iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 Issue: III Month of publication: March 2017

DOI: http://doi.org/10.22214/ijraset.2017.3100

Volume 5 Issue III, March 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Speech Recognition in ATMs: Application of Linear Predictive Coding and Support Vector Machines

Dr. Sonia Sunny[1], Sreekala M[2]
[1]Dept. of Computer Science, Prajyoti Niketan College, Pudukad, Thrissur, Kerala
[2]Dept. of Computer Science, St. Thomas Autonomous College, Thrissur, Kerala

Abstract*: Today, Automated Teller Machines (ATMs) are extensively used by people for financial transactions. It provides a convenient, fast and easy way for customers to access cash. In this paper, a speech recognition system is developed for financial transactions in ATMs using Linear Predictive Coding (LPC) and Support Vector Machines (SVM). Voice signals are sampled directly from the microphone and then they are processed using LPC for extracting the features. Training, testing and pattern recognition are performed using Support Vector Machines. The proposed method is implemented for 200 speakers uttering 10 spoken digits in English. This hybrid architecture of LPC and SVM produced rather good recognition accuracy of 80.25%.
Keywords: Speech Recognition, Automated Teller Machines, Feature Extraction, Linear Predictive Coding, Support Vector Machines, Recognition Accuracy.

I. INTRODUCTION

Speech is a complex signal which is difficult to analyse and is produced as a result of several transformations occurring at different levels. Speech processing is a diverse field which has many applications in our day to day life in which speech recognition is an important one [1]. Automatic Speech Recognition (ASR) has tremendous growth over the last five decades due to the advances in signal processing, new algorithms, new architectures and hardware [2]. Since the human vocal tract and articulators are biological organs with nonlinear properties, these are affected by factors ranging from gender, emotional state etc. So there is wide difference in terms of their accent, pronunciation, articulation, roughness, nasality, pitch, volume, and speed. Moreover, speech patterns are also distorted by background noise and echoes, as well as electrical characteristics. Thus a speech recognition problem is a very complex task [3].

Automated Teller Machines (ATMs) enable customers to perform financial transactions like cash withdrawal, check balances or credit mobile phones with the help of the machine and without any human intervention. Now a days, people depend on ATMs for financial transactions. A plastic ATM card with a magnetic stripe or a plastic smart card with a chip and a unique card number and security information is used for the transaction. Authentication is done by using the Personal Identification Number (PIN). ATM machines also allow the conversion of one currency into another with the best possible exchange rates. Most cash machines are connected to interbank networks. This enables people to withdraw and deposit money from machines not belonging to the bank where they have their accounts or in the countries where their accounts are held enabling cash withdrawals in local currency [4].

Voice recognition service can be used to authenticate customers based on their speech patterns or can recognize the speech samples which allows them to execute banking transactions through the ATM. It provides a quick, secure and convenient way of doing transactions. With the voice recognition technology in place, customers are no longer required to enter their card numbers, PIN and answer security questions to authenticate themselves. Their voice will now act as the password for banking transactions. ATMs provide audible instructions so that persons who cannot read an ATM screen can independently use the machine.

In this work, we have divided the speech recognition process into three stages. The first stage is the creation of the database. Next one is the feature extraction stage wherein short time temporal or spectral parameters of speech signals are extracted. Minimizing the number of parameters in the speech model increases the speed of speech processing. The third module is the classification stage wherein the derived parameters are compared with stored reference parameters and decisions are made based on some kind of a minimum distortion rule. Among these three stages, feature extraction is a key, because better features are necessary for improving the recognition rate.

This paper is organized as follows. The spoken digits database used in this system is given in section 2. In section 3, the transaction

Volume 5 Issue III, March 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

process in ATMs is illustrated. The theory of feature extraction followed by the concepts of LPC employed during this stage is described in section 4. Section 5 describes the pattern classification stage using SVM. Section 6 presents the detailed analysis of the experiments done and the results obtained after classification. Conclusions are given in the last section.

II. SPEECH CORPUS

A spoken digits database is created for English language using 200 speakers uttering 10 digits. We have used 90 male speakers, 90 female speakers and 20 children for creating the database. The samples stored in the database are recorded by using a high quality studio-recording microphone at a sampling rate of 8 kHz (4 kHz band limited). Recognition has been made on the ten digits from 0 to 9 under the same configuration. The database consists of a total of 2000 utterances of the digits. The spoken digits are numbered and stored in the appropriate classes in the database. The spoken digits database is shown in table 1.

Table 1 spoken digits database

DIGITS	DIGITS IN ENGLISH
0	Zero
1	One
2	Two
3	Three
4	Four
5	Five
6	Six
7	Seven
8	Eight
9	Nine

III. METHODOLOGY USED IN ATM TRANSACTIONS

A hands free approach to ATM transactions wherein spoken voice messages replace key strokes is adopted. A customer undertaking an ATM transaction is required to insert his ATM card in the ATM. Then a voice message prompts the customer to enter the PIN. At the backend, the PIN entered undergoes validation and the following options are provided on successful validation:

A. Cash withdrawal
B. Balance enquiry
C. Mini statement
D. Funds transfer

If a customer selects option – 1 which is cash withdrawal through voice message, he/she is prompted to enter the amount to be withdrawn. After entering the required amount, the customer is requested to re-enter the amount in order to confirm it. If both the amounts entered matches, cash is made available. Otherwise customer has to redo the transaction once again. If customer selects option - 2, balance enquiry – balance in the account is displayed. If the customer selects option 3 which is Mini statement, a print of the mini statement is provided. If customer selects option 4, Funds transfer, he/she is required to enter the account number and amount to be transferred. Then this has to be reconfirmed by entering the amount and if found correct, funds are transferred. After completing the transaction, a Thank you screen is displayed. A new transaction can now be initiatedAll paragraphs must be indented.

IV. FEATURE EXTRACTION USING LINEAR PREDICTIVE CODING

Feature extraction is one of the important components of any type of recognition systems because the recognition accuracy depends on the features extracted. There are two fundamental operations in a speech recognition system namely signal modeling and pattern matching [5]. Feature extraction is an important part of signal modeling where speech signal is converted into a set of parameters called feature vectors. Pattern matching is the task of finding parameter set from memory which closely matches the parameter set obtained from the input speech signal also known as classification.

LPC is one of the most powerful methods used in audio and speech signal processing which extracts speech parameters like pitch formants and spectra. It is a powerful technique used for feature extraction purpose of speech signals since it characterizes the vocal tract well. It provides a good model of the speech signal and LPC front-end processing has been used in a large number of speech recognition problems [6]. LPC is a relatively efficient method for computation since it allows encoding of good quality speech at a low bit rate and it also provides extremely accurate estimates of speech parameters. The LPC analysis estimates the vocal tract resonance from a signal's waveform, removing their effects from the speech signal in order to get the source signal. The most important aspect of LPC is the linear predictive filter which allows the value of the next sample to be determined by a linear combination of previous samples [7]. At a particular time, k, the speech sample s(k) is represented as a linear sum of the n previous samples. This can be represented by the equation

$$S(k) = ak\text{-}1\ s(k-1) + ak\text{-}2\ s(k-2) + \ldots\ldots + ak\text{-}n\ s(k-n) \quad (1)$$

Where S (k) is the value of the signal at time (k). The coefficients aki are called the Linear Predictive Coding Coefficients. The coefficients can be analyzed to provide insight to the nature of the signal. Another important feature of LPC is that it minimizes the sum of the squared differences between the original speech and estimated speech signal over a finite duration. It produces a unique set of predictor coefficients which are normally estimated with every frame which is of usually 20 ms to 50 ms long. The predictor coefficients are represented by ak. Another important parameter is the gain (G). The transfer function of the time-varying digital filter is given by equation 2.

$$\mathbf{H}(z) = \frac{G}{1-\sum a_k z^{-k}} \qquad (2)$$

Summation is computed starting at k = 1 up to p. Here LPC 10 is used. This means that only the first 10 coefficient are transmitted to the LPC synthesizer. Auto correlation formulation is used to compute the coefficients.

V. PATTERN CLASSIFICATION USING SUPPORT VECTOR MACHINES

SVMs have proved to be a powerful technique for pattern recognition and classification [8]. It is a classifier which performs classification methods by constructing hyper planes in a multidimensional space that separates different class labels based on statistical learning theory [9][10]. Though SVM is inherently a binary nonlinear classifier, we can extend it to multiclass classification since ASR is a multiclass problem. There are two major strategies for multiclass classification namely One-against-All [9] and One-against-One or pair wise classification [11]. The conventional way is to decompose the M-class problem into a series of two-class problems and construct several binary classifiers. In this work, we have used One-against-One method in which there is one binary SVM for each pair of classes to separate members of one class from members of the other. This method allows us to train all the system, with a maximum number of different samples for each class, with a limited computer memory [12].

For speech recognition in ATMs, the main task is to identify what the customer speaks. For classification, the features obtained after extracting the feature vectors using LPC are divided into training set and testing set. 80% of the data set is used for training and the remaining 20% is used for testing.

VI. EXPERIMENTS AND RESULTS

Different criteria can be used to measure the performance of the speech recognition system developed. A confusion matrix is a simple and easy table which is used to measure the performance of a pattern classification problem. It gives the number of correctly classified and wrongly classified patterns. In this work also, we have used a confusion matrix for measuring the performance of the speech recognition system. The confusion matrix obtained after the classification of the 10 spoken digits is shown in table 2 given below.

TABLE 2 CONFUSION MATRIX

Class	0	1	2	3	4	5	6	7	8	9
0	163	3	5	6	3	6	7	4	2	1
1	6	152	5	4	7	8	9	4	3	2
2	8	7	135	7	7	6	8	9	7	6
3	3	5	4	177	2	4	1	0	2	2
4	5	9	6	4	141	7	9	6	5	8
5	1	3	4	2	2	170	4	5	3	6
6	6	3	4	2	5	6	154	7	8	5
7	2	1	4	5	3	1	2	172	4	6
8	5	7	3	8	4	8	2	1	159	3
9	2	4	1	0	3	5	1	2	0	182

Volume 5 Issue III, March 2017
ISSN: 2321-9653

The results obtained after pattern classification is given in table 3 given below.

TABLE 3 CLASSIFICATION USING SVM

No of speakers	Total samples	Correctly classified	Recognition accuracy
200	2000	1605	80.25%

VIII. CONCLUSIONS

The research in the domain of audio and speech recognition is still going on and this work is an application of speech recognition for financial transactions in ATMs. In this work, a novel speech recognition system is developed using Linear Predictive Coding and Support Vector Machines for recognizing speaker independent spoken digits for ATM transactions. Out of 2000 samples, 1605 samples are recognized correctly and an overall recognition accuracy of 80.25% is obtained from this work. In this experiment, we have used a limited number of samples. The vocabulary size can be increased to obtain more recognition accuracy. Support Vector Machines are found to be suitable for speech recognition and it provides good accuracies. As an extension of this work, alternate classifiers like Artificial Neural Networks, Genetic algorithms, Fuzzy set approaches etc. can also be used.

REFERENCES

[1] Joseph P Campbell, JR, Speaker Recognition: A Tutorial, Proceedings of the IEEE, Vol. 85, No. 9, 1997..
[2] Lawrence R., Applications of Speech Recognition in the Area of Telecommunications, Proceedings of IEEE Workshop on Automatic Speech Recognition and Understanding, pp. 501-510., 1997
[3] .recognition.http://www.learnartificialneuralnetworks.com/speechrecognition.html
[4] https://en.wikipedia.org/wiki/Cash_machine
[5] Picone J.W, Signal Modelling Technique in Speech Recognition, Proc. of the IEEE, Vol. 81, No.9, pp.1215-1247., 1993
[6] Rabiner L., Juang B. H., Fundamentals of Speech Recognition, Prentice-Hall, Englewood Cliffs, NJ, 1993.
[7] Jeremy Bradbury, Linear Predictive Coding, 2000,.
[8] W.M. Campbell *, J.P. Campbell, D.A. Reynolds, E. Singer, P.A. Torres-Carrasquillo, Support vector machines for speaker and language recognition, Computer Speech and Language 20, 210–229, 2006
[9] V.N. Vapnik., Statistical Learning Theory, J. Wiley, N.Y, 1998.
[10] N. Cristianini, J. Shawe-Taylor, An introduction to Support Vector Machines‖; Cambridge University Press, Cambridge, U.K, 2000.
[11] Ulrich H.-G. Kreßel., Pairwise Classification and Support Vector Machines, Advances inKernel Methods Support Vector Machine Learning‖; Cambridge, MA, MIT press, pp. 255-268, 1999.
[12] C.W. Hsu, C.J. Lin. 2002 ―A Comparison of Methods for Multiclass Support Vector Machines‖; IEEE Transactions on Neural Networks, Vol. No. 13, Issue No. 2, 415–425.