticates an authorized individual. A true reject (TR) occurs when the system correctly rejects an unauthorized individual. Error types I & II can also occur from the hypothesis. A Type I error occurs when the authorized individual is falsely rejected (FR). A Type II error occurs when an unauthorized individual is falsely accepted (FA). The four possible results are listed in Table 2.2 [14].

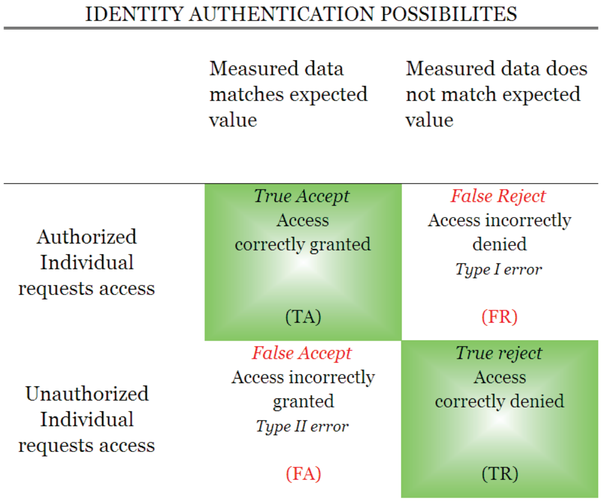
There is a tradeoff in the FR and FA rates. NIST provides a detection cost model for measurement of speaker detection performance. It is given by Equation (2.6) [45]:

$$C_{Det} = C_{FR} \times P_{(FR|Y_0 = Y_M)} \times P_{(Y_M)} + C_{FA} \times P_{(FA|Y_0 \neq Y_M)} \times \left(1 - P_{(Y_M)}\right) \tag{2.6}$$

In Bayesian decision theory, an optimal decision is found at the minimum of Equation (2.6) [46]. C_{FR} and C_{FA} are the costs of a FR and FA respectively. $P_{(FR|Y_0 = Y_M)}$ and $P_{(FA|Y_0 \neq Y_M)}$ are, respectively, the probability of a FR given the real user and FA given an imposter. The *a priori* probability of the specified speaker Y$_M$ is $P_{(Y_M)}$. Minimizing the cost model equation can generate the Bayesian optimal decision rule [46]:

$$\frac{P(Y_0 \mid Y_M)}{P(Y_0 \mid Y_{\overline{M}})} \begin{cases} \geq \theta, & \text{Accept} \\ < \theta, & \text{Reject} \end{cases} \tag{2.7}$$

Table 2.2: Four possible results of identity authentication



IDENTITY AUTHENTICATION POSSIBILITES

	Measured data matches expected value	Measured data does not match expected value
Authorized Individual requests access	*True Accept* Access correctly granted (TA)	*False Reject* Access incorrectly denied *Type I error* (FR)
Unauthorized Individual requests access	*False Accept* Access incorrectly granted *Type II error* (FA)	*True reject* Access correctly denied (TR)

The numerator is simply λ_0, where λ_0 is the likelihood that Y_0 is from Y_M. The denominator is λ_1, where λ_1 is the likelihood that Y_0 is *not* from the modeled speaker Y_M. Thus the overall likelihood, λ, equals λ_0/λ_1 [20]. In this thesis, the log of the likelihood ratio (LR) is used because it is less computationally intensive. The likelihood of L_x observations compared to a single component of the mixture model is given by (2.8).

$$L(\mu)=\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^{L_X}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{L_X}(x_i-\mu)^2\right) \tag{2.8}$$

Taking the log of the likelihood gives:

$$\log[L(\mu)]=-\frac{L_X}{2}\cdot\log(2\pi\sigma^2)-\frac{1}{2\sigma^2}\sum_{i=1}^{L_X}(x_i-\mu)^2 \tag{2.9}$$

The model used to determine λ_0 is simply the Gaussian mixture model developed in the enrollment phase. The model used to determine λ_1 is not as clear. The difficulty is in modeling who a person is not. The difficulty in an open set application leaves an unbounded set of possible speakers, which can be difficult to model. The approach that has been used in much literature is the universal background model (UBM). The UBM is a voice feature model generated from a collection of other speakers [47].

The process of tuning the detection threshold (θ) is one of the more difficult tasks in designing a speaker recognition system. In part, the difficulty is due to the tradeoff between false accept and false reject rates and the need of the particular application. A system requiring high security may weight the cost of a false accept much greater than the cost of a false reject. Higher false reject rates with lower false accept rates may be an inconvenience to user's who are more often falsely rejected, but allows fewer unauthorized individuals from gaining access to the system [14]. The high security system described works great for a nuclear arms facility but may be less effective for a system that allows fast food workers access to the freezer, which requires regular and speedy entrances. To assist in threshold setting, a detection error tradeoff (DET) curve can be developed [48]. A DET curve plots the FR and FA error rates as a function of the threshold level. In this thesis, as λ approaches zero, the more likely the voice sample Y_0 is from speaker model Y_M. Therefore, when the threshold is a large negative number, many FA can be expected. As the threshold tightness is increased ($\theta \rightarrow 0$), less and less false accepts are expected and more false rejects would be expected. At $\theta = 0$, all users, including authorized ones would be rejected making the FR rate = 1. In Figure 2.9, when θ is -4, the FR rate is 100% while the FA rate is 0%. Increasing θ to -16 yields a FR rate

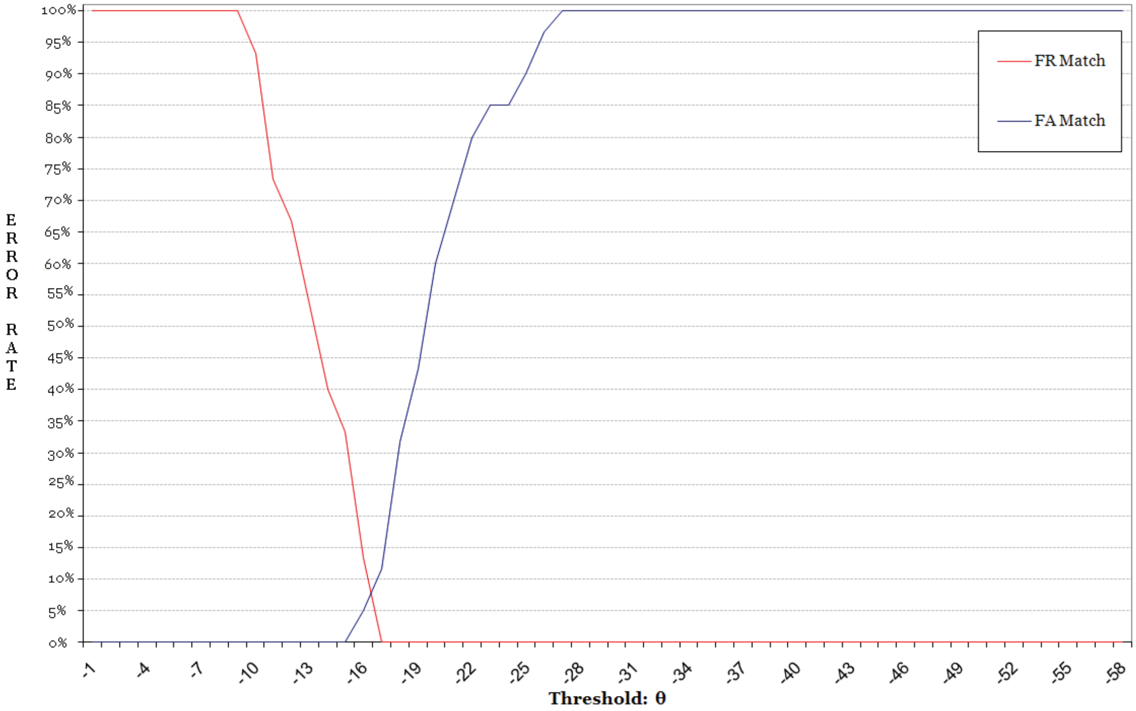


Figure 2.9: An example Detection Error Tradeoff (DET) curve

of 13.3% and a FA rate of 5%. The threshold was loosened, allowing a few people to be falsely accepted. The same threshold loosening reduces the amount of Type I errors (FR). With an even looser threshold of -23, the FA rate is up to 85% and the FR is at 0%.

In review, the task of speaker recognition can be split into two phases, enrollment and testing. The basic setup of a speaker recognition system includes speech signal collection, feature extraction, feature modeling, signal comparison and decision making. Other common tasks in speaker recognition systems include pre-emphasis to mimic the outer ear, pre-signal filtering to mitigate background noise, cepstral domain filtering for removal of static channel effects (such as cepstral mean subtraction and RASTA filtering), and score normalization for mitigation of intra-speaker variations, handset enrollment/testing mismatches and other environmental variations [49-54].

Section 4 – Historical Review of Speaker Recognition

The modern system described in Section 3 is an accumulation of advancements made over the last 50 years. Today's automatic speaker recognition systems verify user access rights, identifying personnel in a group, and they even have some use in forensic applications. Early research in speaker recognition was in the realm of human abilities. War time research in the 1940's allowed for significant advances, producing a tool to allow visual inspection of voice. Advances in signal processing techniques and the rise of the computer permitted true automated systems to be developed. The first automated system was created in the 1970's. From that point forward, the main thrust of research has been in independent speaker recognition. Today's speaker recognition research focuses on lowering error rates, capabilities in identification and creating systems robust in the presence of environmental variations.

The problem of recognizing an individual by their voice is an age old issue. The book of Genesis records Isaac's dilemma in speaker identification when Jacob acts as an imposter to Esau. Isaac's confusion was with contradictory results from two different biometrics. "The voice is Jacob's voice, but the hands are the hands of Esau." Jacob trusted tactility over auditory "and he discerned him not" [55]. The problem of recognizing an individual by their voice arose throughout history and even appears in a recorded judicial case as early as 1660 [56]. It was much later before academic research would begin a scholarly investigation of this topic.

In March of 1932, Charles and Anne Lindbergh's baby boy was abducted and subsequently killed. The investigation led to a clandestine payoff in a cemetery where a Lindbergh operative met with an anonymous male claiming to be the kidnapper. Charles Lindbergh sat in a nearby car. Lindbergh overheard the anonymous man say "Hey Doctor, Over here, over here". The event was the second time Charles Lindberg had heard the man's voice without seeing his face. Two and a half years later at the trial of the accused kidnapper, Bruno Hauptmann, Lindberg claimed to be able to identify Hauptmann's voice as the same voice heard in the cemetery [56].

The Lindberg claim spurred Frances McGehee to initiate the first academic research of reliability of earwitnesses. Her research led to the publication of two significant articles on the topic [57, 58]. Since McGehee, research into speaker recognition has been continuous in forensics and psychology. The later development of the automatic speaker recognition system can also trace its roots to the work of McGehee.

In 1962 the first article on an automated (semi-automated) method for speaker recognition was published in Nature by a Bell Laboratories Physicist, Lawrence G. Kersta. The paper was entitled “Voiceprint Identification” [59]. Two years previous, Bell Laboratories had been approached by law enforcement agencies about the possibility of identifying callers who had made several verbal bomb threats over telephone lines [60]. The task was given to Kersta. After the two years of research he claimed he had a method to identify individuals with very high success rates. His method utilized earlier work on speaker recognition performed by three other Bell Laboratories’ scientists, Potter, Kopp and Green who were working on voice identification for military applications during World War II. They had developed a visual representation of speech called a spectrogram. A spectrogram records the frequency and intensity of a speech signal with respect to time. Kersta’s claims of identifying speech via spectrograms sparked several research projects over the next year. In fact, his article sparked an entire field of research. There were several dissenting views in the next few years and it seemed no other researcher was able to duplicate the high claims Kersta had made [60].

To help settle the matter, a substantial research project was undertaken by Oscar Tosi, a professor at Michigan State University who had doubts about Kersta’s so called “voiceprint”. In conjunction with the Michigan State Police and sponsored by the Federal Department of Justice, Tosi’s research yielded promising results. Tosi’s results tended to support Kersta and lent validity to the field. Tosi’s results were refuted by Bolt a year later as he illustrated holes in the Tosi experimental methodology [60]. Tosi’s experiment lacked scientific basis in practical applications. The FBI, being interested in the forensic application of speaker identification, requested another study be performed by the National Academy of Sciences. The results from the study showed that the technical uncertainties in forensic applications were substantial enough to claim the use of voiceprints were unreliable in real applications. However, voiceprints are still useful in certain circumstances. In fact the FBI has utilized a form of Kersta's spectrographic analysis as late as 2002 [60].

The Kersta method is an aural-visual method. From a voice sample a spectrograph is produced. The spectrogram is then inspected visually for pattern matching and scored by the interpreter. Success rates with the Kersta, spectrogram method, given an expert interpreter and proper environmental circumstances, can be very high. Despite success, the Kersta method requires human interaction, limiting its use in automated security applications. Also, "the good performance reported in Kersta's paper has not been observed in subsequent evaluations simulating real-life conditions" [7].

Though the Kersta method is still utilized in some forensic applications, such as with the FBI, it has not materialized into a practical autonomous speaker recognition system. The reasons are many, human interpretation being a major factor. Other techniques have since been employed allowing for low computing costs with high success rates.

It was in the 1960's when several developments made autonomous automatic speech recognition possible. These developments covered a broad range of disciplines and for the most part were independent of speaker recognition research. For instance, Gunnar Fant produced the first physiological model of human speech production in 1960 [29]. This and similar research that followed, became the basis for understanding how to analyze speech for both speaker recognition as well as automatic speech recognition. It led to the understanding of voice as a linear source-filter model, which allowed for a better understanding of identifiable characteristics in an individual's voice.

As computers became more accessible to more scientists, problems of implementation of continuous-domain mathematical solutions in a discrete machine arose more and more often. The issue was critical to digital signal processing. In 1965 Cooley and Tukey published their method of digital implementation for the Fourier transform. It is now known as the Cooley-Tukey Fast Fourier Transform (FFT) [61]. The FFT gave scientists an efficient method of frequency analysis in computer based systems. It was a major advance and it coincided with other investigations at the time. Two years earlier in 1963 Bogert, Healy and Tukey had published a study on echo detection in seismic signals titled "The Quefrency Analysis of the Time Series for Echoes: Cepstrum, Pseudo-Auto-Covariance, Cross-Cepstrum, and Saphe Cracking" [34]. The oddly titled paper described a method of echo detection by taking the "spectrum" of a log-magnitude spectrum. During the same period, Oppenheim's research into homomorphic signal separation, such as decovolution, led to him defining the complex cepstrum, which is the Fourier transform of the log spectrum, i.e. a spectrum of the spectrum [62]. The cepstrum is described in section 3 of this chapter. The complex cepstrum has become a standard method used in speaker recognition systems.

In another completely unrelated study in the late 1960's Leonard E. Baum and others developed a stochastic model for Markov processes. The process attempts to determine hidden parameters of a statistical model from observable features in the model and is called the Hidden Markov Model (HMM) [63]. The HMM statistical model would find broader application in the parallel studies of speech recognition. The HMM also has a smaller role in speaker recognition.

The fortuitous developments of the 1960's have become the basis for modern speaker recognition systems. It was also during this period that parallel investigations into automatic speaker recognition system began. For instance, Pruzansky, a Bell Laboratories Engineer, investigated early systems for automatic speaker recognition utilizing spectral pattern matching techniques [64, 65]. The spectral pattern matching systems had a measure of success. However, the first completely autonomous speaker recognition system was a multimodal system which utilized voice and signature analysis. It was developed by a team led by George Doddington at Texas Instruments in 1977 [21, 64]. The Doddington system used digital filter banks to do spectral analysis. It was a text-dependent system that prompts the user for the correct verification phrase. The output vector of a 14-channel filter bank is used in a 'Euclidian distance' based algorithm to make a verification decision [7]. Over many years, the Doddington system had a false rejection rate of less than 1% and a false acceptance rate of less than 1% [7].

The early recognition features used as measures included spectral resonance, filter banks vectors and linear predictive coefficients. As shown above, these features had a good level of success. The early successful systems were all text-dependent. Since that time research has been able to improve on the early text-dependent successes. Investigations into text-independent methods since those early days have continued. Today, text-independent research constitutes the brunt of the speaker recognition research. Text-independent research differs from the text-dependent research as scientists look for underlying indentifying attributes, as opposed to spectral pattern matching or phonetic event measurements. Text-independent research is more frequently applied to speaker identification, as opposed to the simpler task of verification.

The application of Bogert and company's brainchild, the cepstrum, to speaker recognition gave a marked improvement in recognition systems. Cepstrum based features have now become standard in recognition systems [21]. Modern recognition systems use the homomorphic deconvolution capabilities of the cepstrum to separate the vocal fold attributes from the vocal tract attributes in the linear source-filter model of human speech production. As of today, the cepstrum and cepstrum coefficients play an integral role in speaker recognition.

As important to an accurate recognition system as voice feature selection / extraction is, the pattern matching and decision making algorithm is equally important. The Hidden Markov Model, developed in the late 1960's, was employed widely in speech and speaker recognition systems during the 1980's. Also, a method of vector quantization (VQ), compressing a speaker feature vectors down to a small set, also had some success in modeling voice features. However, later research showed that with enough enrollment data the HMM and the VQ was about as effective as the less computationally demanding Gaussian Mixture Model (GMM) [21, 40]. Though the HMM has wide application in speech recognition, it is found less often in low-level speaker recognition systems.

The field of speaker recognition study has also made significant improvement from the simple Euclidian distance method found in the TI system. The system that has evolved throughout the early years of recognition research includes a few basic tasks. These tasks are, feature extraction, feature modeling and classification or decision making. The feature matching algorithm which computes the likelihood that one user's voice sample matches the modeled enrollment samples. The classification methods have also made significant improvement from the simple Euclidian distance method. A fairly straightforward, simple decision algorithm may be a likelihood test of the Gaussian Mixture Model of cepstrum features. Though there are many enhancements to the simple authentication system as described, several other methods are being investigated. The basic system presented represents a wide range of modern speaker recognition systems.

Several advances have been realized in the system detailed above. For instance, squaring the log-magnitude spectrum prior to taking the cepstrum can magnify the voice signal while mitigating the effect of background noise. One major area of interest has been score-normalization [20]. Any given speaker has a measure of variability between his own samples. Intrapersonal variations are due to many factors including, emotional state, throat illnesses, phonetic content and background noise. One objective of score-normalization is to mitigate the intra-speaker variability effect. Another objective is to mitigate channel and other environmental effects. Throughout the 1990’s and 2000’s a significant amount of speaker recognition system research has been focused on score-normalization [20].

Score normalization research has largely been based on the work of Li and Porter which presented a method of using imposter score normalization [53]. Their research led to the UBM or “world-model” approach a few years later where a model, often derived from a cohort of imposters is used in the statistical model of the speaker’s enrollment model. The log-likelihood between the speaker and world models error rates are measured against a threshold in order to make an authentication decision. The way these models are derived have advanced throughout the last few years and have led to advanced world-models and score normalizations.

Research continues in various fields today. One topic in speaker recognition research is the continued research into feature selection. Notably, Reynolds, Campbell and others have undertaken the collaborative investigation into the usage of high level information [22-23, 26-27]. Use of multimodal biometric and multimodal user authentication , obtaining confidence levels in a specific systems recognition accuracy and identification applications are all current topics in literature [8, 10, 25, 37, 43].

Section 5 – Outstanding Issues in Speaker Authentication Systems

One specific area of research continues to be environmental variability, such as background noise, intrapersonal variations and handset variability. Environmental concerns become a major factor in applications where unknown conditions exist, such as in distributed systems. With the advent of the internet and security applications over the internet, such as internet banking, security needs in unknown conditions have become more and more relevant. Therefore, research into environmental concerns has gained an increased focus in speaker recognition [66]. There have several compensation techniques presented that have had success in filtering environmental noise. Background noise has been dealt with primarily through filtering [20]. Current research seeks to improve these methods [67, 68].

Handset mismatches refers to differences in the voice capture system used in the enrollment phase verses the system used in testing for authentication phase. When a user is enrolled with one system and attempts recognition with another, it gives significant error rates. Early on Doddington discovered that such mismatches could produce errors in the range of 50% [7]. Differing transducers can affect a voice spectrum by changing spectral characteristics such as band-limiting and shaping [15, 69]. Much has been done to attempt to mitigate the enrollment / testing mismatch obstacle. Cepstral mean subtraction, RASTA filtering, and use of delta coefficients have all been used in attempts to mitigate handset mismatch effect. Each of the listed methods have had a degree of success. However, they need to be greatly improved. It has been proposed on several occasions that more research into the effect of microphone variation is needed [15, 20, 66]. Some research has been conducted to understand the mismatched condition [70-73]. Some attempts at solutions have also been made [70, 71]. One shortcoming of these studies has been their focus on telephone systems. Today, the need for a variety of networked systems is required. This thesis approaches the problem of a mismatched condition from a distributed environment standpoint. Also, little or no research has been performed to understand general performance of microphones compared to one another in similar environments. Another objective of this thesis is to investigate the effect of microphone selection on speaker recognition performance.

CHAPTER III

MEASURING EFFECTS ON SPEAKER RECOGNITION

Section 1 – Objective of the Experiment

The objective of the experiment is to determine the effect of equipment variations on error rates in speaker recognition systems. The first effect that is analyzed is the degradation or improvement of FA / FR rates when enrollment microphones and testing microphones are mismatched. The mismatch effect has been assumed in the past [70]. This study attempts to quantify that the enrollment / training mismatch has a specific effect on speaker recognition FA and FR rates. The second effect that is analyzed is the change in FA and FR rates from one system to the next in similar environments. The analysis of the second effect includes direct comparisons between each system's performance under matched and mismatched conditions in varying background noise levels. The comparison would indicate, in a given environment, that 1) some microphones perform better than most others, 2) some microphones perform worse than most others, 3) all microphones perform about equal or, 4) microphone performance is distributed with some performing better than most, others performing worse than most and some in-between. The characteristics of the performance distribution would indicate, in each environment, the sensitivity of a system's performance to the variation in recording equipment. Further, these distributions in each environment will be compared to determine if a particular microphone(s) is (are) generally better suited for the task of speaker recognition.

Section 2 – Experimental Setup

Ten system setups were investigated, including seven different microphones. Each microphone is a common, real world, device. The selected systems represent a small sampling of typical home and office equipment that is currently available on the market. The focus of the study is security in distributed systems such as the internet. Therefore, microphone selection was based on was on common equipment likely to be used in distributed systems. The selection includes several PC interfacing microphones as well as handheld devices. The full list is found in Table 3.1.

Table 3.1: List of microphones used in the experiment

System#	Short Description	Setup	Manuf.	Model#
1	Desktop Microphone	1" from Speaker		
2	"Quick Cam" Webcam	1" from Speaker	Logitech	960-000247
3	"Quick Cam" Webcam	12" from Speaker	Logitech	960-000247
4	Hands free Microphone	1" from Ear	GE	
5	Hands free Microphone	On Ear as Designed	GE	
6	Logitech Gaming headset	On Head as Designed	Logitech	
7	Digital Voice Recorder	1" from mouth	Olympus	WS-100
8	Sys#4 / Sys#7	sys#4 Plugged into sys#7	GE/Olympus	WS-100
9	Digital Voice Recorder	Same model, different unit as System #7	Olympus	WS-100
10	MP3 Player	1" from Speaker		

Five speakers were selected, 3 male, 2 female. Each spoke three phrases (See Appendix E) into each of the ten systems. The process was repeated in three various background noise levels. A fourth phrase was spoken into all ten systems by all five users on a different date. The systems are the items under investigation, not the speakers. The number of users is not statistically significant for investigation into speaker discrepancies but is designed to give a variety for the testing of the equipment. Likewise the phrase usage is not for understanding phrase discrepancies, but rather to give a variety for the testing of the equipment. Utilizing various individuals and phrases throughout the experiment mitigates how various users or specific phrases affect the results.

Table 3.2: List of controlled variables in the experiment

Parameters	**Quantity**	**Description**
Systems	10 systems (7 Microphones)	See Appendix
Speakers	5 users	3 male, s female
Background Noise Levels	3	Zone1 (<45dB), Zone2 (55-65dB), Zone3 (80-95dB)
Phrases	4	Phrase1 (~3sec), Phrase2 (~3sec), Phrase3 (9sec), Phrase4 (~30sec)
Decision Algorithm	2	

The phrase used for enrollment, was Phrase 3 in each case. The data from the enrollment phrase was put into the recognition system developed for MATLAB. The components of the implemented speaker recognition system, as illustrated in Figure 3.1 include feature extraction via mel-cepstrum MFCCs, feature modeling with a GMM that utilizes an expectation maximization algorithm [41], a likelihood comparison, and then a decision algorithm. There are two decision algorithms used in the experiment. The first is a 'Nearest-To', or shortest distance algorithm. The person with the log-likelihood closest to zero is accepted and everyone else rejected. The second algorithm is the threshold decision algorithm. The threshold algorithm sets an initial threshold θ. Which side of the θ the log-likelihood score falls determines whether one is accepted or rejected. If $\lambda < \theta$ then Y_0 is hypothesized to come from Y_M and the speaker is accepted. Else, if $\lambda > \theta$ then the speaker is rejected. In the algorithm the threshold θ is varied over an appropriate range. As θ is varied FA and FR rates are recorded. The outcomes of the two algorithms are described in Chapter IV. To determine how error rates are affected, each system is evaluated with the enrollment and testing systems matching and mismatching. An individual system is enrolled and then tested against each of the ten systems. Resulting error rates are evaluated for both matched and mismatched enrollment/testing conditions.

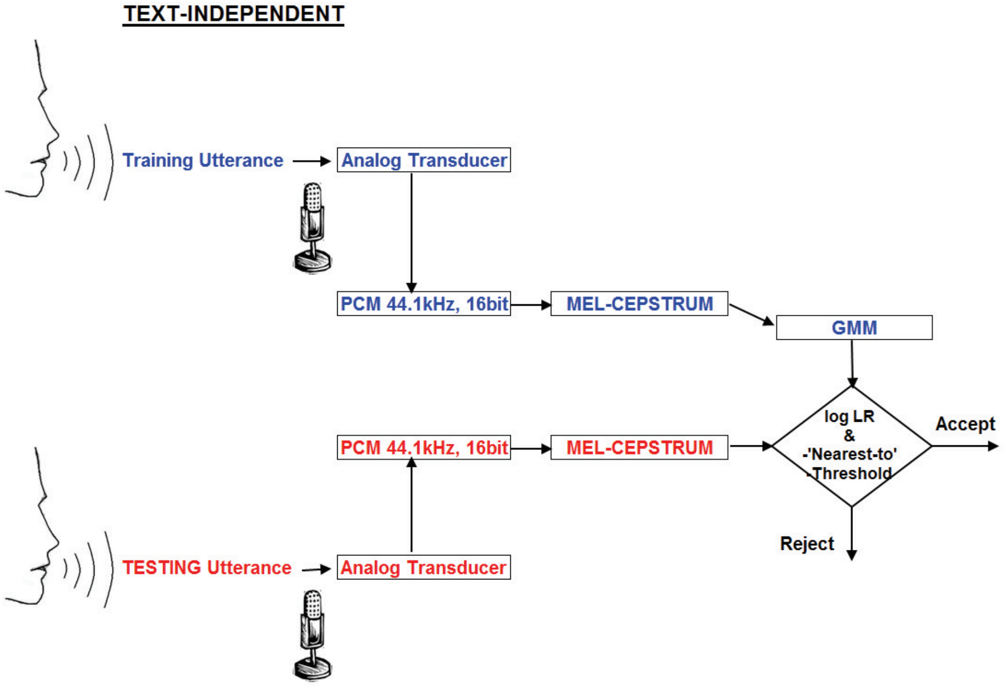


Figure 3.1: Outline of speaker recognition system used in experiment

The background noise levels were controlled within the specified decibel ranges. In Zone1, as measured at the microphone at the beginning of each session, the background noise level was less than 45dB. Zone 2 static noise was added and decibel level was controlled between 55-65dB. Zone 3 had an increase in static noise. Zone 3 also had an addition of dynamically changing, non-voice noise. The Zone 3 background noise was controlled in a range of 80-95dB. The frequency range of the background noise was not controlled.

Section 3 – Description of Recording Environment

The physical environment was an isolated area with precautions made to mitigate outside noise. In Zone1 (<45dB), precaution was taken to mitigate sound by isolating the PC and turning off all other devices in the room (such as the air conditioner). Zone2 (55-65dB) the noise was increased by turning on fans, the air conditioner, having the PC near the recording area and having low magnitude level static from a radio at a given distance from the recording area. Zone3 (80-95dB) was the same as Zone2 with an increase in the radio static and the addition of a given portion of the first movement of Beethoven's fifth symphony. Background noise levels were taken at the beginning of each user's session (a session includes one speaker uttering a set of three phrases into 10 systems on the first day and one phrase into 10 systems on the second day). The background noise level was recorded with RadioShack's "7-range Analog Display Sound Level Meter". Measurements were taken within a few inches of the user's mouth. The database generated in this research is specific to common distributed systems. Devices and background levels were selected for a distributed system. The database varies from available commercial voice sample databases.

Section 4 – General Discussion of Results

A total of 500 voice samples were collected. Average file size was 840kB. All voice samples were saved as .wav files in a PCM Stereo format at a 44.1KHz sample rate and a 16 bit AD conversion. Voice samples taken with both handheld voice recorders (items 7&8) as well as the MP3 player (item 9) save their data files in an MP3 format. In order to stay consistent and to analyze these files in the MATLAB speaker recognition system, the MP3 files were converted to WAV files with formatting consistent with the rest of the experiment. Research into the effects of speech compression algorithms on speaker recognition has been conducted [74]. Results from the research indicated that these algorithms had little to no effect on the error rates of the system. Phrases 1-3 were recorded 113 days prior to Phrase 4. Phrase 4 was not uttered in Zones 2 or 3. Voice samples were analyzed via a speaker recognition program developed for MATLAB. The decision algorithms were coded in VBA. A flow chart for the MATLAB portion is found in the appendices.

Table 3.3 gives a time domain representation of each speaker's voice as they say Phrase 3 in each zone as they spoke into System 1. Table 3.4 shows the frequency domain of the same samples. Table 3.5 is Table 3.6 except the y-axes are adjusted to illustrate the spectrum shape. Zone 3 noise is readily seen. Zone 2 noise is not as apparent unless viewed with the adjusted y-axes. The voice sample from Speaker 1 in Zone 2 was low magnitude at all frequencies. Noise was low as well. The low noise level could be due to user variability such as the direction of the microphone in relation to the speaker and noise sources. It could also be due to equipment malfunctions such as a loose microphone connection. The entire recording session with Speaker 1 in Zone 2 on System 1 had attenuated amplitudes. This appears to be an anomaly as it was not noted in other sessions.

Table 3.3: Time domain of voice signals – System 1 Phrase 1

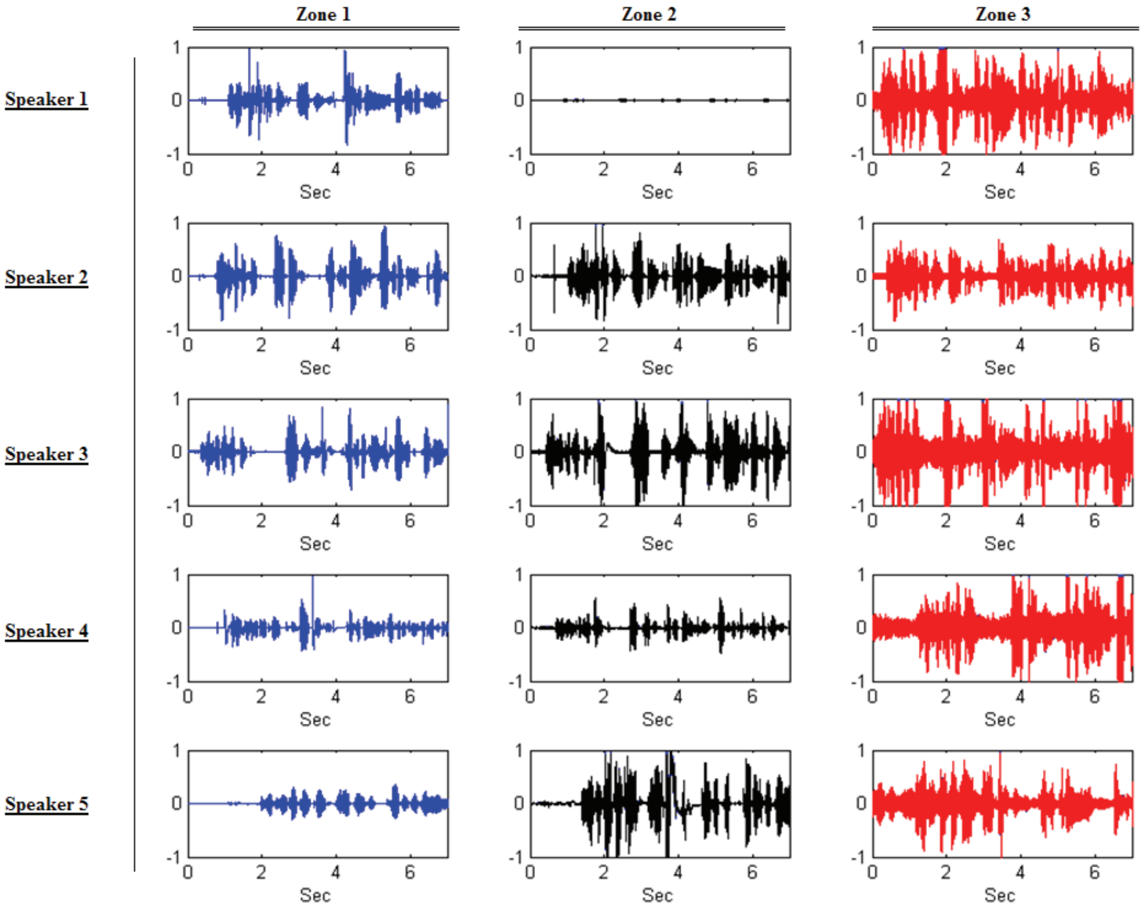


Table 3.4: Frequency domain of voice signals – System 1 Phrase 1

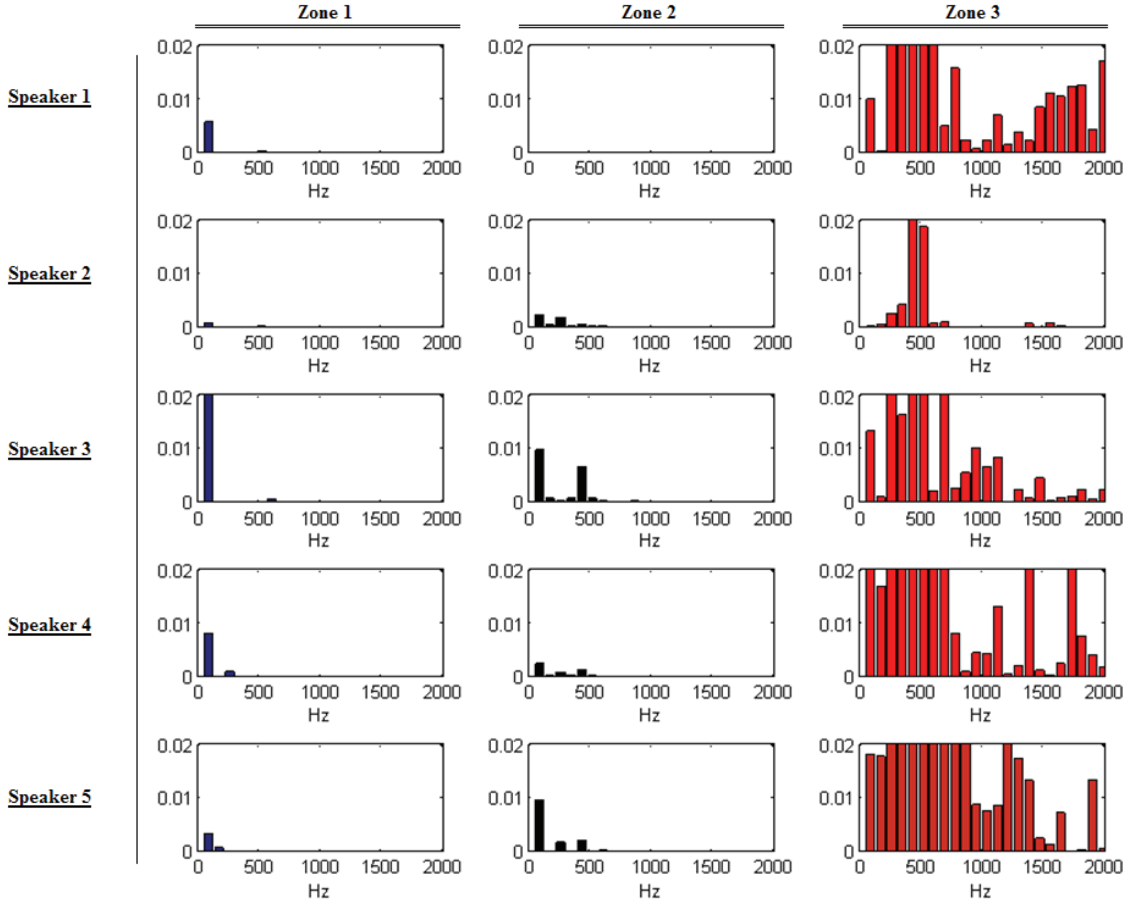
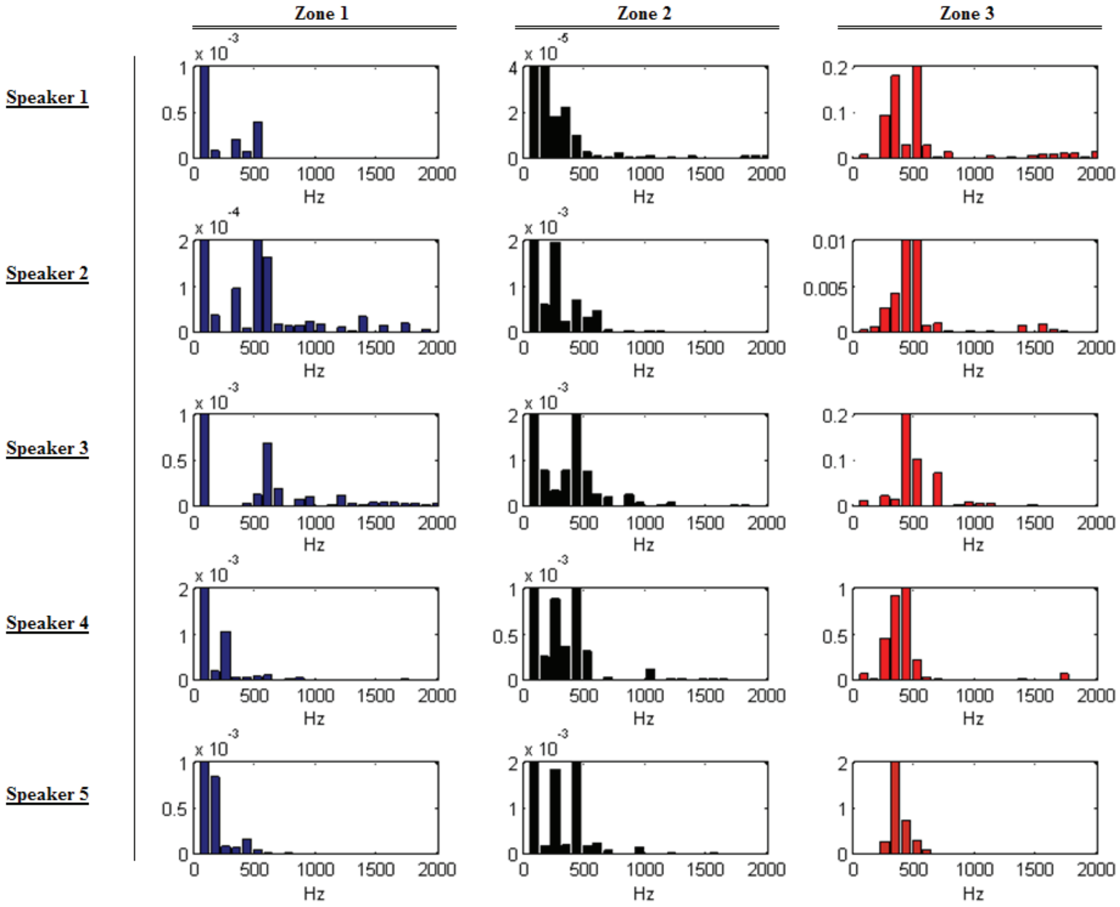


Table 3.5: Scale shifted frequency domain of voice signal – System 1 Phrase 1



CHAPTER IV

RESULTS AND ANALYSIS

Section 1 – Description of Analysis Techniques

The four possible results for identity authentication are a true accept (TA), a true reject (TR), a false accept (FA) and a false reject (FR). The FA and FR rates will be used for analysis. The FR rate is defined as the number of individuals falsely rejected divided by the total number of people who should be accepted. Likewise, the FA rate is defined as the number of individuals falsely accepted divided by the total number of people who should be rejected. By plotting these two rates as a function of the threshold, a DET curve is developed. The point at which the FR rate and the FA rate cross is called the equal error rate (EER) [20]. The EER holds information about a system's susceptibility to a security breach as well as information about a system's usability. Though the equal error rate is not associated to any specific threshold setting, it can be used as a comparative measure of performance between systems. The EER is an arbitrary point which is used to indicate a system's ability to authenticate authorized individuals and decline imposters. The EER is not necessarily the minimum error point. It has traditionally been used as a relative measure between systems. The EER would be a good relative measure between systems if the FA slope and FR slopes of each system was identical. For identical slopes, a linear shift in θ would not indicate a change in the system's performance. A high threshold value would be just as valid as a low threshold value as long as the EER was sufficiently low. However, in real systems an EER at a higher threshold can be an indication of a poorly performing system. A system that can properly classify voice signals, in respect to the speaker, with a high level of success will naturally have a λ close to zero for a TA and a significantly more negative λ for a TR. With a large disparity in the average λ for the cases of TA and TR, error rates can be mitigated. It is proposed that a dead-band could be injected into the decision algorithm. The decision algorithm output for the dead-band would supplement the 'accept' and 'reject' possibilities by a third '*undetermined'* value. What is done in a system when an *undetermined* is found would be subject to system design. The system could prompt the user to re-enter a voice sample or in multimodal systems, a separate identity authentication method could be used to further verify authorization.

$$\frac{P(Y_0 \mid Y_M)}{P(Y_0 \mid Y_{\overline{M}})} = \lambda_0 \qquad \begin{cases} \lambda_0 \geq \theta_A, & \text{Accept} \\ \theta_A > \lambda_0 \geq \theta_R, & \textit{undetermined} \\ \lambda_0 < \theta_R, & \text{Reject} \end{cases} \tag{4.1}$$

When analyzing speaker recognition systems, traditionally, the DET curves are used as design tools and the EER is a loose method of comparing performance. The DET curves give more information than the point at which the two error rates cross. For instance, the slope of the curve in the region of concern is a measure of the system's robustness to changes in the threshold (θ). To illustrate this point, examine Figure 4.1. Systems 1 and 2 in the graph have the same EER. However, these are two distinct systems that do not behave similarly. System 1 has few false rejects except at the most stringent thresholds. The EER is 5% at θ = -5. By loosening the threshold to -7, the FA rate increases up to 35% while the FR rate is down to 1%. A slight change in thresholds generates a significant change in error rates for both error types.

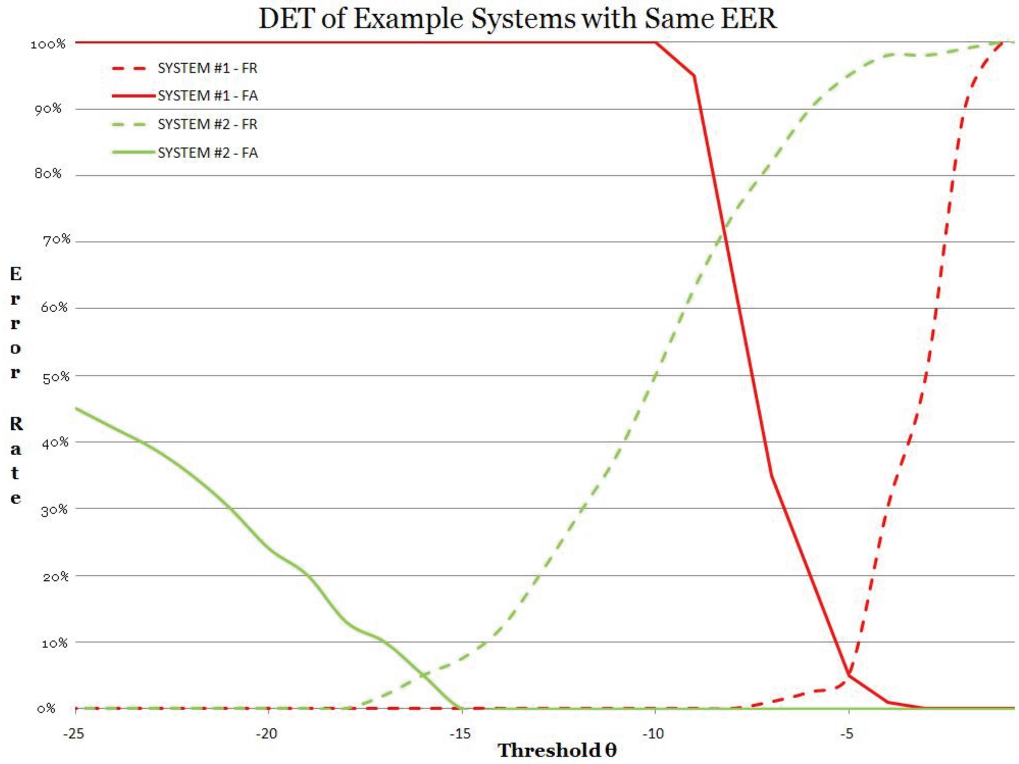


Figure 4.1: A DET curve of 2 systems with equivalent EER

For System 2, a slight change is not as detrimental to either rate. The EER is again at 5%. The threshold θ is -16. A threshold change of -2 in this case leads to a FA rate of 13% and a FR rate near 0%. This example demonstrates the robustness to changes in θ. Further studies are needed to evaluate if the slope of the DET curve could be an indicator of system robustness to environmental changes as well, such as background noise, channel effects and equipment variations. In this study, performance is measured with the standard EER and the accompanying threshold level. The EER as well as the

EER threshold level will be used in comparing systems in the analysis of the experiment utilizing the threshold decision algorithm.

The ‘Nearest-To’ Algorithm designates the speaker in a test set with the λ nearest to zero as the ‘Accepted’ user. All other speakers in the test set are rejected. For the ‘Nearest-To’ Algorithm, a test set is defined as testing each speaker in the group once (all in the same zone with the same microphone saying the same phrase). Each test set produces likelihood ratios λ_i for each speaker in the set. From each test set, the λ that is closest to 0 will be accepted while all others will be rejected. The ‘Nearest-To’ decision process utilizes a specific case of the DET curve. When the correct speaker is accepted, all others are rejected. Thus, in the ‘Nearest-To’ algorithm, with the correct speaker accepted, FA = FR = 0%. If an imposter is accepted and all others are rejected, those being rejected will include the correct speaker. Therefore a false accept equals a false reject, FA = FR = 100%. For example, if the speaker model, Y_M, came from Speaker 1 while Speaker 5 speech sample Y_5 had the greatest likelihood of coming from Y_M, then Speaker 5 is accepted and all other speakers, including Speaker 1, is rejected. Thus the FA rate equals the FR rate in every case. For simplicity, the error rate used is where *error rate = FA = FR.*

Section 2 – Results from ‘Nearest-To’ Decision Algorithm

Table 4.1 gives the overall average results of the ‘Nearest-To’ decision algorithm experiments grouped by systems. It is a brief summary of Appendix A. One result made clear in Table 4.1 is the wide disparity in error rates of matched systems versus miss-matched systems.

Table 4.1: Breakdown by System of 'Nearest-To' results

	AVE	Matched	Mismatched
System #1	*54%*	13%	58%
System #2	*66%*	11%	72%
System #3	*66%*	27%	71%
System #4	*62%*	18%	67%
System #5	*68%*	20%	73%
System #6	*61%*	11%	66%
System #7	*48%*	16%	51%
System #8	*50%*	2%	55%
System #9	*51%*	4%	56%
System #10	*74%*	18%	80%
AVE	***60%***	**14%**	**65%**

The Chart in Figure 4.2 shows how much of a role the miss-matched systems have in system performance.

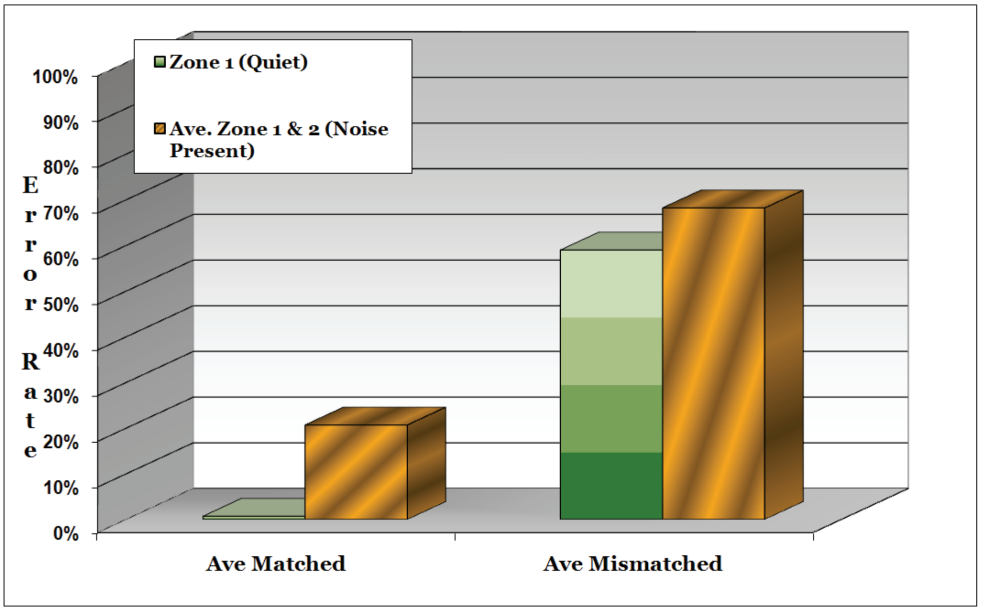


Figure 4.2: 'Nearest-To' match / mismatch comparison chart

Variation in error rates per speakers increased with the background noise level (see Table 4.2). Error rate variation as dependent on phrase was negligible. The average of the standard deviation of error rates per phrase per zone was 0.024. Table 4.3 gives an overall summary of the results of the 'Nearest-To' algorithm experiment.

Table 4.2: Standard deviation of scores per speaker per Zone

	σ
Zone 1	.042
Zone 2	.093
Zone 3	.173

Table 4.3: Summary of 'Nearest-To' algorithm results

	Zone1	Zone2	Zone3	
Total Ave	53%	62%	65%	*60%*
Ave Matched	1%	28%	13%	14%
Ave Mismatched	59%	66%	71%	65%

In a low-noise environment, the matched system error rate was one percent. In the same environment the mismatched error rate was 58% higher. In both Systems noise had an effect. From Zone 1 to Zone 3 a total error rate increase of 12% was observed. The noise effect was insignificant when compared to the mismatched system error rates.

Section 3 – Results from Threshold Decision Algorithm

This section gives the threshold decision algorithm results. Each System in each zone has 2 DET curves (see Appendix C for all DET curves). One curve is for matched

Systems and the other for mismatched Systems. Figure 4.3 shows a baseline measurement of the speaker recognition system enrolled on System 5. Figure 4.3 includes voice samples from all three background noise levels, all speakers, and from all microphones. The graph is an example of what a system's results would be in a system without controls or constraints on the testing phase.

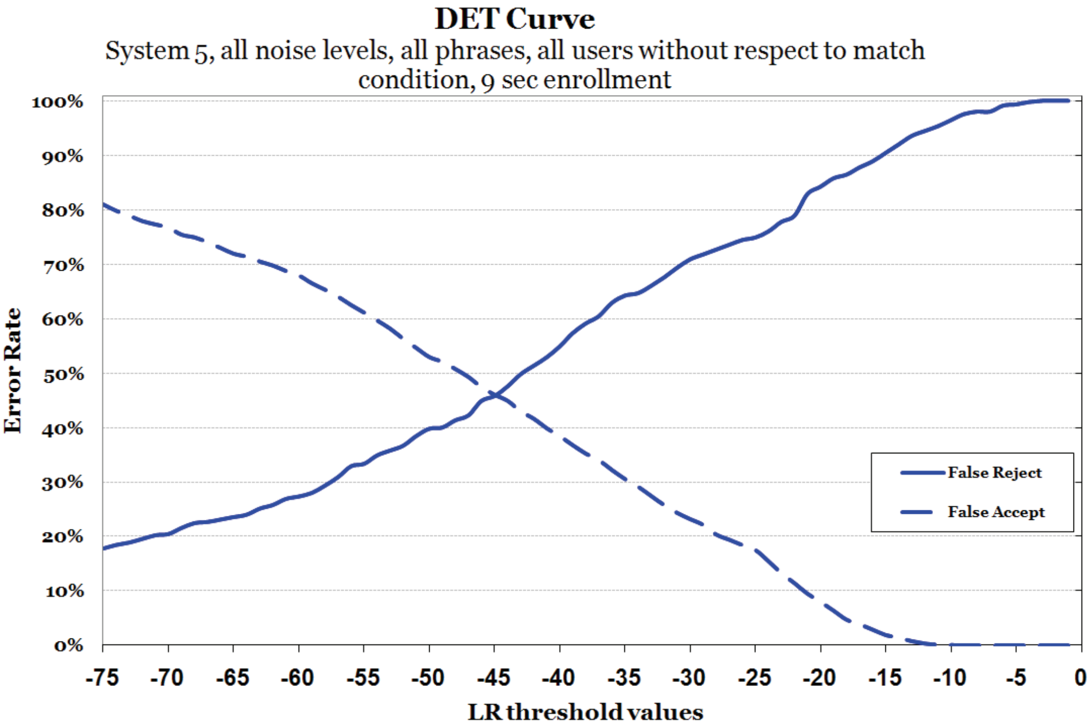


Figure 4.3: DET curves for System #5 with no constraints in the testing phase

In order to review system error rates of the ten systems in a comprehensive manner the DET curves are summarized by discussing the equal error rate (EER). The graph in Figure 4.4 is a summary of all of the EER for each System in matched conditions. Each point represents the EER of 5 speakers uttering 4 phrases in a single zone and on a single system. They are grouped by system. Each system has three points. The points represent the error rates in the three background noise-level zones. In all but one case, the point furthest to the left (most negative) is the Zone 3 EER. The exception is System 1 where the Zone 2 point (center point) is further to the left. Furthest to the right is Zone 1 in each case except System 6. System 6 has the Zone 2 furthest to the right. The Zone 2 EER in 7 of 10 systems is the highest of the three system values. Further discussion of the EER behavior is found in Section 4 of this chapter.

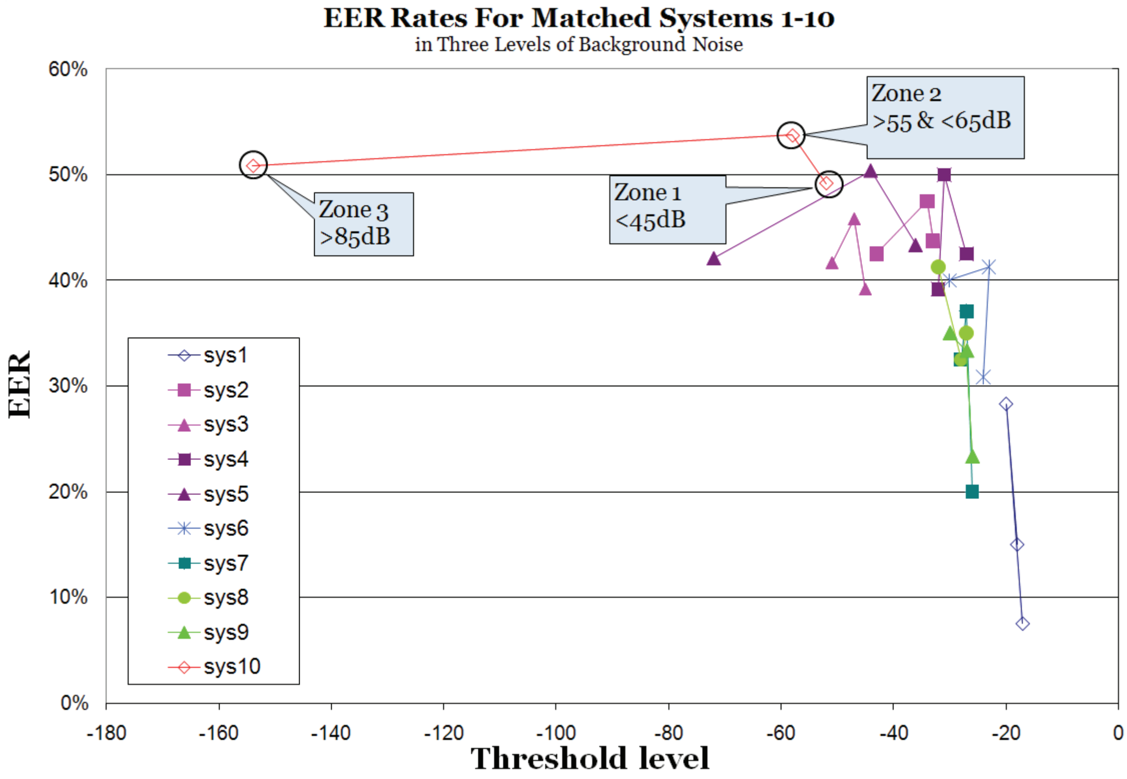


Figure 4.4: EER for all Systems in all zones under matched conditions

Of note in Figure 4.4, the EER generally increases as the threshold point of the EER becomes more negative. Recall that if each system had the exact same curve that a shift along the x-axis would be insignificant and that the EER alone would be sufficient to rate a system's capability. However, in real systems, when a threshold is relatively large, the EER is likely to increase. The EER increase is illustrated more clearly in

Figure 4.5, which shows the general EER threshold decrease with EER increase by displaying a linear trend line for data sets in each zone. In a system that produces a λ close to zero for a TA and a λ much more negative for a TR (a system that can

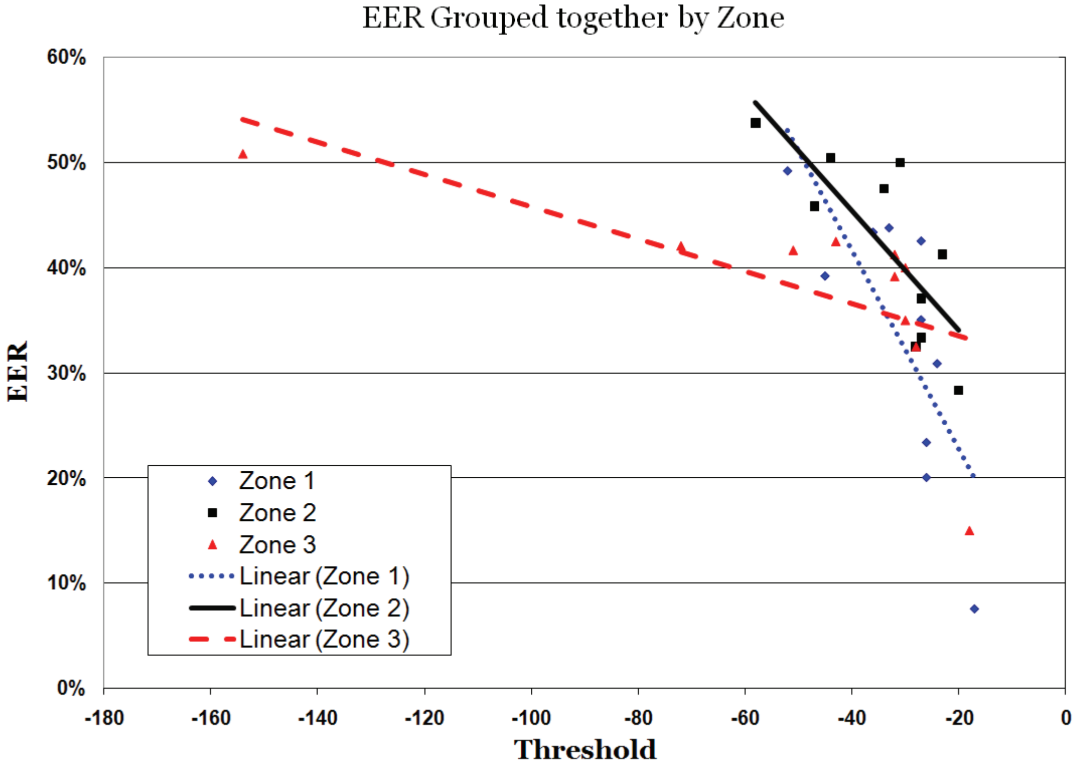


Figure 4.5: EER and linear trends for Zones 1-3

distinguish well between an imposter and an authorized individual) the point at which the FA and FR meet will be low. Ideally, as the threshold is loosened (becomes more negative) the FR rate will be zero before the FA rate curve can begin to increase, leaving an EER of zero. The same scenario, where λ is close to zero for TAs and much more negative for TRs, indicates that a FA will not occur until θ is much more negative. Likewise, false rejects will not occur until the likelihood is near zero. The increase of FR as the likelihood approaches zero is because of the system's ability to detect an authorized individual at a λ close to zero. Therefore, as θ increases along the x-axis, the EER increases along with it. In Figure 4.4, the best performing system are in the lower right corner (close to zero on both the x- and y-axis) and the worst performing system are in the upper left portion of the graph. Most systems' equal error rates ranged between 30-45%. However, the range of System 1 was well below most systems at 7-28% and System 10 was well above most systems at 50-55%.

For systems with a mismatched condition, EER were greater. The distribution of EER from the mismatched systems was for the most part tighter with the exception of System 10 (See Figure 4.6). System 10's threshold level for Zone 1 was -75, for Zone 2 was -77 and Zone 3 was -325, with EERs of 0.511, 0.495, and 0.540 respectively. All

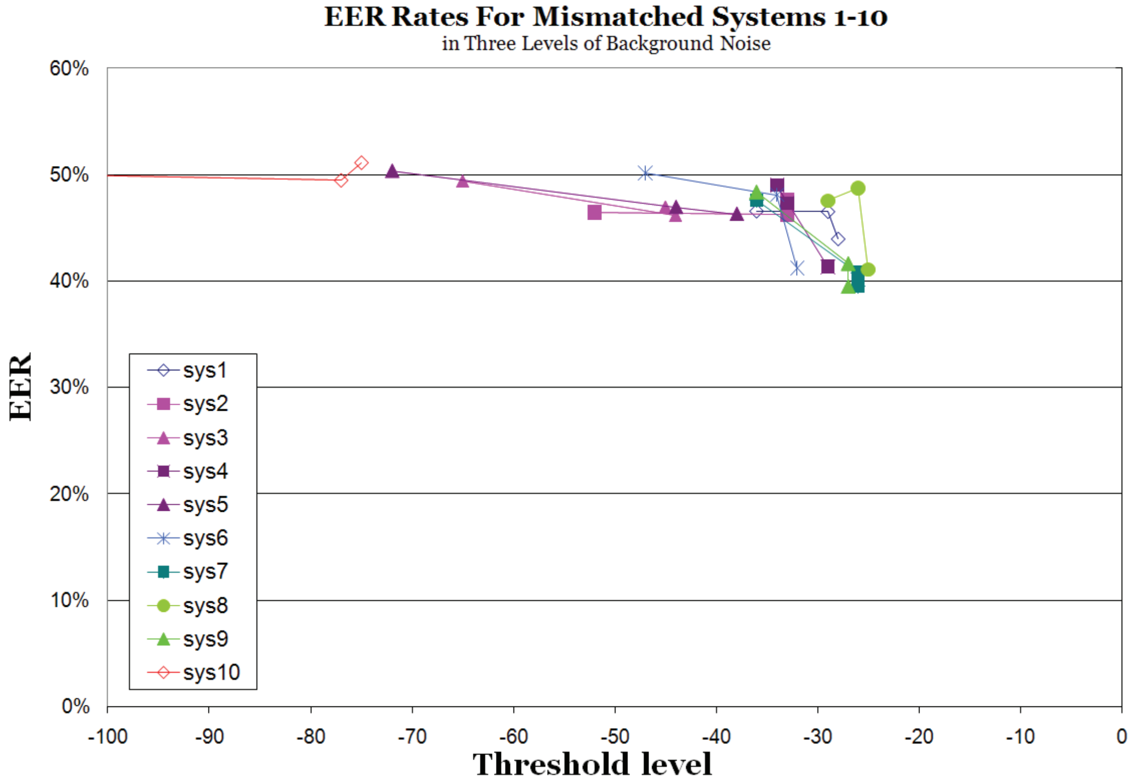


Figure 4.6: EER for all Systems in all zones under mismatched conditions

other threshold levels were between 0.395 and 0.504. Performance change due to noise was greatest when the background noise level was increased from Zone 1 to Zone 2. From Zone 2 to Zone 3, performance variations were less with some performance improving.

Section 4 – Discussion

The results of the 'Nearest-To' decision algorithm further validates the notion that the condition of mismatched enrollment and testing microphones has a major effect on speaker recognition system performance [70-72]. Noise had an effect on the error rates. When the noise level increased from <45dB in Zone 1 to 55-65dB in Zone 2, the EER generally increased and the threshold level generally became more negative. This EER / threshold relationship was true for both the matched and mismatched systems when moving from Zone 1 to Zone 2. The relationship held true as well for the mismatched systems moving from Zone 2 to Zone 3. However, a different phenomenon was observed in the matched systems when increasing the noise from Zone 2 to Zone 3. In these cases, θ continued to increase, however, the EER decreased in 80% of the cases (see Figure 4.4 and Table 4.4). Recall that it was stated that the EER would generally increase as the

Table 4.4: Percent of Systems with EER & θ Looseness Increases with Increased Noise

	MISMATCHED CONDITION		MATCHED CONDITION	
	THRESHOLD θ	EER	THRESHOLD θ	EER
Zone1 to Zone2	60% ↑	70% ↑	90% ↑	90% ↑
Zone2 to Zone3	90% ↑	80% ↑	90% ↑	20% ↑

threshold increased. The increase did not occur when going from Zone 2 to Zone 3 with matched conditions. One explanation for this is that the FA / FR ratio changes. As the noise increases to a loud level, the FR rate will naturally increase (the curve will move left) as it is harder to isolate and extract clean voice features. However, the increase in noise also makes it hard for an imposter to match the voice sample Y_M. The imposter's inability to be falsely accepted moves the FA curve left. The change in ratio explains an increased θ even with a decreased EER. It is not that the system improved, just that the ratio of errors was altered. The lower EER with higher thresholds illustrate the idea that θ is required to describe system performance. It is important to note that the EER is not a design point but a point of simple comparisons on the DET curves.

Even in the presence of noise, the mismatched condition was the variable in this study that had the most prominent effect on system performance. This effect is further illustrated in the threshold decision experiment. The probability distributions are plotted in Figure 4.7. The dotted lines represent the mismatched zone. Note that in each case the mean EER of the mismatched condition is significantly higher than the matched condition. Also of note is the variances of the matched cases are significantly larger than that of the mismatched cases. The difference in variance denotes a shift in system performance of a few systems. The EER in mismatched conditions never approach the low rates seen in the matched condition. However, on the high error rate end of both conditions, the EERs are relatively close in value.

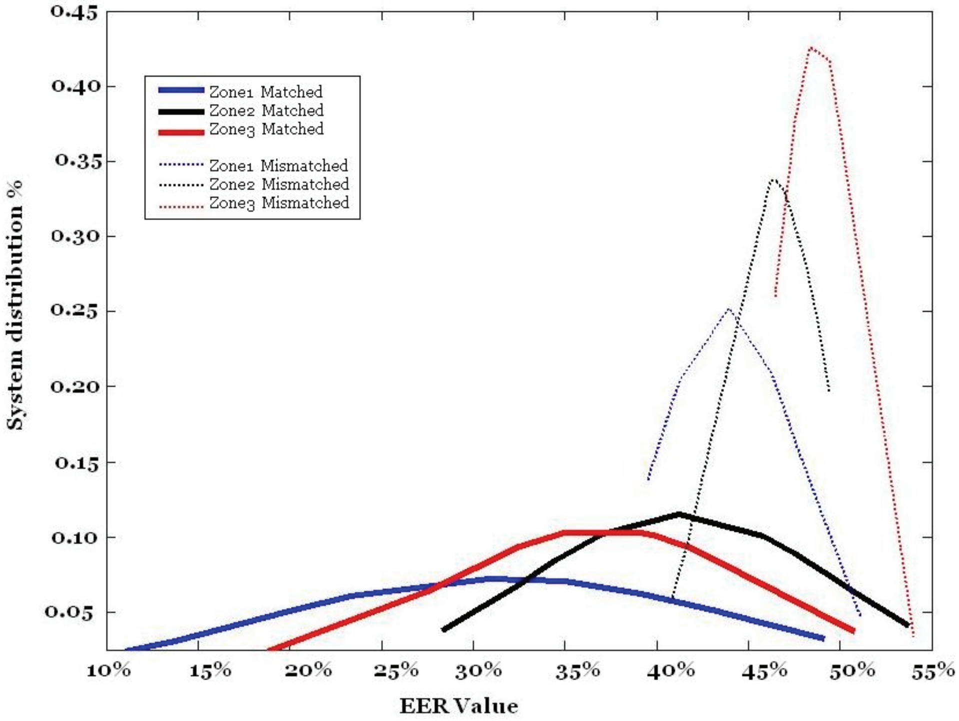


Figure 4.7: Zones 1-3 for matched and mismatched conditions - Probability distribution of Systems' EER

By overlaying the graph in Figure 4.4 with the graph in Figure 4.6, as is done in Figure 4.8, the shift in EER and threshold can be seen. The mismatched condition equal error points had a tighter distribution than the matched condition. The average threshold level shifts up 28.1% and the EER has a 22.5% increase when a system goes from matched to mismatched conditions. It is clear that the mismatched condition has a significant effect on system performance.

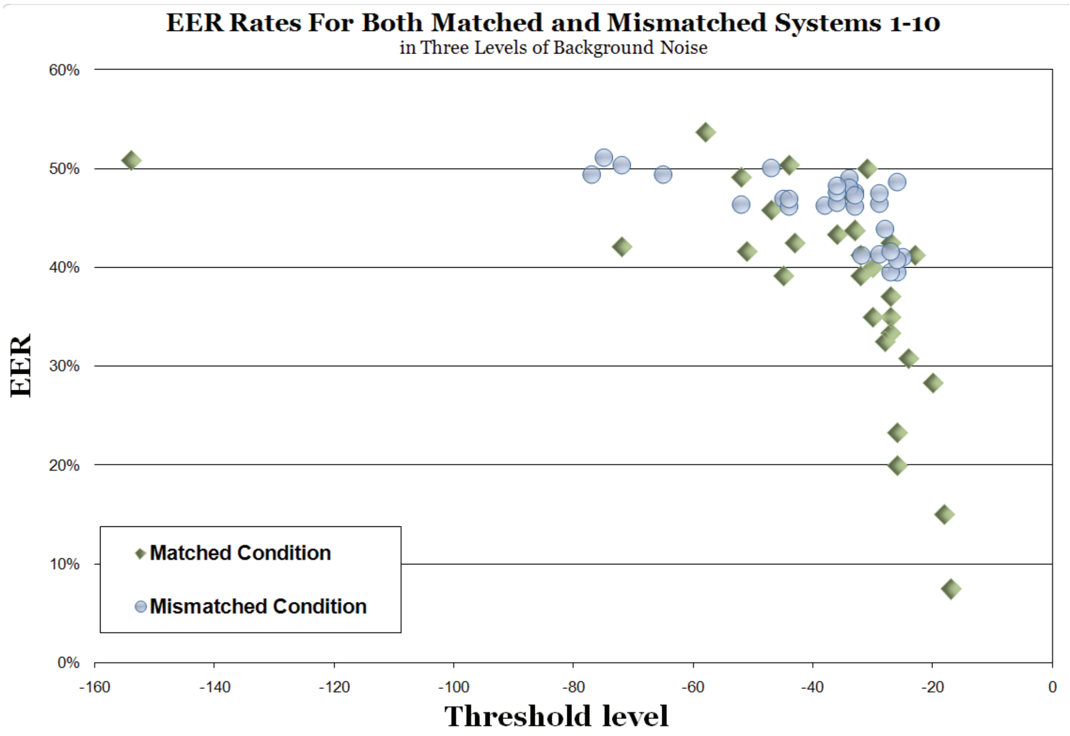


Figure 4.8: Overlay of matched and mismatched EERs

The second objective was to compare each system against each other to determine if some recording systems perform better than others in speaker recognition systems. The comparison between systems is illustrated in Figure 4.4 and 4.6. The graph in Figure 4.9 shows the EER of each zone stacked on top of one another. The stacked plots allow one to look at each system in each zone and compare. The stacked plots represent the sum of EERs in the three zones and illustrates overall relative performance. When under mismatched conditions, systems had a summed EER of 1.25 to 1.55. This range broadened to 0.5-1.54 for matched conditions. Several systems performed notably better in matched condition, especially when in Zone 1. System 1 in Zone 1 had an EER of 7%. Even in its worst performing zone (Zone 2) System 1 had an EER of 28%. System 1 performed better that other systems given matched conditions. It is also of note that

Systems 7 and 9, which were the same model microphone, performed similarly in most cases. By reviewing Figures 4.4, 4.6 and 4.10, it can be seen that System 10 was by far the worst performing system in all cases. System 10 was one of the four systems that required a file format change. However, the format change did not appear to play a major role as Systems 7, 8 and 9 also had format changes and were among the top performers. In matched conditions it is clear that some systems are significantly better suited for speaker recognition than others. In the mismatched condition, it is not as clear how much performance depends on the recording system. System 10 still performed significantly worse (see Figure 4.6), however the rest of the systems had a significantly smaller deviation from the norm. In matched conditions, the microphone plays a significant role in speaker performance. Choosing the proper microphone for the authentication system is important. Or in the case of uncontrolled microphone usage, such as in many distributed systems, it would be important to consider the recording system and design accordingly. One such method for design may include utilizing a proper score normalization technique. A method of evaluating microphone error rates, such as the method used in this thesis, would be useful in making system design decisions.

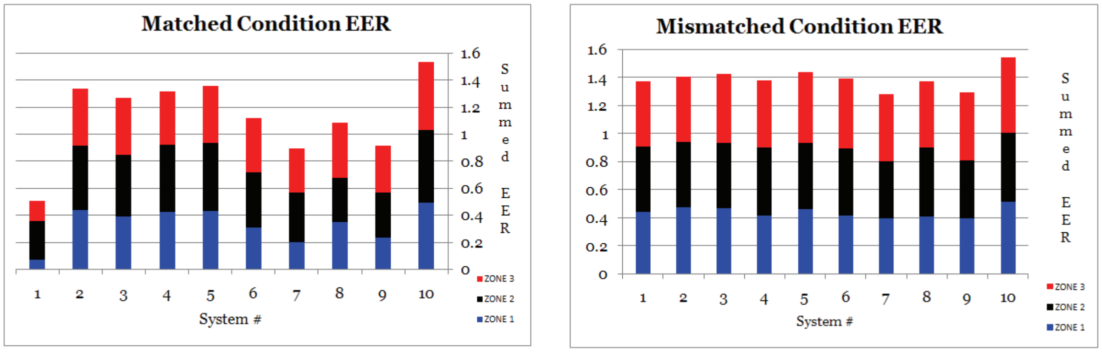


Figure 4.9: EER Summary

CHAPTER V

CONCLUSIONS

Section 1 – Summary

In distributed systems automatic identity authentication is a difficult aspect to control. Often identity authentication systems can decrease susceptibility to a security breach by adding extra elements to the authentication of one's identity. There are three main divisions in the methods of identity authentication; what you know (example: password or login), what you have (example: debit card or key) and what you are (biometrics). Biometrics is a measure of what you are or what you do. Speaker recognition, the biometric of voice, utilizes one's voice as a metric to detect a specific speaker.

Over the past five decades great strides toward wide-spread commercial speaker recognition systems have been made. Early research in speaker recognition was in the realm of human abilities. Later war time research allowed for significant advances, producing a tool to allow visual inspection of voice. Advances in signal processing techniques and the rise of the computer permitted true automated recognition systems to be developed. Early systems such as Doddington's text-dependent system found measures of success spurring on the research for automatic text-independent systems. One of the outstanding issues in the field of speaker recognition is handling environmental variations such as channel effects, background noise, intrapersonal variation and microphone variability. Much progress has been made in these areas. Yet, environmental variations remain as one significant dilemma to real world speaker recognition, especially in distributed systems.

For speaker recognition systems in a distributed application microphones are apt to vary. Frequency response to different microphone transducers can vary widely. Microphone variation can produce two non-matching signals for the exact same recording. The objective of the research was to discover whether or not the varying of microphones has an effect on the ability of a speaker recognition system to perform identity authentication.

The task of speaker recognition is divided into two phases, enrollment and testing. An enrollment voice sample was taken, features extracted, and a model generated. During the testing phase a voice sample was taken, features extracted, the extracted features were measured against the model and a decision was made, accept or reject. The system used in this research used the common feature of MFCCs and common modeling method of GMM. A log-likelihood ratio comparison was used in the decision process. There were two decision algorithms used in the experiment. The first was a 'Nearest-To' algorithm where the person with the log-likelihood ratio closest to zero was accepted and everyone else rejected. The second algorithm was a threshold algorithm. The FA and FR rates were measured as the threshold θ was varied. Which side of the θ the log-likelihood score fell on determined if one was accepted or rejected. The specific objective of the research was to determine how error rates vary with respect to a variation in microphones. Two types of microphone variation were investigated. The first type of variation study was to understand effects on system performance when the microphone differs from the enrollment to the testing phase. The second type of variation study was to understand how a system performs in relation to other microphones when in similar environments and setups.

Section 2 – Conclusions

The research in this thesis demonstrated and utilized a method of evaluating microphones for use in speaker recognition systems via error rates. The experimental results show the effect that microphone variability has on the error rates of speaker recognition systems. Two types of microphone variation that alter the error rates of a speaker recognition system are illustrated in the results. The first type of variation analyzed was when the enrollment and testing microphones were different. This is referred to as a mismatched condition. The second type of variation analyzed was how different microphone error rates vary without regard to environmental conditions such as matched and mismatched conditions. First, mismatched systems are responsible for significantly higher FA & FR rates. The mismatched-transducer effect has been seen as well in past studies, though the past studies have focused on telephone applications. The research presented in this study concurs with previous assumptions, that the mismatch condition has a significant effect on speaker recognition error rates. Noise affected error rates as well. However, the noise effect was insignificant when compared to the effect of a mismatched condition.

The second analysis showed that some microphones had better speaker recognition error rates than other microphones. The EER for System #1 had lower error rates with more stringent thresholds. System #10 error rates are significantly higher than the majority of systems and are more than 40% higher than System #1. The rate variations illustrate that speaker recognition system design must account for microphone variability in order to be viable in distributed environments.

A method of evaluating microphones for use in speaker recognition systems was successfully demonstrated and utilized. In the case that a false accept has a higher associated cost the method can be used to assist in threshold setting. If a false reject is more important, the method is just as useful for threshold setting. Further studies are needed to evaluate if the slope of the DET curve can be used as an indicator of system robustness to environmental variations. The system utilized in the research had typical classifications of accept or reject. A future study ought to be undertaken to evaluate error rates in systems with a third ‘undetermined’ classification.

REFERENCES

[1] J. M. Acken, "How Watermarking Adds Value to Digital Content," *Communications of the ACM*, Vol. 41 Number 7, July 1998.

[2] W. E. Burr, D. F. Dodson, and W. T. Polk, "NIST Special Publication 800-63 Ver. 1", *Recommendations of the National Institute of Standards and Technology*, Gaithersburg, MD, June 2004.

[3] L. O'Gorman. "Comparing passwords, tokens, and biometrics for user authentication," In *Proc. of the IEEE*, 91(12):2021--2040, 2003.

[4] T.Y.C. Woo, S.S. Lam, "Authentication for distributed systems," *Computer*, vol.25, no.1, pp.39-52, Jan 1992.

[5] J. Ortega-Garcia, J. Bigun, D. Reynolds, J. Gonzalez-Rodriguez, "Authentication gets personal with biometrics," *Signal Processing Magazine, IEEE* , vol.21, no.2, pp. 50-62, Mar 2004.

[6] J. M. Acken and L. E. Nelson, "Statistical Basics for Testing and Security of Digital Systems for Identity Authentication," *Proc. 6th Int. Conf. on Computing, Communications and Control Technologies: CCCT 2008*, Florida, pp. 122-128, 2008.

[7] G. Doddington, "Speaker Recognition – Identifying People by their Voices," *Proceedings of the IEEE*, IEEE, pp. 1651-1664, November 1984.

[8] I. Masatsugu, H. Sakan, and N. Komatsu, "Multimodal Biometrics of Lip Movements and Voice Using Kernel Fisher Discriminant Analysis Con. Auto. Robotics", *ICARV 2006*, pp. 1-6, December 2006.

[9] G. Richard, Y. Menguy, I. Guis, N. Suaudeau, J. Boudy, P. Lockwood, C. Fernandez, F. Fernández, C. Kotropoulos, A. Tefas, Pitas, R. Heimgartner, P. Ryser, C. Beumier, P. Verlinde, S. Pigeon, G. Matas, J. Kittler, J. Bigün, Y. Abdeljaoued, E. Meurville, L. Besacier, M. Ansorge, G. Maitre, J. Luettin, S. Ben-Yacoub, B. Ruiz, K. Aldama, and J. Cortes, "Multi Modal Verification for Teleservices and Security Applications (M2VTS) ," In *Proceedings of the IEEE international Conference on Multimedia Computing and Systems* – Vol. 2 (June 07 - 11, 1999). ICMCS. IEEE Computer Society, Washington, DC, June 1999

[10] A. Ross and A. K. Jain, “Multimodal Biometrics: An Overview,” In *Proc. Of European Signal Processing Conference (EUSIPCO)*, pp. 1221-1224, Vienna, Austria, September 2004.

[11] J.A. Halderman, B. Waters, and E.W. Felten, “A Convenient Method for Securely Managing Passwords,” In *Proceedings of the 14th international Conference on World Wide Web*, Chiba, Japan, May 10 - 14, 2005.

[12] M. Pawlewski, J. Jones, “Speaker verification: Part 1,” *Biometric Technology Today*, Volume 14, Issue 6, pp. 9-11, June 2006.

[13] S. Tsuge, M. Fukumi, M. Shishibori, F. Ren, K. Kita, S. Kuroiwa, “Study of Relationships between Intra-speaker's Speech Variability and Speech Recognition Performance,” *Intelligent Signal Processing and Communications*, *International Symposium on* , vol., no., pp.41-44, 12-15 Dec. 2006.

[14] S. Sohoni, C.D. Shaver, J.M. Acken, D. Mertz, L. E. Nelson, J. Remington, B. Sadr, and G. Sundararajan, “Evaluation Criteria for Biometric Based Identity Authentication Systems”, *Procedings of ISSSIS2009*, Coimbatore, India, pp.8-10 January 2009.

[15] D. A. Reynolds, “An Overview of Automatic Speaker Recognition Technology”, *Proceedings of ICASSP*, IEEE, Vol. 4, Orlando, Florida, pp.4072-4075, 13-17 May 2002.

[16] J. P. Campbell, “Speaker Recognition: A Tutorial” *Proceedings of the IEEE,* IEEE*,* pp. 1437-1462, September 1997.

[17] J.P. Campbell, “*Features and Measures for Speaker Recognition*,” Doctoral Thesis, Oklahoma State University, 1992.

[18] J. T. Foil and D. H. Johnson, “Text Independent Speaker Recognition,” IEEE Communications Magazine, IEEE, pp. 22-25, December 1983.

[19] S. Furui, “Recent Advances in Speaker Recognition,” Invited Paper, In *Proc. of the First international Conference on Audio- and Video-Based Biometric Person Authentication*, pp. 237-252, London, March 12 - 14, 1997.

[20] F. Bimbot, J.F. Bonastre, C. Fredouille, G. Gravier, I. Chagnoleau, S. Meignier, T. Merlin, J. Ortega-Garcia, D. Petrovska-Delacretaz, and D. Reynolds, “A Tutorial on Text-Independent Speaker Verification,” *Journal on Applied Signal Processing*, EURASIP, Issue 4, pp. 431-450, 2004.

[21] S. Furui, “50 Years of Progress in Speech and Speaker Recognition,” In *Proceedings of SPECOM*, pp.1-9, Patras, Greece, 2005.

[22] D. Reynolds, W. Andrews, J. Campbell, J. Navratil, B. Peskin, A. Adami, Q. Jin, D. Klusacek, J. Abramson, R. Mihaescu, J. Godfrey, D. Jones, and B. Xiang, "The SuperSID Project: Exploiting High-level Information for High-accuracy Speaker Recognition," *Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03). 2003 IEEE International Conference on*, vol.4, no., pp. IV-784-7 vol.4, 6-10 April 2003.

[23] G. Doddington, G. "Speaker Recognition based on Idiolectal Differences Between Speakers," *Proc. of Eurospeech*, ISCA, Vol. 4, pp. 2517-2520, 2001.

[24] W.D. Andrews, M.A. Kohler, J.P. Campbell, J.J. Godfrey, J. Hernandez-Cordero, "Gender-dependent phonetic refraction for speaker recognition," *Acoustics, Speech, and Signal Processing, 2002. Proceedings. (ICASSP '02). IEEE International Conference on*, vol.1, no., pp. I-149-I-152 vol.1, 2002.

[25] Q. Jin, T. Schultz, and A. Waibel, "Phonetic Speaker Identification," *Swedish Phonetics Conference*, Fonetik 2008, Gothenburg, Sweden, 11. June 2008.

[26] Q. Jin, J. Navratil, D. Reynolds, J. Campbell, W. Andrews, and J. Abramson, "Combining Cross-Stream and Time Dimensions in Phonetic Speaker Recognition," *ICASSP*, 2003.

[27] J. Campbell, D. Reynolds, and R. Dunn, "Fusing High and Low-Level Features for Speaker Recognition," Proceedings In *European Conference on Speech Communication and Technology*, Geneva, Switzerland, pp. 2665-2668, 2003.

[28] H. Gish, M. Schmidt, "Text-Independent Speaker Identification," *Signal Processing Magazine, IEEE*, vol.11, no.4, pp.18-32, Oct 1994.

[29] G. Fant, "Acoustic Theory of Speech Production", *Mouton and Co*., The Hague, Netherlands, 1962.

[30] M. Iseli, Y.-L. Shue, A. Alwan, "Age, Sex, and Vowel Dependencies of Acoustical Measures Related to the Voice Source," *Journal of the Acoustic Society of America*, Vol. 121, Issue 4, pp. 2283-2295, April 2007.

[31] B.S. Atal, "Automatic recognition of speakers from their voices," *Proceedings of the IEEE*, vol.64, no.4, pp. 460-475, April 1976.

[32] N. Fakotakis, J. Sirigos, G. Kokkinakis, "High performance text-independent speaker recognition system based on voiced/unvoiced segmentation and multiple neural nets", In *Proc of EUROSPEECH'99*, pp. 979-982, 1999.

[33] J. P. Campbell, T.E. Tremain, “Voiced/Unvoiced Classification of Speech with Applications to the U.S. Government LPC-10E Algorithm,” In *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing,* Tokyo, Japan, pp. 473–476, April 1986.

[34] Bogert, Healy, and Tukey, “The Quefrency Analysis of the Time Series for Echoes: Cepstrum, Pseudo-Auto-Covariance, Cross-Cepstrum, and Saphe Cracking” in *Time Series Analysis*, M. Rosenblatt, Ed., ch.15, pp. 209-243, 1963.

[35] S. Molau, M. Pitz, R. Schluter, H. Ney, “Computing Mel-frequency cepstral coefficients on the power spectrum,” *Acoustics, Speech, and Signal Processing*, In *Proc. (ICASSP '01). 2001 IEEE International Conference on,* vol.1, no., pp.73-76, 2001.

[36] H. Weiping, and R. Linggard, “Speech Signal Deconvolution Using Wavelet Filter Banks,” In *Proc. of the Second international Conference on Wavelet Analysis and Its Applications*, December 18 - 20, 2001.

[37] S. Furui, “Cepstral analysis technique for automatic speaker verification,” *Acoustics, Speech and Signal Processing, IEEE Transactions on,* vol.29, no.2, pp. 254-272, Apr 1981.

[38] M. N. Nazar, “Speaker Identification Using Cepstral Analysis,” *Students Conference 2002,* Proc. IEEE, vol. 1, pp. 139-143, August 2002.

[39] S. Furui, “Speaker-dependent-feature extraction, recognition and processing techniques,” In *SCST-1990*, 10-27, 1990.

[40] D.A. Reynolds, R.C. Rose, “Robust text-independent speaker identification using Gaussian mixture speaker models,” *Speech and Audio Processing, IEEE Transactions on* , vol.3, no.1, pp.72-83, Jan 1995.

[41] J. Bilmes. “A Gentle Tutorial on the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models,” *Technical Report TR-97-021*, ICSI, 1997.

[42] D. Burton, “Text-Dependent Speaker Verification Using Vector Quantization Source Coding,” *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol.35, no.2, pp. 133-143, Feb 1987.

[43] W. M. Campbell, D. A. Reynolds, J. P. Campbell, and K. J. Brady, “Estimating and evaluating confidence for forensic speaker recognition,” in *Proc. ICASSP*, pp. 717–720, 2005.

[44] S. M. Matyas Jr, J. Stapleton, “A Biometric Standard for Information Management and Security,” *Computers & Security*, Volume 19, Issue 5, Pages 428-441, 1 July 2000.

[45] “The NIST Year 2008 Speaker Recognition Evaluation Plan”, http://www.nist.gov/speech/tests/sre/2008/index.html, 2008.

[46] K. Chen, “Towards better making a decision in speaker verification,” *Pattern Recognition*, Volume 36, Issue 2, Pages 329-346, February 2003.

[47] D. A. Reynolds, “Universal Background Models, Encyclopedia of Biometric Recognition,” Springer, Journal Article, February 2008.

[48] A. Martin, G. Doddington, T. Kamm, M. Ordowski, M. Przybocki, “The DET curve in assessment of detection task performance”, In *EUROSPEECH-1997*, pp. 1895-1898, 1997.

[49] R. Auckenthaler, M. Carey, H. Lloyd-Thomas, “Score Normalization for Text-Independent Speaker Verification Systems,” *Digital Signal Processing*, Volume 10, Issues 1-3, Pages 42-54, January 2000.

[50] A. Solomonoff, W.M. Campbell, I. Boardman, “Advances In Channel Compensation For SVM Speaker Recognition,” *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on*, vol.1, no., pp. 629-632, March 18-23, 2005.

[51] F. Zhonghua; X. Lei; Z. Rongchun, “Channel robust speaker verification via extended feature mapping,” *Signal Processing, 2004. Proceedings. ICSP '04. 2004 7th International Conference on*, vol.3, no., pp. 2417-2420 vol.3, 31 Aug.-4 Sept. 2004.

[52] D. Hardt, K. Fellbaum, “Spectral subtraction and RASTA-filtering in text-dependent HMM-based speaker verification,” *Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on*, vol.2, no., pp.867-870 vol.2, 21-24 Apr 1997.

[53] K. Li, J. Porter, “Normalizations and Selection of Speech Segments for Speaker Recognition Scoring”, *IEEE Conference on Acoustics, Speech, Signal Processing,* Vol. 1, pp. 595-598.

[54] D.A. Reynolds, “Channel robust speaker verification via feature mapping,” In *ICASSP-03*, vol. 2, pp. 53-56, 2003.

[55] Genesis 27:22-23

[56] A.D. Yarmey, M.J. Yarmey, L. Todd, “Frances McGehee (1912-2004) The First Earwitness Researcher” *Perceptual and Motor Skills*, 106: pp. 387-394, 2008.

[57] F. McGehee, “The Reliability of the Identification of the Human Voice,” *Journal of General Psychology*, vol. 17, 249–271, 1937.

[58] F. McGehee, “An experimental investigation of voice recognition,” *Journal of General Psychology*, vol. 31, 53–65, 1944.

[59] L.G. Kersta, “Voiceprint identification,” *Nature*, 196: 1253-1257, 1962.

[60] J. Lindh, “Handling the Voiceprint Issue”, in Proceedings FONETICK, pp. 72-75, 2004.

[61] J.W. Cooley, and J.W. Tukey, “An Algorithm for the Machine Computation of Complex Fourier Series,” *Math Computation*, Vol. 19, pp.297-301, April 1965.

[62] A. Oppenheim, and R. Shafer, “From Frequency to Quefrency: A History of the Cepstrum”, *IEEE Signal Processing*, September 2004.

[63] L. Baum, T. Petrie, G. Soules, and N. Weiss, “A Maximization Technique in the Statistical Analysis of Probabilistic Functions of Markov Chains”, *The Annals of Mathematical Statistics*, Vol. 41, No.1, 1970.

[64] J. Woodard, N. Orlans, and P. Higgins, Biometrics, McGraw-Hill, 2003.

[65] S. Pruzansky, “Pattern-Matching Procedure for Automatic Talker Recognition,” *J. Acoust. Soc. Amer.*, Vol. 35, pp. 354-358, 1963.

[66] D. A. Reynolds, and L.P. Heck, “Automatic Speaker Recognition: Recent Progress, Current Applications and Future Trends,” *American Association for the Advancement of Science (AAAS) Symposium*, February 2000.

[67] K. Kumar, Q. Wu, Y. Wang, and M. Savvides, “Noise Robust Speaker Identification Using Bhattacharyya Distance In Adapted Gaussian Models Space,” Pittsburgh, Carnegie Mellon University, 2008.

[68] Q. Jin, Y. Pan and T. Schultz, "Far-field Speaker Recognition", International Conference on Acoustic, Speech, and Signal Processing (ICASSP) 2006.

[69] G. Boré and S. Peus, "Microphones for Studio and Home-Recording Applications: Operation Principles and Type Examples," Druck-Centrum Fürst GmbH, Berlin, 4th ed., 1999.

[70] D.A. Reynolds, “HTIMIT and LLHDB : Speech Corpora for the Study of Handset Transducer Effects,” *ICASSP-97*, IEEE, pp. 1535-1538, Apr 1997.

[71] D.A. Reynolds, “The Effects of Handset Variability on Speaker Recognition Performance: Experiments on the Switchboard Corpus,” *ICASSP*, IEEE, Atlanta, Georgia, Vol. 1, pp. 113-116, 7-10 May 1996.

[72] A. Alexander, F. Botti, D. Dessimoz, and A. Drygajlo, "The effect of Mismatched recording conditions on human and automatic speaker recognition in forensic applications," Forensic Science International, Vol. 146, Pages S95-S99, 2 December 2004.

[73] J.M. Naik, L.P. Netsch and G.R. Doddington, “Speaker Verification Over Long Distance Telephone Lines,” *Proceedings on ICASSP*, IEEE, Vol. 1, Glasgow, UK, pp. 524-527, 23-26 May 1989.

[74] R.B. Dunn, T.F. Quatieri, D.A. Reynolds, J.P. Campbell, “Speaker recognition from coded speech and the effects of score normalization,” *Signals, Systems and Computers*, *Conference Record of the 35th Asilomar Conference on* , vol.2, no., pp.1562-1567 vol.2, 2001.

APPENDICES

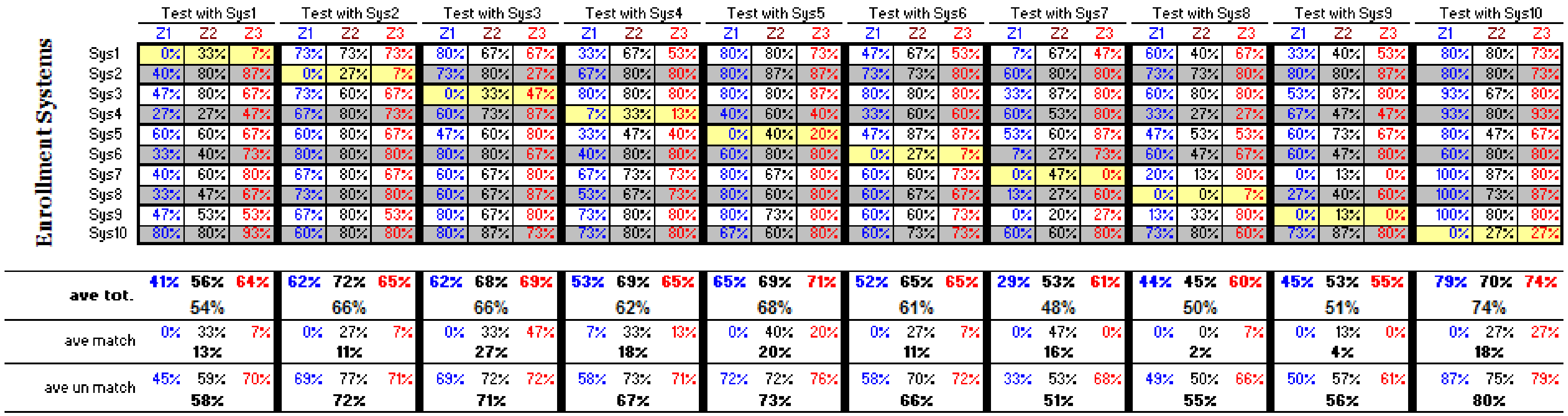
Appendix Page

APPENDIX A

Error Rates for ‘Nearest-To’ Algorithm

Appendix A provides a table of results from the ‘Nearest-To’ decision algorithm. Rows of the table dictate the system used for enrollment. In the top portion of the table the columns of the table indicate which system was tested and in which zone. Each cell’s percentage rate is the percentage rate of 15 samples including all 5 speakers saying Phrases 1-3. The bottom section of the table is a summary of the top portion. This includes matched (same microphone used in enrollment and testing phases) and mismatched conditions (microphone used in enrollment phase is differnet than microphone used in testing phase) for each of the zones, total error rates and total error rates for each zone and both conditions. This graph is further summarized in Table 4.2.

APPENDIX A: Error rates for 'Nearest-to' algorithm



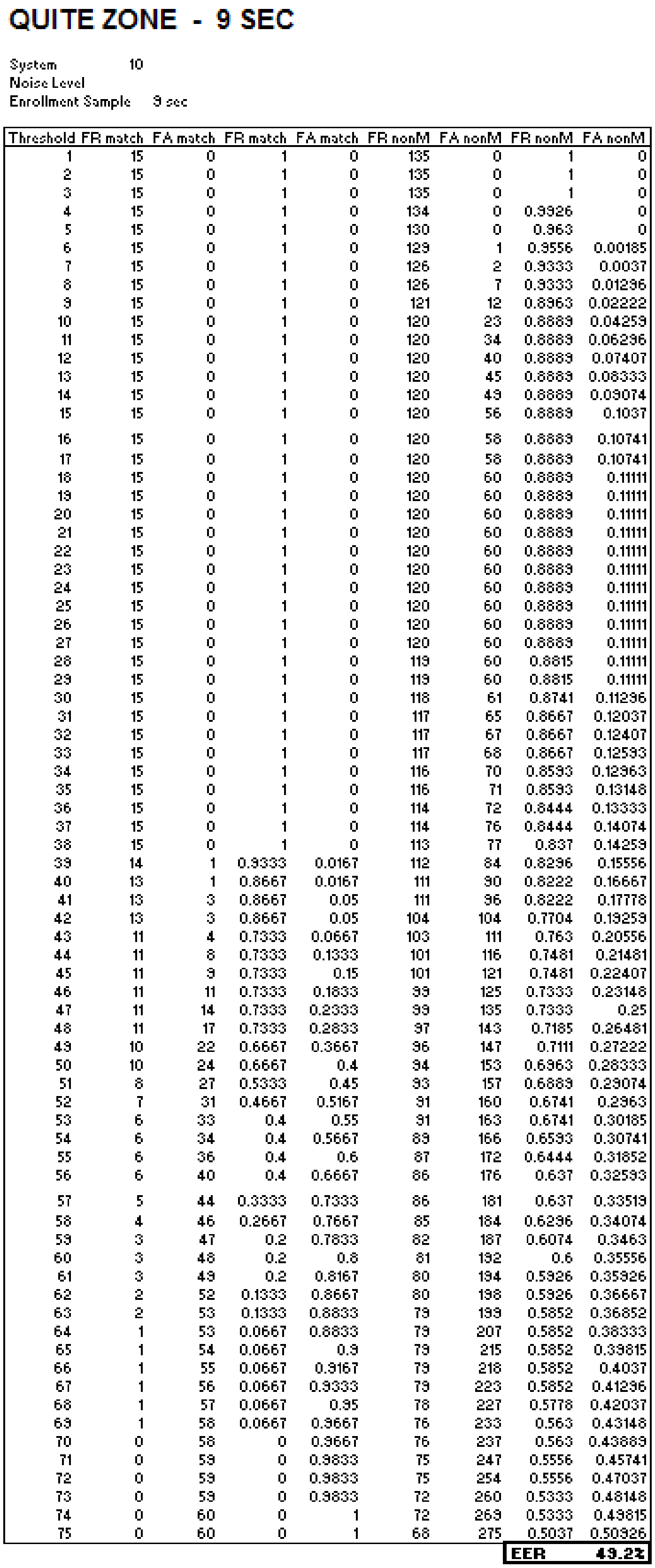
Enrollment Systems	Test with Sys1			Test with Sys2			Test with Sys3			Test with Sys4			Test with Sys5		
	Z1	Z2	Z3	Z1	Z2	Z3	Z1	Z2	Z3	Z1	Z2	Z3	Z1	Z2	Z3
Sys1	0%	33%	7%	73%	73%	73%	80%	67%	67%	33%	67%	53%	80%	80%	73%
Sys2	40%	80%	87%	0%	27%	7%	73%	80%	27%	67%	80%	80%	80%	87%	87%
Sys3	47%	80%	67%	73%	60%	67%	0%	33%	47%	80%	80%	80%	80%	80%	87%
Sys4	27%	27%	47%	67%	80%	73%	60%	73%	87%	7%	33%	13%	40%	60%	40%
Sys5	60%	60%	67%	60%	80%	67%	47%	60%	80%	33%	47%	40%	0%	40%	20%
Sys6	33%	40%	73%	80%	80%	80%	80%	80%	67%	40%	80%	80%	60%	80%	80%
Sys7	40%	60%	80%	67%	80%	67%	60%	67%	80%	67%	73%	73%	80%	67%	80%
Sys8	33%	47%	67%	73%	80%	80%	60%	67%	87%	53%	67%	73%	80%	60%	80%
Sys9	47%	53%	53%	67%	80%	53%	80%	67%	80%	73%	80%	80%	80%	73%	80%
Sys10	80%	80%	93%	60%	80%	80%	80%	87%	73%	73%	80%	80%	67%	60%	80%
ave tot.	41%	56%	64%	62%	72%	65%	62%	68%	69%	53%	69%	65%	65%	69%	71%
		54%			66%			66%			62%			68%	
ave match	0%	33%	7%	0%	27%	7%	0%	33%	47%	7%	33%	13%	0%	40%	20%
		13%			11%			27%			18%			20%	
ave un match	45%	59%	70%	69%	77%	71%	69%	72%	72%	58%	73%	71%	72%	72%	76%
		58%			72%			71%			67%			73%	

Enrollment Systems	Test with Sys6			Test with Sys7			Test with Sys8			Test with Sys9			Test with Sys10		
	Z1	Z2	Z3	Z1	Z2	Z3	Z1	Z2	Z3	Z1	Z2	Z3	Z1	Z2	Z3
Sys1	47%	67%	53%	7%	67%	47%	60%	40%	67%	33%	40%	53%	80%	80%	73%
Sys2	73%	73%	80%	60%	80%	80%	73%	73%	80%	80%	80%	87%	80%	80%	73%
Sys3	80%	80%	80%	33%	87%	80%	60%	80%	80%	53%	87%	80%	93%	67%	80%
Sys4	33%	60%	60%	60%	53%	80%	33%	27%	27%	67%	47%	47%	93%	80%	93%
Sys5	47%	87%	87%	53%	60%	87%	47%	53%	53%	60%	73%	67%	80%	47%	67%
Sys6	0%	27%	7%	7%	27%	73%	60%	47%	67%	60%	47%	80%	60%	80%	80%
Sys7	60%	60%	73%	0%	47%	0%	20%	13%	80%	0%	13%	0%	100%	87%	80%
Sys8	60%	67%	67%	13%	27%	60%	0%	0%	7%	27%	40%	60%	100%	73%	87%
Sys9	60%	60%	73%	0%	20%	27%	13%	33%	80%	0%	13%	0%	100%	80%	80%
Sys10	60%	73%	73%	60%	60%	80%	73%	80%	60%	73%	87%	80%	0%	27%	27%
ave tot.	52%	65%	65%	29%	53%	61%	44%	45%	60%	45%	53%	55%	79%	70%	74%
		61%			48%			50%			51%			74%	
ave match	0%	27%	7%	0%	47%	0%	0%	0%	7%	0%	13%	0%	0%	27%	27%
		11%			16%			2%			4%			18%	
ave un match	58%	70%	72%	33%	53%	68%	49%	50%	66%	50%	57%	61%	87%	75%	79%
		66%			51%			55%			56%			80%	

APPENDIX B

False Accept and False Reject Rates

Appendix B provides includes a sample table utilized for calculating and plotting DET curves and EER in the threshold algorithm. The FR and FA rates are shown for both the matched (same microphone used in enrollment and testing phases) and mismatched condition (microphone used in enrollment phase is differnet than microphone used in testing phase) for all users saying Phrases 1-3 on a particular enrollment system in a specific zone. The entire data set may be supplied upon request.



QUITE ZONE - 9 SEC

System	10
Noise Level	
Enrollment Sample	9 sec

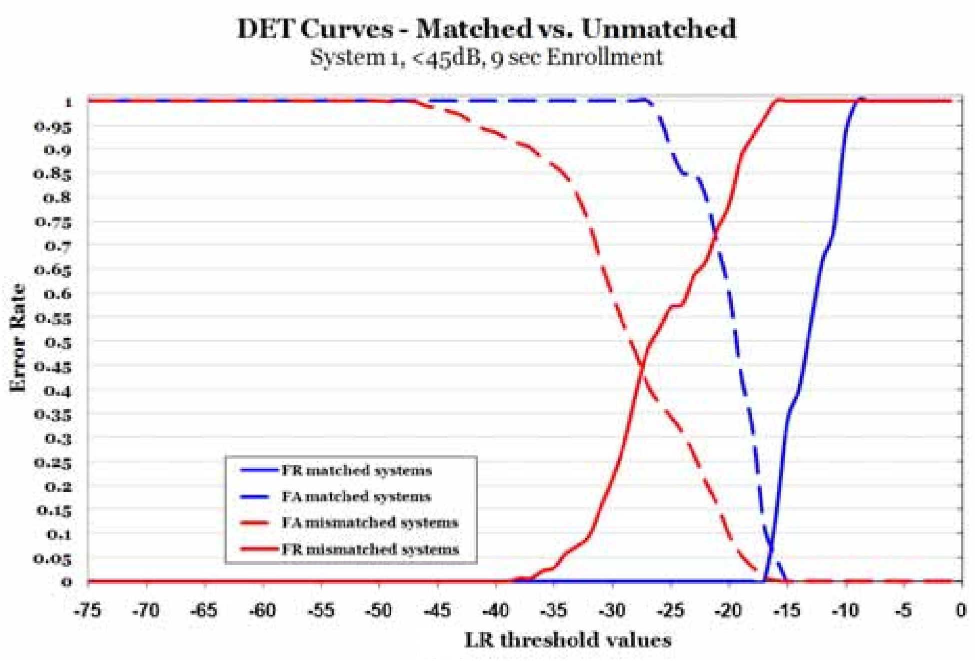
Threshold	FR match	FA match	FR match	FA match	FR nonM	FA nonM	FR nonM	FA nonM
1	15	0	1	0	135	0	1	0
2	15	0	1	0	135	0	1	0
3	15	0	1	0	135	0	1	0
4	15	0	1	0	134	0	0.9926	0
5	15	0	1	0	130	0	0.963	0
6	15	0	1	0	129	1	0.9556	0.00185
7	15	0	1	0	126	2	0.9333	0.0037
8	15	0	1	0	126	7	0.9333	0.01296
9	15	0	1	0	121	12	0.8963	0.02222
10	15	0	1	0	120	23	0.8889	0.04259
11	15	0	1	0	120	34	0.8889	0.06296
12	15	0	1	0	120	40	0.8889	0.07407
13	15	0	1	0	120	45	0.8889	0.08333
14	15	0	1	0	120	49	0.8889	0.09074
15	15	0	1	0	120	56	0.8889	0.1037
16	15	0	1	0	120	58	0.8889	0.10741
17	15	0	1	0	120	58	0.8889	0.10741
18	15	0	1	0	120	60	0.8889	0.11111
19	15	0	1	0	120	60	0.8889	0.11111
20	15	0	1	0	120	60	0.8889	0.11111
21	15	0	1	0	120	60	0.8889	0.11111
22	15	0	1	0	120	60	0.8889	0.11111
23	15	0	1	0	120	60	0.8889	0.11111
24	15	0	1	0	120	60	0.8889	0.11111
25	15	0	1	0	120	60	0.8889	0.11111
26	15	0	1	0	120	60	0.8889	0.11111
27	15	0	1	0	120	60	0.8889	0.11111
28	15	0	1	0	119	60	0.8815	0.11111
29	15	0	1	0	119	60	0.8815	0.11111
30	15	0	1	0	118	61	0.8741	0.11296
31	15	0	1	0	117	65	0.8667	0.12037
32	15	0	1	0	117	67	0.8667	0.12407
33	15	0	1	0	117	68	0.8667	0.12593
34	15	0	1	0	116	70	0.8593	0.12963
35	15	0	1	0	116	71	0.8593	0.13148
36	15	0	1	0	114	72	0.8444	0.13333
37	15	0	1	0	114	76	0.8444	0.14074
38	15	0	1	0	113	77	0.837	0.14259
39	14	1	0.9333	0.0167	112	84	0.8296	0.15556
40	13	1	0.8667	0.0167	111	90	0.8222	0.16667
41	13	3	0.8667	0.05	111	96	0.8222	0.17778
42	13	3	0.8667	0.05	104	104	0.7704	0.19259
43	11	4	0.7333	0.0667	103	111	0.763	0.20556
44	11	8	0.7333	0.1333	101	116	0.7481	0.21481
45	11	9	0.7333	0.15	101	121	0.7481	0.22407
46	11	11	0.7333	0.1833	99	125	0.7333	0.23148
47	11	14	0.7333	0.2333	99	135	0.7333	0.25
48	11	17	0.7333	0.2833	97	143	0.7185	0.26481
49	10	22	0.6667	0.3667	96	147	0.7111	0.27222
50	10	24	0.6667	0.4	94	153	0.6963	0.28333
51	8	27	0.5333	0.45	93	157	0.6889	0.29074
52	7	31	0.4667	0.5167	91	160	0.6741	0.2963
53	6	33	0.4	0.55	91	163	0.6741	0.30185
54	6	34	0.4	0.5667	89	166	0.6593	0.30741
55	6	36	0.4	0.6	87	172	0.6444	0.31852
56	6	40	0.4	0.6667	86	176	0.637	0.32593
57	5	44	0.3333	0.7333	86	181	0.637	0.33519
58	4	46	0.2667	0.7667	85	184	0.6296	0.34074
59	3	47	0.2	0.7833	82	187	0.6074	0.3463
60	3	48	0.2	0.8	81	192	0.6	0.35556
61	3	49	0.2	0.8167	80	194	0.5926	0.35926
62	2	52	0.1333	0.8667	80	198	0.5926	0.36667
63	2	53	0.1333	0.8833	79	199	0.5852	0.36852
64	1	53	0.0667	0.8833	79	207	0.5852	0.38333
65	1	54	0.0667	0.9	79	215	0.5852	0.39815
66	1	55	0.0667	0.9167	79	218	0.5852	0.4037
67	1	56	0.0667	0.9333	79	223	0.5852	0.41296
68	1	57	0.0667	0.95	78	227	0.5778	0.42037
69	1	58	0.0667	0.9667	76	233	0.563	0.43148
70	0	58	0	0.9667	76	237	0.563	0.43889
71	0	59	0	0.9833	75	247	0.5556	0.45741
72	0	59	0	0.9833	75	254	0.5556	0.47037
73	0	59	0	0.9833	72	260	0.5333	0.48148
74	0	60	0	1	72	269	0.5333	0.49815
75	0	60	0	1	68	275	0.5037	0.50926
							EER	49.2%

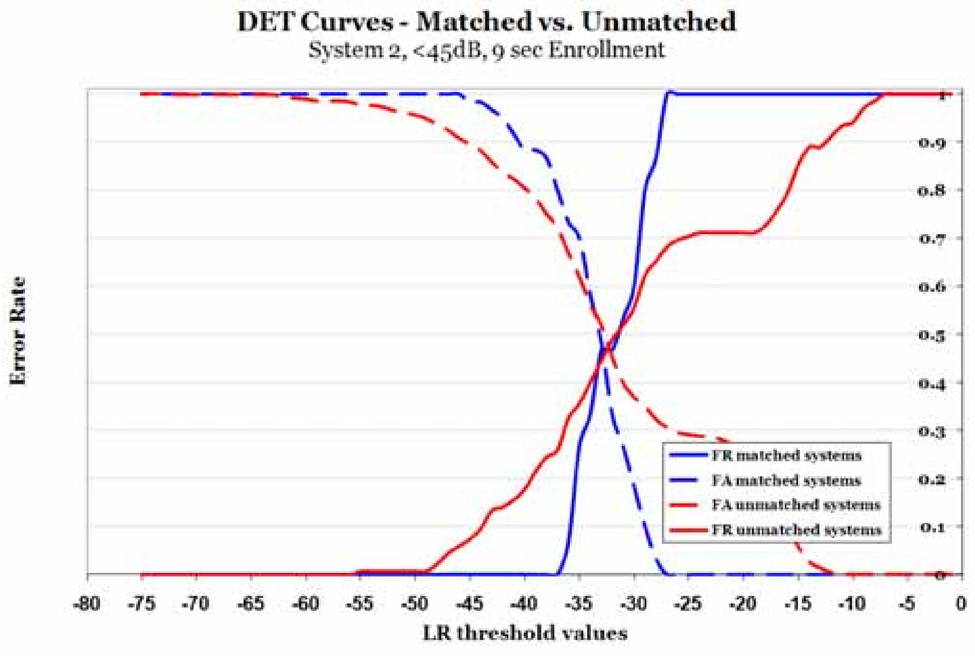
APPENDIX C

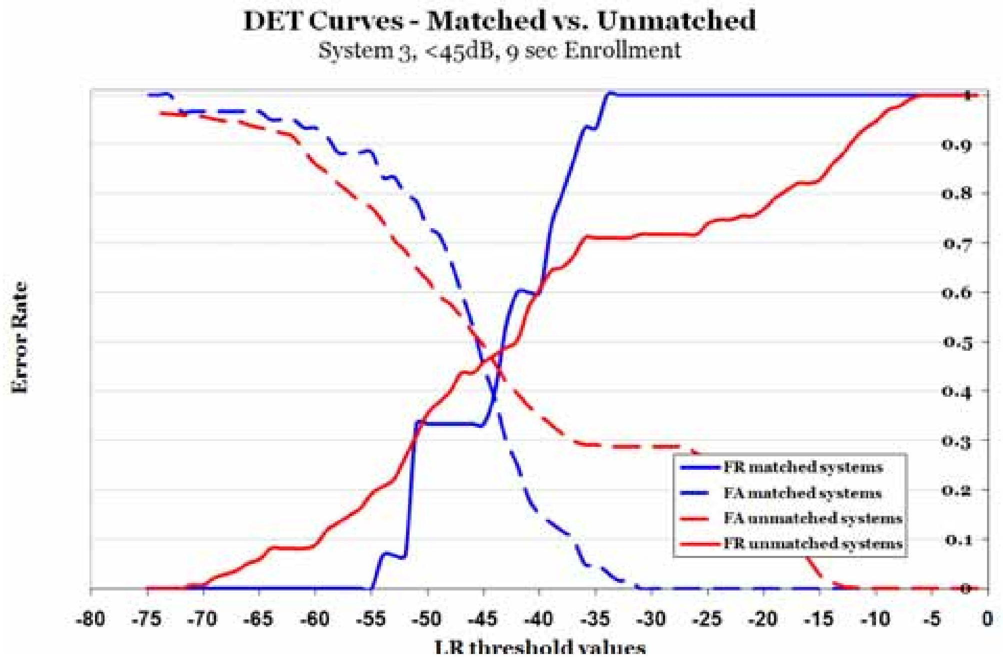
Detection Error Tradeoff Curves

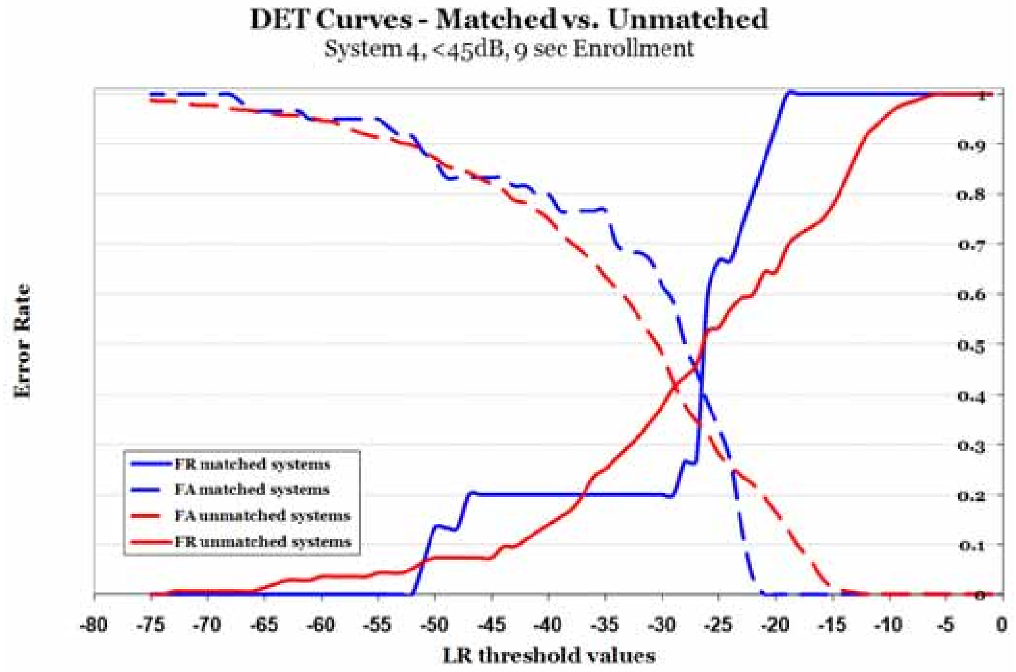
This Appendix provides detection error tradeoff (DET) curves for each enrollment system. Each DET curve represents all speakers saying Phrases 1-3 in a specific zone. The dashed lines represent false accept (FA) rates and the solid lines represent false reject (FR) rates. The X-axis in the DET curves are the log likelihood ratio, abbreviated LR.

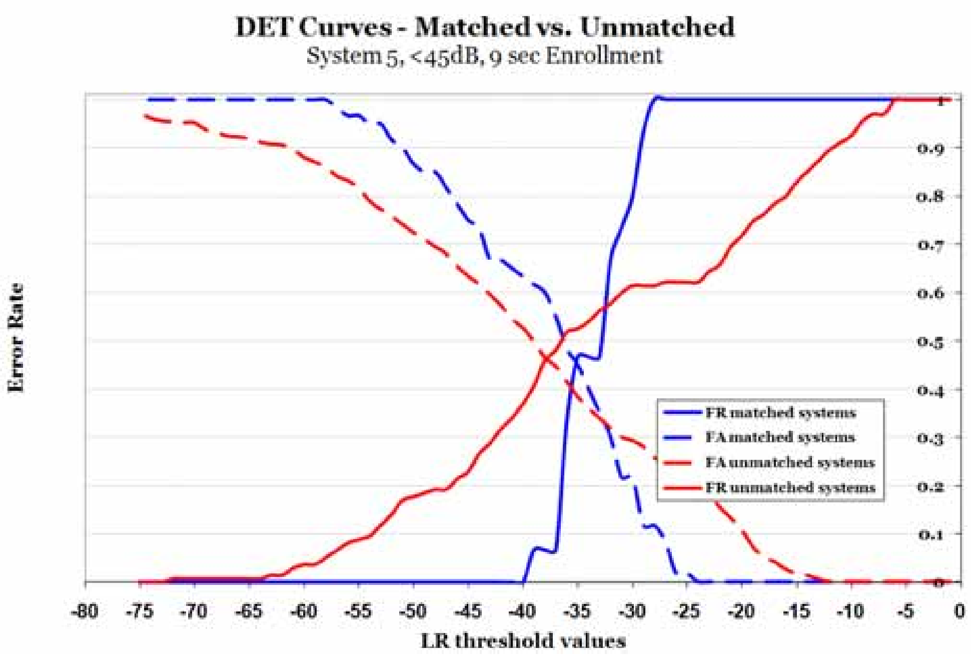
ZONE 1

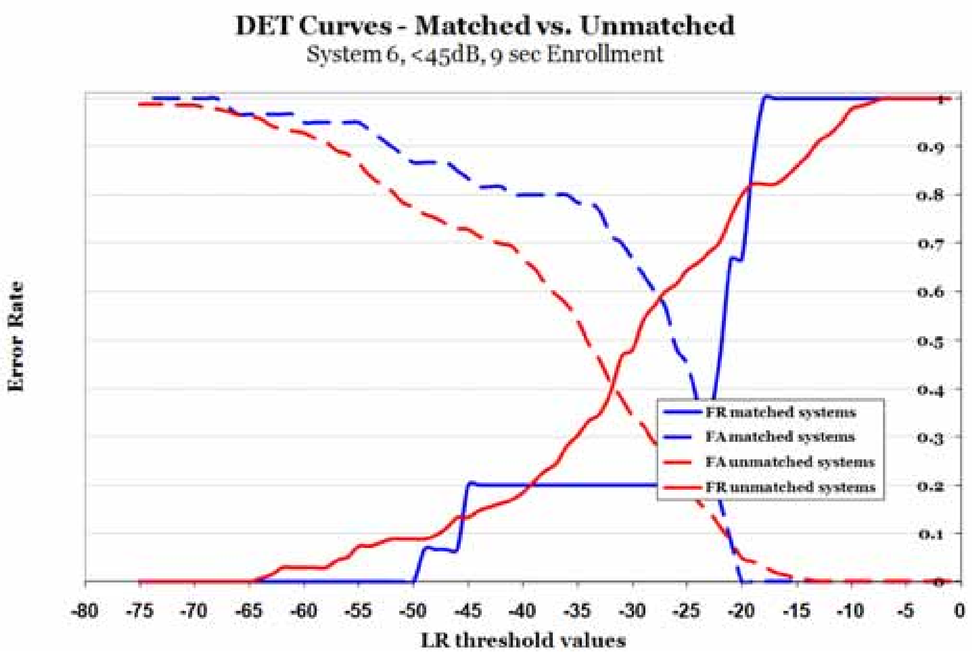

DET Curves - Matched vs. Unmatched
System 1, <45dB, 9 sec Enrollment
Error Rate
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-75
-70
-65
-60
-55
-50
-45
-40
-35
-30
-25
-20
-15
-10
-5
0
LR threshold values
FR matched systems
FA matched systems
FA mismatched systems
FR mismatched systems

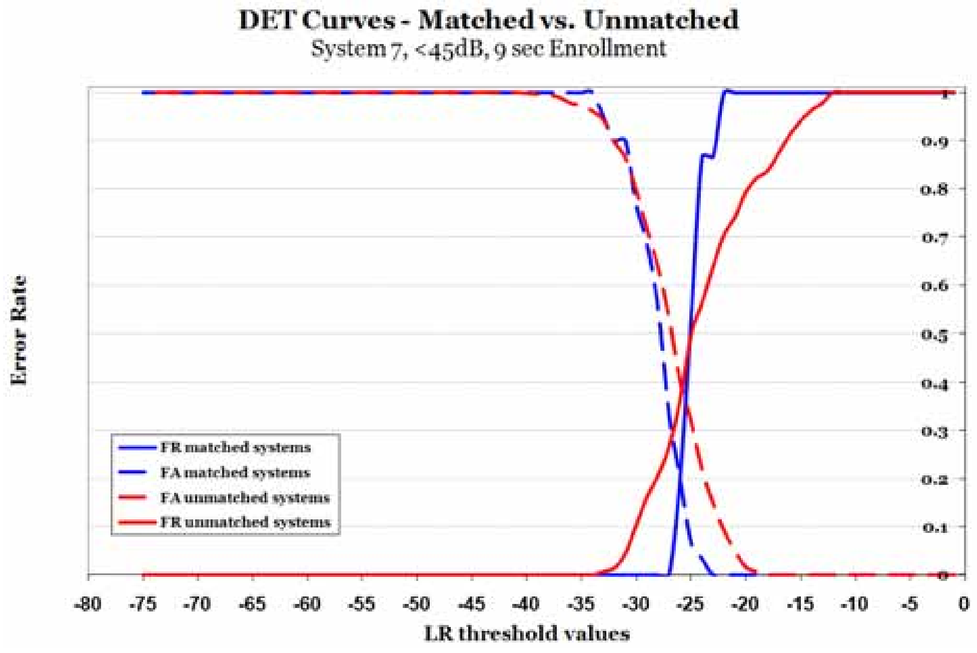

DET Curves - Matched vs. Unmatched
System 2, <45dB, 9 sec Enrollment
Error Rate
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-80
-75
-70
-65
-60
-55
-50
-45
-40
-35
-30
-25
-20
-15
-10
-5
0
LR threshold values
FR matched systems
FA matched systems
FA unmatched systems
FR unmatched systems

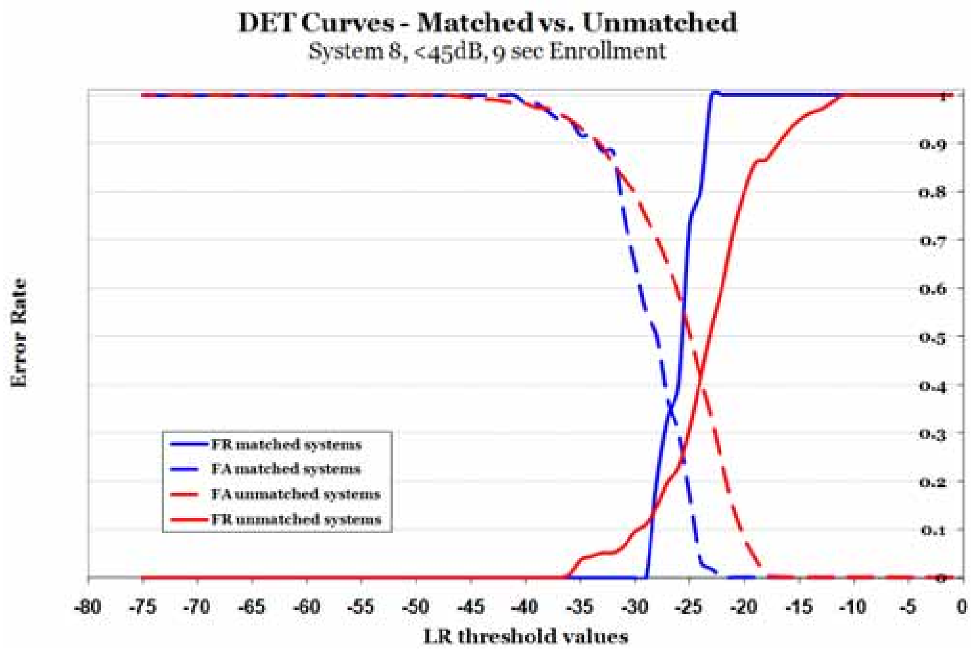

DET Curves - Matched vs. Unmatched
System 3, <45dB, 9 sec Enrollment
Error Rate
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-80
-75
-70
-65
-60
-55
-50
-45
-40
-35
-30
-25
-20
-15
-10
-5
0
LR threshold values
FR matched systems
FA matched systems
FA unmatched systems
FR unmatched systems

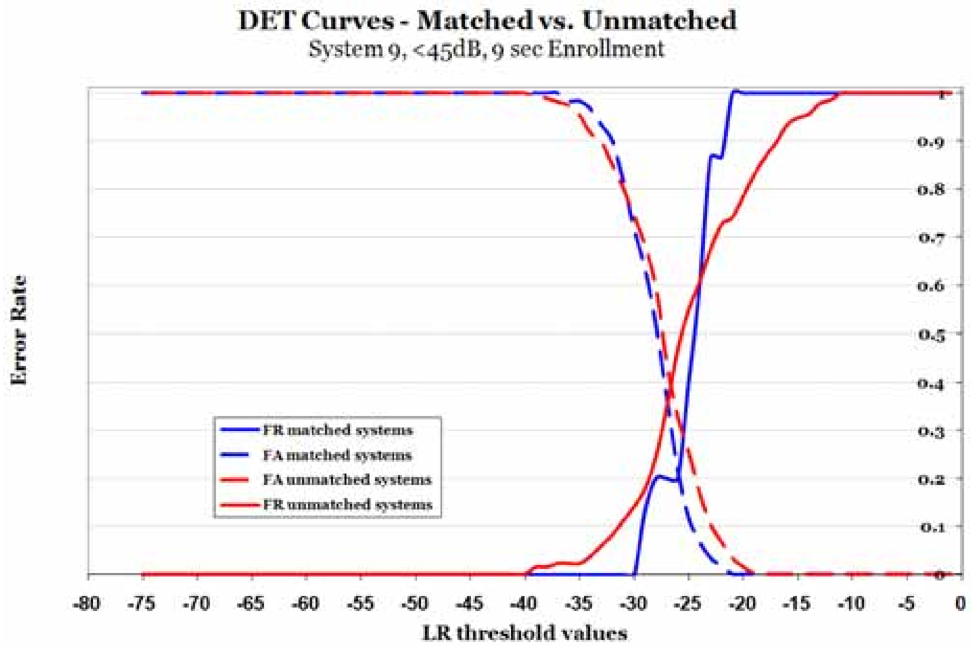

DET Curves - Matched vs. Unmatched
System 4, <45dB, 9 sec Enrollment
Error Rate
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-80
-75
-70
-65
-60
-55
-50
-45
-40
-35
-30
-25
-20
-15
-10
-5
0
LR threshold values
FR matched systems
FA matched systems
FA unmatched systems
FR unmatched systems

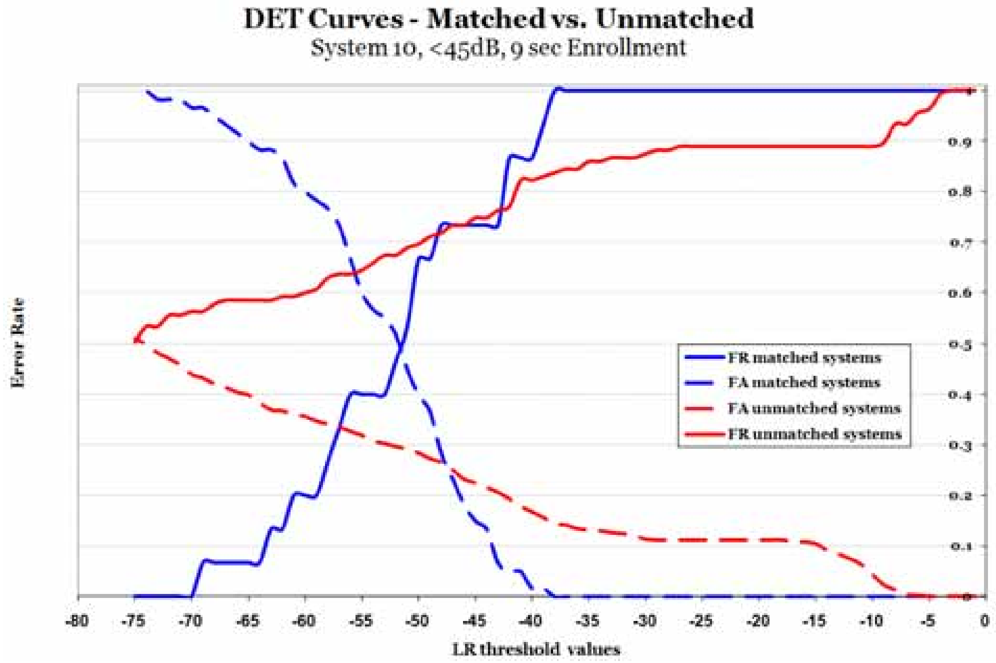

DET Curves - Matched vs. Unmatched
System 5, <45dB, 9 sec Enrollment
Error Rate
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
-80
-75
-70
-65
-60
-55
-50
-45
-40
-35
-30
-25
-20
-15
-10
-5
0
LR threshold values
FR matched systems
FA matched systems
FA unmatched systems
FR unmatched systems


DET Curves - Matched vs. Unmatched
System 6, <45dB, 9 sec Enrollment
Error Rate
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
-80
-75
-70
-65
-60
-55
-50
-45
-40
-35
-30
-25
-20
-15
-10
-5
0
LR threshold values
FR matched systems
FA matched systems
FA unmatched systems
FR unmatched systems

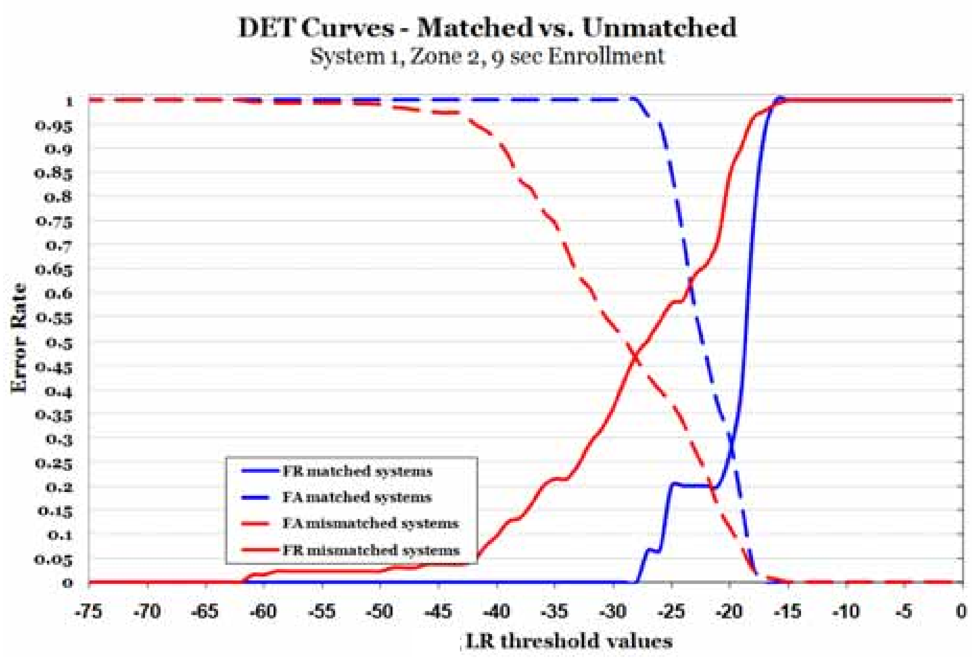

DET Curves - Matched vs. Unmatched
System 7, <45dB, 9 sec Enrollment
Error Rate
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
FR matched systems
FA matched systems
FA unmatched systems
FR unmatched systems
-80 -75 -70 -65 -60 -55 -50 -45 -40 -35 -30 -25 -20 -15 -10 -5 0
LR threshold values

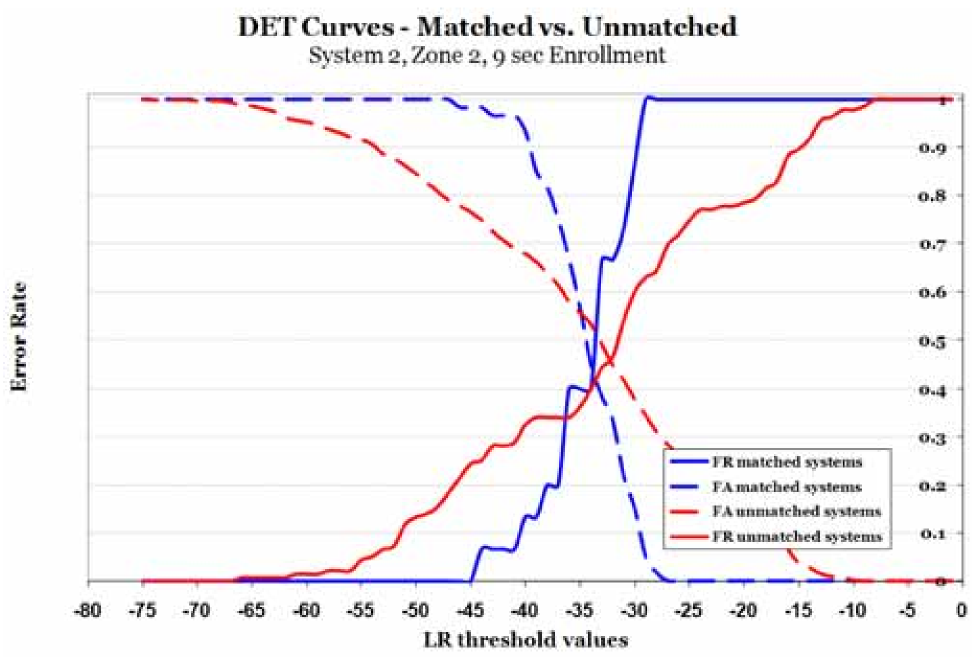

DET Curves - Matched vs. Unmatched
System 8, <45dB, 9 sec Enrollment
Error Rate
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
FR matched systems
FA matched systems
FA unmatched systems
FR unmatched systems
-80 -75 -70 -65 -60 -55 -50 -45 -40 -35 -30 -25 -20 -15 -10 -5 0
LR threshold values

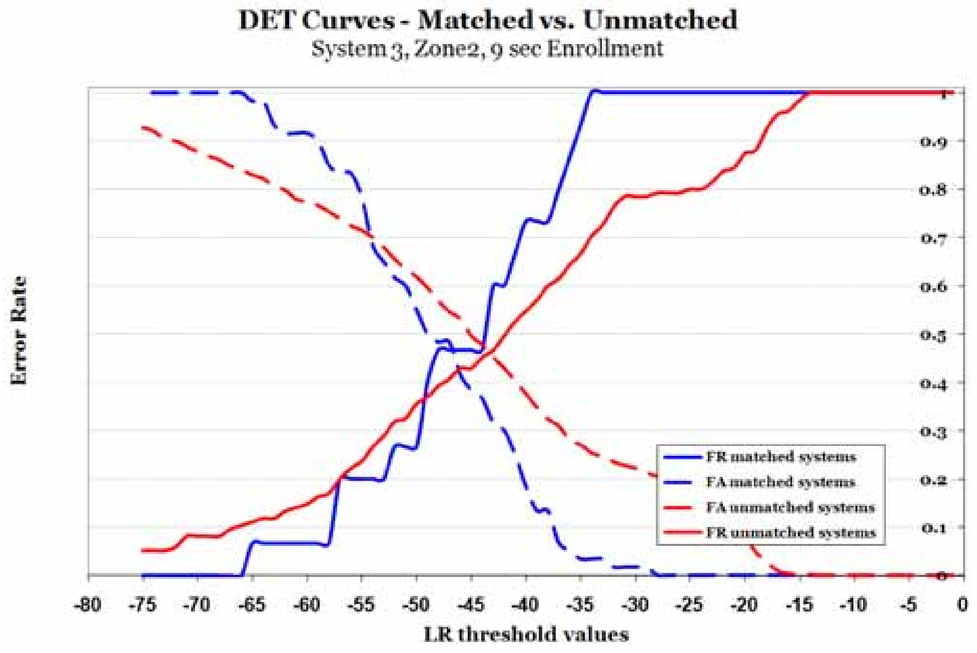

DET Curves - Matched vs. Unmatched
System 9, <45dB, 9 sec Enrollment
Error Rate
LR threshold values
FR matched systems
FA matched systems
FA unmatched systems
FR unmatched systems
-80
-75
-70
-65
-60
-55
-50
-45
-40
-35
-30
-25
-20
-15
-10
-5
0
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0


DET Curves - Matched vs. Unmatched
System 10, <45dB, 9 sec Enrollment
Error Rate
LR threshold values
FR matched systems
FA matched systems
FA unmatched systems
FR unmatched systems
-80
-75
-70
-65
-60
-55
-50
-45
-40
-35
-30
-25
-20
-15
-10
-5
0
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0

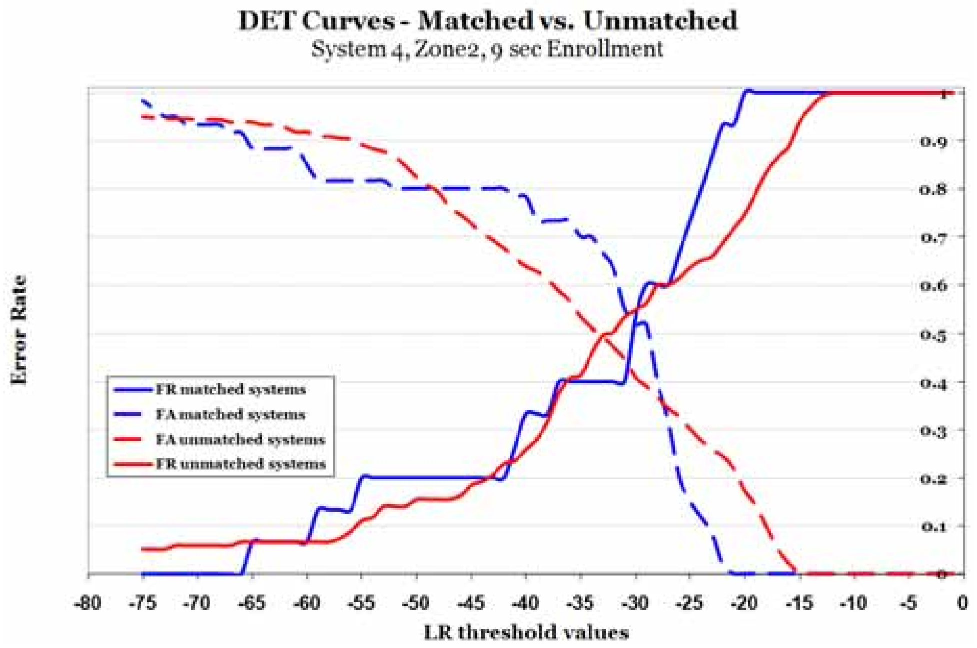
ZONE 2


DET Curves - Matched vs. Unmatched
System 1, Zone 2, 9 sec Enrollment
Error Rate
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
FR matched systems
FA matched systems
FA mismatched systems
FR mismatched systems
-75
-70
-65
-60
-55
-50
-45
-40
-35
-30
-25
-20
-15
-10
-5
0
LR threshold values

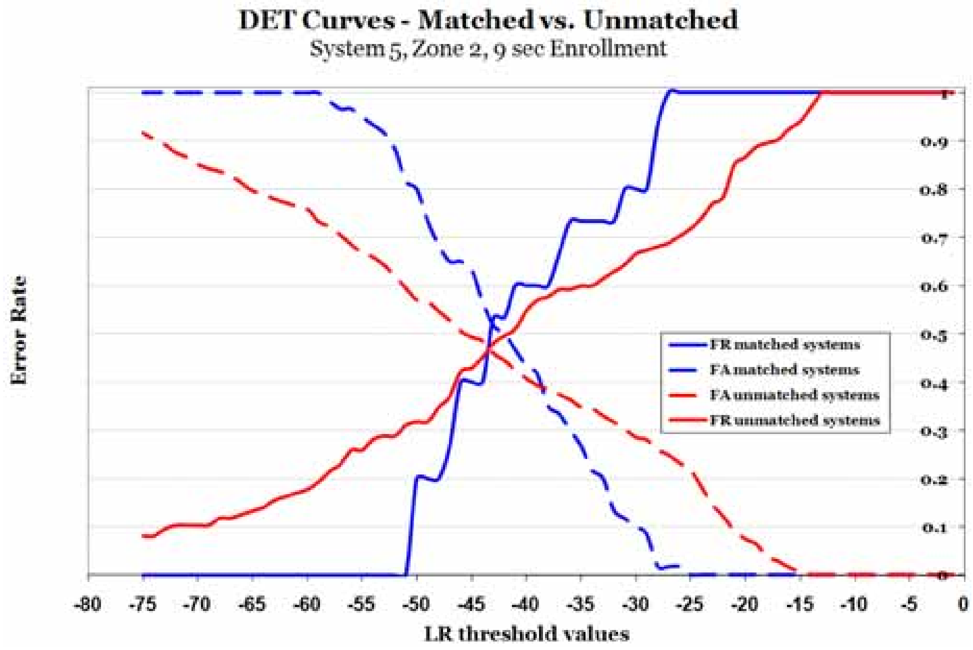

DET Curves - Matched vs. Unmatched
System 2, Zone 2, 9 sec Enrollment
Error Rate
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
FR matched systems
FA matched systems
FA unmatched systems
FR unmatched systems
-80
-75
-70
-65
-60
-55
-50
-45
-40
-35
-30
-25
-20
-15
-10
-5
0
LR threshold values

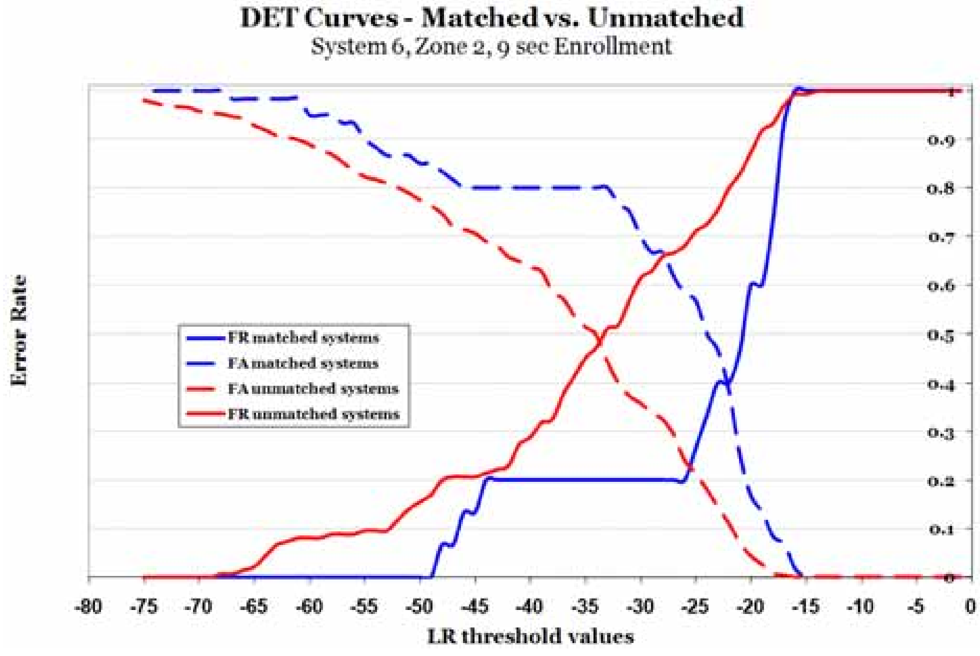


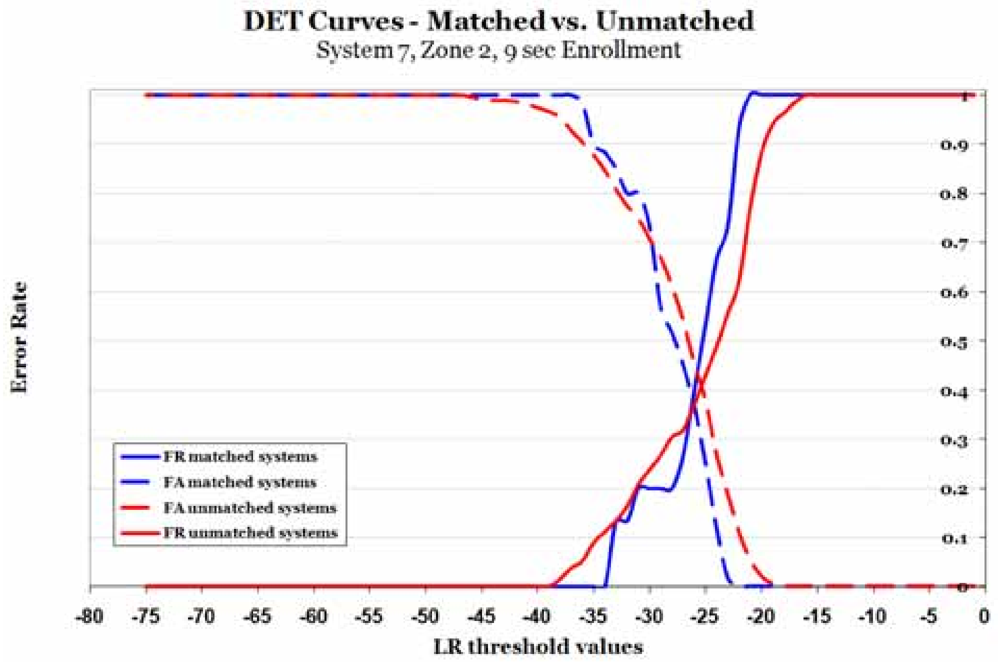
DET Curves - Matched vs. Unmatched
System 3, Zone2, 9 sec Enrollment

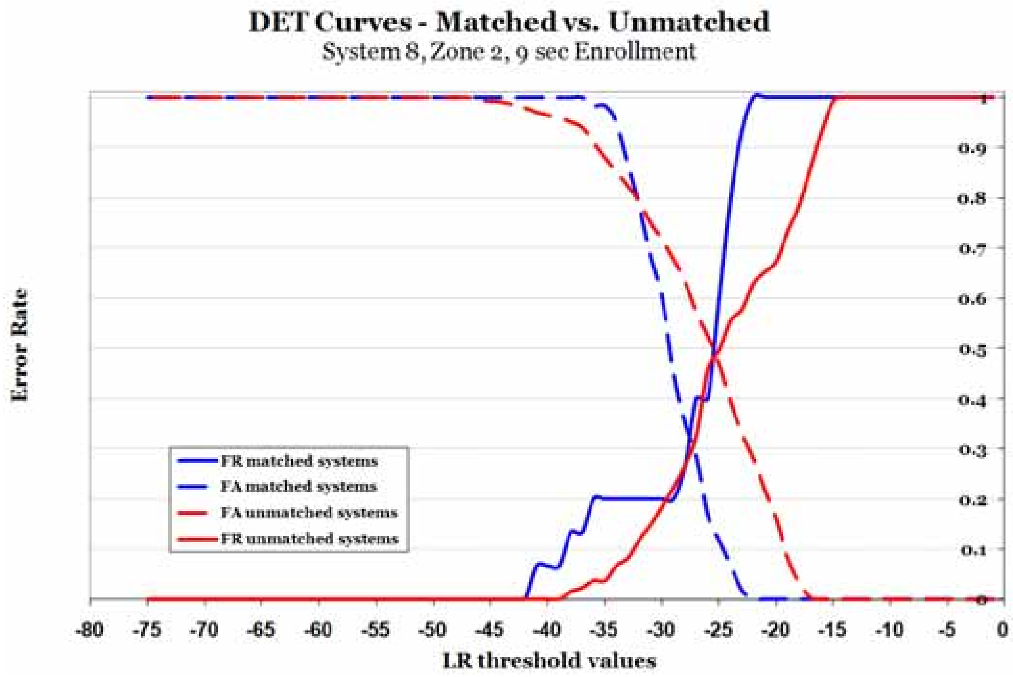


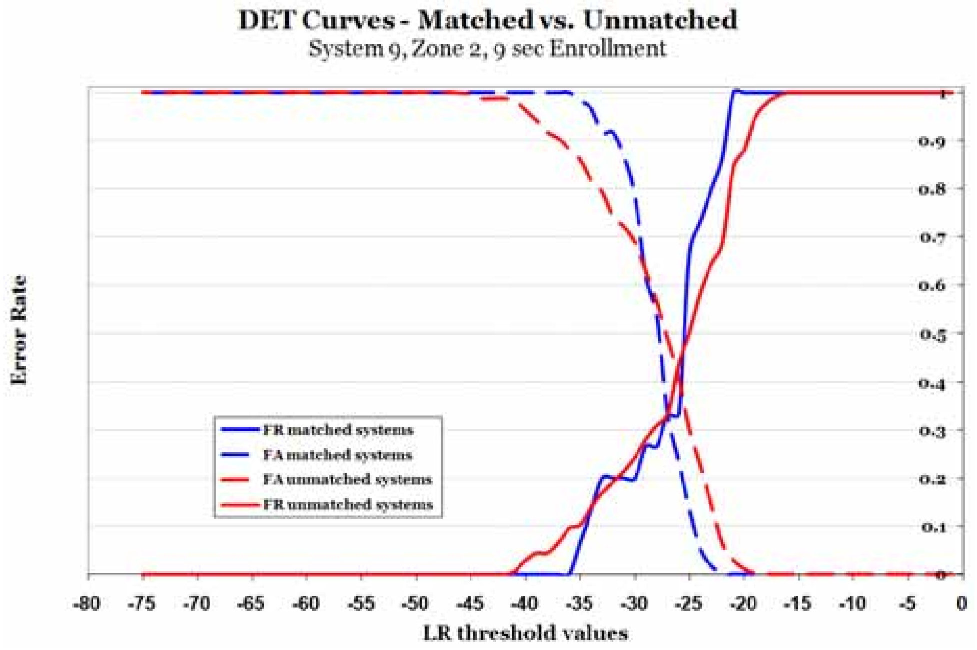
DET Curves - Matched vs. Unmatched
System 4, Zone2, 9 sec Enrollment

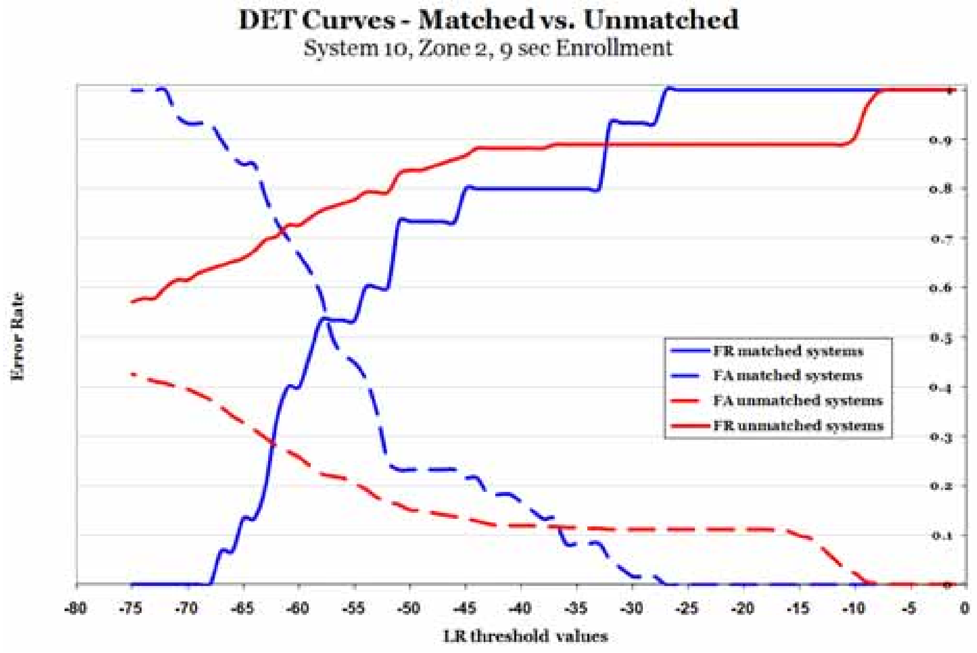

DET Curves - Matched vs. Unmatched
System 5, Zone 2, 9 sec Enrollment
Error Rate
LR threshold values
-80
-75
-70
-65
-60
-55
-50
-45
-40
-35
-30
-25
-20
-15
-10
-5
0
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
FR matched systems
FA matched systems
FA unmatched systems
FR unmatched systems


DET Curves - Matched vs. Unmatched
System 6, Zone 2, 9 sec Enrollment
Error Rate
LR threshold values
-80
-75
-70
-65
-60
-55
-50
-45
-40
-35
-30
-25
-20
-15
-10
-5
0
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
FR matched systems
FA matched systems
FA unmatched systems
FR unmatched systems

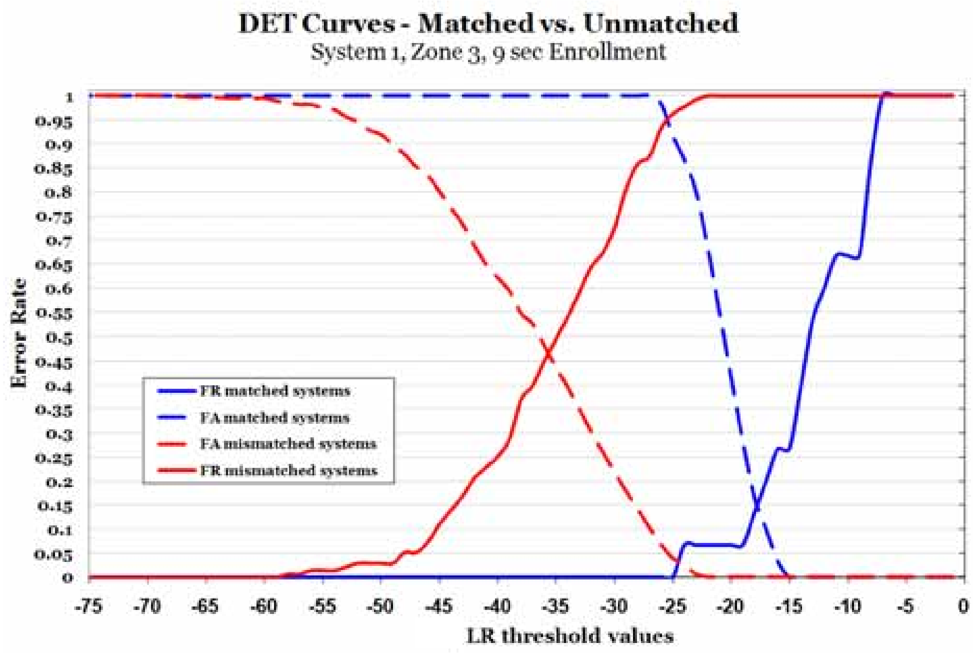

DET Curves - Matched vs. Unmatched
System 7, Zone 2, 9 sec Enrollment
Error Rate
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
FR matched systems
FA matched systems
FA unmatched systems
FR unmatched systems
-80
-75
-70
-65
-60
-55
-50
-45
-40
-35
-30
-25
-20
-15
-10
-5
0
LR threshold values


DET Curves - Matched vs. Unmatched
System 8, Zone 2, 9 sec Enrollment
Error Rate
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
FR matched systems
FA matched systems
FA unmatched systems
FR unmatched systems
-80
-75
-70
-65
-60
-55
-50
-45
-40
-35
-30
-25
-20
-15
-10
-5
0
LR threshold values

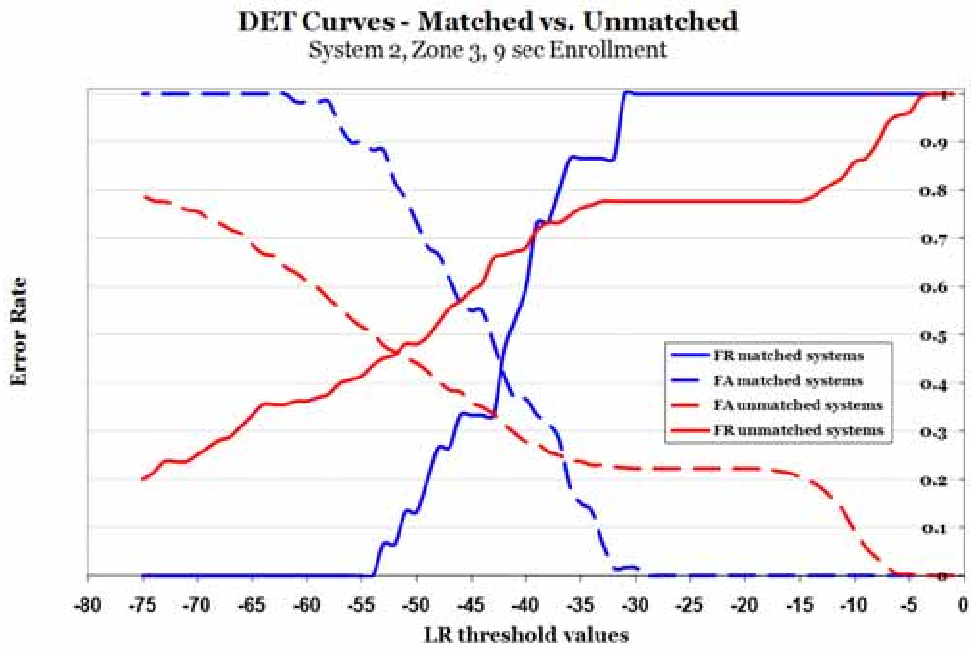

DET Curves - Matched vs. Unmatched
System 9, Zone 2, 9 sec Enrollment
Error Rate
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-80
-75
-70
-65
-60
-55
-50
-45
-40
-35
-30
-25
-20
-15
-10
-5
0
LR threshold values
FR matched systems
FA matched systems
FA unmatched systems
FR unmatched systems


DET Curves - Matched vs. Unmatched
System 10, Zone 2, 9 sec Enrollment
Error Rate
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-80
-75
-70
-65
-60
-55
-50
-45
-40
-35
-30
-25
-20
-15
-10
-5
0
LR threshold values
FR matched systems
FA matched systems
FA unmatched systems
FR unmatched systems

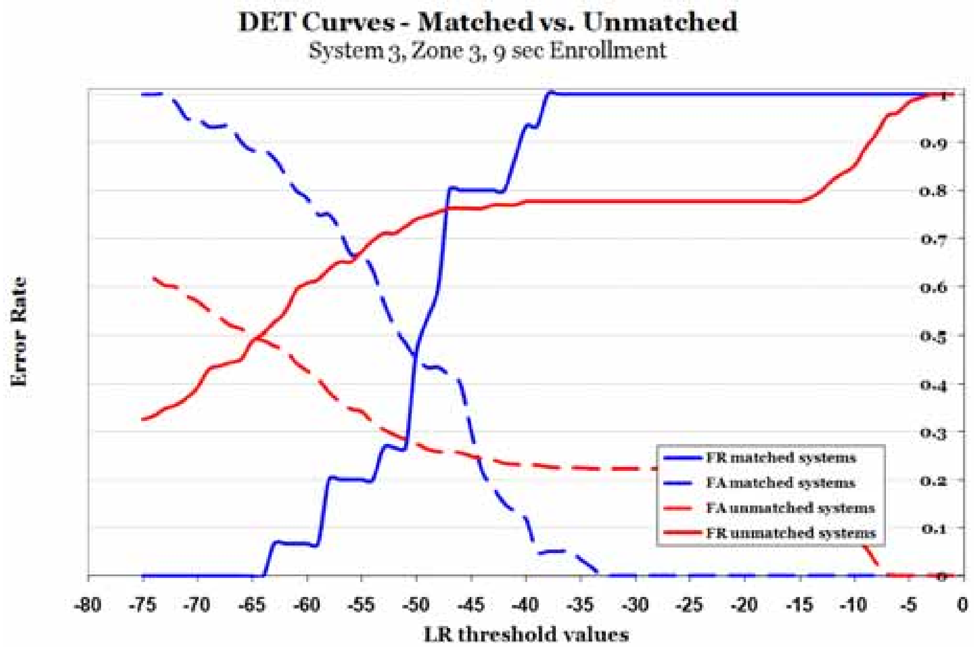
ZONE 3

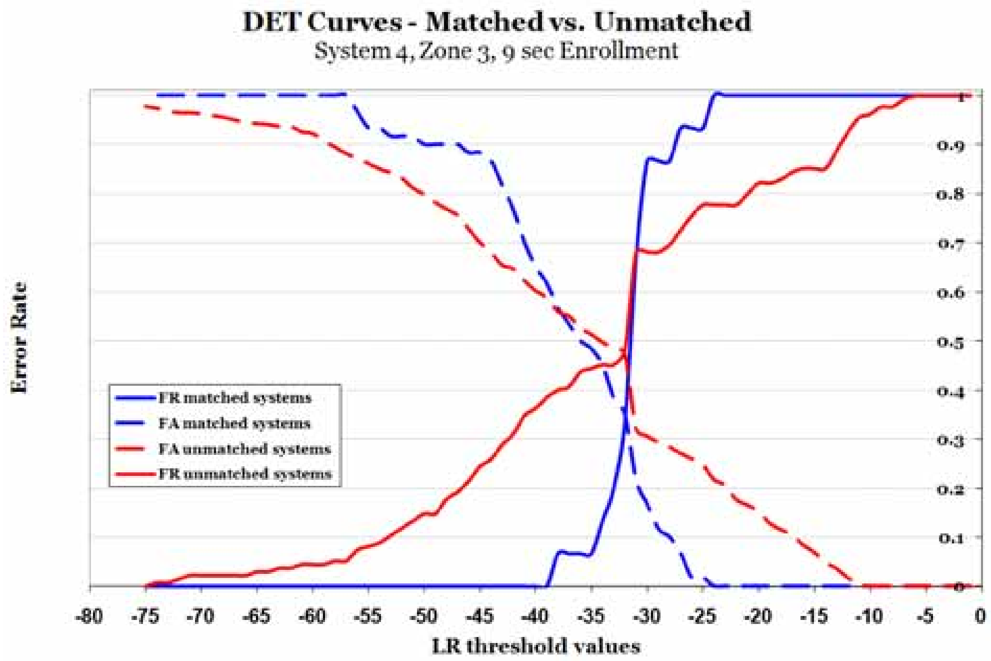


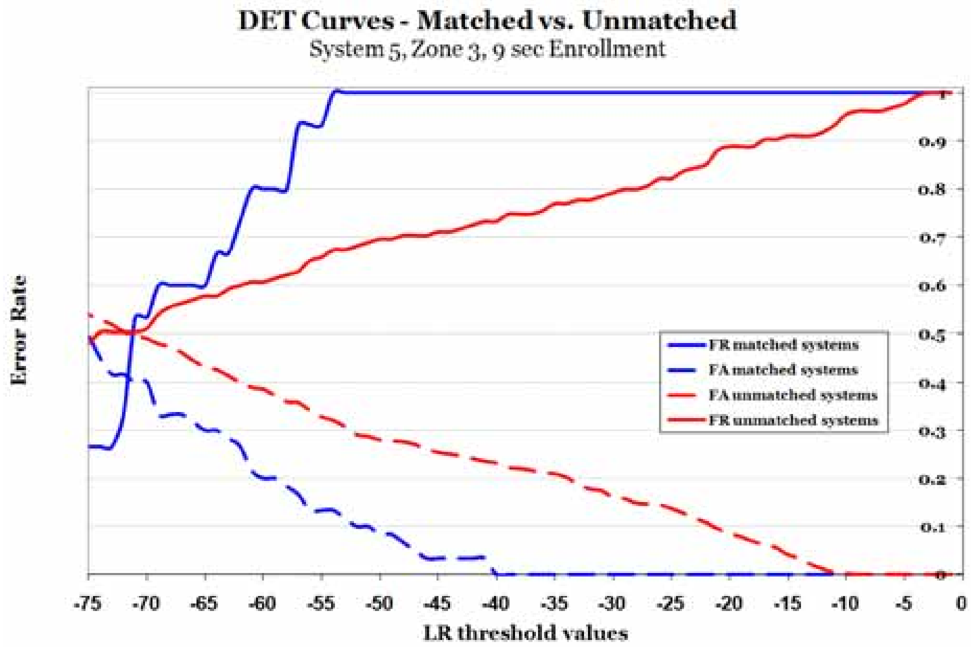
DET Curves - Matched vs. Unmatched
System 1, Zone 3, 9 sec Enrollment
Error Rate
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
FR matched systems
FA matched systems
FA mismatched systems
FR mismatched systems
-75
-70
-65
-60
-55
-50
-45
-40
-35
-30
-25
-20
-15
-10
-5
0
LR threshold values

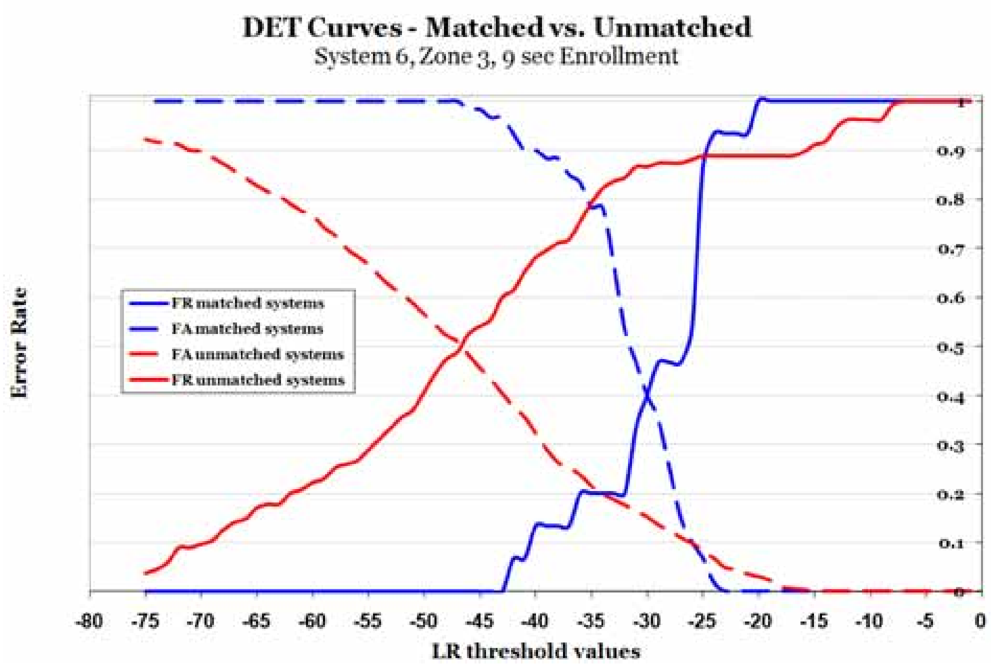


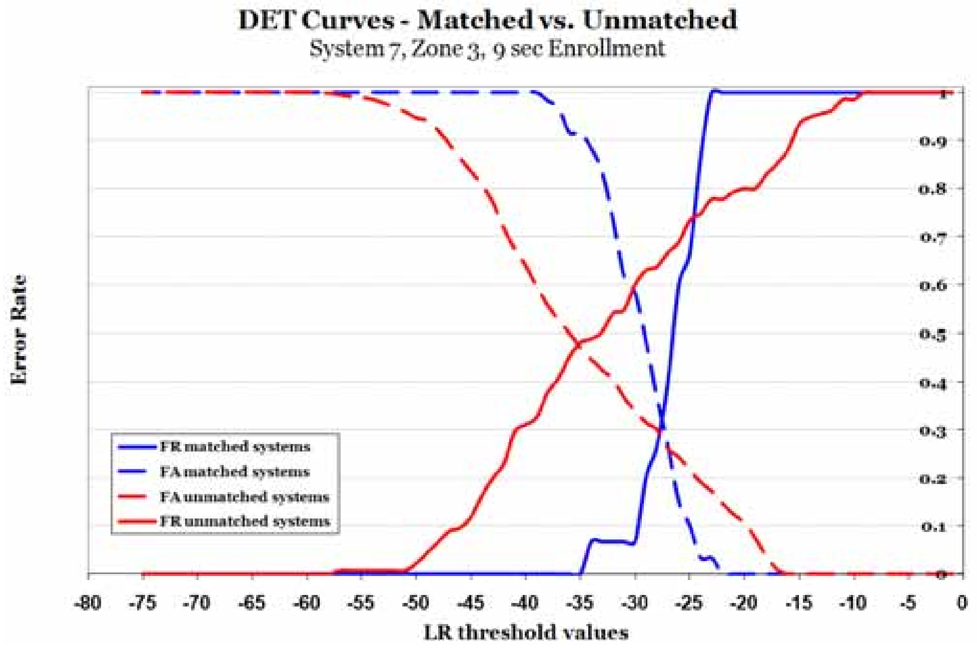
DET Curves - Matched vs. Unmatched
System 2, Zone 3, 9 sec Enrollment
Error Rate
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
FR matched systems
FA matched systems
FA unmatched systems
FR unmatched systems
-80
-75
-70
-65
-60
-55
-50
-45
-40
-35
-30
-25
-20
-15
-10
-5
0
LR threshold values

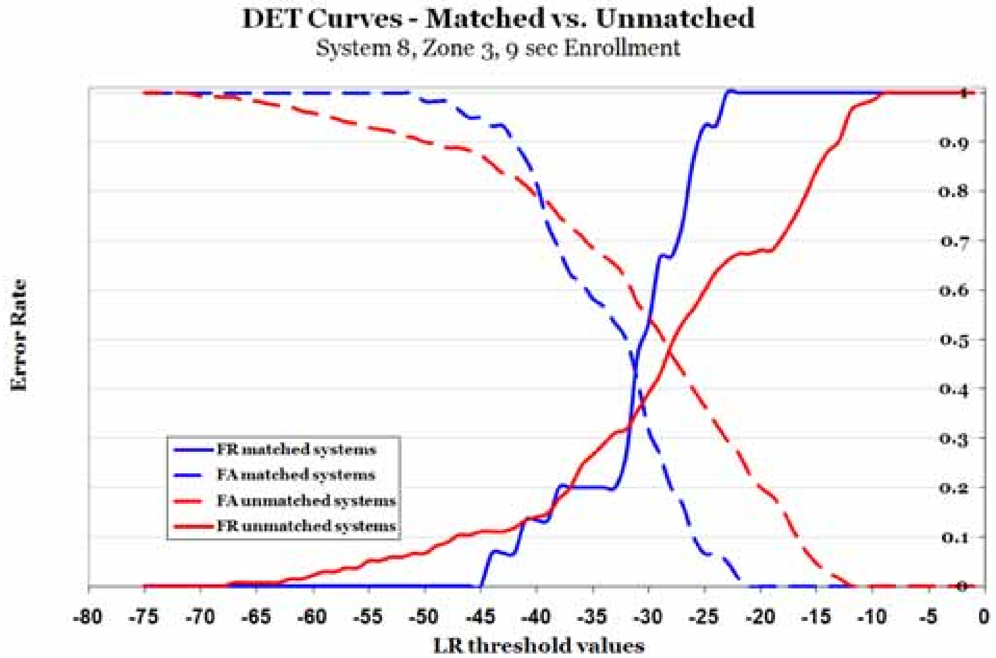

DET Curves - Matched vs. Unmatched
System 3, Zone 3, 9 sec Enrollment
Error Rate
LR threshold values
-80
-75
-70
-65
-60
-55
-50
-45
-40
-35
-30
-25
-20
-15
-10
-5
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
FR matched systems
FA matched systems
FA unmatched systems
FR unmatched systems

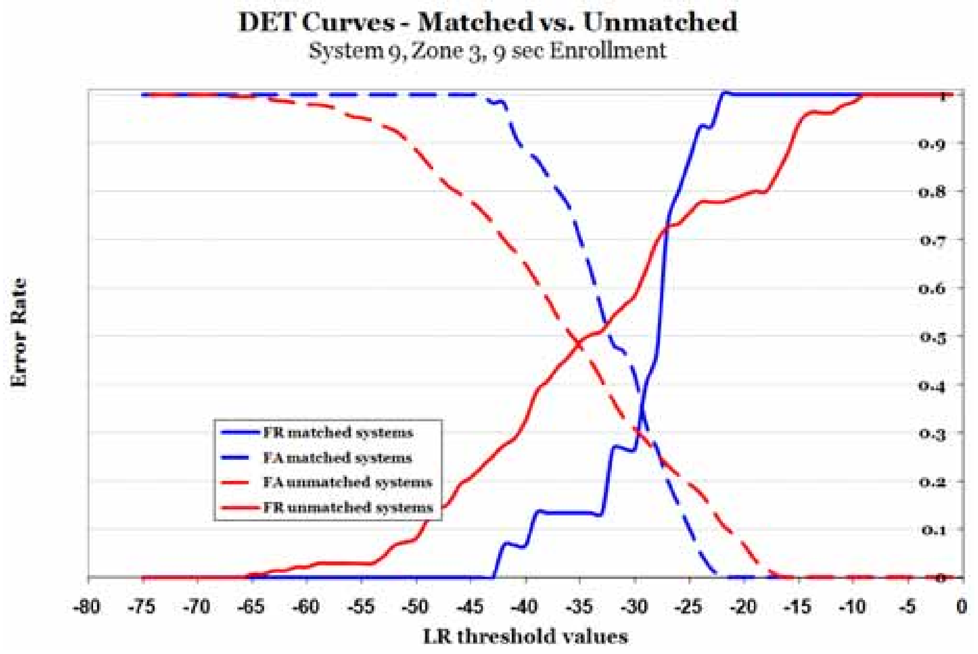

DET Curves - Matched vs. Unmatched
System 4, Zone 3, 9 sec Enrollment
Error Rate
LR threshold values
-80
-75
-70
-65
-60
-55
-50
-45
-40
-35
-30
-25
-20
-15
-10
-5
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
FR matched systems
FA matched systems
FA unmatched systems
FR unmatched systems

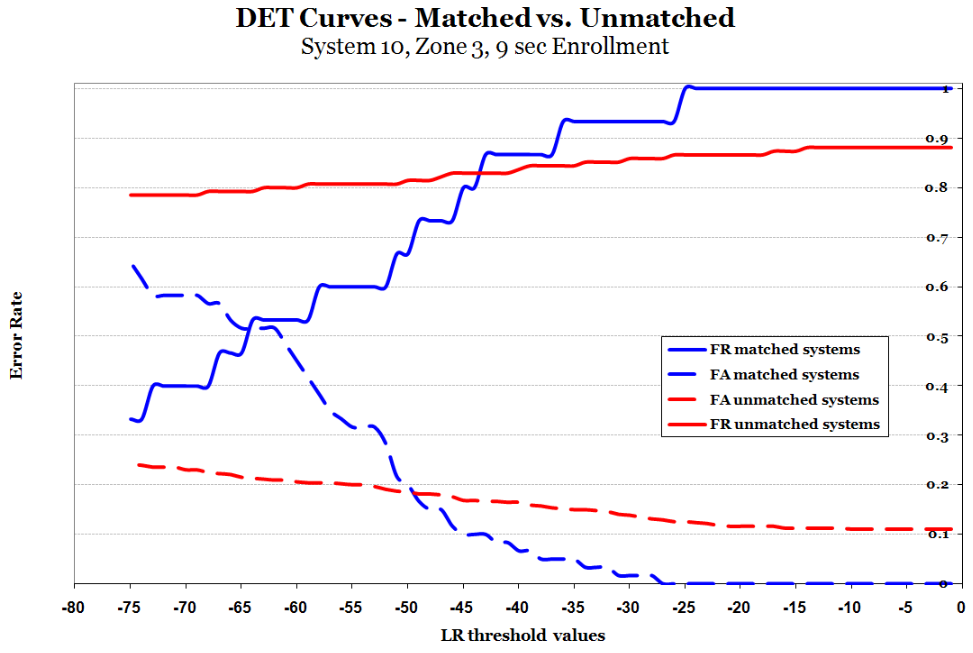

DET Curves - Matched vs. Unmatched
System 5, Zone 3, 9 sec Enrollment
Error Rate
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
FR matched systems
FA matched systems
FA unmatched systems
FR unmatched systems
-75
-70
-65
-60
-55
-50
-45
-40
-35
-30
-25
-20
-15
-10
-5
0
LR threshold values


DET Curves - Matched vs. Unmatched
System 6, Zone 3, 9 sec Enrollment
Error Rate
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
FR matched systems
FA matched systems
FA unmatched systems
FR unmatched systems
-80
-75
-70
-65
-60
-55
-50
-45
-40
-35
-30
-25
-20
-15
-10
-5
0
LR threshold values


DET Curves - Matched vs. Unmatched
System 7, Zone 3, 9 sec Enrollment
Error Rate
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
FR matched systems
FA matched systems
FA unmatched systems
FR unmatched systems
-80
-75
-70
-65
-60
-55
-50
-45
-40
-35
-30
-25
-20
-15
-10
-5
0
LR threshold values


DET Curves - Matched vs. Unmatched
System 8, Zone 3, 9 sec Enrollment
Error Rate
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
FR matched systems
FA matched systems
FA unmatched systems
FR unmatched systems
-80
-75
-70
-65
-60
-55
-50
-45
-40
-35
-30
-25
-20
-15
-10
-5
0
LR threshold values

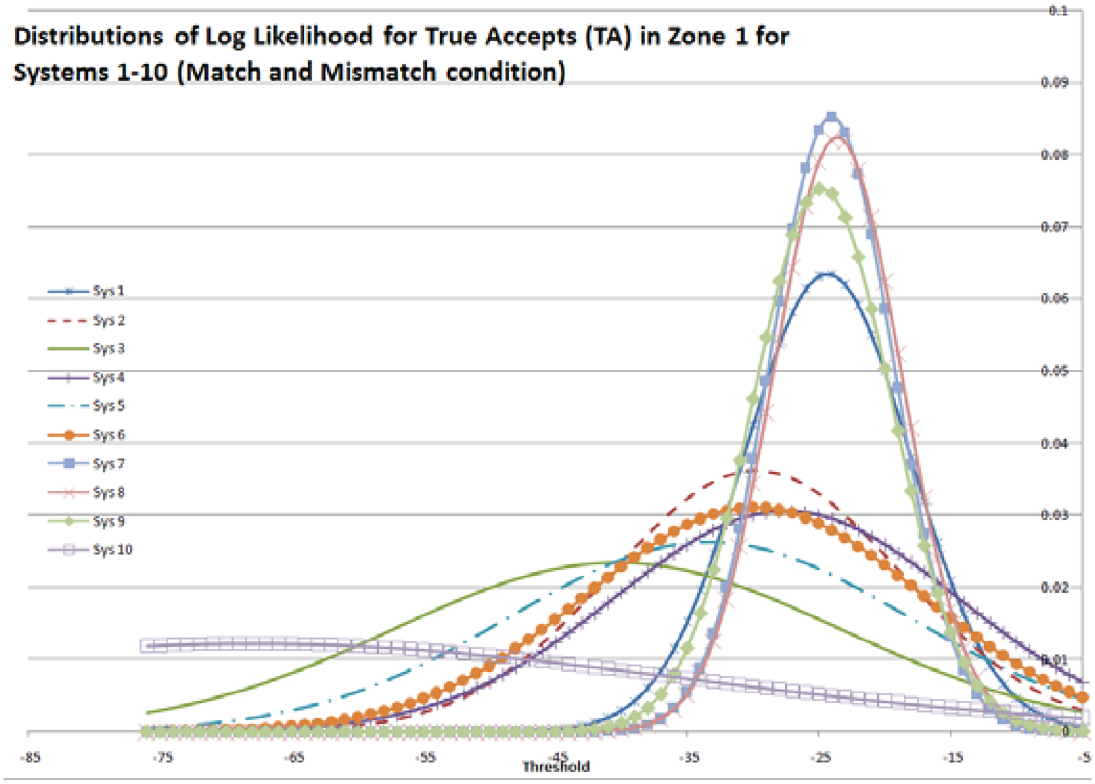

DET Curves - Matched vs. Unmatched
System 9, Zone 3, 9 sec Enrollment
Error Rate
FR matched systems
FA matched systems
FA unmatched systems
FR unmatched systems
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-80
-75
-70
-65
-60
-55
-50
-45
-40
-35
-30
-25
-20
-15
-10
-5
0
LR threshold values

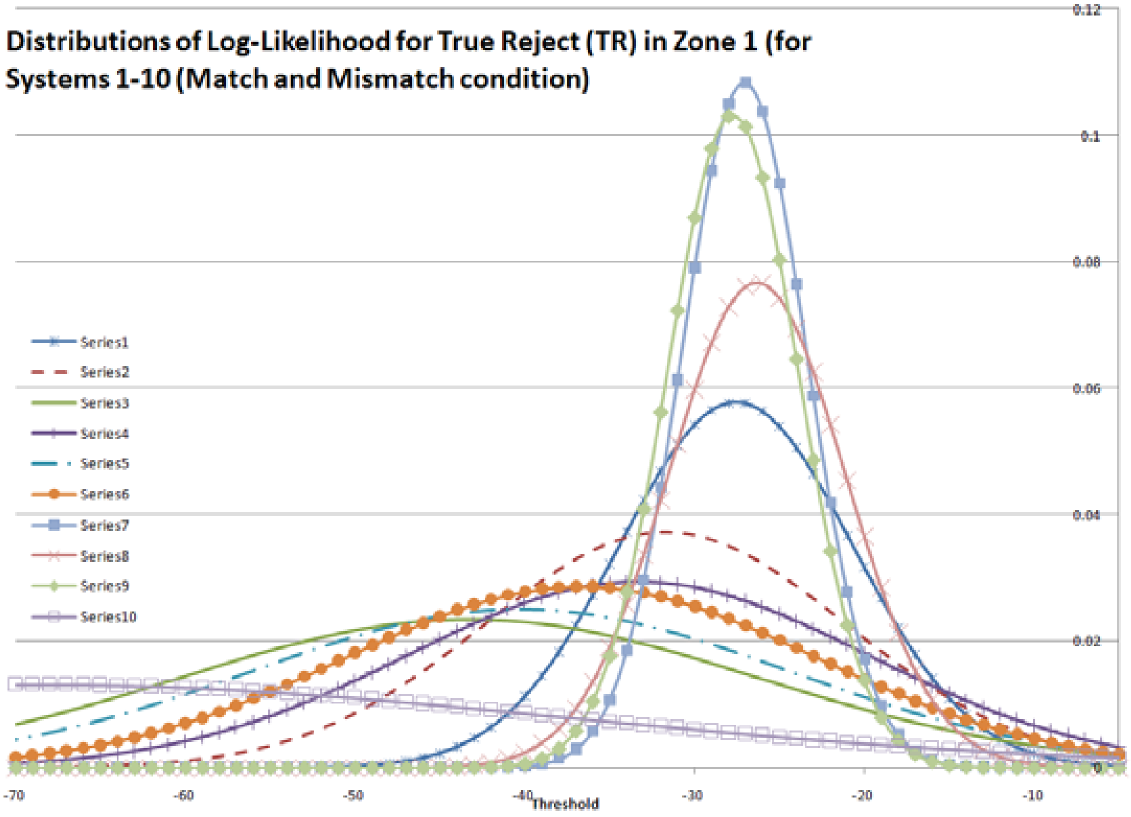

DET Curves - Matched vs. Unmatched
System 10, Zone 3, 9 sec Enrollment
Error Rate
FR matched systems
FA matched systems
FA unmatched systems
FR unmatched systems
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-80
-75
-70
-65
-60
-55
-50
-45
-40
-35
-30
-25
-20
-15
-10
-5
0
LR threshold values

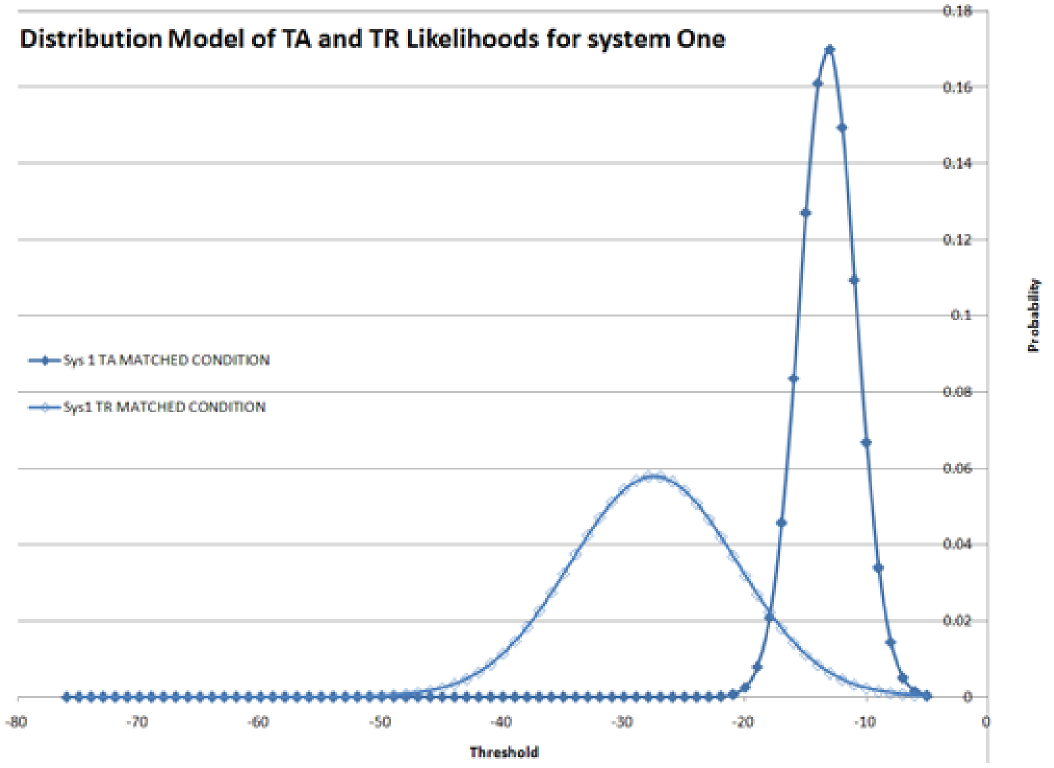
APPENDIX D

Distribution of Speaker Likelihoods for True Accepts (TA) and True Rejects (TR)

Appendix D plots the distribution models of the likelihood scores of Systems for true accepts, true rejects and a sample of matching a true accept with a true reject in a matched condition.




Distributions of Log-Likelihood for True Reject (TR) in Zone 1 (for Systems 1-10 (Match and Mismatch condition)
Series1
Series2
Series3
Series4
Series5
Series6
Series7
Series8
Series9
Series10
0.12
0.1
0.08
0.06
0.04
0.02
0
-70
-60
-50
-40
Threshold
-30
-20
-10


Distribution Model of TA and TR Likelihoods for system One
Sys 1 TA MATCHED CONDITION
Sys1 TR MATCHED CONDITION
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
Probability
-80
-70
-60
-50
-40
-30
-20
-10
0
Threshold

APPENDIX E

Phrases 1-4 Used For Voice Samples

Appendix E contains a list of the phrases used in the voice sample collection stage of the thesis. Phrase 1-3 were utilized during sample collection on October 18th 2008. Phrase 4 was utilized during sample collection on February 8th 2009. Phrase 4 was only spoken in Zone 1. Phrases 1-3 were spoken in each zone.

Phrase 1:

"Hello my name is (user states first and last name)"

Typical time duration: ~2-3 seconds

Phrase 2:

"Can you tell me how to get to Sesame Street?"

Typical time duration: ~3 seconds

Phrase 3:

"A Winston Churchill Quote: I like pigs. Cats look down on us, dogs look up to us, but pigs, they treat us as equals."

Typical time duration: ~9 seconds

Phrase 4:

“ I, Nephi, having been born of goodly parents, therefore I was taught somewhat in all the learning of my father; and having seen many afflictions in the course of my days, nevertheless, having been highly favored of the Lord in all my days; yea, having had a great knowledge of the goodness and the mysteries of God, therefore I make a record of my proceedings in my days. Yea, I make a record in the language of my father, which consists of the learning of the Jews and the language of the Egyptians.”

Typical time duration: ~28-33 seconds

Quote from 1 Nephi 1:1-2 in “The Book of Mormon”

APPENDIX E.1

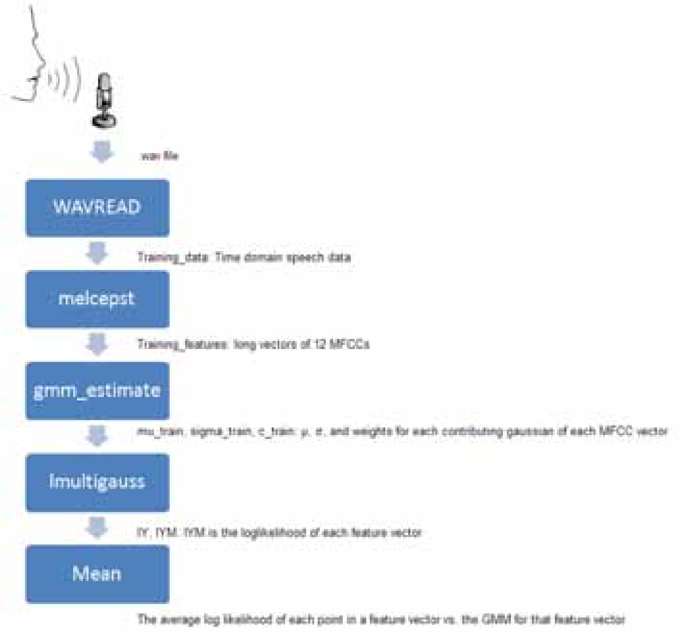
The Answer

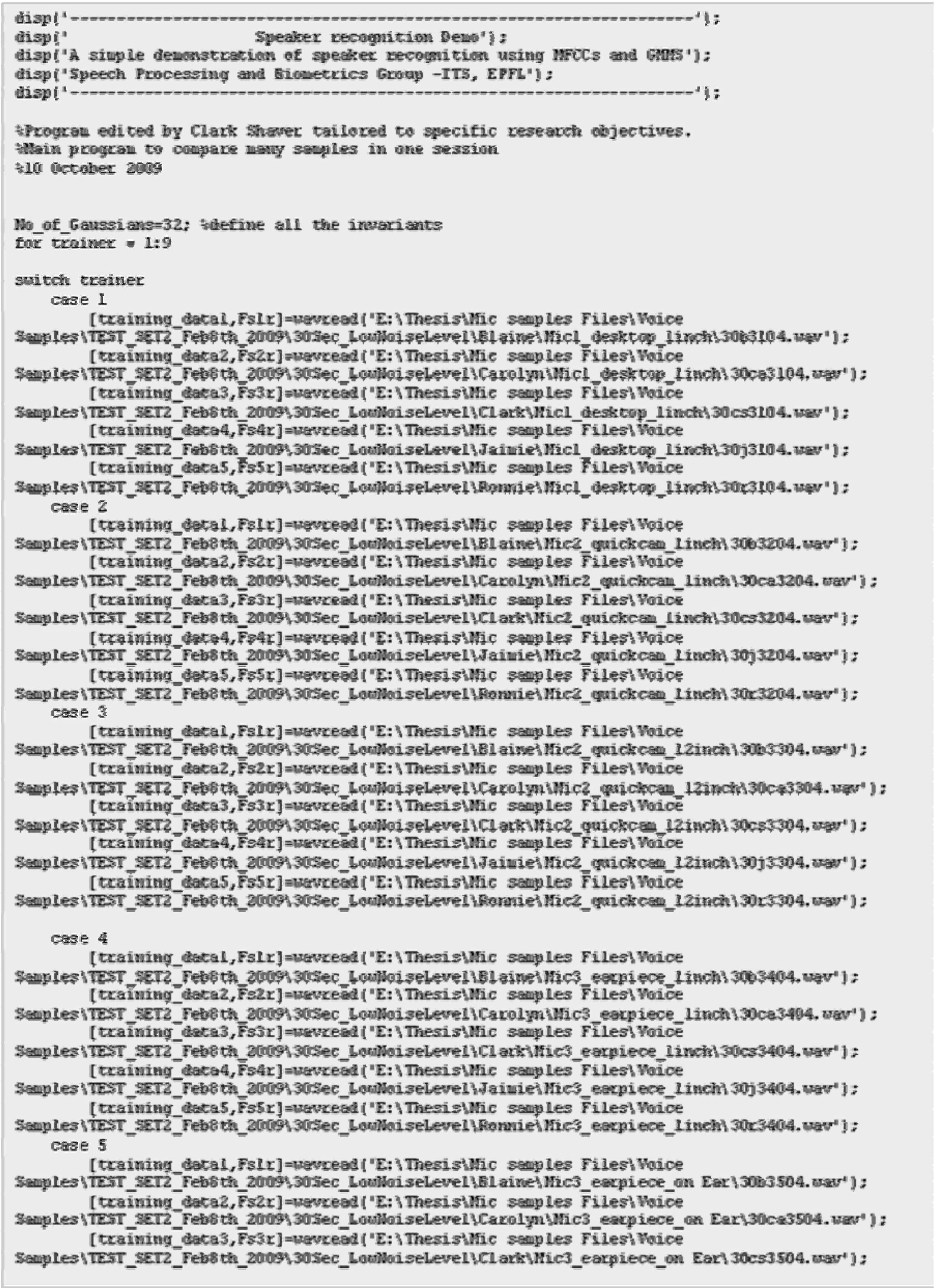
42

APPENDIX F

Program Listings

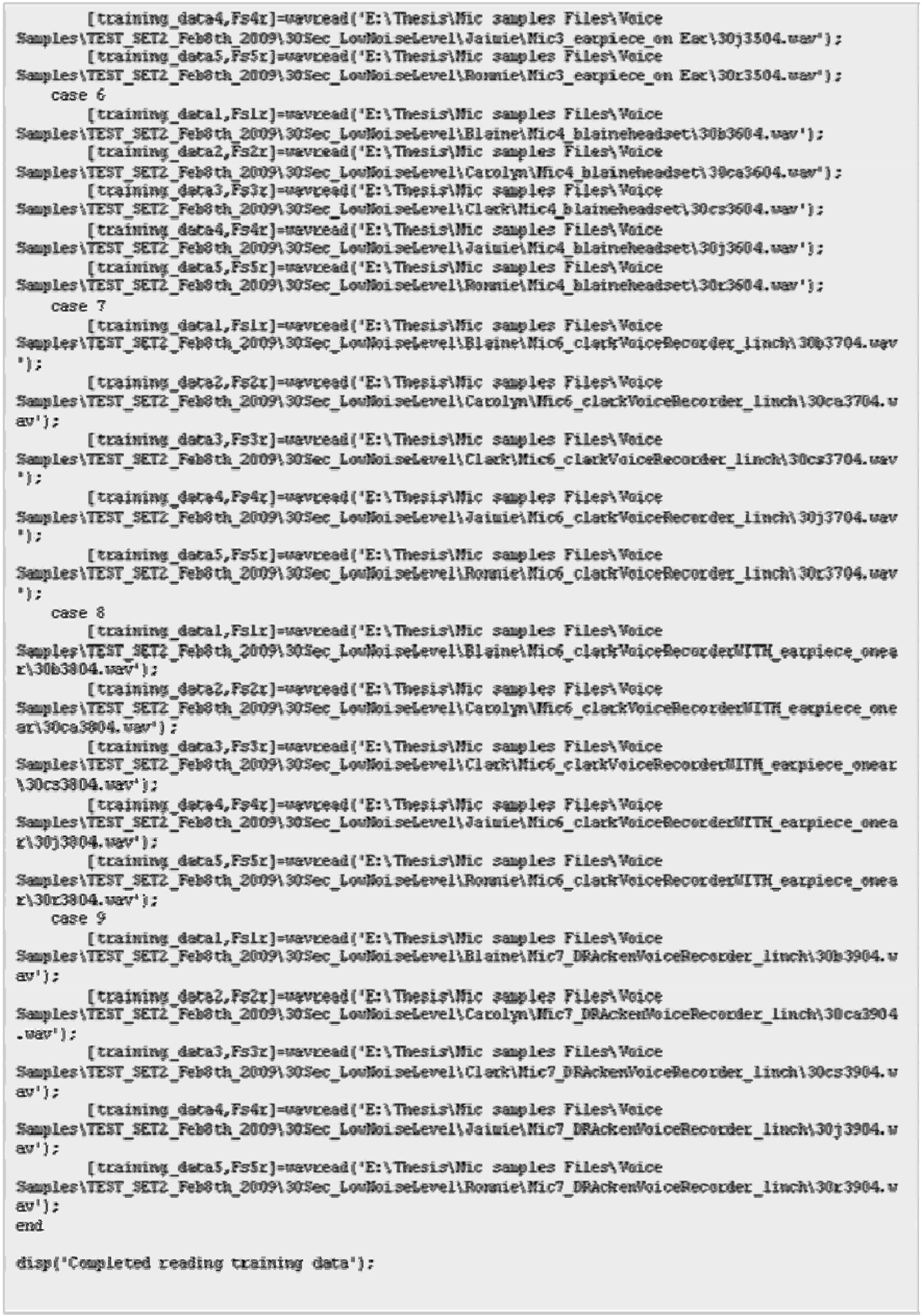
Appendix F contains the program listings for all customized programs utilized throughout the thesis. The first section is the MATLAB portion of the speaker recognition. The second section includes MSExcel VBA programming. The VBA code is the two decision algorithms, ‘nearest to’ and threshold.


wav file
WAVREAD
Training_data: Time domain speech data
melcepst
Training_features: long vectors of 12 MFCCs
gmm_estimate
mu_train, sigma_train, c_train: μ, σ, and weights for each contributing gaussian of each MFCC vector
lmultigauss
lY, lYM: lYM is the loglikelihood of each feature vector
Mean
The average log likelihood of each point in a feature vector vs. the GMM for that feature vector



```
disp('-----------------------------------------------------------------');
disp('                      Speaker recognition Demo');
disp('A simple demonstration of speaker recognition using MFCCs and GMMS');
disp('Speech Processing and Biometrics Group -ITS, EPFL');
disp('-----------------------------------------------------------------');

%Program edited by Clark Shaver tailored to specific research objectives.
%Main program to compare many samples in one session
%10 October 2009


No_of_Gaussians=32; %define all the invariants
for trainer = 1:9

switch trainer
    case 1
        [training_data1,Fs1r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Blaine\Mic1_desktop_1inch\30b3104.wav');
        [training_data2,Fs2r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Carolyn\Mic1_desktop_1inch\30ca3104.wav');
        [training_data3,Fs3r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Clark\Mic1_desktop_1inch\30cs3104.wav');
        [training_data4,Fs4r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Jaimie\Mic1_desktop_1inch\30j3104.wav');
        [training_data5,Fs5r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Ronnie\Mic1_desktop_1inch\30r3104.wav');
    case 2
        [training_data1,Fs1r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Blaine\Mic2_quickcam_1inch\30b3204.wav');
        [training_data2,Fs2r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Carolyn\Mic2_quickcam_1inch\30ca3204.wav');
        [training_data3,Fs3r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Clark\Mic2_quickcam_1inch\30cs3204.wav');
        [training_data4,Fs4r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Jaimie\Mic2_quickcam_1inch\30j3204.wav');
        [training_data5,Fs5r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Ronnie\Mic2_quickcam_1inch\30r3204.wav');
    case 3
        [training_data1,Fs1r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Blaine\Mic2_quickcam_12inch\30b3304.wav');
        [training_data2,Fs2r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Carolyn\Mic2_quickcam_12inch\30ca3304.wav');
        [training_data3,Fs3r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Clark\Mic2_quickcam_12inch\30cs3304.wav');
        [training_data4,Fs4r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Jaimie\Mic2_quickcam_12inch\30j3304.wav');
        [training_data5,Fs5r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Ronnie\Mic2_quickcam_12inch\30r3304.wav');

    case 4
        [training_data1,Fs1r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Blaine\Mic3_earpiece_1inch\30b3404.wav');
        [training_data2,Fs2r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Carolyn\Mic3_earpiece_1inch\30ca3404.wav');
        [training_data3,Fs3r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Clark\Mic3_earpiece_1inch\30cs3404.wav');
        [training_data4,Fs4r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Jaimie\Mic3_earpiece_1inch\30j3404.wav');
        [training_data5,Fs5r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Ronnie\Mic3_earpiece_1inch\30r3404.wav');
    case 5
        [training_data1,Fs1r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Blaine\Mic3_earpiece_on Ear\30b3504.wav');
        [training_data2,Fs2r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Carolyn\Mic3_earpiece_on Ear\30ca3504.wav');
        [training_data3,Fs3r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Clark\Mic3_earpiece_on Ear\30cs3504.wav');
```

```
        [training_data4,Fs4r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Jaimie\Mic3_earpiece_on Ear\30j3504.wav');
        [training_data5,Fs5r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Ronnie\Mic3_earpiece_on Ear\30r3504.wav');
    case 6
        [training_data1,Fs1r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Blaine\Mic4_blaineheadset\30b3604.wav');
        [training_data2,Fs2r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Carolyn\Mic4_blaineheadset\30ca3604.wav');
        [training_data3,Fs3r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Clark\Mic4_blaineheadset\30cs3604.wav');
        [training_data4,Fs4r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Jaimie\Mic4_blaineheadset\30j3604.wav');
        [training_data5,Fs5r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Ronnie\Mic4_blaineheadset\30r3604.wav');
    case 7
        [training_data1,Fs1r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Blaine\Mic6_clarkVoiceRecorder_1inch\30b3704.wav
');
        [training_data2,Fs2r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Carolyn\Mic6_clarkVoiceRecorder_1inch\30ca3704.w
av');
        [training_data3,Fs3r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Clark\Mic6_clarkVoiceRecorder_1inch\30cs3704.wav
');
        [training_data4,Fs4r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Jaimie\Mic6_clarkVoiceRecorder_1inch\30j3704.wav
');
        [training_data5,Fs5r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Ronnie\Mic6_clarkVoiceRecorder_1inch\30r3704.wav
');
    case 8
        [training_data1,Fs1r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Blaine\Mic6_clarkVoiceRecorderWITH_earpiece_onea
r\30b3804.wav');
        [training_data2,Fs2r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Carolyn\Mic6_clarkVoiceRecorderWITH_earpiece_one
ar\30ca3804.wav');
        [training_data3,Fs3r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Clark\Mic6_clarkVoiceRecorderWITH_earpiece_onear
\30cs3804.wav');
        [training_data4,Fs4r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Jaimie\Mic6_clarkVoiceRecorderWITH_earpiece_onea
r\30j3804.wav');
        [training_data5,Fs5r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Ronnie\Mic6_clarkVoiceRecorderWITH_earpiece_onea
r\30r3804.wav');
    case 9
        [training_data1,Fs1r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Blaine\Mic7_DRAckenVoiceRecorder_1inch\30b3904.w
av');
        [training_data2,Fs2r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Carolyn\Mic7_DRAckenVoiceRecorder_1inch\30ca3904
.wav');
        [training_data3,Fs3r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Clark\Mic7_DRAckenVoiceRecorder_1inch\30cs3904.w
av');
        [training_data4,Fs4r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Jaimie\Mic7_DRAckenVoiceRecorder_1inch\30j3904.w
av');
        [training_data5,Fs5r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Ronnie\Mic7_DRAckenVoiceRecorder_1inch\30r3904.w
av');
end

disp('Completed reading training data');
```



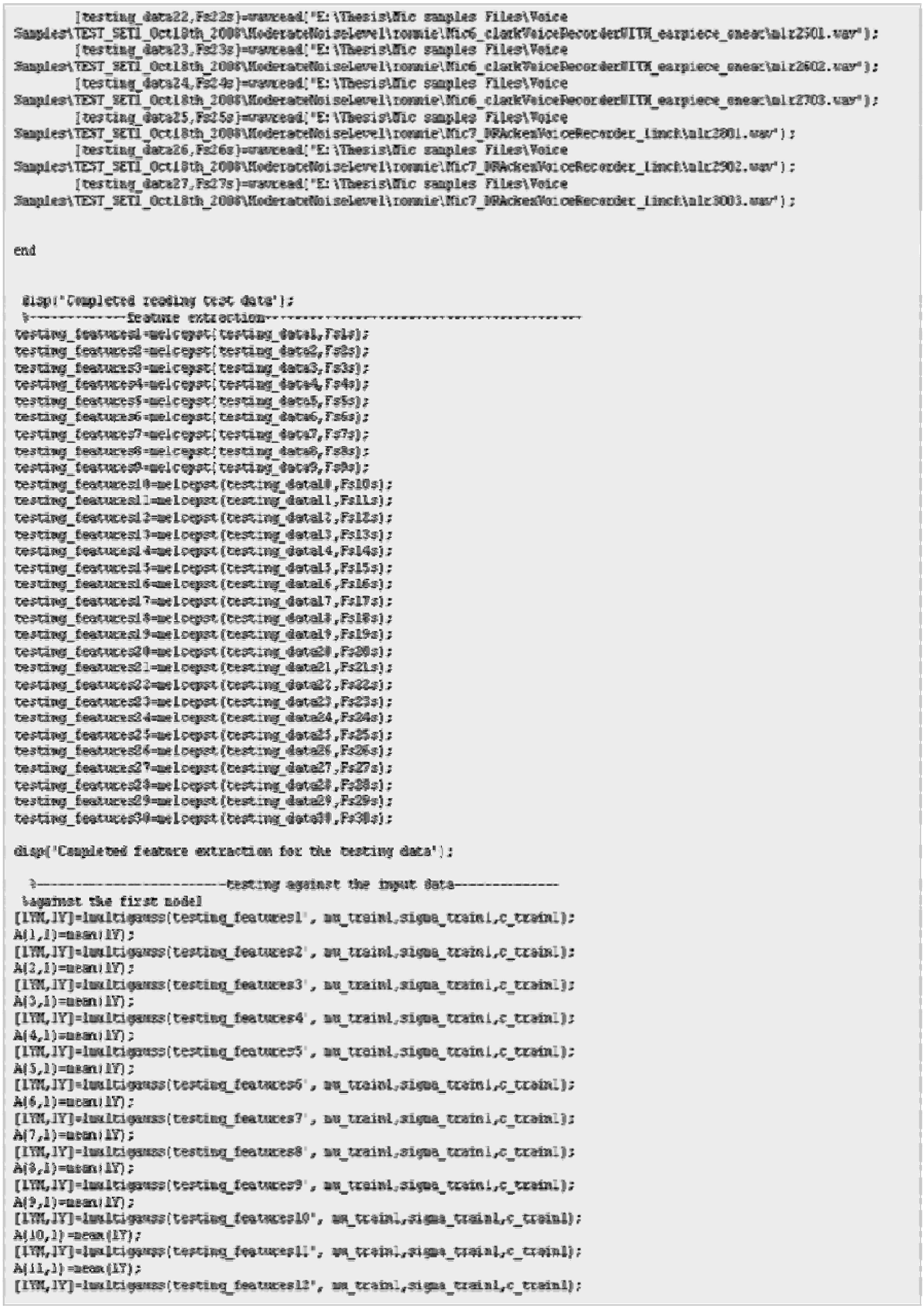
```
        [training_data4,Fs4r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Jaimie\Mic3_earpiece_on Ear\30j3504.wav');
        [training_data5,Fs5r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Ronnie\Mic3_earpiece_on Ear\30r3504.wav');
    case 6
        [training_data1,Fs1r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Blaine\Mic4_blaineheadset\30b3604.wav');
        [training_data2,Fs2r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Carolyn\Mic4_blaineheadset\30ca3604.wav');
        [training_data3,Fs3r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Clark\Mic4_blaineheadset\30cs3604.wav');
        [training_data4,Fs4r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Jaimie\Mic4_blaineheadset\30j3604.wav');
        [training_data5,Fs5r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Ronnie\Mic4_blaineheadset\30r3604.wav');
    case 7
        [training_data1,Fs1r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Blaine\Mic6_clarkVoiceRecorder_linch\30b3704.wav
');
        [training_data2,Fs2r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Carolyn\Mic6_clarkVoiceRecorder_linch\30ca3704.w
av');
        [training_data3,Fs3r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Clark\Mic6_clarkVoiceRecorder_linch\30cs3704.wav
');
        [training_data4,Fs4r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Jaimie\Mic6_clarkVoiceRecorder_linch\30j3704.wav
');
        [training_data5,Fs5r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Ronnie\Mic6_clarkVoiceRecorder_linch\30r3704.wav
');
    case 8
        [training_data1,Fs1r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Blaine\Mic6_clarkVoiceRecorderWITH_earpiece_onea
r\30b3804.wav');
        [training_data2,Fs2r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Carolyn\Mic6_clarkVoiceRecorderWITH_earpiece_one
ar\30ca3804.wav');
        [training_data3,Fs3r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Clark\Mic6_clarkVoiceRecorderWITH_earpiece_onear
\30cs3804.wav');
        [training_data4,Fs4r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Jaimie\Mic6_clarkVoiceRecorderWITH_earpiece_onea
r\30j3804.wav');
        [training_data5,Fs5r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Ronnie\Mic6_clarkVoiceRecorderWITH_earpiece_onea
r\30r3804.wav');
    case 9
        [training_data1,Fs1r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Blaine\Mic7_DRAckenVoiceRecorder_linch\30b3904.w
av');
        [training_data2,Fs2r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Carolyn\Mic7_DRAckenVoiceRecorder_linch\30ca3904
.wav');
        [training_data3,Fs3r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Clark\Mic7_DRAckenVoiceRecorder_linch\30cs3904.w
av');
        [training_data4,Fs4r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Jaimie\Mic7_DRAckenVoiceRecorder_linch\30j3904.w
av');
        [training_data5,Fs5r]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET2_Feb8th_2009\30Sec_LowNoiseLevel\Ronnie\Mic7_DRAckenVoiceRecorder_linch\30r3904.w
av');
end

disp('Completed reading training data');
```

```
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\blaine\Mic2_quickcam_1inch\mb0502.wav');
        [testing_data6,Fs6s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\blaine\Mic2_quickcam_1inch\mb0603.wav');
        [testing_data7,Fs7s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\blaine\Mic2_quickcam_12inch\mb0701.wav');
        [testing_data8,Fs8s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\blaine\Mic2_quickcam_12inch\mb0802.wav');
        [testing_data9,Fs9s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\blaine\Mic2_quickcam_12inch\mb0903.wav');
        [testing_data10,Fs10s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\blaine\Mic3_earpiece_1inch\mb1001.wav');
        [testing_data11,Fs11s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\blaine\Mic3_earpiece_1inch\mb1102.wav');
        [testing_data12,Fs12s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\blaine\Mic3_earpiece_1inch\mb1203.wav');
        [testing_data13,Fs13s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\blaine\Mic3_earpiece_on Ear\mb1301.wav');
        [testing_data14,Fs14s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\blaine\Mic3_earpiece_on Ear\mb1402.wav');
        [testing_data15,Fs15s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\blaine\Mic3_earpiece_on Ear\mb1503.wav');
        [testing_data16,Fs16s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\blaine\Mic4_Blaineheadset\mb1601.wav');
        [testing_data17,Fs17s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\blaine\Mic4_Blaineheadset\mb1702.wav');
        [testing_data18,Fs18s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\blaine\Mic4_Blaineheadset\mb1803.wav');
        [testing_data28,Fs28s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\blaine\Mic8_CurtisMP3player_1inch\mb1901.WAV');
        [testing_data29,Fs29s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\blaine\Mic8_CurtisMP3player_1inch\mb2002.WAV');
        [testing_data30,Fs30s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\blaine\Mic8_CurtisMP3player_1inch\mb2103.WAV');
        [testing_data19,Fs19s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\blaine\Mic6_clarkVoiceRecorder_1inch\mb2201.wav');
        [testing_data20,Fs20s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\blaine\Mic6_clarkVoiceRecorder_1inch\mb2302.wav');
        [testing_data21,Fs21s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\blaine\Mic6_clarkVoiceRecorder_1inch\mb2403.wav');
        [testing_data22,Fs22s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\blaine\Mic6_clarkVoiceRecorderWITH_earpiece_onear\mb2
501.wav');
        [testing_data23,Fs23s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\blaine\Mic6_clarkVoiceRecorderWITH_earpiece_onear\mb2
602.wav');
        [testing_data24,Fs24s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\blaine\Mic6_clarkVoiceRecorderWITH_earpiece_onear\mb2
703.wav');
        [testing_data25,Fs25s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\blaine\Mic7_DRAckenVoiceRecorder_1inch\mb2801.wav');
        [testing_data26,Fs26s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\blaine\Mic7_DRAckenVoiceRecorder_1inch\mb2902.wav');
        [testing_data27,Fs27s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\blaine\Mic7_DRAckenVoiceRecorder_1inch\mb3003.wav');

    case 2
        [testing_data1,Fs1s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\carolyn\Mic1_desktop_1inch\mca0101.wav');
        [testing_data2,Fs2s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\carolyn\Mic1_desktop_1inch\mca0202.wav');
        [testing_data3,Fs3s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\carolyn\Mic1_desktop_1inch\mca0303.wav');
        [testing_data4,Fs4s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\carolyn\Mic2_quickcam_1inch\mca0401.wav');
        [testing_data5,Fs5s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\carolyn\Mic2_quickcam_1inch\mca0502.wav');
        [testing_data6,Fs6s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\carolyn\Mic2_quickcam_1inch\mca0603.wav');
        [testing_data7,Fs7s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\carolyn\Mic2_quickcam_12inch\mca0701.wav');
        [testing_data8,Fs8s]=wavread('E:\Thesis\Mic samples Files\Voice
```

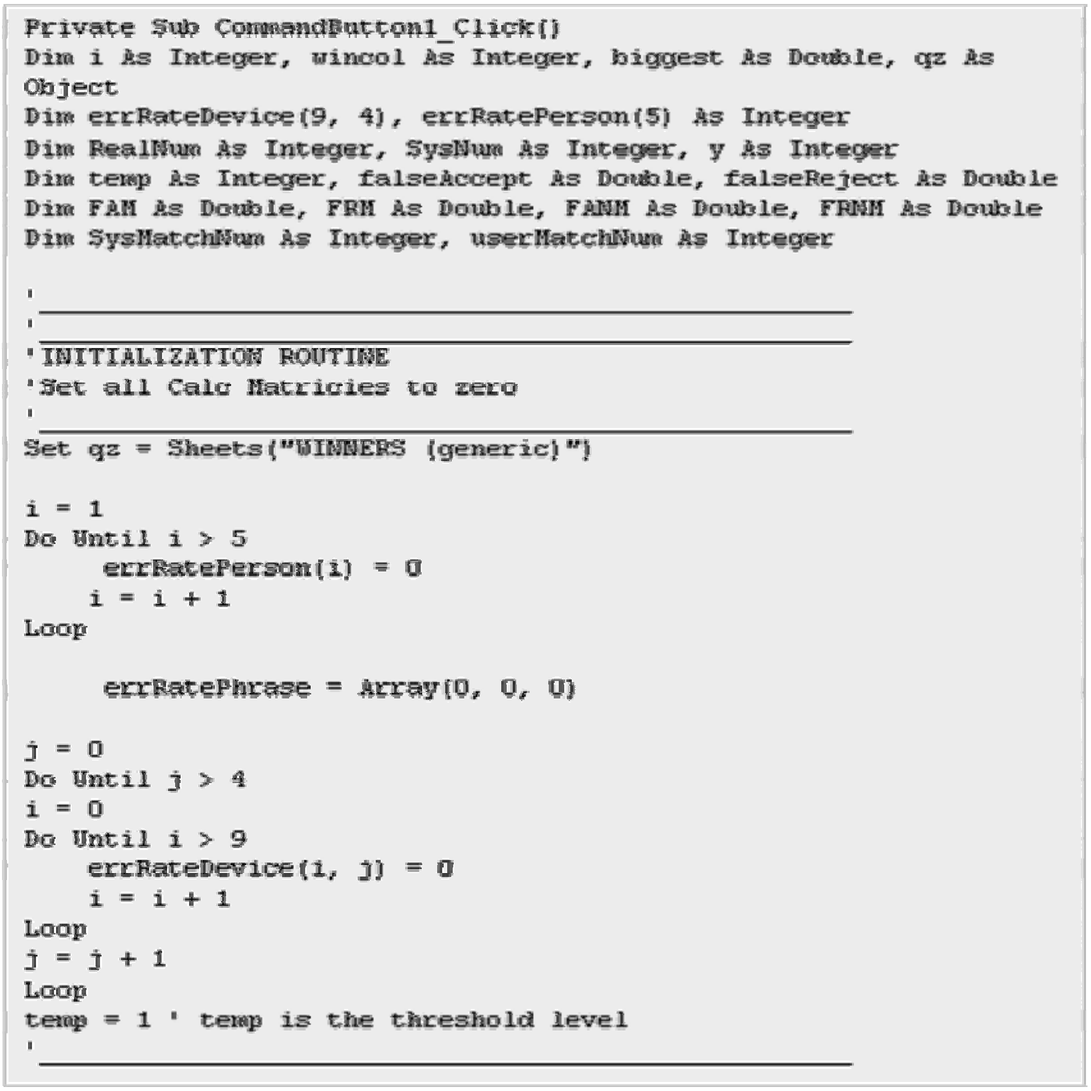
```
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\carolyn\Mic2_quickcam_12inch\mca0802.wav');
        [testing_data9,Fs9s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\carolyn\Mic2_quickcam_12inch\mca0903.wav');
        [testing_data10,Fs10s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\carolyn\Mic3_earpiece_1inch\mca1001.wav');
        [testing_data11,Fs11s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\carolyn\Mic3_earpiece_1inch\mca1102.wav');
        [testing_data12,Fs12s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\carolyn\Mic3_earpiece_1inch\mca1203.wav');
        [testing_data13,Fs13s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\carolyn\Mic3_earpiece_on Ear\mca1301.wav');
        [testing_data14,Fs14s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\carolyn\Mic3_earpiece_on Ear\mca1402.wav');
        [testing_data15,Fs15s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\carolyn\Mic3_earpiece_on Ear\mca1503.wav');
        [testing_data16,Fs16s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\carolyn\Mic4_Blaineheadset\mca1601.wav');
        [testing_data17,Fs17s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\carolyn\Mic4_Blaineheadset\mca1702.wav');
        [testing_data18,Fs18s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\carolyn\Mic4_Blaineheadset\mca1803.wav');
        [testing_data28,Fs28s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\carolyn\Mic8_CurtisMP3player_1inch\mca1901.WAV');
        [testing_data29,Fs29s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\carolyn\Mic8_CurtisMP3player_1inch\mca2002.WAV');
        [testing_data30,Fs30s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\carolyn\Mic8_CurtisMP3player_1inch\mca2103.WAV');
        [testing_data19,Fs19s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\carolyn\Mic6_clarkVoiceRecorder_1inch\mca2201.wav');
        [testing_data20,Fs20s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\carolyn\Mic6_clarkVoiceRecorder_1inch\mca2302.wav');
        [testing_data21,Fs21s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\carolyn\Mic6_clarkVoiceRecorder_1inch\mca2403.wav');
        [testing_data22,Fs22s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\carolyn\Mic6_clarkVoiceRecorderWITH_earpiece_onear\mca2501.wav');
        [testing_data23,Fs23s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\carolyn\Mic6_clarkVoiceRecorderWITH_earpiece_onear\mca2602.wav');
        [testing_data24,Fs24s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\carolyn\Mic6_clarkVoiceRecorderWITH_earpiece_onear\mca2703.wav');
        [testing_data25,Fs25s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\carolyn\Mic7_BRAckenVoiceRecorder_1inch\mca2801.wav');
        [testing_data26,Fs26s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\carolyn\Mic7_BRAckenVoiceRecorder_1inch\mca2902.wav');
        [testing_data27,Fs27s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\carolyn\Mic7_BRAckenVoiceRecorder_1inch\mca3003.wav');

    case 3
        [testing_data1,Fs1s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\clark\Mic1_desktop_1inch\mlc0101.wav');
        [testing_data2,Fs2s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\clark\Mic1_desktop_1inch\mlc0202.wav');
        [testing_data3,Fs3s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\clark\Mic1_desktop_1inch\mlc0303.wav');
        [testing_data4,Fs4s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\clark\Mic2_quickcam_1inch\mlc0401.wav');
        [testing_data5,Fs5s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\clark\Mic2_quickcam_1inch\mlc0502.wav');
        [testing_data6,Fs6s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\clark\Mic2_quickcam_1inch\mlc0603.wav');
        [testing_data7,Fs7s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\clark\Mic2_quickcam_12inch\mlc0701.wav');
        [testing_data8,Fs8s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\clark\Mic2_quickcam_12inch\mlc0802.wav');
        [testing_data9,Fs9s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\clark\Mic2_quickcam_12inch\mlc0903.wav');
        [testing_data10,Fs10s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\clark\Mic3_earpiece_1inch\mlc1001.wav');
        [testing_data11,Fs11s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\clark\Mic3_earpiece_1inch\mlc1102.wav');
        [testing_data12,Fs12s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\clark\Mic3_earpiece_1inch\mlc1203.wav');
```

```
        [testing_data13,Fs13s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\clark\Mic3_earpiece_on Ear\mlc1301.wav');
        [testing_data14,Fs14s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\clark\Mic3_earpiece_on Ear\mlc1402.wav');
        [testing_data15,Fs15s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\clark\Mic3_earpiece_on Ear\mlc1503.wav');
        [testing_data16,Fs16s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\clark\Mic4_Blaineheadset\mlc1601.wav');
        [testing_data17,Fs17s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\clark\Mic4_Blaineheadset\mlc1702.wav');
        [testing_data18,Fs18s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\clark\Mic4_Blaineheadset\mlc1803.wav');
        [testing_data28,Fs28s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\clark\Mic5_CurtisMP3player_1inch\mlc1901.WAV');
        [testing_data29,Fs29s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\clark\Mic5_CurtisMP3player_1inch\mlc2002.WAV');
        [testing_data30,Fs30s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\clark\Mic5_CurtisMP3player_1inch\mlc2103.WAV');
        [testing_data19,Fs19s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\clark\Mic6_clarkVoiceRecorder_1inch\mlc2201.wav');
        [testing_data20,Fs20s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\clark\Mic6_clarkVoiceRecorder_1inch\mlc2302.wav');
        [testing_data21,Fs21s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\clark\Mic6_clarkVoiceRecorder_1inch\mlc2403.wav');
        [testing_data22,Fs22s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\clark\Mic6_clarkVoiceRecorderWITH_earpiece_onear\mlc2501.wav');
        [testing_data23,Fs23s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\clark\Mic6_clarkVoiceRecorderWITH_earpiece_onear\mlc2602.wav');
        [testing_data24,Fs24s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\clark\Mic6_clarkVoiceRecorderWITH_earpiece_onear\mlc2703.wav');
        [testing_data25,Fs25s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\clark\Mic7_DRAckenVoiceRecorder_1inch\mlc2801.wav');
        [testing_data26,Fs26s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\clark\Mic7_DRAckenVoiceRecorder_1inch\mlc2902.wav');
        [testing_data27,Fs27s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\clark\Mic7_DRAckenVoiceRecorder_1inch\mlc3003.wav');

    case 4
        [testing_data1,Fs1s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\jaimie\Mic1_desktop_1inch\mlj0101.wav');
        [testing_data2,Fs2s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\jaimie\Mic1_desktop_1inch\mlj0202.wav');
        [testing_data3,Fs3s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\jaimie\Mic1_desktop_1inch\mlj0303.wav');
        [testing_data4,Fs4s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\jaimie\Mic2_quickcam_1inch\mlj0401.wav');
        [testing_data5,Fs5s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\jaimie\Mic2_quickcam_1inch\mlj0502.wav');
        [testing_data6,Fs6s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\jaimie\Mic2_quickcam_1inch\mlj0603.wav');
        [testing_data7,Fs7s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\jaimie\Mic2_quickcam_12inch\mlj0701.wav');
        [testing_data8,Fs8s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\jaimie\Mic2_quickcam_12inch\mlj0802.wav');
        [testing_data9,Fs9s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\jaimie\Mic2_quickcam_12inch\mlj0903.wav');
        [testing_data10,Fs10s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\jaimie\Mic3_earpiece_1inch\mlj1001.wav');
        [testing_data11,Fs11s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\jaimie\Mic3_earpiece_1inch\mlj1102.wav');
        [testing_data12,Fs12s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\jaimie\Mic3_earpiece_1inch\mlj1203.wav');
        [testing_data13,Fs13s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\jaimie\Mic3_earpiece_on Ear\mlj1301.wav');
        [testing_data14,Fs14s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\jaimie\Mic3_earpiece_on Ear\mlj1402.wav');
        [testing_data15,Fs15s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\jaimie\Mic3_earpiece_on Ear\mlj1503.wav');
        [testing_data16,Fs16s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\jaimie\Mic4_Blaineheadset\mlj1601.wav');
        [testing_data17,Fs17s]=wavread('E:\Thesis\Mic samples Files\Voice
```

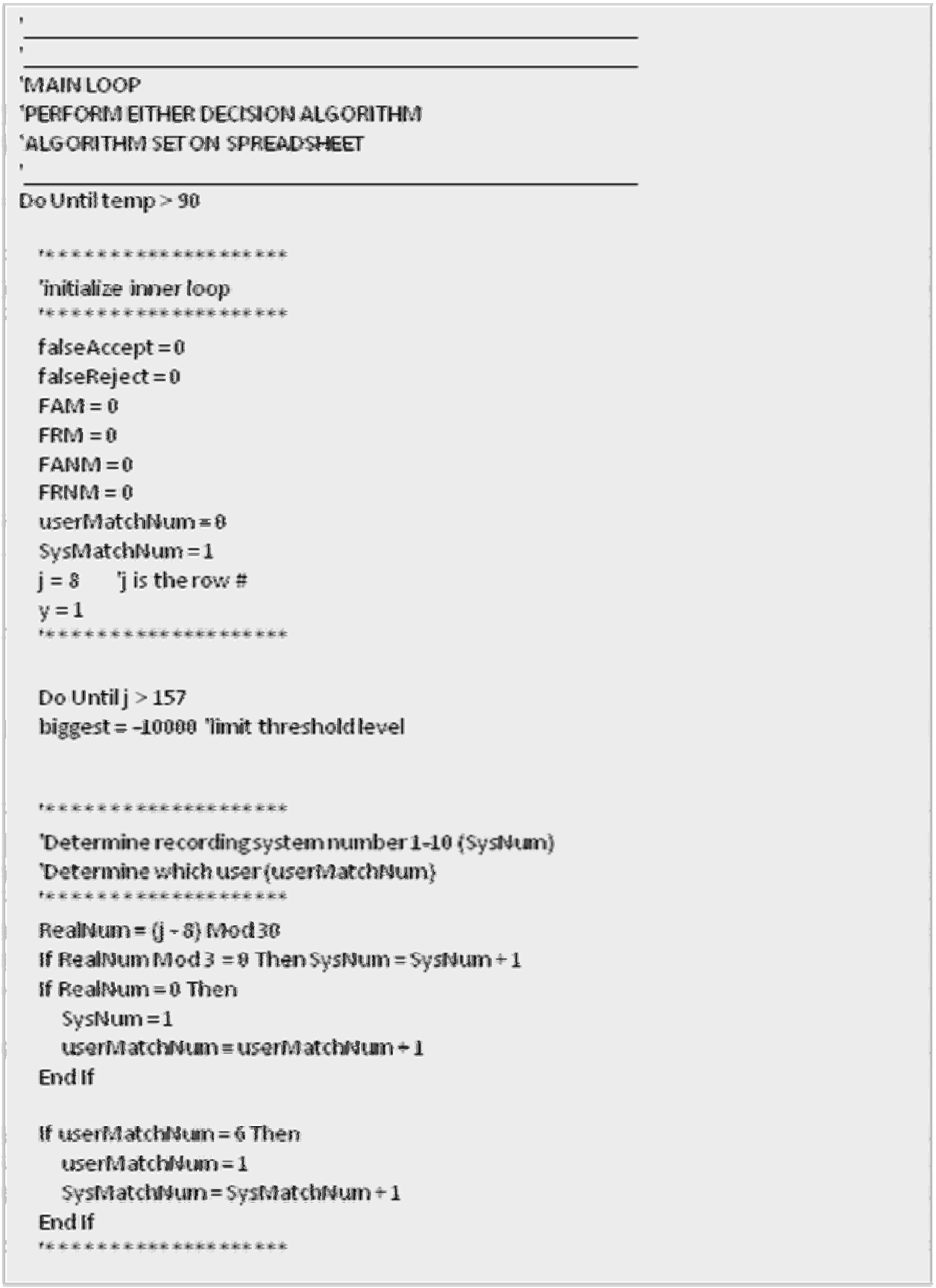
```
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\jaimie\Mic4_Blaineheadset\alj1702.wav');
        [testing_data18,Fs18s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\jaimie\Mic4_Blaineheadset\alj1803.wav');
        [testing_data28,Fs28s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\jaimie\Mic8_CurtisMP3player_1inch\alj1901.WAV');
        [testing_data29,Fs29s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\jaimie\Mic8_CurtisMP3player_1inch\alj2002.WAV');
        [testing_data30,Fs30s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\jaimie\Mic8_CurtisMP3player_1inch\alj2103.WAV');
        [testing_data19,Fs19s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\jaimie\Mic6_clarkVoiceRecorder_1inch\alj2201.wav');
        [testing_data20,Fs20s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\jaimie\Mic6_clarkVoiceRecorder_1inch\alj2302.wav');
        [testing_data21,Fs21s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\jaimie\Mic6_clarkVoiceRecorder_1inch\alj2403.wav');
        [testing_data22,Fs22s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\jaimie\Mic6_clarkVoiceRecorderWITH_earpiece_onear\alj2501.wav');
        [testing_data23,Fs23s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\jaimie\Mic6_clarkVoiceRecorderWITH_earpiece_onear\alj2602.wav');
        [testing_data24,Fs24s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\jaimie\Mic6_clarkVoiceRecorderWITH_earpiece_onear\alj2703.wav');
        [testing_data25,Fs25s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\jaimie\Mic7_DRAckerVoiceRecorder_1inch\alj2801.wav');
        [testing_data26,Fs26s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\jaimie\Mic7_DRAckerVoiceRecorder_1inch\alj2902.wav');
        [testing_data27,Fs27s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\jaimie\Mic7_DRAckerVoiceRecorder_1inch\alj3003.wav');

    case 5
        [testing_data1,Fs1s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\ronnie\Mic1_desktop_1inch\alr0101.wav');
        [testing_data2,Fs2s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\ronnie\Mic1_desktop_1inch\alr0202.wav');
        [testing_data3,Fs3s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\ronnie\Mic1_desktop_1inch\alr0303.wav');
        [testing_data4,Fs4s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\ronnie\Mic2_quickcam_1inch\alr0401.wav');
        [testing_data5,Fs5s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\ronnie\Mic2_quickcam_1inch\alr0502.wav');
        [testing_data6,Fs6s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\ronnie\Mic2_quickcam_1inch\alr0603.wav');
        [testing_data7,Fs7s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\ronnie\Mic2_quickcam_12inch\alr0701.wav');
        [testing_data8,Fs8s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\ronnie\Mic2_quickcam_12inch\alr0802.wav');
        [testing_data9,Fs9s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\ronnie\Mic2_quickcam_12inch\alr0903.wav');
        [testing_data10,Fs10s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\ronnie\Mic3_earpiece_1inch\alr1001.wav');
        [testing_data11,Fs11s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\ronnie\Mic3_earpiece_1inch\alr1102.wav');
        [testing_data12,Fs12s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\ronnie\Mic3_earpiece_1inch\alr1203.wav');
        [testing_data13,Fs13s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\ronnie\Mic3_earpiece_on Ear\alr1301.wav');
        [testing_data14,Fs14s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\ronnie\Mic3_earpiece_on Ear\alr1402.wav');
        [testing_data15,Fs15s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\ronnie\Mic3_earpiece_on Ear\alr1503.wav');
        [testing_data16,Fs16s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\ronnie\Mic4_Blaineheadset\alr1601.wav');
        [testing_data17,Fs17s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\ronnie\Mic4_Blaineheadset\alr1702.wav');
        [testing_data18,Fs18s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\ronnie\Mic4_Blaineheadset\alr1803.wav');
        [testing_data28,Fs28s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\ronnie\Mic8_CurtisMP3player_1inch\alr1901.WAV');
        [testing_data29,Fs29s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\ronnie\Mic8_CurtisMP3player_1inch\alr2002.WAV');
        [testing_data30,Fs30s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\ronnie\Mic8_CurtisMP3player_1inch\alr2103.WAV');
        [testing_data19,Fs19s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\ronnie\Mic6_clarkVoiceRecorder_1inch\alr2201.wav');
        [testing_data20,Fs20s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\ronnie\Mic6_clarkVoiceRecorder_1inch\alr2302.wav');
        [testing_data21,Fs21s]=wavread('E:\Thesis\Mic samples Files\Voice
Samples\TEST_SET1_Oct18th_2008\ModerateNoiseLevel\ronnie\Mic6_clarkVoiceRecorder_1inch\alr2403.wav');
```



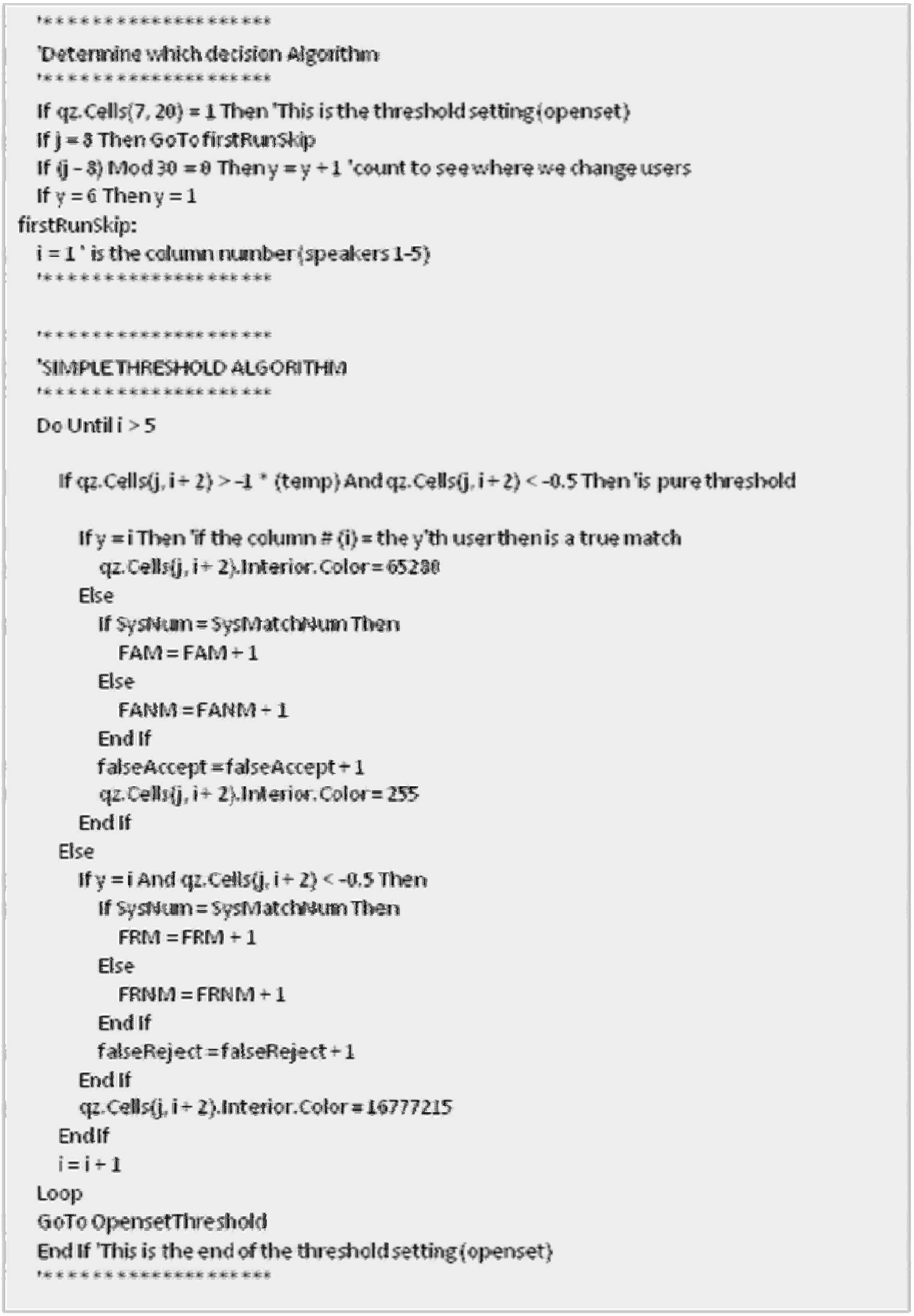
```
        [testing_data22,Fs22s]=wavread('E:\Thesis\Mic samples Files\Voice Samples\TEST_SET1_Oct18th_2008\ModerateNoiselevel\rommie\Mic6_clarkVoiceRecorderWITH_earpiece_sneat\mlr2501.wav');
        [testing_data23,Fs23s]=wavread('E:\Thesis\Mic samples Files\Voice Samples\TEST_SET1_Oct18th_2008\ModerateNoiselevel\rommie\Mic6_clarkVoiceRecorderWITH_earpiece_sneat\mlr2602.wav');
        [testing_data24,Fs24s]=wavread('E:\Thesis\Mic samples Files\Voice Samples\TEST_SET1_Oct18th_2008\ModerateNoiselevel\rommie\Mic6_clarkVoiceRecorderWITH_earpiece_sneat\mlr2703.wav');
        [testing_data25,Fs25s]=wavread('E:\Thesis\Mic samples Files\Voice Samples\TEST_SET1_Oct18th_2008\ModerateNoiselevel\rommie\Mic7_DRAckenVoiceRecorder_linck\mlr2801.wav');
        [testing_data26,Fs26s]=wavread('E:\Thesis\Mic samples Files\Voice Samples\TEST_SET1_Oct18th_2008\ModerateNoiselevel\rommie\Mic7_DRAckenVoiceRecorder_linck\mlr2902.wav');
        [testing_data27,Fs27s]=wavread('E:\Thesis\Mic samples Files\Voice Samples\TEST_SET1_Oct18th_2008\ModerateNoiselevel\rommie\Mic7_DRAckenVoiceRecorder_linck\mlr3003.wav');


end


 disp('Completed reading test data');
 %-------------feature extraction-----------------------------------------
testing_features1=melcepst(testing_data1,Fs1s);
testing_features2=melcepst(testing_data2,Fs2s);
testing_features3=melcepst(testing_data3,Fs3s);
testing_features4=melcepst(testing_data4,Fs4s);
testing_features5=melcepst(testing_data5,Fs5s);
testing_features6=melcepst(testing_data6,Fs6s);
testing_features7=melcepst(testing_data7,Fs7s);
testing_features8=melcepst(testing_data8,Fs8s);
testing_features9=melcepst(testing_data9,Fs9s);
testing_features10=melcepst(testing_data10,Fs10s);
testing_features11=melcepst(testing_data11,Fs11s);
testing_features12=melcepst(testing_data12,Fs12s);
testing_features13=melcepst(testing_data13,Fs13s);
testing_features14=melcepst(testing_data14,Fs14s);
testing_features15=melcepst(testing_data15,Fs15s);
testing_features16=melcepst(testing_data16,Fs16s);
testing_features17=melcepst(testing_data17,Fs17s);
testing_features18=melcepst(testing_data18,Fs18s);
testing_features19=melcepst(testing_data19,Fs19s);
testing_features20=melcepst(testing_data20,Fs20s);
testing_features21=melcepst(testing_data21,Fs21s);
testing_features22=melcepst(testing_data22,Fs22s);
testing_features23=melcepst(testing_data23,Fs23s);
testing_features24=melcepst(testing_data24,Fs24s);
testing_features25=melcepst(testing_data25,Fs25s);
testing_features26=melcepst(testing_data26,Fs26s);
testing_features27=melcepst(testing_data27,Fs27s);
testing_features28=melcepst(testing_data28,Fs28s);
testing_features29=melcepst(testing_data29,Fs29s);
testing_features30=melcepst(testing_data30,Fs30s);

disp('Completed feature extraction for the testing data');

  %-------------------------testing against the input data--------------
 %against the first model
[lYM,lY]=lmultigauss(testing_features1', mu_train1,sigma_train1,c_train1);
A(1,1)=mean(lY);
[lYM,lY]=lmultigauss(testing_features2', mu_train1,sigma_train1,c_train1);
A(2,1)=mean(lY);
[lYM,lY]=lmultigauss(testing_features3', mu_train1,sigma_train1,c_train1);
A(3,1)=mean(lY);
[lYM,lY]=lmultigauss(testing_features4', mu_train1,sigma_train1,c_train1);
A(4,1)=mean(lY);
[lYM,lY]=lmultigauss(testing_features5', mu_train1,sigma_train1,c_train1);
A(5,1)=mean(lY);
[lYM,lY]=lmultigauss(testing_features6', mu_train1,sigma_train1,c_train1);
A(6,1)=mean(lY);
[lYM,lY]=lmultigauss(testing_features7', mu_train1,sigma_train1,c_train1);
A(7,1)=mean(lY);
[lYM,lY]=lmultigauss(testing_features8', mu_train1,sigma_train1,c_train1);
A(8,1)=mean(lY);
[lYM,lY]=lmultigauss(testing_features9', mu_train1,sigma_train1,c_train1);
A(9,1)=mean(lY);
[lYM,lY]=lmultigauss(testing_features10", mu_train1,sigma_train1,c_train1);
A(10,1)=mean(lY);
[lYM,lY]=lmultigauss(testing_features11", mu_train1,sigma_train1,c_train1);
A(11,1)=mean(lY);
[lYM,lY]=lmultigauss(testing_features12", mu_train1,sigma_train1,c_train1);
```



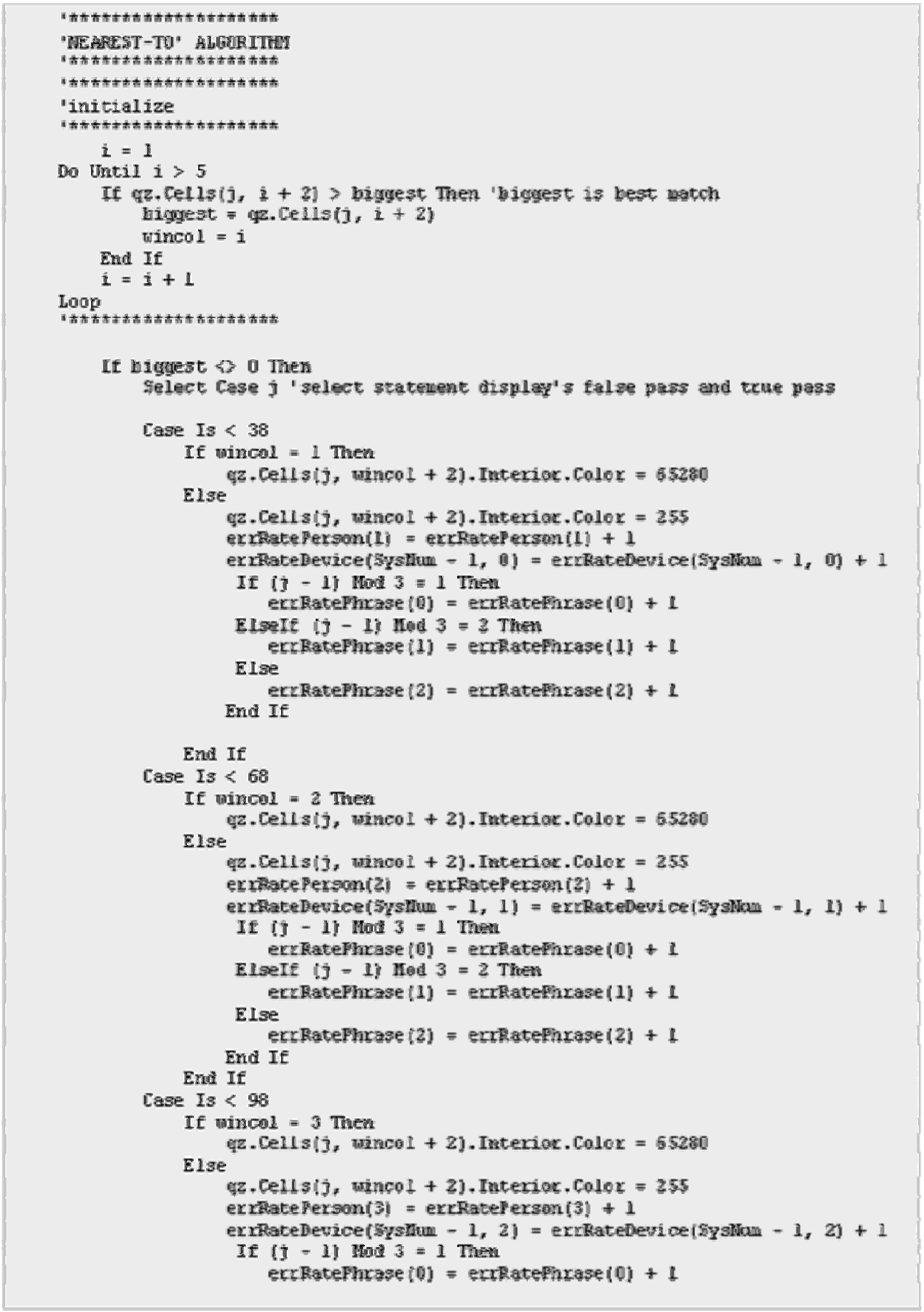
```
Private Sub CommandButton1_Click()
Dim i As Integer, wincol As Integer, biggest As Double, qz As
Object
Dim errRateDevice(9, 4), errRatePerson(5) As Integer
Dim RealNum As Integer, SysNum As Integer, y As Integer
Dim temp As Integer, falseAccept As Double, falseReject As Double
Dim FAM As Double, FRM As Double, FANM As Double, FRNM As Double
Dim SysMatchNum As Integer, userMatchNum As Integer

'___________________________________________________
'___________________________________________________
'INITIALIZATION ROUTINE
'Set all Calc Matricies to zero
'___________________________________________________
Set qz = Sheets("WINNERS (generic)")

i = 1
Do Until i > 5
     errRatePerson(i) = 0
    i = i + 1
Loop

     errRatePhrase = Array(0, 0, 0)

j = 0
Do Until j > 4
i = 0
Do Until i > 9
    errRateDevice(i, j) = 0
    i = i + 1
Loop
j = j + 1
Loop
temp = 1 ' temp is the threshold level
'___________________________________________________
```



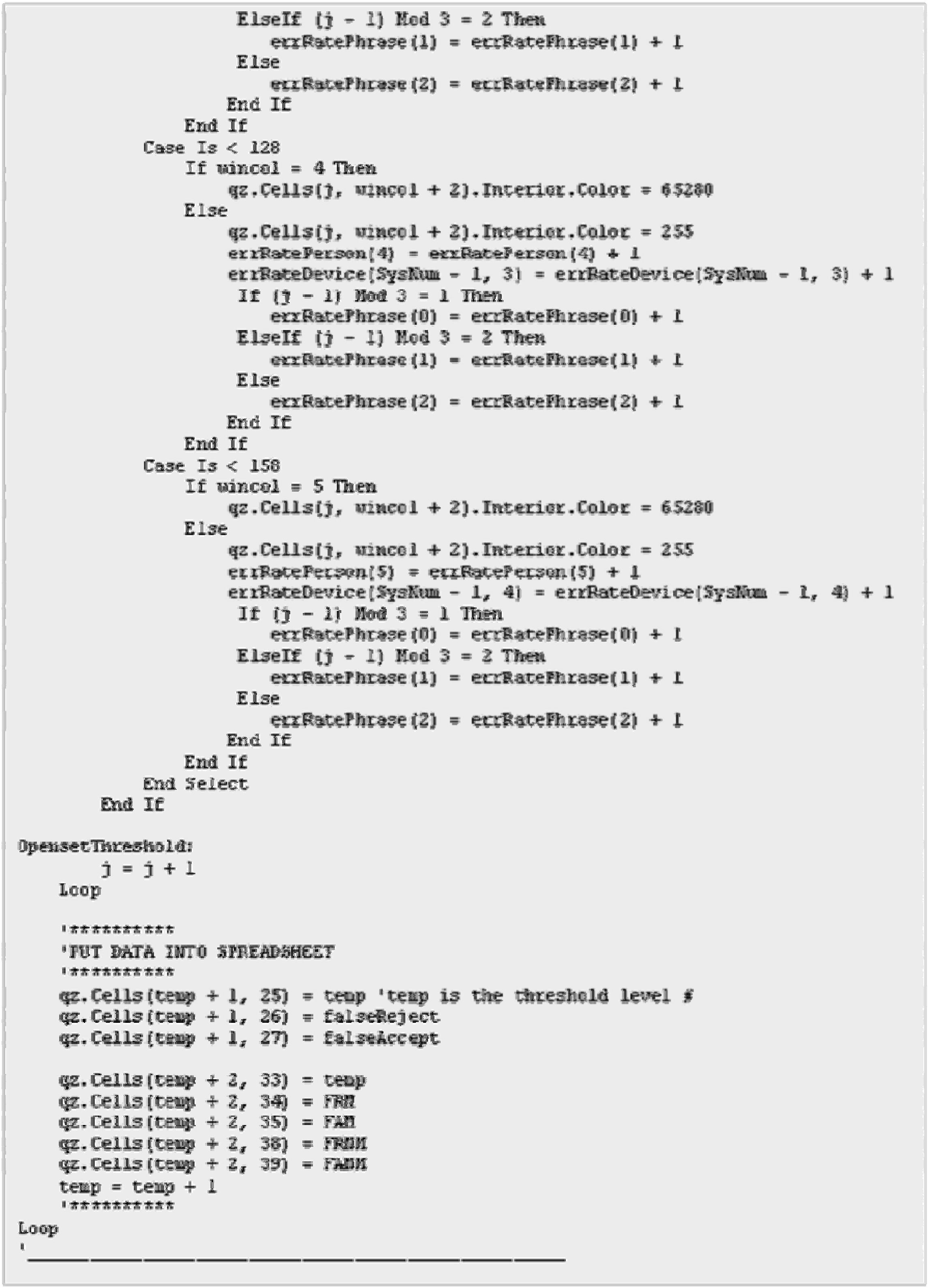
```
'____________________________________________________
'____________________________________________________
'MAIN LOOP
'PERFORM EITHER DECISION ALGORITHM
'ALGORITHM SET ON SPREADSHEET
'____________________________________________________
Do Until temp > 90

  '********************
  'initialize inner loop
  '********************
  falseAccept = 0
  falseReject = 0
  FAM = 0
  FRM = 0
  FANM = 0
  FRNM = 0
  userMatchNum = 0
  SysMatchNum = 1
  j = 8      'j is the row #
  y = 1
  '********************

  Do Until j > 157
  biggest = -10000 'limit threshold level


  '********************
  'Determine recording system number 1-10 (SysNum)
  'Determine which user (userMatchNum)
  '********************
  RealNum = (j - 8) Mod 30
  If RealNum Mod 3 = 0 Then SysNum = SysNum + 1
  If RealNum = 0 Then
    SysNum = 1
    userMatchNum = userMatchNum + 1
  End If

  If userMatchNum = 6 Then
    userMatchNum = 1
    SysMatchNum = SysMatchNum + 1
  End If
  '********************
```



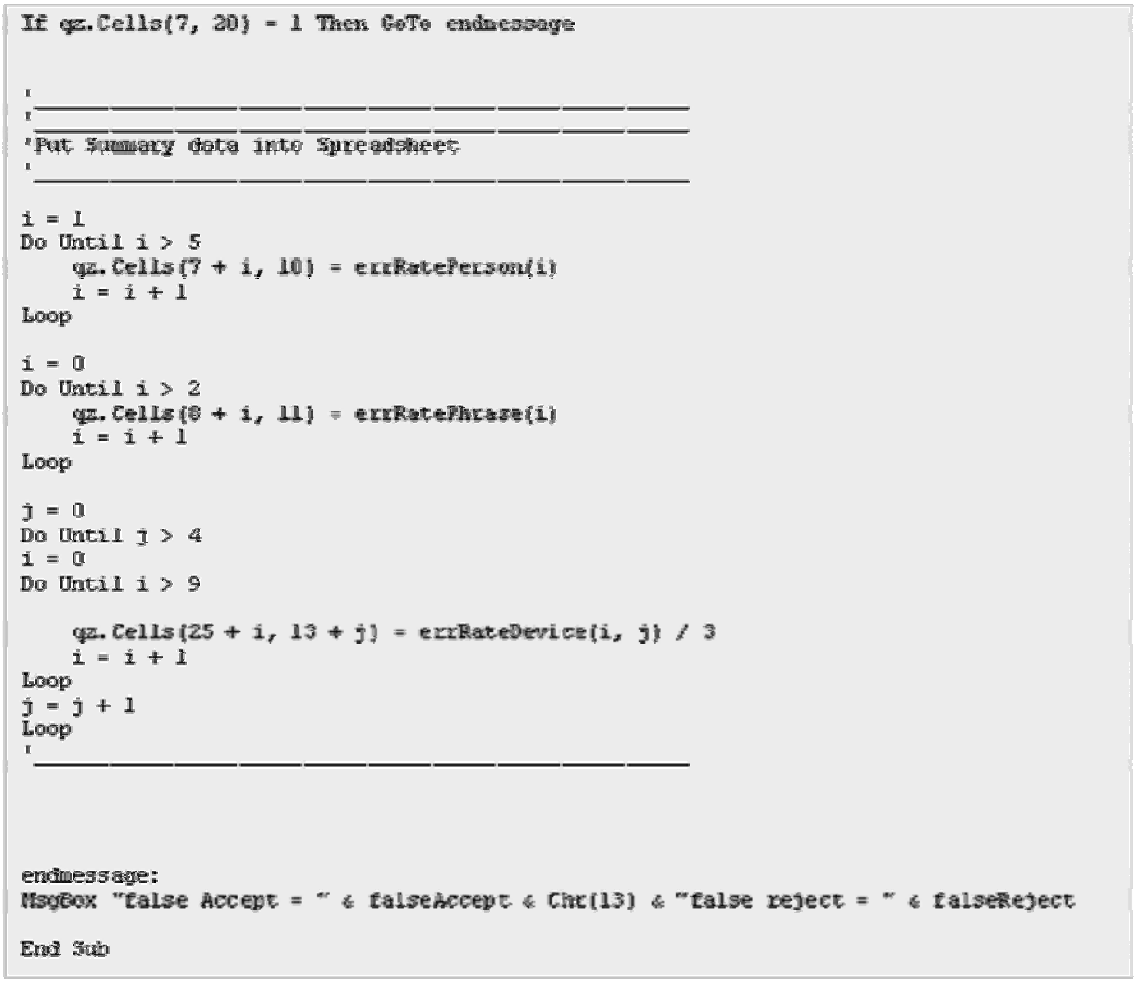
```
  '*************** *** ***
  'Determine which decision Algorithm
  '*************** *** ***
  If qz.Cells(7, 20) = 1 Then 'This is the threshold setting(openset)
  If j = 8 Then GoTo firstRunSkip
  If (j - 8) Mod 30 = 0 Then y = y + 1 'count to see where we change users
  If y = 6 Then y = 1
firstRunSkip:
  i = 1 ' is the column number(speakers 1-5)
  '*************** *** ***


  '*************** *** ***
  'SIMPLE THRESHOLD ALGORITHM
  '*************** *** ***
  Do Until i > 5

    If qz.Cells(j, i + 2) > -1 * (temp) And qz.Cells(j, i + 2) < -0.5 Then 'is pure threshold

      If y = i Then 'if the column # (i) = the y'th user then is a true match
        qz.Cells(j, i + 2).Interior.Color = 65280
      Else
        If SysNum = SysMatchNum Then
          FAM = FAM + 1
        Else
          FANM = FANM + 1
        End If
        falseAccept = falseAccept + 1
        qz.Cells(j, i + 2).Interior.Color = 255
      End If
    Else
      If y = i And qz.Cells(j, i + 2) < -0.5 Then
        If SysNum = SysMatchNum Then
          FRM = FRM + 1
        Else
          FRNM = FRNM + 1
        End If
        falseReject = falseReject + 1
      End If
      qz.Cells(j, i + 2).Interior.Color = 16777215
    EndIf
    i = i + 1
  Loop
  GoTo OpensetThreshold
  End If 'This is the end of the threshold setting(openset)
  '*************** *** ***
```



```
'********************
'NEAREST-TO' ALGORITHM
'********************
'********************
'initialize
'********************
    i = 1
Do Until i > 5
    If qz.Cells(j, i + 2) > biggest Then 'biggest is best match
        biggest = qz.Cells(j, i + 2)
        wincol = i
    End If
    i = i + 1
Loop
'********************

    If biggest <> 0 Then
        Select Case j 'select statement display's false pass and true pass

        Case Is < 38
            If wincol = 1 Then
                qz.Cells(j, wincol + 2).Interior.Color = 65280
            Else
                qz.Cells(j, wincol + 2).Interior.Color = 255
                errRatePerson(1) = errRatePerson(1) + 1
                errRateDevice(SysNum - 1, 0) = errRateDevice(SysNum - 1, 0) + 1
                 If (j - 1) Mod 3 = 1 Then
                    errRatePhrase(0) = errRatePhrase(0) + 1
                 ElseIf (j - 1) Mod 3 = 2 Then
                    errRatePhrase(1) = errRatePhrase(1) + 1
                 Else
                    errRatePhrase(2) = errRatePhrase(2) + 1
                End If

            End If
        Case Is < 68
            If wincol = 2 Then
                qz.Cells(j, wincol + 2).Interior.Color = 65280
            Else
                qz.Cells(j, wincol + 2).Interior.Color = 255
                errRatePerson(2) = errRatePerson(2) + 1
                errRateDevice(SysNum - 1, 1) = errRateDevice(SysNum - 1, 1) + 1
                 If (j - 1) Mod 3 = 1 Then
                    errRatePhrase(0) = errRatePhrase(0) + 1
                 ElseIf (j - 1) Mod 3 = 2 Then
                    errRatePhrase(1) = errRatePhrase(1) + 1
                 Else
                    errRatePhrase(2) = errRatePhrase(2) + 1
                End If
            End If
        Case Is < 98
            If wincol = 3 Then
                qz.Cells(j, wincol + 2).Interior.Color = 65280
            Else
                qz.Cells(j, wincol + 2).Interior.Color = 255
                errRatePerson(3) = errRatePerson(3) + 1
                errRateDevice(SysNum - 1, 2) = errRateDevice(SysNum - 1, 2) + 1
                 If (j - 1) Mod 3 = 1 Then
                    errRatePhrase(0) = errRatePhrase(0) + 1
```



```
                     ElseIf (j - 1) Mod 3 = 2 Then
                        errRatePhrase(1) = errRatePhrase(1) + 1
                     Else
                        errRatePhrase(2) = errRatePhrase(2) + 1
                    End If
                End If
            Case Is < 128
                If wincol = 4 Then
                    qz.Cells(j, wincol + 2).Interior.Color = 65280
                Else
                    qz.Cells(j, wincol + 2).Interior.Color = 255
                    errRatePerson(4) = errRatePerson(4) + 1
                    errRateDevice(SysNum - 1, 3) = errRateDevice(SysNum - 1, 3) + 1
                     If (j - 1) Mod 3 = 1 Then
                        errRatePhrase(0) = errRatePhrase(0) + 1
                     ElseIf (j - 1) Mod 3 = 2 Then
                        errRatePhrase(1) = errRatePhrase(1) + 1
                     Else
                        errRatePhrase(2) = errRatePhrase(2) + 1
                    End If
                End If
            Case Is < 158
                If wincol = 5 Then
                    qz.Cells(j, wincol + 2).Interior.Color = 65280
                Else
                    qz.Cells(j, wincol + 2).Interior.Color = 255
                    errRatePerson(5) = errRatePerson(5) + 1
                    errRateDevice(SysNum - 1, 4) = errRateDevice(SysNum - 1, 4) + 1
                     If (j - 1) Mod 3 = 1 Then
                        errRatePhrase(0) = errRatePhrase(0) + 1
                     ElseIf (j - 1) Mod 3 = 2 Then
                        errRatePhrase(1) = errRatePhrase(1) + 1
                     Else
                        errRatePhrase(2) = errRatePhrase(2) + 1
                    End If
                End If
            End Select
        End If

OpensetThreshold:
        j = j + 1
    Loop

    '**********
    'PUT DATA INTO SPREADSHEET
    '**********
    qz.Cells(temp + 1, 25) = temp 'temp is the threshold level #
    qz.Cells(temp + 1, 26) = falseReject
    qz.Cells(temp + 1, 27) = falseAccept

    qz.Cells(temp + 2, 33) = temp
    qz.Cells(temp + 2, 34) = FRM
    qz.Cells(temp + 2, 35) = FAM
    qz.Cells(temp + 2, 38) = FRMM
    qz.Cells(temp + 2, 39) = FAMM
    temp = temp + 1
    '**********
Loop
'____________________________________________________
```



```
If qz.Cells(7, 20) = 1 Then GoTo endmessage


'____________________________________________________
'____________________________________________________
'Put Summary data into Spreadsheet
'____________________________________________________

i = 1
Do Until i > 5
    qz.Cells(7 + i, 10) = errRatePerson(i)
    i = i + 1
Loop

i = 0
Do Until i > 2
    qz.Cells(8 + i, 11) = errRatePhrase(i)
    i = i + 1
Loop

j = 0
Do Until j > 4
i = 0
Do Until i > 9

    qz.Cells(25 + i, 13 + j) = errRateDevice(i, j) / 3
    i = i + 1
Loop
j = j + 1
Loop
'____________________________________________________



endmessage:
MsgBox "false Accept = " & falseAccept & Chr(13) & "false reject = " & falseReject

End Sub
```

VITA

Clark Damon Shaver

Candidate for the Degree of

Master of Science

Thesis: EFFECTS OF EQUIPMENT VARIATIONS ON SPEAKER RECOGNITION ERROR RATES

Major Field: Electrical Engineering

Biographical:

Education:
Completed the requirements for the Master of Science in electrical Engineering at Oklahoma State University, Stillwater, Oklahoma in December, 2009.
Received a Bachelor of Science in Electrical Engineering, specializing in computer engineering in May of 2006.

Experience: Automation Engineer, Blue Bell Creameries, 2000 – 2006. R&D Engineer for Baker Hughes-Centrilift, 2006-present

Professional Memberships: Institute of Electrical and Electronics Engineering, Foundation of Ancient Research and Mormon Studies, American Society of Church History

Name: Clark D. Shaver Date of Degree: December, 2009

Institution: Oklahoma State University Location: Stillwater, Oklahoma

Title of Study: EFFECTS OF EQUIPMENT VARIATIONS ON SPEAKER RECOGNITION ERROR RATES

Pages in Study:112 Candidate for the Degree of Master of Science

Major Field: Electrical Engineering

Scope and Method of Study: The purpose of this study was to examine the effects that equipment variation has on speaker recognition performance. Specifically microphone variation is investigated. The study examines the error rates of a speaker recognition system when microphones vary between the enrollment and testing phases. The study also examines the error rates of a speaker recognition system when microphones differ in similar environments and conditions. The metric for evaluation of effect is the false identity acceptance and the false identity rejection error rates.

Findings and Conclusions: The results of the research demonstrate that microphone variation has a major effect on speaker recognition error rates. Error rates include the rates of false acceptance and false rejection of identity. The effect of a training / enrollment microphone mismatched was significant. Mismatched conditions produce significantly greater false accept and false reject rates as compared to matched conditions. In fact, the mismatched condition had a more significant impact on error rates than noise. The research also demonstrates that speaker recognition error rates are microphone dependant. The microphone dependency was seen in both the matched and mismatched condition. However, the microphone dependency was more prevalent in the matched condition. Microphone selection has an effect on error rates in different environments and in matched and mismatched conditions. The research provides a method to evaluate the direct effects of microphone selection on speaker recognition systems.

ADVISER'S APPROVAL: Dr. John M. Acken