ACCESS CONTROL BY MEANS OF SPEECH RECOGNITION AND ITS IMPACT ON THE AUDITOR

BY

JOHAN HENDRIK OTTO VAN GRAAN

ESSAY

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE

MASTER OF COMMERCE

IN

COMPUTER AUDITING

IN THE

FACULTY OF ECONOMIC AND BUSINESS SCIENCES

AT THE

RAND AFRIKAANS UNIVERSITY.

STUDY LEADER : MR A. DU TOIT

JOHANNESBURG

NOVEMBER 1990

ACKNOWLEDGMENT

To First National Bank of Southern Africa Limited for making documentation on its Voice Recognition Pilot available for use as an example of a speech recognition application.

DECLARATION

I declare that this essay hereby submitted to the Rand Afrikaans University for the degree Master of Commerce is my own unaided work, except to the extent acknowledged in the text, and has not been submitted previously for any degree or examination at this or any other university.

Johan Hendrik Otto van Graan.

TABLE OF CONTENTS

TOEGANGSBEHEER DEUR SPRAAKHERKENNING

EN DIE INVLOED DAARVAN OP DIE OUDITEUR

DEUR

JOHAN HENDRIK OTTO VAN GRAAN

OPSOMMING VAN VERHANDELING INGEDIEN

VIR DIE GEDEELTELIKE VOLDOENING AAN DIE VEREISTES

VIR DIE GRAAD MAGISTER IN REKENAAROUDITERING

IN DIE FAKULTEIT EKONOMIESE EN BESTUURSWETENSKAPPE

BY DIE RANDSE AFRIKAANSE UNIVERSITEIT

STUDIELEIER: MNR A. DU TOIT

JOHANNESBURG

NOVEMBER 1990

OPSOMMING

TOEGANGSBEHEER

Toegangsbeheer het in die laaste tyd al hoe belangriker geword vanwee die toename in misdaad en politieke onrus. Die toename in data op rekenaars het ook die behoorlike toegangsbeheer tot data meer noodsaaklik gemaak.

Identifisering van 'n gebruiker deur 'n rekenaar kan op drie maniere gedoen word. Die identifisering is gebaseer op iets wat die gebruiker besit; iets wat die gebruiker weet of 'n persoonlike eienskap van die gebruiker:

* Die iets wat die gebruiker besit is gewoonlik 'n kaart met 'n magneetstrokie daarin. Die rekenaar kontroleer of die kaart 'n geldige kaart is. Hierdie metode word gewoonlik gebruik in die toegangsbeheer tot geboue.

* Die iets wat 'n gebruiker weet is gewoonlik 'n kodewoord. Hierdie kodewoord word ingesleutel deur die gebruiker en die rekenaar gaan dan na of die regte kodewoord ingesleutel is.

* Die rekenaar moet die persoonlike eienskappe van die gebruiker eien en die gebruiker positief identifiseer. Voorbeelde van eienskappe is vingerafdrukke, spraak en handskrif. Die gebruik van persoonlike eienskappe bied die grootste graad van veiligheid in toegangsbeheer.

OORSIG OOR SPRAAKHERKENNING

Die proses van spraakherkenning deur die rekenaar werk soos volg: klankgolwe word omskep deur die rekenaar na 'n digitale voorstelling daarvan. Die digitale voorstelling word dan as 'n rekord gestoor. Wanneer met die rekenaar gepraat word, word die klanke vergelyk met die rekords wat reeds voorheen deur die rekenaar gestoor is. Indien dit ooreenstem kan die rekenaar daarop reageer.

Daar is twee metodes van spraakherkenning:

Spreker onafhanklike herkenning.

Spreker afhanklike herkenning.

By spreker onafhanklike herkenning het die rekenaar 'n aantal rekords gestoor. Hierdie rekords kan gesien word as die rekenaar se woordeskat. Die rekords is die digitale voorstelling van 'n spesifieke woord van die gemiddelde spreker. Die woordeskat van die rekenaar is 'n voorstelling van die gemiddelde spreker, daarom moet daar vir 'n graadverskil voorsiening gemaak word. Dit is dus moontlik dat die rekenaar die manier waarop 'n sekere spreker die woord "nege" sê, as "sewe" mag hoor.

By spreker afhanklike herkenning word die rekenaar se woordeskat opgebou uit rekords van 'n spesifieke woord vir 'n spesifieke persoon. Hier is dus geen moontlikheid vir 'n fout in die herkenning deur die rekenaar nie. Vir spreker afhanklike herkenning moet egter baie meer rekords gestoor word as by spreker onafhanklike herkenning.

TOEGANGSBEHEER DEUR SPRAAKHERKENNING

Spraakherkenning kan in beide logiese sowel as fisiese toegangsbeheer gebruik word.

Fisiese toegangsbeheer het betrekking op die beheer van toegang tot eiendom. Logiese toegangsbeheer verwys na die beheer van toegang tot data en funksies op 'n rekenaar.

'n Voorbeeld van fisiese toegangsbeheer is die toegang tot 'n gebou of gedeeltes daarvan. Die gemagtigde persone se kodewoorde word as rekords deur die rekenaar gestoor. Wanneer 'n persoon toegang tot die gebou wil verkry, word hy vir sy kodewoord gevra. Die kodewoord word omskep in digitale vorm en vergelyk met die gestoorde rekords. Indien dit ooreenstem word toegang verleen.

Wanneer daar van afhanklike herkenning gebruik gemaak word, kan die kodewoord enige woord of woorde wees. By onafhanklike herkenning is dit gewoonlik 'n aantal syfers. Die rekenaar sal elke syfer van die kodewoord vergelyk met sy rekord (woordeskat) van syfers. Die kodewoord word dan vergelyk met 'n gemagtigde tabel voordat toegang verleen word.

INVLOED OP DIE OUDITEUR

Wanneer die ouditeur 'n stelsel wat van spraakherkenning gebruik maak, oudit, moet hy die volgende vasstel:

* Word van 'n hoofraam, persoonlike rekenaar, plaaslike netwerk of 'n mini gebruik gemaak?
* Word spreker afhanklike herkenning, spreker onafhanklike herkenning of beide gebruik?
* Word die stelsel vir fisiese toegangsbeheer, logiese toegangsbeheer of beide gebruik?
* Word die stelsel gebruik vir toegangsbeheer tot hoe sekuriteitsareas/-data, lae sekuriteitsareas/-data of vir beide?

Die antwoorde op bogenoemde vrae sal die stelsel kategoriseer in die risiko matrys wat volg.

Risiko Matrys

	SPREKER ONAFHANKLIKE HERKENNING	SPREKER AFHANKLIKE HERKENNING	BEIDE
LAE SEKURITEIT	1	2	3
HOE SEKURITEIT	4	5	6
BEIDE	7	8	9

Oudit Riglyn

Uit 'n oudit oogpunt is 1 die laagste en 9 die hoogste risiko.

Vir kategorie 1 tot 3 behoort die ouditeur 'n oorsig te doen van die toegang tot 'n gebou of gedeeltes daarvan, asook die data wat deur die stelsel beheer word. Die ouditeur kan dan besluit of verdere oudit stappe nodig is.

Die ouditeur behoort die volgende te oudit vir kategorie 4 tot 9:

* Standaarde van die stelsel.
* Die prosedures vir die onderhoud van gebruikers van die stelsel.
* Die beheer oor die rekenaar waarvan die stelsel gebruik maak.
* Die toegangsprofiel van persone.
* Die aanvaarbaarheid en reputasie van die spraakherkenning apparatuur en programatuur wat gebruik word.
* Die prosedures wat gevolg word om gebruikers te laai op die stelsel waar daar van afhanklike herkenning gebruik gemaak word.
* Die riglyne vir waar afhanklike en waar onafhanklike herkening gebruik moet word.

Die ouditeur moet verseker dat onafhanklike herkenning slegs gebruik word vir lae risiko areas/data en afhanklike herkenning gebruik word vir ho risiko areas/data.

GEVOLGTREKKING

Spraakherkenning het gekom om te bly. Die gebruik daarvan sal toeneem met die tyd en in verskillende soorte stelsels gebruik word.

Die ouditeur hoef hom op die oomblik nie te veel te bekommer oor spraakherkenning nie. Die rede daarvoor is dat die gebruik van spraakherkenning beperk is tot toegangsbeheer. Beide fisiese en logiese toegangsbeheer kan maklik geoudit word deur normale ouditprosedures te gebruik saam met 'n basiese kennis van spraakherkenning.

Die toekoms belowe egter opwindende toepassings van spraakherkenning. Daar is selfs 'n moontlikheid dat 'n persoon met die rekenaar sal kan praat asof dit 'n ander persoon is. Die ouditeur sal op hoogte van sake moet bly en saam met tegnologie groei.

SUMMARY

In recent times access control has become more and more important, largely as a result of changes in society and an increase in the quantity and sensitivity of information being stored on computers.

Speech recognition is nothing but communication which occurs when two persons have a conversation and one understands what the other says and means. This process consists of sound waves (analogue signals) that are carried through the air. The sound is converted (digitized) by the ear to impulses. The brain matches these impulses to a meaning (template) to which the person responds by an action.

Speaker independent recognition involves converting the spoken word into an electronic signal. The signal is then compared to the computer's vocabulary, which consists of a set of templates which have been chosen to represent the average speaker.

Speaker dependent recognition consists of training the computer to recognize a specific word spoken by an individual. This is done by having the speaker say the word several times. The computer then creates an average template for that word for that speaker which is then used for reference.

For any speech recognition system that an auditor needs to audit, the following have to be established:

* What does the system reside on? A mainframe, Mini, PC or LAN.
* Is the system speaker independent, speaker dependent or both?
* Is the system used for control of physical access, logical access or both?
* Is the system used for control of access to high security area/data, low security area/data or both?

The answers to the above will place the system in one of the categories of the following risk matrix.

Risk Matrix

	SPEAKER INDEPENDENT RECOGNITION	SPEAKER DEPENDENT RECOGNITION	BOTH
LOW SECURITY	1	2	3
HIGH SECURITY	4	5	6
BOTH	7	8	9

The risks for the auditor are ranked from 1 to 9, where 1 represents the lowest risk and 9 the highest risk.

For categories 1 to 3 the auditor should review the data and areas of access controlled by the application. He should also consider whether any further work is necessary.

The auditor should review the following for categories 4 to 9:

* Standards for the system.
* Procedures for maintenance of the users of the system.
* Control over the PC/Mini/Mainframe on which the system resides.
* The type of profiles that a person can be allocated.
* The market's acceptance and recognition of the card and independent template used in the system.
* The procedures for training the system to provide dependent recognition.
* Guidelines on where dependent and independent recognition should be used.

At the moment the auditor need not be excessively concerned about speech recognition, as it is mainly confined to access control. Both physical and logical access control can easily be audited using normal audit techniques, with a basic knowledge of speech recognition.

The future promises exciting applications for speech recognition, which may even include the ability to communicate with the computer in the same way as one speaks to another human being. The auditor will have to grow with technology and keep up to date with developments.

CHAPTER 1

INTRODUCTION

The objective of this chapter is to provide reasons for the choice of this subject, as well as describe the approach that was used in the research. This chapter is divided into the following areas:

1 INTRODUCTION AND FORMULATION OF THE PROBLEM

2 OBJECTIVES OF THIS ESSAY

3 RESEARCH APPROACH

4 CONSTRAINTS AND EXCLUSIONS

1 INTRODUCTION AND FORMULATION OF THE PROBLEM

Speech Recognition is a new technology which is taking off very rapidly. While more and more companies are beginning to use it, at present it is used only for access control in most cases. In due course, however, other applications will be found for speech recognition, which presents a problem to the auditor as he or she needs to keep up to date with, and learn about, this new technology.

In this country not much has been published on speech recognition and my research revealed no publication which directly addresses the impact of speech recognition on the auditor.

The problem is that the auditor needs to audit speech recognition applications and must understand the underlying principles of speech recognition. By using normal auditing techniques, he must assess the risk and audit the application of speech technology.

2 OBJECTIVES OF THIS ESSAY

This essay has two objectives. Firstly it explains what speech recognition is and the principles on which it is based. Some applications which use speech recognition are discussed subsequently. The discussion includes pointers on the audit concerns and how to audit the application.

The second objective is to develop a model which the auditor can use as a guideline for auditing speech recognition applications.

3 RESEARCH APPROACH

The research approach used can be divided into three distinct areas. The research was begun with a literature survey. This was followed by a description of some speech recognition applications. The research was concluded by creating a model based on the information derived from the first two research areas.

A literature survey was done on the following:

Access control;

Speech recognition;

Digital signal processing;

Hardware and Software used in speech recognition.

4 CONSTRAINTS AND EXCLUSIONS

Speech recognition is a complex subject which draws from the fields of mathematics, electronic engineering and computer science. This is further complicated by the fact that speech recognition can be used in various applications. Therefore, some constraints and exclusions are necessary.

The constraints are summarized as follows:

- No attempt is made to explain the mathematical formulas on which digital signal processing is based.
- The electronic devices used in speech recognition are not described in detail.

The following overall control objectives are excluded:

- Completeness of input, processing, updating of files and output.
- Accuracy of input, processing, updating of files and output.
- Maintenance of data while transient and static.

They, however, do apply to an access control system using speech recognition i.e. the control over the input of the recorded voice of the person that will use the system.

Authoritative literature dealing with the subject of access control using speech recognition could not be found, thus various articles and pamphlets had to be used.

CHAPTER 2

LITERATURE SURVEY

This chapter is divided into the following areas:

1 ACCESS CONTROL

2 OVERVIEW OF SPEECH RECOGNITION

3 ANALYSIS OF SPEECH RECOGNITION

4 HARDWARE AND SOFTWARE USED IN SPEECH RECOGNITION

1 ACCESS CONTROL

1.1 Introduction

In recent times access control has become more and more important, as a result of changes in society and the increase in information being stored on computers.

The increase in crime and political unrest has necessitated tightening the security of physical access to buildings, in order to stop the planting of bombs and to prevent thieves gaining access to valuable assets.

The competitiveness and growth in business has forced companies to rely more and more on computers. Vast amounts of data are stored on computers and managers rely on this information to make decisions. It is thus important that the access to data be controlled, in order to stop intentional or unintentional manipulation of data that will destroy its integrity.

1.2 Definition of Access Control

According to Vallabhaneni (1983:262), identification to a computer system can be classified into three categories based on the following:

* Something that the user has in his/her possession.
* Something known to the user.
* Something to do with the user's personal characteristics.

The first category uses keys and magnetic or optical cards to allow access to a system, whereby the system verifies that the item is valid. The system cannot, however, verify whether the item is being used by the authorized user. This method is normally used for physical access. It is relatively inexpensive, but unauthorized access can be gained by theft or duplication of the key or the card.

The second category consists of the use of passwords as an access control facility. A combination of a magnetic card and a password or a PIN (Personal Identification Number) offers a satisfactory degree of security, and is most often used by banks in automatic teller systems. The success of this method depends on the password being known only to the authorized user.

The third category positively identifies the user via body geometry, fingerprints, speech recognition, signature analysis, and various other techniques which use generic characteristics. This method makes access by an unauthorized person virtually impossible.

2 OVERVIEW OF SPEECH RECOGNITION

2.1 Introduction

Speech recognition is nothing but communication which occurs when two persons have a conversation and one understands what the other says and means. This process consists of sound waves (analogue signals) that are carried through the air. The sound is converted (digitized) by the ear to impulses. The brain matches these impulses to a meaning (template) to which the person responds by an action.

Since the invention of the first computer, man has wanted to talk directly to computers and to be understood in the same way as when talking to another person. While the communication process between man and computer has been the subject of research for decades, success has been hampered because computers are unable to recognize the spoken word.

Speech recognition has recently became jargon in Information Technology, because it has become possible to talk to computers and to be understood to a limited extent.

Communicating with computers through speech recognition works as follows:

The analogue signal is converted by the computer to a digital representation of the sound wave, which is known as a template. The computer has several templates stored in its memory. The computer tries to match the new template to the stored templates. If a match is made, the computer reacts by performing an action or talking back via the screen or speech synthesizer.

2.2 Biometric recognition

Speech recognition is a subset of Biometric recognition. Some of the other Biometric methods are:

Thumbprint, fingerprints, handprints;

Retina scanning;

Hair analysis.

All Biometric methods are based on prerecorded templates. Various types of prints, retina scanning and speech recognition have been used mainly in access control to ensure physical security.

2.3 Types of speech recognition

According to Furui (1989:227) there are two basic different types of speech recognition. They are:

* Speaker independent recognition
* Speaker dependent recognition.

2.3.1 Speaker independent recognition

Speaker independent recognition consists of converting the spoken word into an electronic signal, which is then compared to the computer's vocabulary. This vocabulary consists of a set of templates which have been chosen to represent the average speaker.

For speaker independent recognition to function effectively, there should be a pause between words. This is necessary as computers cannot yet understand a conversation and will only understand those words that form part of its vocabulary.

Since the template stored for a word is that of the average speaker, allowance for a tolerance level must be made. This means there is scope for error. A speaker may, for example, pronounce "five" in such a way that when matched to the templates the computer will understand it to mean "nine".

2.3.2 Speaker dependent recognition

Speaker dependent recognition is where the computer is trained to recognize a specific word spoken by an individual. This is done by having the speaker say the word several times. The computer then creates an average template for that word as spoken by that speaker which is then used for reference.

In speaker dependent recognition a conversation is also not possible. There is, however, very little scope for error, as a word offered for recognition by a speaker is compared against an average template for that speaker.

2.4 Uses for speech recognition

	USE	DESCRIPTION	TYPE
1	Aircraft Weapon Systems	The USA Air Force has a system installed in the attack helicopters that controls the aiming and firing of the weapon systems. This is a good example of the use of speech recognition in a high background noise area. Coler (1982)	Dependent

	USE	DESCRIPTION	TYPE
2	Access control in buildings:		
	High security area	A person wanting to gain access will enter a PIN that identifies that person's template. The person will then be asked three or more passwords in a random order with a set response time.	Dependent
	Low security area	A person wanting to gain access will speak a password (normally a number) on request. This will be compared against a table of valid passwords.	Independent
3	Logical access control	A person will enter his user number via a keyboard (the user number can also be entered by way of independent recognition). This will identify the person's template. The password will then be spoken. If there is not a 100% match, access will be denied.	Dependent

3 ANALYSIS OF SPEECH RECOGNITION

3.1 Introduction

To understand how speech recognition works, the following should be examined:

* There must be an understanding of speech.
* The process of digital coding of speech must be understood.

3.2 Speech

3.2.1 The Speech Process

According to Yannakoudakis (1987:15) speech consists of a continuously varying sound wave which links the speaker to the listener. Sound moves through a medium. The most common medium is air.

When we make a sound we disturb molecules of air close to our mouths. These molecules oscillates about their point of rest and collide with adjacent molecules and create a chain reaction.

The following definitions of the characteristics of the speech process will be used in this paper:

* The "Amplitude" of vibration is the distance which the molecules move from their point of rest.

* The number of times that molecules move in a complete cycle in one second is called the "frequency".

3.2.2 The Human Speech Process

Yannakoudakis (1987:19) goes on to describe the human speech process.

The fundamental frequencies are produced by the vocal cords, which are then transformed by the various resonating chambers to form the various sounds we utter.

When speaking we use a combination of the following:

Larynx (vibration source);
Lungs (energy source);
Vocal tract (resonance source);
Nasal cavity (resonance source);
Articulatory organs (to change the "shape" of the resonant cavities).

3.2.3 The building blocks of speech

Phones are the individual sounds of speech, i.e. the minimum units of identifiable speech.

Phones grouped together for organizational convenience are phonemes.

Allophones are variants of phonemes which characterize the precise nature in a given context.

Allophones grouped together produce morphemes or words.

3.2.4 Problems associated with the human speech process

Different sound waves emitted from individuals produce the same word. There are significant differences in pitch, intonation and stress. Sounds in speech are context dependent and, even in the same context, are subject to a wide variety of acoustic interpretation.

Phones are pronounced at various speeds, depending on the context and the speaker.

There are also physical differences between speakers i.e. the vocal chords of a man are different to those of a woman. Therefore the frequencies for the same word are different. This complicates recognition by a computer. While man can understand what is said by placing words into context and anticipating what he will hear, the computer is unable to do so.

3.3 Techniques for Digital Coding of Speech

3.3.1 Introduction

Speech can be stored by the computer in either analogue or digital form.

3.3.1.1 Analogue storage

This method stores the original analogue representation of the sound wave. It is unsuitable for computer use as it is difficult to access quickly and tape recorders are prone to mechanical breakdown.

3.3.1.2 Digital Storage

Digital storage is divided into two types; waveform and parametric. In both techniques a series of binary digits have to be sorted to represent the incoming sound wave. This is done by sampling the wave. An analysis is performed to obtain some form of binary representation of the changing shape of the wave.

Digital signal processing is twofold; obtaining a discrete representation of the signal, and through various mathematical methods, the processing and implementation of results.

In converting analogue to digital a record of both amplitude and associated time has to be made. The analogue signal is first sampled and the samples are then converted to digital form.

3.3.2 Sampling and Digitalization

When taking a sample, there should not be any sound waves with frequencies greater than half the sample frequency.

After taking a sample, the amplitude has to be digitized. One way to digitize is to split the amplitude range into equal regions, where points are termed "levels"., Each level has a binary number associated with it. The amplitude of the sample is then compared with the levels, rounded down to the nearest level and stored.

The problem is, if one assumes a sample frequency of 10 000 samples per second (this is not unreasonable for speech) with a 8-bit converter, then 80 000 bits per second are needed. As 10 to 15 phonemes are produced in a second of normal speech, a considerable amount of storage is needed.

Various techniques of digitizing attempt to overcome the storage problem by reducing the number of bits that need to be stored. This is done by removing redundant information without affecting the quality of the reproduced speech signal.

3.3.3 Waveform Coding

Waveform coding aims to preserve all the necessary features present in the original analogue speech signal by digitizing the entire sound wave. According to Furai (1989:71) the following are the most important techniques:

* Pulse code modulation;
* Differential pulse code modulation;
* Adaptive pulse code modulation;
* Delta modulation;
* Adaptive delta modulation;
* The Mozer method.

The disadvantage of these techniques is that storage requirements are high. Approximately 17000 to 70000 bits per second of speech are needed.

3.3.3.1 Pulse code modulation (P.C.M.)

This technique uses the most bits in the digitizing process. It is, however, the most established and most widely applied of all digital coding techniques. Advantages are that it is instantaneous and is not signal specific.

P.C.M. analyses and reconstructs a waveform by using only the bandwidth of the waveform and ignoring the underlying structure, with the result that complete messages can be generated. This is, however, not always possible by concatenating the one held in memory to create new phrases, messages, etc.

3.3.3.2 Differential pulse code modulation (D.P.C.M.)

This technique uses prediction techniques to anticipate the incoming signal. Although storage space is saved, the technique use more resources to implement the prediction algorithm.

In D.P.C.M. the difference between the incoming and the predicted signal is split in steps of amplitude and encoded for transmission. Owing to the predicting of the incoming signal, the technique has to be speech dependent, because knowledge of the speech process is needed for the prediction.

As the difference in amplitude is split into fixed steps, two basic reconstruction errors can occur: slope overload and granular noise. Slope overload occurs when the step size is too small for the incoming wave, while granular noise arises when the step size is too large.

3.3.3.3 Adaptive pulse code modulation (A.P.C.M.)

In A.P.C.M. the predictor or step size can be altered to consider the 'shape' of the signal. This means that the contours of the speech signal can be followed more readily.

3.3.3.4 Delta modulation (D.M.)

D.M. is a subclass of D.P.C.M. and uses only one-bit quantizisers. With D.M. the changes in the signal amplitude between consecutive samples are transmitted in place of the absolute signal amplitude.

3.3.3.5 Adaptive delta modulation (A.D.M.)

Both the amplitude and the sample frequency can be altered.

3.3.3.6 The Mozer method.

Only the relevant information is extracted from the signal and is used only with speech signals. The speech wavefrom is divided into basic segments called "analysis periods" that are equal to the period of the vocal chord vibration.

3.3.4 Parametric coding

Parametric coding aims to removes redundancies in the acoustic signal by encoding only the core signals. The core signals are based on known human sounds, derived from the study of the human vocal tract and articulatory organs. According to Furui (1989:55) examples of this technique are channel coding, formant coding and linear predictive coding.

3.3.4.1 The vocoder

The vocoder is the oldest form of channel coding and usually operates at speeds less than 6 kilobits per second. Channel coding digitizes speech by passing the source signal through a series of filters in parallel. Each filter processes the amplitude and its associated frequency band to distinguish between voiced and non-voiced signals.

The main advantage of the vocoder is the quality of signal reproduction.

3.3.4.2 Format coding

Format coding reduces the redundancy in the transmitted signal even further by making use of the format structure of voiced speech sounds.

3.3.4.3 Linear predictive coding (L.P.C.)

A linear predictor compares the input and output signals to extract the fundamental frequencies and formant trajectories.

A linear prediction coder consists of three components:

* A transmitter, which carries out the L.P.C. analysis and pitch detection as well as coding the parameters.
* A channel, which sends the parameters.
* A receiver, which can decode the parameters to produce synthesized speech.

3.4 Speech Recognition

3.4.1 Introduction

There are several major problem areas in speech recognition:

* Continuous speech has to be segmented.
* Speech patterns vary between speakers as well as for an individual speaker.
* A word can vary in volume, pitch, stress and pronunciation rate.
* The speaker's geographical origin has an impact.
* Different words sound similar.
* Background noise can distort the signal.
* Individual elements lose their identity in the speech process i.e. word flow into another when talking fast.

There are two basic different types of speech recognition:

* Speaker dependent recognition
* Speaker independent recognition.

There is also a further differentiation in speech recognition; that of isolated word recognition and continuous speech recognition.

Speaker independent recognition involves converting the spoken word into an electronic signal, which is then compared to the computer's vocabulary. This vocabulary consists of a set of templates which has been chosen to represent the average speaker. The set of templates is normally for the numbers "0" to "9" and the words "Yes" and "No".

In speaker dependent recognition, the computer is trained to recognize a specific word spoken by an individual. This is done by having the speaker saying the word several times. The computer then creates an average template of that word for that speaker which is then used for reference.

3.4.2 Basic steps in speech recognition

Yannakoudakis (1987:68) define the basic steps in speech recognition as the following:

* Sampling;
* Detection of start and end of incoming signal;
* Calculation of speech spectra;
* Pitch contour evaluation;
* Segmentation;
* Word recognition;
* Responding to a message.

See Diagram 1

3.4.2.1 Sampling

The speech signal is sampled and then digitized using an analogue to digital converter (techniques for digitizing are described in section 3.3).

3.4.2.2 Detecting the beginning and end of an incoming signal

The signal processor can detect the difference between voice signals and non-voice signals. The algorithms used for digitizing also incorporate the necessary logic to detect the beginning and the end of the incoming signals.

STEPS IN SPEECH RECOGNITION

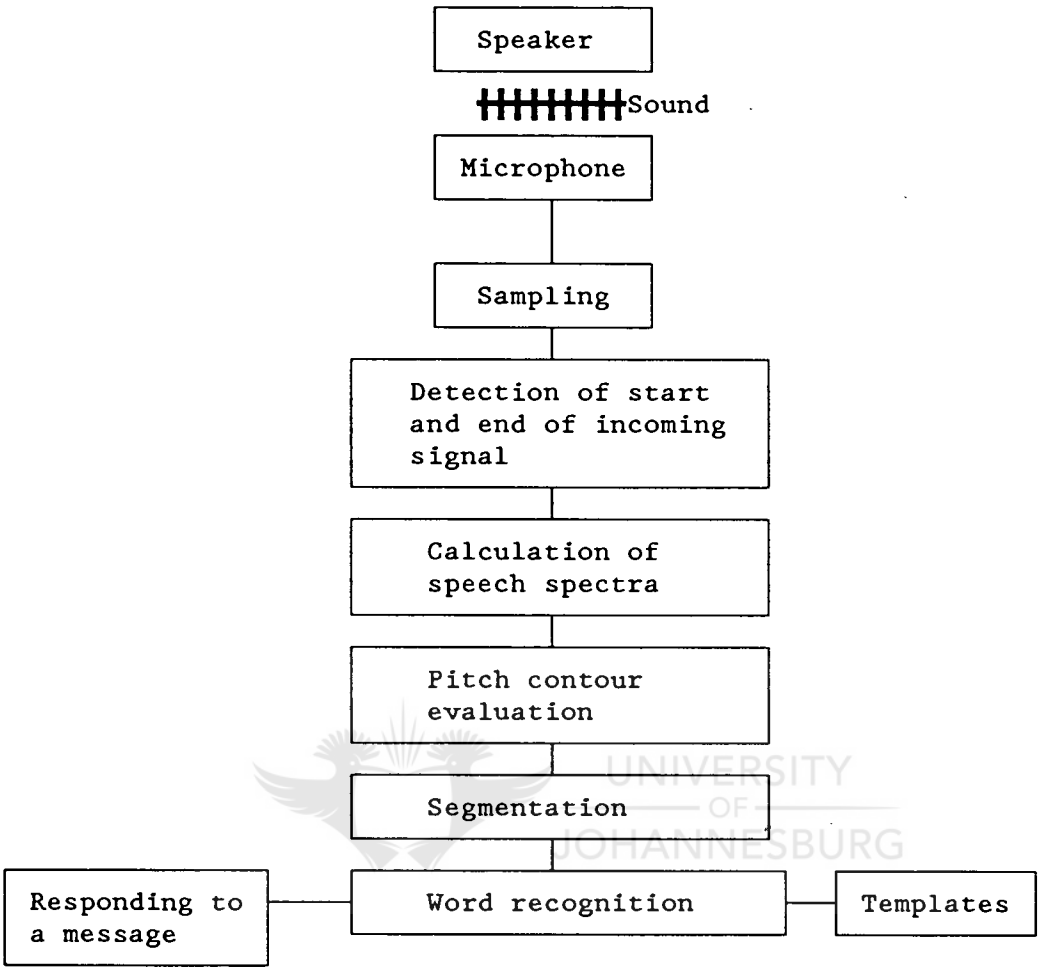

Diagram 1

3.4.2.3 Calculation of speech spectra

Here the amplitude of a signal is examined as a function of the frequency to allow the identifying features to be extracted.

3.4.2.4 Pitch contour evaluations

The contours of the fundamental frequency (as it rises and falls) are used as an indication of major syntactic boundaries.

3.4.2.5 Segmentation

Segmentation is done either on the basis of phonetic analysis, or according to certain parameters (voiced/unvoiced), or on a time basis.

Segmentation on a time basis

Speech signals are chopped into time segments ignoring phones, phonemes, syllables, etc. All signals are treated equally. The major disadvantages are the storage space needed and the number of different templates needed for the same word to accommodate varying speech rates of individual speakers.

Segmentation according to various parameters

There are various parameters that characterize different speech sounds. These parameters are used for phonetic processing to determine the point where the speech signal is to be segmented.

Some of the parameters are:

* Voiced/unvoiced
* Sharp variations in the intensity of the signal as a function of time
* Formants
* Gross spectral shape
* Zero-crossing density

Segmentation on phonetic basis

Segmentation can take place at allophone, phoneme, diphone, syllable and word level.

3.4.2.6 Word recognition

This is where incoming signals are matched to the stored templates.

3.4.2.7 Responding to the message

This is system dependent i.e. allow access, run a programme, give a balance of a bank account etc.

4 HARDWARE AND SOFTWARE USED IN SPEECH RECOGNITION

4.1 Introduction

All speech recognition systems consist of four basic components: a microphone, a central processing unit (CPU), a storage device and a loudspeaker. These hardware components are all controlled by software.

See diagram 2.

COMPONENTS OF SPEECH RECOGNITION SYSTEM

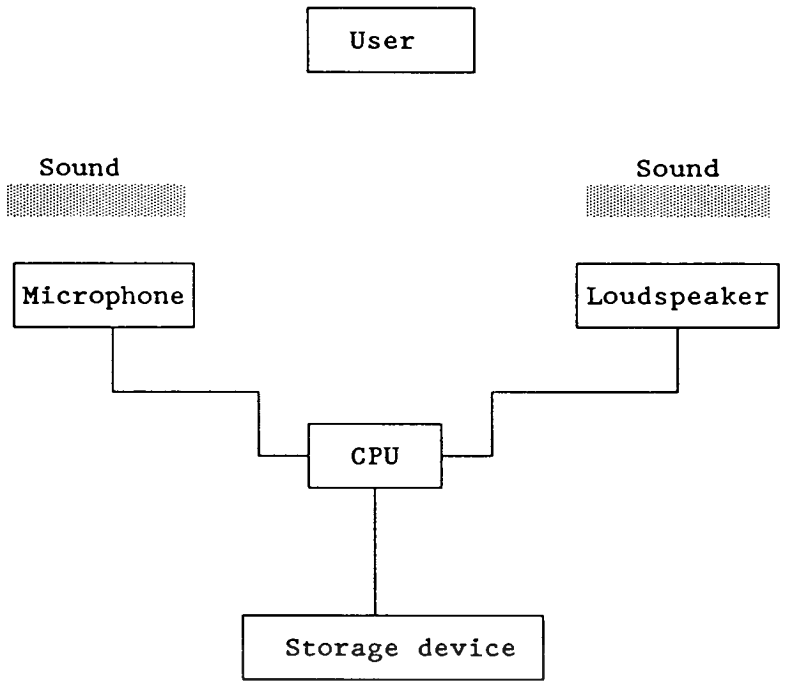

Diagram 2

The process of communicating with the system is as follows:

The user speaks into the microphone.

The CPU digitizes the analog signal.

The CPU compares the digital presentation with the templates stored in the storage device.

Depending on the outcome of the comparison, the system responds to the user via the loudspeaker.

As can be seen from the above, the crux of the system lies within the CPU, with secondary importance attached to the storage device.

4.2 Cards in the CPU

The major functionality of speech recognition, the digitizing of the analogue signal and the comparison of the digital representation to templates, is provided by a speech card that slots into a motherboard of a CPU. The majority of speech recognition systems are designed to work on a personal computer (PC).

A full function speech card will be in the form of a single printed circuit board that offers the following: speaker dependent recognition, speaker independent recognition, voice response, voice store and forward, vocoding and speaker verification.

According to Votan (1987:2) voice response is where a previous vocal input is "reconstituted" for audio playback on command. Voice store and forward is similar to an answering machine that stores a message and plays the message back on request. The difference is, however, that the message is stored in digital instead of analogue format. An example of such a speech card is the Votan V5000 card.

4.3 Software

The speech cards run under the Disk Operating System (DOS). Some software is needed to act as interface between the speech card and DOS. For the V5000 card the software is called VOS (Voice Operating System). The main function of VOS is to convert the functionality provided by the speech card to a DOS format enabling manipulation using the standard DOS commands.

CHAPTER 3

EXAMPLES OF SPEECH RECOGNITION APPLICATIONS

This chapter is divided into the following:

1 INTRODUCTION AND OBJECTIVE

2 PHYSICAL ACCESS CONTROL

3 LOGICAL ACCESS CONTROL

4 HOME BANKING USING SPEECH RECOGNITION

1 Introduction and objective

The most common uses of speech recognition are all concerned with the verification of a person's identity. If a person is recognized by the system, that person will be given access to the building, programme or information.

The type of system that uses speech recognition is important because it will have an impact on the audit implications.

The auditor first has to establish the following:

* Type of application, i.e. physical access control.
* Does the system run on PC, LAN, Mini or Mainframe?
* Is the system speaker dependent or speaker independent.

When the auditor has established the above, the audit risk can be assessed and the audit approach established.

The objective of this chapter is to give examples of systems that use speech recognition. For each example the suggested audit objective and approach is given which will provide the auditor with a starting point when faced with a similar system.

2 Physical access control

2.1 Example of an access system

The example is VASS, Voice Activated Security System, as described by CM Professional Advisors (Pty) Ltd (1989).

2.1.1 Requirements

The system resides on an AT-type personal computer fitted with a Voice Recognition and Response Interface Card.

The system makes use of standard security booths fitted with microswitch status sensors and solenoid locks. A microphone, loudspeaker and an operational numeric keypad are installed in each booth, all of which are linked to the personal computer.

2.1.2 Functions

- Authorized users are enrolled in the system by recording their voiceprints in the form of a single or multiple password. This is used for future comparison each time access to a controlled area is required.

The system offers two modes of operation. These are high throughput and high security.

The events that will occur for high throughput are the following:

* The user will type a unique four to six digit code on the keypad installed in the booth.
* Using the code, the system recovers the template for that user.
* A tone is emitted from the loudspeaker and the user speaks his/her password.
* The password is digitized and compared with the template. When a match is made the solenoid lock is activated and entry given.

The events for high security are:

* The user speaks his/her first password. The password is digitized and compared to templates of all first passwords.
* When a matched is made, the template of the second password is recovered from storage.
* The user is then requested to speak his/her second password.
* The second password is digitized and compared to the template. Access is given if the password matches.
* Depending on the level of security a third, fourth or fifth password can be asked.
* To prevent an intruder from using a recorder the passwords are asked in a random order.

Visitor access works as follows:

* The visitor enters a prerecorded code for visitors when in the booth.
* The person being visited will enter his code and password by telephone.
* If the password matches, the visitor will be given access.
* Attempts by the employee to leave before his visitor will be blocked and logged on a report.

2.2 Audit objective

The objective is to ensure that one of the main internal controls, that of restricting access to assets and records of a company, is in operation.

2.3 Audit approach

The following should be reviewed:

* Standards for the degree of access granted to which individuals.
* Procedures for recording a template.
* The control over the PC on which the system resides.
* The maintenance of the templates of personnel that leave the company or change in access level.
* Possibility of bypassing the recognition points by gaining access to a building/room/warehouse by alternate doors or windows.

Establish where speaker dependent and speaker independent recognition is used.

Test to ensure access can only be gained as set out in the standards. This can be done by obtaining a template for various audit personnel to attempt access to the areas controlled by the system.

3 Logical access control

3.1 Example of an access system

This system is still being developed. Its intended use is to control access to the terminals in the branches of a bank. Each branch has a LAN connected to the mainframe.

3.1.1 Requirements

The file server of the LAN must have a Voice Recognition and Response Interface Card installed. Each workstation must be fitted with a microphone.

3.1.2 Functions

Templates with the password(s) for every employee of a branch are made. The templates are linked to an employee by means of his/her personnel number.

The LAN supervisor then links each employee to a table of functions the employee needs in the course of his or her duties.

When a person signs on, the personnel number is entered and the system retrieves the template for that person. The person is asked to speak the password(s). When a match is made, the menu of functions the employee is allowed, is displayed on the screen.

3.1.3 Advantages and Risks

The main advantage is that an employee has access to the mainframe only via the LAN where his/her password is stored. This resolves the problem of ordinary passwords in many large systems, where passwords may not be revoked immediately on an employee's transfer, which would allow the employee access to functions no longer required.

The risk of this system lies in the linking of functions by the LAN supervisor. The supervisor function must be controlled.

3.2 Audit objective

The objective is to ensure that only authorized personnel are allowed access to functions offered by a computer application.

3.3 Audit approach

The following should be reviewed:

* Standards for who should have access to which functions.
* Procedures for recording a template.
* The control over the LAN on which the system resides.
* The maintenance of the templates of personnel who leave the company or change their access level.
* Possibility of gaining access by using system utilities, TSO, etc., thereby bypassing the recognition .

Test to ensure access can only be gained as set out in the standards. This can be done by obtaining a template for various audit personnel to attempt access to the various functions.

4 HOME BANKING USING SPEECH RECOGNITION

4.1 Introduction

There is currently no bank in the RSA that offers home banking using Speech Recognition technology. This facility is offered by various banks in the USA, most of whom provide balance information and transfers to other customer-owned accounts.

The system discussed here is being used as a pilot by First National Bank and is called Ted. This system is the equivalent of the Standard Bank's TONI and Volkskas's BANKTEL. The main difference is that the latter two are voice response systems while Ted is a voice recognition system. Ted will be described under the following headings:

* Functions offered.
* Technology used.
* How the system works.
* Risks of the system.

4.2 Functions offered

4.2.1 Balance enquiry

A client can obtain the balance of any of his or her accounts which are identified as a Ted account.

4.2.2 Mini Statement

The balance plus the last five entries of an account are supplied over the phone.

4.2.3 Transfers

Money is transferred between any of the Ted accounts in the client's name.

4.2.4 Account payments

A client can pay his or her accounts from a Ted current or transmission account. The restriction is that payment can only be made to a nominated creditor who has been set up by means of written instructions from the client. The creditor's account may be with any financial institution.

4.3 Technology used

The system consists of a LAN linked to the mainframe via one PC in the LAN. Each PC in the LAN will be diskless but fitted with a voice communications card. The LAN will be linked to an optical datadisk on which the templates are stored. Each PC will have several phone lines connected to it.

The system will make use of both speaker dependent and speaker independent recognition. For speaker independent recognition the Marconi Bankcall system is used.

4.4 How the system work

4.4.1 Recording the template

A client will go to his or her branch to become a Ted user, where the keywords for each client are recorded. The keywords are Balance, Statement, Payment, Transfer, Current, Savings and the name of the payee i.e. Edgars. A password is also recorded.

The template is then linked to a unique number and to the client's accounts.

4.4.2 Conversation with Ted

The client phones the Ted number and the phone is answered by a PC. The client is requested to give the unique number. The number can be given either by speaking it or using a touch tone telephone. Ted uses the Bankcall system to recognize the number independently. The number is then used to retrieve the template for that client.

The client is asked to speak the password. Using speaker dependent recognition, the password is compared to the template. When a match is made, the client is asked what functions he or she would like to perform. If a match is not made, the client is asked to give his or her unique number.

When a client selects a function, the keyword for that function is compared to the template. If a match is not made, the client is asked to try again. After three unsuccessful attempts, the client is refused access and asked to contact his or her branch.

When amounts are specified in transfers or payments the client can again speak or use the touch tone phone. For spoken amounts the Bankcall system is used to recognize the numbers. Before a transaction is performed, the amount will be voiced by the system and the client asked to confirm the amount.

4.5 Risks of the system

4.5.1 Denying access to valid clients

The Bankcall system uses an average template for speaker independent recognition. Templates are currently only available in English and are based on the average British English speaker. Therefore, the possibility of the system misinterpreting South African English does exist. This problem is overcome by allowing numbers to be entered using a touch tone telephone.

4.5.2 Unauthorized access will be gained

If a person obtains a client's unique number, he or she can attempt to access the system by using the number and mimicking the client's voice.

This risk is covered by the nature of speech recognition. Since the system makes use of speaker dependent recognition and a voice can not be duplicated by another person, access will be denied when the keywords are compared.

Unauthorized access may, however, be gained by recording the password and keywords of the client. Tests have proved that this does not work where the recording is played directly into the microphone attached to the recognition system.

It is, however, a real risk where Ted is used because filters are used to cut out background noise in telephone lines, the background noise of the recording is filtered out by the phone system. The risk is, however, reduced as payments can only be made to nominated accounts.

4.6 Audit objective

The objective is to ensure that only authorized clients are allowed access to functions offered by Ted.

4.7 Audit approach

The following should be reviewed:

* Standards for Ted operation.
* Procedures for recording a template.
* The control over the LAN/Mainframe on which the system resides.
* The risks identified.

The system should be tested to ensure that access can only be gained as set out in the standards. This can be done by using a template for various audit personnel to attempt access to various functions.

CHAPTER 4

MODEL FOR AUDIT OF SPEECH RECOGNITION APPLICATIONS

This chapter is divided into the following areas:

1 INTRODUCTION

2 MODEL

3 CONCLUSION

1 INTRODUCTION

The model provided in this chapter is for the auditing of access control applications using speech recognition. The model consists of three parts: a general information questionnaire, a risk matrix and an audit guideline.

The general information questionnaire is used to develop an understanding and an initial assessment of the importance of the speech recognition system.

The risk matrix, which is dependent on the answers given in the questionnaire, will provide a risk ranking of the system. The risk ranking indicates the guideline which suggests an audit approach for that system.

2 MODEL

2.1 General information questionnaire

Mark the correct answer with an X.

1 The system resides on:

Mainframe	Mini	PC	LAN

2 The system uses:

Speaker independent recognition	Speaker dependent recognition	Both

3 The system is used for control of:

Physical access	Logical access	Both

4 The system is used for control of:

High security area/data	Low security area/data	Both

2.2 Risk Matrix

	SPEAKER INDEPENDENT RECOGNITION	SPEAKER DEPENDENT RECOGNITION	BOTH
LOW SECURITY	1	2	3
HIGH SECURITY	4	5	6
BOTH	7	8	9

The risks for the auditor are ranked from 1 to 9, where 1 represents the lowest risk and 9 the highest risk.

2.3 Audit Guideline

In each of the discussions, 'profiles' refer to:

* areas to which access is permitted, for physical access control;
* the data and functions available, for logical access control.

2.3.1 Speaker independent recognition for low security area/data

The auditor should review the following:

* Standards for the system.
* Procedures for maintenance of the users of the system.
* Control over the PC/Mini/Mainframe on which the system resides.
* The types of profile that a person can be allocated.
* The market's acceptance and recognition of the card and independent template used in the system.

Tests can be performed to ensure that individuals are allocated the correct profile.

2.3.2 Speaker dependent recognition for a low security area/data

The auditor should review the following:

* Standards for the system.
* Procedures for maintenance of the users of the system.
* Control over the PC/Mini/Mainframe on which the system resides.
* The types of profile that a person can be allocated.
* The market's acceptance and recognition of the card used in the system.
* The procedures for training the users on the system to provide dependent recognition.

Tests can be performed to ensure that individuals are allocated the correct profile.

The auditor should consider whether dependent recognition and the associated costs (i.e. diskspace, training of users to form a dependent template) are justified for access control of low security area/data.

2.3.3 Speaker dependent and independent recognition for a low security area/data.

The auditor should review the following:

* Standards for the system.
* Procedures for maintenance of the users of the system.
* Control over the PC/Mini/Mainframe on which the system resides.
* The types of profile that a person can be allocated.
* The market's acceptance and recognition of the card and independent template used in the system.
* The procedures for training the users on the system to provide dependent recognition.
* The guidelines on where dependent and where independent recognition should be used.

Tests can be performed to ensure that individuals are allocated the correct profile.

The auditor should consider whether dependent recognition and the associated costs (i.e. diskspace, training of users to form a dependent template) are justified for access control of low security area/data.

2.3.4 Speaker independent recognition for high security area/data

The auditor should review the following:

* Standards for the system.
* Procedures for maintenance of the users of the system.
* Control over the PC/Mini/Mainframe on which the system resides.
* The types of profile that a person can be allocated.
* The market's acceptance and recognition of the card and independent template used in the system.

Tests can be performed to ensure that individuals are allocated the correct profile.

The auditor should consider whether independent recognition is adequate for restricting access to the high security area/data.

2.3.5 Speaker dependent recognition for a high security area/data

The auditor should review the following:

* Standards for the system.
* Procedures for maintenance of the users of the system.
* Control over the PC/Mini/Mainframe on which the system resides.
* The types of profile that a person can be allocated.
* The market's acceptance and recognition of the card used in the system.
* The procedures for training the users on the system to provide dependent recognition.

Tests can be performed to ensure that individuals are allocated the correct profile.

2.3.6 Speaker dependent and independent recognition for a high security area/data.

The auditor should review the following:

* Standards for the system.
* Procedures for maintenance of the users of the system.
* Control over the PC/Mini/Mainframe on which the system resides.
* The types of profile that a person can be allocated.
* The market's acceptance and recognition of the card and independent template used in the system.
* The procedures for training the users on the system to provide dependent recognition.

* The guidelines on where dependent and where independent recognition should be used.

Tests can be performed to ensure that individuals are allocated the correct profile.

2.3.7 Speaker independent recognition for high and low security area/data

The auditor should review the following:

* Standards for the system.
* Procedures for maintenance of the users of the system.
* Control over the PC/Mini/Mainframe on which the system resides.
* The types of profile that a person can be allocated.
* The market's acceptance and recognition of the card and independent template used in the system.

Tests can be performed to ensure that individuals are allocated the correct profile.

The auditor should consider whether independent recognition is adequate for restricting access to the high security area/data.

2.3.8 Speaker dependent recognition for a high and low security area/data

The auditor should review the following:

* Standards for the system.
* Procedures for maintenance of the users of the system.
* Control over the PC/Mini/Mainframe on which the system resides.
* The types of profile that a person can be allocated.
* The market's acceptance and recognition of the card used in the system.
* The procedures for training the users on the system to provide dependent recognition.

Tests can be performed to ensure that individuals are allocated the correct profile.

The auditor should consider whether dependent recognition and the associated costs (i.e. diskspace, training of users to form a dependent template) are justified for access control of low security area/data.

2.3.9 Speaker dependent and independent recognition for a high and low security area/data.

The auditor should review the following:

* Standards for the system.
* Procedures for maintenance of the users of the system.

* Control over the PC/Mini/Mainframe on which the system resides.
* The types of profile that a person can be allocated.
* The market's acceptance and recognition of the card and independent template used in the system.
* The procedures for training the users on the system to provide dependent recognition.
* The guidelines on where dependent and where independent recognition should be used.

Tests can be performed to ensure that individuals are allocated the correct profile.

The auditor should ensure that independent recognition is used for the low security area/data and dependent recognition for the high security area/data.

3 CONCLUSION

Speech recognition has become a permanent feature and its use will therefore increase as more and more applications are found for it.

At the moment the auditor need not be overly concerned about speech recognition, as it is mainly confined to access control. Both physical and logical access control can easily be audited using normal audit techniques, with a basic knowledge of speech recognition.

The future promises exciting applications of speech recognition, even the ability to communicate with the computer in the same way as one speaks to another human being. The auditor will consequently have to grow with technology and keep up to date with developments.

BIBLIOGRAPHY.

1 Yannakoudakis, E. J., Hutton, P. J. "Speech Synthesis and Recognition Systems" (John Wiley & Sons, Inc. 1987)

2 Cadzow, James. A.,"Foundations of Digital Signal Processing and Data Analysis" (Macmillan Publishing Company. 1987)

3 VOTAN "Press Background Information" (Publisher and date unknown)

4 Coler, Clayron. R., "Helicopter Speech-Command Systems" (Speech Technology, Volume 1, Number 3, September October 1982)

5 CM Professional Advisors (Pty) Ltd, "Introduction Voice Activated Security System" (1989)

6 Information Systems Division, First National Bank, "SDLC documentation on Ted" (1989)

7 Vallabhaneni, S.R. "Information system audit review manual" (Schaunburg : EDP Auditors Foundation. 1983)

8 Furui, S. "Digital Speech Processing, Synthesis, and Recognition" (Marcel Deker, Inc. 1989)