

## Financial Transactions in ATM Machines using Speech Signals

Dr. Sonia Sunny\*

\*(Department of Computer Science, Prajyoti Niketan College, Pudukad, Thrissur)

### ABSTRACT

Speech is the natural and simplest way of communication and Speech Recognition is a fascinating application of Digital Signal Processing which has many real-world applications. In this paper, a speech recognition system is developed for Automated Teller Machines (ATMs) using Wavelet Packet Decomposition (WPD) and Artificial Neural Networks (ANN). Speech signals are one-dimensional and are random in nature. ATM machines communicate with the customers using the stored speech samples and the user communicates with the machine using spoken digits. Daubechies wavelets are employed here. A multilayer neural network trained with back propagation training algorithm is used for classification purpose. The proposed method is implemented for 500 speakers uttering 10 spoken digits in English. The experimental results show good recognition accuracy of 87.38% and the efficiency of combining these two techniques.

**Keywords:** Automated Teller Machines, Feature Extraction, Pattern Recognition, Recognition Accuracy, Speech Recognition.

### I. INTRODUCTION

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. 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 [1].

A voice recognition system can be used for different purposes. Machines can identify the speaker through the voice thus enhancing the security and privacy of the transactions. The main objective of this work is to allow illiterate people to operate the ATMs without reading the instructions for the cash transactions. Speech recognition is still an intensive field of research in the area of digital signal processing due to its versatile applications. Since speech is the primary means of communication between people, research in automatic speech recognition and speech synthesis by machine has attracted a great deal of attention over the past five decades [2]. Speech recognition system mostly performs two fundamental operations: signal modelling and pattern matching [3]. During signal modelling, speech signal is

converted into a set of parameters using different feature extraction methods. The extracted features are compared with the parameter set from memory of the ATM machine and the pattern which closely matches the parameter set obtained from the input speech signal is classified during the classification stage. Better feature is good for improving recognition rate and the performance of a speech recognition system is measured in terms of recognition accuracy.

The main advantage of this system is that the user can interact with the ATM machine through speech that too using only spoken digits. The paper is organized as follows. In section 2, the methodology employed in this work is illustrated. The spoken digits database used in this system is given in section 3. The theory of feature extraction followed by the concepts of wavelet packet decomposition employed during this stage is described in section 4. Section 5 describes the pattern classification stage using ANN. Section 6 presents the detailed analysis of the experiments done and the results obtained. Conclusions are given in the last section.

### II. METHODOLOGY

The transactions in ATM machines can be performed using spoken voice messages. A customer enters the ATM cabin and inserts his ATM card in the ATM machine. The customer is then prompted to enter his/her PIN number through voice message. The PIN is validated and then the customer has to select from the options provided.

- 1) Cash withdrawal
- 2) Balance enquiry
- 3) Mini statement
- 4) 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 initiated.

### III. SPOKEN DIGITS DATABASE

For this experiment, a spoken digit database is created for English language using 500 speakers. Voice has a physiological trait. Every person has different pitch, frequency and so the voice recognition is mainly based on the study of the way a person speaks, commonly classified as behavioral. We have used 200 male speakers, 200 female speakers and 100 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 5000 utterances of the digits. The spoken digits are preprocessed, numbered and stored in the appropriate classes in the database. The spoken digits database is shown in Table 1.

**Table 1** Spoken Digits Database

Number digit	Digits in English
0	Zero
1	One
2	Two
3	Three
4	Four
5	Five
6	Six
7	Seven
8	Eight
9	Nine

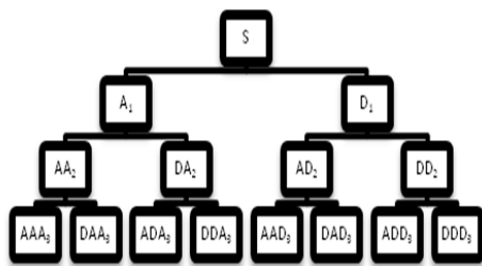
### IV. SPEECH FEATURE EXTRACTION

Feature extraction plays a vital role in the speech recognition process. During feature extraction step, the original speech signal is converted into a sequence of feature vectors. The feature vectors thus obtained are the inputs to the

classification stage of the speech recognition system. The technique selected for feature extraction plays a vital role in the speech recognition rate. Researchers have experimented with many different types of methods for use in speech recognition. Most of the speech-based studies are based on Fourier Transforms (FTs), Short Time Fourier Transforms (STFTs), Mel-Frequency Cepstral coefficients (MFCCs), Linear predictive Coding (LPCs), and prosodic parameters. Literature on various studies reveals that in the case of the above said parameters, the feature vector dimensions and computational complexity are higher to a greater extent. Moreover, many of these methods accept signals stationary within a given time frame. So, it is difficult to analyse the localized events correctly. By using wavelets, the size of the feature vector can be reduced when compared to other methods and thus the computational complexity also can be successfully reduced. Daubechies wavelets are used here because of its orthogonality property and efficient filter implementation [4].

#### 4.1 The Wavelet Packet Decomposition

In WPD, the speech signals are decomposed into low frequency components and high frequency components at each level called the approximation coefficients and detail coefficients. These are again decomposed to get new low resolution approximation and detail coefficients [5]. The wavelet packet decomposition applies the transform step to both the low pass and the high pass result. It allows simultaneous use of long-time interval for low-frequency information and short-time interval for high-frequency information [6]. In wavelet packet decomposition, each detail coefficient vector is also decomposed into two parts using the same approach as in approximation vector splitting. When a signal is decomposed into sub-bands using wavelet transform, approximation components contain the characteristics of a signal and high frequency components are related with noise and disturbance in a signal [7]. It allows simultaneous use of long-time interval for low-frequency information and short-time interval for high-frequency information [8] Though removing the high frequency contents retain the features of the signal, sometimes it may contain useful features of the signal. So both the high and low frequency components are decomposed in WPD. The decomposition tree for WPD is given in fig.1.



**Fig.1** Wavelet Packet Decomposition tree

**V. NEURAL NETWORKS CLASSIFIER**

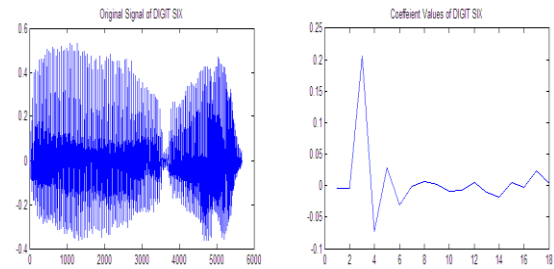
Artificial Neural Networks have been investigated for many years in the hope that speech recognition can be done similar to human beings. A Neural Network is a massively parallel-distributed processor made up of simple processing units. It can store experimental knowledge and make it available for use. Inspired by the structure of the brain, a neural network consists of a set of highly interconnected entities, called nodes designed to mimic its biological counterpart, the neurons. Each neuron accepts a weighted set of inputs and produces an output [9]. Neural Networks have become a very important method for pattern recognition because of their ability to deal with uncertain, fuzzy, or insufficient data. Algorithms based on neural networks are well suitable for addressing speech recognition tasks.

The architecture of the Multi Layer Perceptron (MLP) network, which consists of an input layer, one or more hidden layers, and an output layer, is used here. The algorithm used is the back propagation training algorithm which is a systematic method for training multi-layer neural networks. This is a multi-layer feed forward, supervised learning network based on gradient descent learning rule. In this type of network, the input is presented to the network and moves through the weights and nonlinear activation functions towards the output layer, and the error is corrected in a backward direction using the well-known error back propagation correction algorithm. After extensive training, the network will eventually establish the input-output relationships through the adjusted weights on the network [10]. In most networks, the principle of learning a network is based on minimizing the gradient of error [11][12]. After training the network, it is tested with the dataset used for testing.

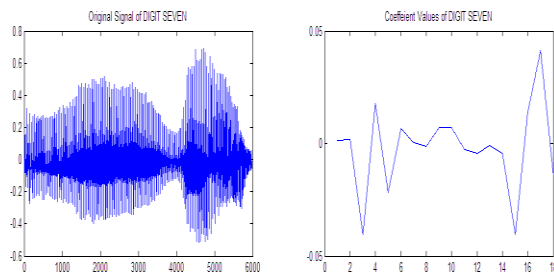
**VI. EXPERIMENTS AND RESULTS**

The recognition accuracy depends on the wavelet family chosen and the mother wavelet used. The most popular wavelets called the Daubechies wavelets with the db4 type of mother wavelet is used for feature extraction. Daubechies wavelets are found to perform better than the other wavelet

families based on recognition accuracy [13]. The speech samples in the database are successively decomposed into approximation and detailed coefficients. The speech signal is decomposed up to 12 levels. The original signal and the coefficient values after decomposition of digits 6 and 7 are shown in figure 2 and figure 3.



**Fig. 2** Decomposition of digit six



**Fig. 3** Decomposition of digit seven

Since speech recognition is a multi class classification problem, the developed feature vectors are given to an ANN, which can handle multi class parameter classification. We have divided the database into three. 70% of the data is used for training, 15% for validation and 15% for testing. MLP architecture is used for the classification scenario. It uses one input layer, one hidden layer and one output layer. Using this network, the feature vector set obtained is trained first and then they are tested. From the results obtained, it is found that the MLP structure could successfully recognize the spoken digits. After testing, the corresponding accuracy of each spoken digit is obtained.

The feature vectors obtained after wavelet packet decomposition are given as input to the ANN for classification and an overall recognition accuracy of 87.38% is obtained for the spoken digits by using the combination of Wavelet Packet Decomposition and Artificial Neural Networks. The confusion matrix obtained using this classification is given in Table 2 given below.

**Table 2** Confusion Matrix

Class	0	1	2	3	4	5	6	7	8	9
0	421	8	10	7	2	11	6	9	13	13
1	4	457	8	11	3	1	9	2	5	0
2	9	12	398	5	16	9	12	10	18	11
3	4	6	5	464	1	9	0	3	5	3
4	8	7	10	9	410	8	11	14	14	9
5	11	4	7	8	12	417	13	14	7	7
6	1	9	6	7	10	8	449	3	4	3
7	4	5	3	1	0	6	3	472	4	2
8	5	7	8	10	4	3	8	5	445	5
9	4	10	9	9	6	10	7	5	4	436

The results obtained after classification is given in Table 3.

**Table 3** Classification results using ANN

No of speakers	Total samples	Correctly classified	Recognition accuracy
500	5000	4369	87.38%

## VII. CONCLUSION

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 consisting of wavelet packet decomposition and Artificial Neural Networks for recognizing speaker independent spoken digits for ATM transactions. The computational complexity and feature vector size is successfully reduced to a great extent by using wavelet packet decomposition. Out of 5000 samples, 4369 samples are recognized correctly and an overall recognition accuracy of 87.38% 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. Neural network classifiers are well suited for speech recognition and it provides good accuracies. As an extension of this work, alternate classifiers like Support Vector Machines, Genetic algorithms, Fuzzy set approaches etc. can also be used and a comparative study of these can be performed.

## REFERENCES

- [1]. [https://en.wikipedia.org/wiki/Cash\\_machine](https://en.wikipedia.org/wiki/Cash_machine)
- [2]. Rabiner L., Juang B. H., *Fundamentals of Speech Recognition* ( Prentice-Hall, Englewood Cliffs, NJ., 1993).
- [3]. Picone J.W., Signal Modelling Technique in Speech Recognition, (Proc. of the IEEE, Vol. 81, No.9, pp.1215-1247, 1993.)
- [4]. Hu Dingyin, Li Wei, Chen Xi , “Feature Extraction of Motor Imagery EEG Signals based on Wavelet Packet Decomposition”, Proceedings of the 2011 IEEE International Conference on Complex Medical Engineering , 694-697 , 2011.
- [5]. S. Chan Woo, C.Peng Lin, R. Osman. Development of a Speaker Recognition System using Wavelets and Artificial Neural Networks; *Proc. of Int. Symposium on Intelligent Multimedia, Video and Speech processing*, 413-416.
- [6]. S. Kadambe, P. Srinivasan. 1994 “Application of Adaptive Wavelets for Speech, Optical Engineering”; Vol. No. 33, Issue No. 7, 2204-2211.
- [7]. B. C. Li, J. S. Luo. *Wavelet Analysis and Its Applications* ( Electronics Engineering Press, Beijing, China, 2003).
- [8]. W. Ting, Y. G. Zheng, Y. Bang-hua, and S. Hong, “EEG feature extraction based on wavelet packet decomposition for brain computer interface,” *Measurement*, vol. 41, no. 6, pp. 618-625, July 2008
- [9]. Freeman J. A, Skapura D. M., *Neural Networks, Algorithm, Application and Programming Techniques*(Pearson Education, 2006.)
- [10]. Economou K., Lymberopoulos D., “A New Perspective in Learning Pattern Generation for Teaching Neural Networks”, Volume 12, Issue 4-5, pp. 767-775, 1999.
- [11]. Anil K. Jain, Robert P.W. Duin, Jianchang Mao, “Statistical Pattern Recognition: A Review, *IEEE Transactions on Pattern Analysis and Machine Intelligence*”, Vol. 22., pp.4-37, 2000
- [12]. Eiji Mizutani and James W. Demmel, On Structure-exploiting Trust Region Regularized Nonlinear Least Squares Algorithms for Neural-Network Learning, *Neural Networks*, Volume 16, Issue 5-6, 2003, 745-753.
- [13]. Sonia Sunny, David Peter S and K Poulouse Jacob, Optimal Daubechies Wavelets for Recognizing Isolated Spoken Words with Artificial Neural Networks Classifier; *International Journal of Wisdom Based Computing*, Vol. No. 2, Issue No. 1, 2012, 35-41.