Enhanced voice recognition to reduce fraudulence in ATM machine

[1]Hridya Venugopal, Hema.U, Kalaiselvi.S, Mahalakshmi.M
Department of Information Technology
Alpha college of Engineering
Email:hridya.nbr@gmail.com,hemau5490@gmail.com,kalaika3@gmail.com, mahamuthu.91@gmail.com

Abstract

The aim of voice recognition in ATM machine is to achieve secured transaction. The focus here is mainly for disabled people to perform transaction at ATM centre. The security measures are introduced to reduce cases of fraud and theft due to its methods used in identification of individuals. In this paper, we present a security based implementation of Hidden markov model algorithm (HMM) to calculate speech rate, frequency and modulation pitch detection algorithm (PDA) for pitch calculation of voiceprints and Accent Classification (AC) for the accent analysis in voice. The combination of these algorithms allows us to provide a much more secured voice recognition system in ATM machine. This voice recognition system is proven to provide security based access control.

Index Terms – VRS, ATM, HMM, PDA, AC

1. INTRODUCTION

Voice recognition is the ability of a machine or program to receive and interpret dictation, or to understand and carry out spoken commands. It is generally regarded as one of the convenient and safe recognition technique [1]. Due to the advancement in technology this system becomes more secured. Voice recognition system (VRS) is used in several applications by many people. The main application of VRS is used in secured door system, calling cards, military, mobile banking and medical transcription. The VRS functions not by pressing buttons or interacting with a computer screen, users must speak to the computer, and this means there will be a level of uncertainty associated with their input, as automatic speech recognition only returns probabilities, not certainties. The analog audio must be converted into digital signals. This requires analog-to-digital conversion technique. The VRS is basically of two types: One is voice dependent which is less efficient and not accurate. It has high error rate if it is accented. Another one is voice independent system which is efficient and the accuracy level is about 90%. If the accent is recognized the error rate is minimized.

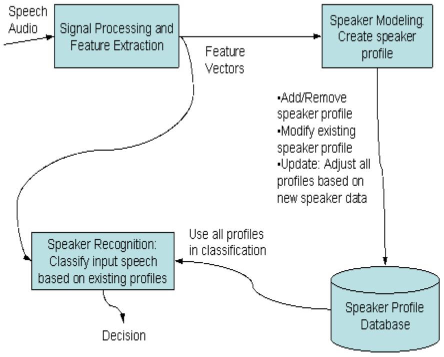

Speech Audio
Signal Processing and Feature Extraction
Feature Vectors
Speaker Modeling: Create speaker profile
•Add/Remove speaker profile
•Modify existing speaker profile
•Update: Adjust all profiles based on new speaker data
Speaker Recognition: Classify input speech based on existing profiles
Use all profiles in classification
Speaker Profile Database
Decision

Figure 1.Simple voice recognition system

The main objective of the paper is mainly based on secured transaction for disabled person. It involves the implementation of certain algorithms combined together to get much more reliable and robust voice recognition system. The Hidden Markov Model (HMM) algorithm is used for speech rate, frequency and modulation calculation; pitch detection algorithm (PDA) is used for pitch calculation and accent analysis is used for accent calculation. We briefly discuss about the combination of the above mentioned algorithm for secured transaction. The advantages of VRS are: (1) It is mainly designed for less fortunate like disabled person those who cannot use the existing ATM machines (2) It is much secured than other system (3) Effective communication and increased accessibility.

A. Related work

Voice recognition in secured door system is used for access control. One of the important security systems is for building security in door access control[2]. The ability to verify the identity of a person by analyzing his/her speech, or speaker verification provides security for admission into an important or secured place. Spectrogram is the tool used to identify the voice recognition for door system. The voice of the person is saved as .wave files in the database. The objective of door system is to achieve the highest possible classification accuracy. It is speaker dependent voice recognition system. Three different feature extractions they are Liner Prediction Cepstral Coefficients (LPCCs), Mel Frequency Cepstral Coefficients (MFCCs) and Perceptual Linear Prediction (PLP) coefficients. LPCCs, MFCCs and PLP coefficients are used as features. Moreover, SVM is adopted and evaluated to model the authorized person base on feature extracted from the authorized person's voice[2]. The existing system makes use of the following algorithms individually are shown below:

Hidden Markov Model(HMM) algorithm:

- Forward and backward algorithm
- Viterbi algorithm
- Baum-Welch algorithm
- Expectation algorithm

Pitch Detection algorithm:

- Pitch detection algorithm 1
- Pitch detection algorithm 2

Accent classification algorithm:

- Stochastic Trajectory Model (STM)
- Parametric Trajectory Model (PTM)
- Likelihood Score and Duration Distribution

Disadvantages:

- Voice recognition system does not have accuracy.
- VRS is based on the environmental factors like background noises, interpretation of voice, etc.
- Even after hours of training your voice this system tends to make mistake or error.
- VRS works best if the microphone is close to the user. More distant microphone will tend to increase the number of errors.
- VRS cannot understand all the words spoken by the user.

2. PROPOSED WORK

The description of voice recognition system comprises of eight modules: 1) microphone which is used to receive voice signals from the user, (2) channel is used to transmit information from sender to receiver, (3) A/D convertor is used to convert the speech signal from analog form to digital form for security measure, (4) filter bank is a device which is used to avoid distortion in voice (5) character distilling is performed to a voice signal to avoid distortion and background noise, (6) The voice signal should be passed through D/A convertor which converts the digital signal into analog form, (7) The voiceprint after conversion is verified with the voiceprints in the database and the voice is verified, (8) The verified voice is sent to the ATM machine through speaker.

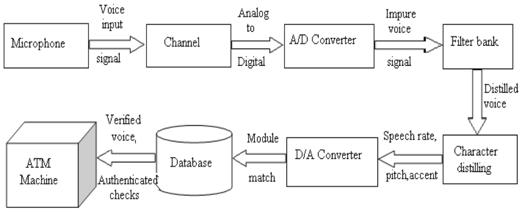
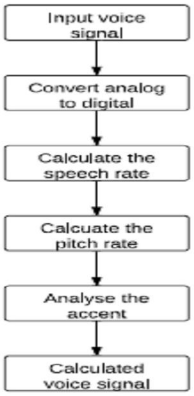


Figure 2. System architecture

Algorithm implementation:



I.Hidden Markov Model

A hidden markov model (HMM) is a statistical Markov model in which the system being modeled is assumed to be a Markov process with hidden states.
Transition probabilities **A** = {a_{ij} = P(q_j at t +1 | q_i at t)}

I (a)Improved forward algorithm

Let $a_t(i)$ be the probability of the partial observation sequence O_t={o(1),o(2),…….o(t)} to be produced by all possible state sequences that end at the i-th state.
$a_t(i)$=P(o(1),o(2),o(3),………o(t) | q(t)=q_i)
Initialization: $\alpha_1(i) = p_i\, b_i(o(1))$, i =1, ... , N
Recursion:

$$\alpha_{t+1}(i) = \left[\sum_{j-1}^{N} \alpha_t(j)\, a_{ji}\right] b_i\big(o(t+1)\big) \quad (1)$$

here i =1, ... , N , t =1, ... , T - 1

Termination:

$$P\big(o(1)o(2)...o(T)\big) = \sum_{j-1}^{N} \alpha_T(j) \quad (2)$$

I(b) Backward Algorithm

A symmetrical backward variable $\beta_t(i)$ as the conditional probability of the partial observation sequence from o(t+1) to the end to be produced by all state sequences that start at i-th state.
$\beta_t(i)$ = P(o(t+1), o(t+2), ... , o(T) | q(t) = q_i).
To find the optimal state sequence and estimating the HMM parameters.
Initialization:
$\beta_T(i)$ = 1 , i =1, ... , N
Recursion:

$$\beta_t(i) = \sum_{j-1}^{N} a_{ij} b_j\big(o(t+1)\big)\beta_{t+1}(j) \quad (3)$$

here i =1, ... , N , t = T - 1, T - 2 , . . . , 1
Termination:

$$P(o(1)o(2)\ldots o(T))=\sum_{j=1}^{N}p_j b_j(o(1))\beta_1(j) \qquad (4)$$

I(c)Posterior decoding

The states are chosen individually at the time when a symbol is emitted. This approach is called posterior decoding.

Let $\lambda_t(i)$ be the probability of the model to emit the symbol o(t) being in the i-th state for the given observation sequence O.

$\lambda_t(i) = P(\ q(t) = q_i \mid O\ )$.

To derive , $\lambda_t(i) = \alpha_t(i)\ \beta_t(i)\ /\ P(\ O\ )$, i =1, ... , N , t =1, ... , T

Then at each time we can select the state q(t) that maximizes $\lambda_t(i)$.

$q(t) = \arg\max\ \{\lambda_t(i)\}$

I(d)Viterbi algorithm

The Viterbi algorithm chooses the best state sequence that maximizes the likelihood of the state sequence for the given observation sequence.

Let $\delta_t(i)$ be the maximal probability of state sequences of the length t that end in state i and produce the t first observations for the given model.

$\delta_t(i) = \max\{P(q(1), q(2), ..., q(t\text{-}1)\ ;\ o(1), o(2), \ldots , o(t) \mid q(t) = q_i\ ).\}$

The Viterbi algorithm is a dynamic programming algorithm that uses the same schema as the Forward algorithm except for two differences:

- ➢ It uses maximization in place of summation at the recursion and termination steps.
- ➢ It keeps track of the arguments that maximize $\delta_t(i)$ for each t and i, storing them in the N by T matrix **ψ**. This matrix is used to retrieve the optimal state sequence at the backtracking step.

Initialization:

$\delta_1(i)= p_i b_i(o(1))$

$\psi_1(i)=0$, i =1,..,N

Recursion:

$\delta_t(\ j) = \max_i [\delta_{t-1}(i)\ a_{ij}]\ b_j\ (o(t))$

$\psi_t(\ j) = \arg\max_i [\delta_{t-1}(i)\ a_{ij}]$

Termination:

$p^* = \max_i [\delta_T(\ i\ )]$

$q^*_T = \arg\max_i [\delta_T(\ i\ )]$

Path (state sequence) backtracking:

$q^*_t = \psi_{t+1}(\ q^*_{t+1})$, t = T - 1, T - 2 , . . . , 1

I(e)Baum-Welch algorithm

Let us define $\xi_t(i, j)$, the joint probability of being in state q_i at time t and state q_j at time t +1 , given the model and the observed sequence:

$\xi_t(i, j) = P(q(t) = q_i, q(t+1) = q_j \mid O, \Lambda)$

we get

$$\xi_t(i,j)=\frac{\alpha_t(i)\,a_{ij}\,b_j(o(t+1))\,\beta_{t+1}(j)}{P(O\mid\Lambda)}$$

The probability of output sequence can be expressed as

$$P(O\mid\Lambda)=\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_t(i)\,a_{ij}\,b_j(o(t+1))\,\beta_{t+1}(j)=\sum_{i=1}^{N}\alpha_t(i)\,\beta_t(i)$$

The probability of being in state q_i at time t:

$$\gamma_t(i)=\sum_{j=1}^{N}\xi_t(i,j)=\frac{\alpha_t(i)\,\beta_t(i)}{P(O\mid\Lambda)}$$

Initial probabilities:

$$\overline{p}_i=\gamma_1(i)$$

Transition probabilities:

$$\overline{a}_{ij}=\frac{\sum_{t=1}^{T-1}\xi_t(i,j)}{\sum_{t=1}^{T-1}\gamma_t(i)}$$

Emission probabilities:

$$\overline{b}_{jk}=\frac{\sum_t^{*}\gamma_t(j)}{\sum_{t=1}^{T}\gamma_t(j)}$$

II. Pitch detection Algorithm

A pitch detection algorithm (PDA**)** is designed to estimate the pitch or fundamental frequency of periodic signal, usually a digital recording of speech or a musical note or tone. This can be done in the time domain or the frequency domain.

II.(a)PDA ALGORITHM 1:

A modified autocorrelation using center clipping and infinite peak clipping for time domain preprocessing is defined as PDA algorithm1.

To identify the center clipped signal,

$S_c(n)=\{s(n)+c_t,\ s(n)\leq -c_t$
$0,\ -c_t\leq s(n)\leq +c_t$
$S(n)-c_t,\ s(n)\geq +c_t$

(1)

Autocorrelation is given by

$R(m)=$ $\sum_{i=0}^{n} S$ m=0,1,……M

(2)

$\check{R}(m)=R(m)/R(0)$

(3)

By computing the energy for each section,

$E= {}^{N}\sum_{n=0} s^2\ (n)$

(4)

II (b) PDA ALGORITHM2:

A modified autocorrelation method using nonlinear transformation and center clipping for time domain preprocessing. In PDA algorithm 1, the setting of the clipping level threshold is very sensitive to pitch detection.

Each signal is then center clipped as in PDA algorithm1 to remove the ripples associated with the formants. It is further weighted by a Hamming window to produce a smooth tapering of the autocorrelation output. By comparing the correlation peak value to a decision threshold and also to distinguish background noise from speech section by comparing the energy of the speech sections to a predetermined noise (silence) level threshold.

III (c) Accent Classification Algorithm.

Accent classification or accent identification can be useful in speaker profiling for call classification, as well as for data mining and spoken document retrieval. English accent can be defined as the patterns of pronunciation features which characterize an individual's speech as belonging to a particular language group. The level of accent depends on the following factors they are: 1) the age at which a speaker learns the second language; 2) the nationality of the speaker's language instructor; and 3) the amount of interactive contact the speaker has with native talkers.

Trajectory models:

The sequence of points reflects movement in the speech production and feature spaces which can be called the trajectory of speech. a speech signal can be represented as a point which moves as the articulatory configuration Changes.

(a) Stochastic Trajectory Model (STM)

An STM represents the acoustic observations of a phoneme as clusters of trajectories in a parametric space. Let X be a sequence of N points:$X=(x_0,x_1,\ldots,x_{N-1})$,where each point is a D-dimensional vector in a speech production space. The probability density function (pdf) of a segment X, given a duration and the segment symbol is written as,

$$p(X|d,s) = \sum_{tk€Ts} p(X \mid t_k,d,s) \quad P_r(t_k|s) \qquad (1)$$

the assumption of frame independent trajectories, the pdf is modeled as

$$p(X \mid t_k,d,s) = {}^{N-1}\boldsymbol{\Pi}_{i=0}\,\text{Gaussian}\,(X;{}^{s}m_{k,i}\,,\ {}^{s}\textstyle\sum_{k,i}) \qquad (2)$$

(b) Parametric Trajectory Model (PTM)

An alternative to the STM is the PTM. The PTM treats each speech unit to be modeled by a collection of curves in the feature space, where the features typically are cepstral based. For the parametric trajectory, we model each speech segment

feature dimension as

$$c(n) = \mu\,(n) + e(n),\text{for } n= 1,\ldots,N$$

The speech segment can be modeled as

$$C=ZB+E$$

(c) Likelihood Score and Duration Distribution

At the classification stage, the likelihood of an unknown speech segment X given segment class s with T_s trajectories can be expressed as

$$p(X,s) = p(X|d,s)^{\alpha}\ .\ P_r(d|s)^{\beta}.$$

Advantages:

- The background noises and distortion in voice can be rectified by using an advanced microphone for better clarity and efficient filtering is done in advanced microphones
- It cannot be accessed by unauthorized users because the voice signal can have a minimum of 15% distortion.
- By combining HMM, PDA, AC the efficiency level of the VRS can be increased.

3. IMPLEMENTATION

This solution was implemented using Open Source Mozilla Firefox1.5 web browser from Mozilla foundation. The modified web browser was successfully built with the help of the build documentation provided on Mozilla web site on Microsoft's Windows Vista using JSP. The Mozilla Firefox web browser executes Scripting language-JavaScript included in web pages with the help of the preventer engine called Voice XML to make it more interactive to the user. It is used to execute Scripting language JavaScript programs included in web pages. The solution needed some major changes in the scripting language-JavaScript engine and some minor changes in the other components of the web browser. The backend used for VRS is Mysql. The Testing tools used for testing the voice recognition software is software test Automation testing.

4.EXPERIMENT RESULTS

The experiments were conducted for the evaluation of the traditional algorithm and proposed algorithm. The speech rate for the system is calculated by,

$\alpha_t= \qquad {}_t\, a_i\, b_j\ / \qquad .$

Pitch is calculated by,

$E={}^{N}\sum_{n=1} S(n)\, S_c(n)\times N.$

The recognition rate is overall estimation of all the metrics. The recognition rate for the proposed algorithm is found to be above 90%. When compared to traditional algorithm above 75%. Thus the accuracy, efficiency of the proposed system is made effective.

Table -1Comparison between traditional and proposed algorithm

Proposed algorithm		Traditional algorithm	
Speech rate	93.2%	Speech rate	78.1%
Pitch	99.2%	Pitch	85%
Frequency	92.4%	Frequency	71.3%
Accent	90.4%	Accent	53.6%
Recognition rate	92.4%	Recognition rate	78.9%

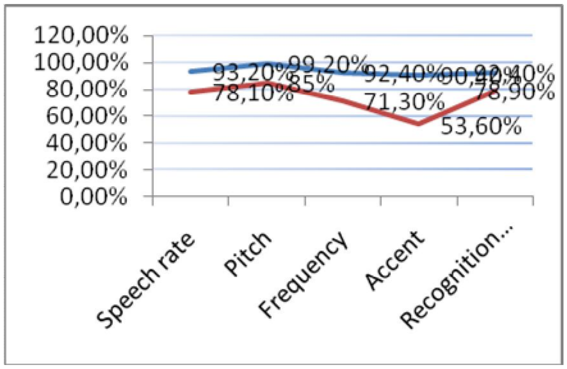


Figure.3-Comparison graph

5.CONCLUSION

We have determined HPA algorithm for improving security, accuracy and robustness in noisy environments. The HPA is based on the calculation of the metrics like frequency, speech rate, modulation, accent using respective algorithm. With all the innovation the proposed voice recognition system overcomes the drawbacks in other existing system and provides better performance, security, accuracy when compared with other voice recognition system. The further enhancement can be made after the research being conducted in this paper.

.Acknowledgement

We wish to express our sincere thanks to all the staff members of I.T Department, Alpha College of Engineering for their help and co-operation.

REFERENCES

[1] Bo Cui, Tongze Xu.” Design and Realization of an Intelligent Access Control System Based on Voice Recognition”. ISECS International colloquium on computing, communication, control and management, press 2009.

[2] Syazilawati Mohamed, Wahyudi Marton. “Design of Post-Mapping Fusion Classifiers for Voice-Based Access Control System”. 12th International Conference on Computer Modeling and Simulation, press 2010.

[3] Rozeha A. Rashid, Nur Hija Mahalin, Mohd Adib Sarijari, Ahmad Aizuddin Abdul Azi. Security System Using Biometric Technology: Design and Implementation of Voice Recognition System (VRS). Proceedings of the International Conference on Computer and Communication Engineering, 2008.

[4] Zeliang Zhang, Xiongfei Li. “A Study on Improved Hidden Markov Models andApplications to Speech Recognition”, Press 1999.

[5] R. Sankar. “PITCH EXTRACTION AUXRITHM FOR VOICE RECOGNITION APPLICATIONS”,0094-2989/88/0000/0384$01.00 © 1988.

[6] Kaibao Nie, Member, IEEE, Ginger Stickney, and Fan-Gang Zeng*, Member, IEEE,” Encoding Frequency Modulation to Improve Cochlear Implant Performance in Noise. IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, VOL. 52, NO. 1, JANUARY 2005

[7] Om Deshmukh, Carol Y. Espy-Wilson, Ariel Salomon, and Jawahar Singh.” Use of Temporal Information: Detection of Periodicity, Aperiodicity, and Pitch in Speech”, IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, VOL.13, NO.5, SEPTEMBER 2005.

[8] Pongtep Angkititrakul, Member, IEEE, and John H. L. Hansen, Senior Member, IEEE,” Advances in Phone-Based Modeling for Automatic Accent Classification”, IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL.14, NO. 2, MARCH 2006.

[9] Alexander Krueger, Student Member, IEEE, and Reinhold Haeb-Umbach, Senior Member, IEEE, “Model-Based Feature Enhancement for Reverberant Speech Recognition”, IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 18, NO. 7, SEPTEMBER 2010.