

Voice Enabled ATMs and Deployment of Wavelets for Recognition and Authentication of Voice Based Inputs

Dr. Sonia Sunny¹

¹Department of Computer Science, Prajyoti Niketan College, Pudukad, Thrissur, Kerala, India.
sonia.deepak@yahoo.co.in

Abstract

Automated Teller Machines (ATMs) enable customers to perform financial transactions like cash withdrawal, balance checking or credit mobile phones with the help of the machine and without any human intervention. In this paper, a speech recognition framework is modelled for performing financial transactions at ATMs. Since speech is the most natural, convenient and easiest mode of communication, communication between the customer and the ATM can be performed easily through speech. This allows ATM machines to communicate with the customers using the stored speech samples and the user communicates with the machine through spoken digits. Here, speech recognition systems are developed by employing two feature extraction techniques namely Discrete Wavelet Transforms (DWT) and Wavelet Packet Decomposition (WPD). The classifier used along with both the methods is Artificial Neural Networks (ANN). Both methods were found efficient and produced good recognition accuracies of 87.575% and 85.5% respectively. But DWT and ANN combination produced better results than WPD and ANN combination.

Key Words and Phrases: Automated Teller Machines, Feature Extraction, Wavelet Packet Decomposition, Pattern Recognition.

1 Introduction

Now a days, people depend on ATMs for financial transactions. Voice recognition services can be used to authenticate customers depending on their speech patterns. It also allows the customers to recognize the speech samples so that they can execute banking transactions through the ATMs easily. It provides a system with faster, secured and convenient way of doing transactions. One of the main advantages of using voice recognition technology is that, customers need not key in their card numbers, PIN and answer security questions to authenticate themselves. Their voice will now function as the password for authenticating banking transactions. This system also allows persons who cannot read an ATM screen to use the machine independently by means of the audible instructions provided by the ATMs.

Through a voice recognition system, machines can identify the speaker through the voice for enhancing the security and privacy of the transactions and can also recognize the words spoken. A speech recognition system identifies, separates and recognizes the sequence of speech sounds and the messages that best matches the input speech signal [1, 2]. Due to the wide range of applications of Automatic Speech Recognition (ASR) in many fields of life, this has become an intensive area of research [3]. This work allows illiterate people to operate the ATMs without reading the instructions for the cash transactions.

The human vocal band and articulators are biological organs with nonlinear properties. Therefore, constituting elements ranging from gender, emotional state etc. influence the speech recognition process. Another difficulty is that the speech waves are distorted by noise in the background, echoes and electrical characteristics [4]. Due to these variations and distortion, designing a speech recognition system is a challenging task [5]. Speech recognition system typically performs two central operations: signal modeling and pattern matching [6]. Here, speech signals are transformed into a set of feature vectors using the feature extraction methods DWT and WPD. The extracted features are compared with the parameter set stored in the memory of the ATM machine. During the classification stage, the pattern which closely resembles the parameter set acquired from the input speech signal is classified using ANN. The main advantage of this system is that the user can interact with the ATM machine through speech, that too using only spoken digits. Here, recognition accuracy and confusion matrix generated are used as the performance criteria.

The development of a speech recognition system requires the building of a Speech Corpus, which is given in section 2. In the following section, an outline of the transaction process in ATMs is illustrated. The theory of feature extraction and the concepts of DWT as well as WPD are described in section 4. Section 5 describes the pattern classification stage using ANN. Section 6 analyses the experiments done in detail. The results thus attained are compared in section 7, followed by conclusions presented in the last section.

2 Spoken Digits Database

We have created the speech corpus for English language using 400 speakers pronouncing 10 digits. Since there is much difference in the frequency rate of male, female and child speakers, we have selected 175 male, 175 female and 50 children for creating the sound signal database. The database is created using speech samples that are recorded at a sampling rate of 8 kHz (4 kHz band limited) by a microphone. Ten digits in English from 0 to 9 that are recorded under the same configuration are used as samples in the database. The database comprises a total of 4000 utterances of the spoken digits. The spoken digits database is shown in table 1.

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

Table 1: Spoken Digits Database.

3 ATM Transaction Process

In this work, only the basic ATM transaction procedures are considered and are 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 through speech.

- 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 say the amount to be withdrawn. After saying the required amount, the customer is requested to re-enter the amount once again through speech in order to confirm it. If both the amounts spoken matches, cash is made available. Otherwise customer has to redo the transaction once again. If a 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 speak the account number and amount to be transferred. Then this has to be reconfirmed by speaking the amount and account number once again and if found correct, funds are transferred. After completing the transaction, a Thank you screen is displayed. A new transaction can now be initiated.

4 Feature Extraction

The feature extraction process removes unwanted and redundant information and retains only the useful information in a speaker independent ASR system. But there are also chances of mislaying or losing some information during feature extraction [7]. The technique chosen for feature extraction has decisive role in the recognition accuracy of the speech samples. Since speech recognition is an important area in research, researchers have experimented with different feature extraction techniques. Researches show that there is difference in the feature vector dimensions and computational complexity among the various feature extraction techniques. Since the wavelets are well localized in time and frequency domain, the feature vector size and the computational complexity can be considerably reduced. Among the different types of wavelets available, Daubechies wavelets which have the highest number of vanishing points, orthogonality

property and efficient filter execution are employed in this work [8]. The feature vectors thus generated are given to the classification stage as inputs in the speech recognition system.

A. Discrete Wavelet Transforms

The DWT is an implementation of the wavelet transform which is a relatively of-late and effective technique for extracting features from non-stationary signals like speech. DWT decomposes a speech signal into mutually orthogonal set of wavelets using a discrete set of the wavelet scales and translations according to some definite rules. Since the traditional Fourier Transforms are localized only in frequency and the Wavelet Transforms are localized in both time and frequency, DWT can address the hitches connected to non-stationary signals in a more efficient way [9]. Both DWT and WPD employ digital filtering methods in order to produce a time-scale demonstration of the signals. DWT is defined by

$$W(j, K) = \sum_j \sum_k X(k) 2^{-j/2} \psi(2^{-j} n - k) \quad (1)$$

In equation (1), $\Psi(t)$ is the transforming function called the mother wavelet. The original signal is decomposed into different frequency bands by passing it through half band low-pass filters and high-pass filters. This reduces the time resolution by half, doubles the frequency resolution and decomposes the speech signal into two sets of coefficients called approximation coefficients and detail coefficients [10]. Then the successive high pass and low pass filtering of the signal is produced by

$$Y_{high}[k] = \sum_n x[n]g[2k - n] \quad (2)$$

$$Y_{low}[k] = \sum_n x[n]h[2k - n] \quad (3)$$

In equations (2) and (3), Y_{high} represents the detail coefficients and Y_{low} represents the approximation coefficients that are attained by sub sampling by 2. During decomposition, a signal is characterized more by the low frequency components $h[n]$ than its high frequency components $g[n]$ [11]. This filtering process through high pass and low pass filters is continued until the preferred level is obtained according to Mallat algorithm [12].

B. Wavelet Packet Decomposition

Like DWT, WPD also generates good time and frequency resolutions. It is a broad view of DWT which has a rich analysis technique than DWT. WPD also decomposes the speech signal into low frequency components and high frequency components at each level. But here, the transformation step is applied to both the approximation and detail information to form a complete binary tree whereas DWT applies the filtering only to low frequency components [13].

5 Classification Using ANN

The feature vectors generated from DWT and WPD are given for classification during classification process. During classification stage, the input data is divided into training and test data so that they are trained using the information relating to known patterns and then tested using the test data set. In this work also, we have employed ANN as the classifier. Neural Networks are extensively used in pattern recognition problems mainly due to the accessibility of effective learning algorithms. In addition, they are domain independent in relation to model-based and rule-based approaches [14]. In this work, a Multi-Layer Perceptron (MLP) network, having an input layer, one hidden layer, and an output layer, is employed along with back propagation training. Here, the inputs are presented to the nodes which pass through the weights and nonlinear activation functions towards the output layer. The error occurred is fed back to the system in order to fine-tune the system using the back propagation correction algorithm. The network finally establishes the input-output relationships through the adjusted weights on the network after sufficient number of training [15]. Learning of network is based on minimizing the gradient of error [16]. Once extensive training is done, the system is tested with the test dataset.

6 Experiments

A large number of wavelets are available and they differ in terms of many features such as number of vanishing points, symmetry, regularity, scaling function etc. Choosing the right wavelet family and employing an apt mother wavelet has an important role in the recognition accuracy.

The most popular wavelet that is suitable for discrete analysis called the Daubechies wavelets with the db4 type of mother wavelet is used for feature extraction. Daubechies wavelets outperform other wavelet families in recognizing speech signals in terms of accuracy rate [17].

6.1 Experiments using DWT

The spoken digits stored in the database are decomposed in sequence into approximation and exhaustive coefficients upto level 8. The feature vectors for each spoken digit are taken by decomposing the less frequency components in each level. The original signal and the 8th level approximation and comprehensive coefficients of spoken digit four obtained after decomposition using DWT is depicted in figure 1.

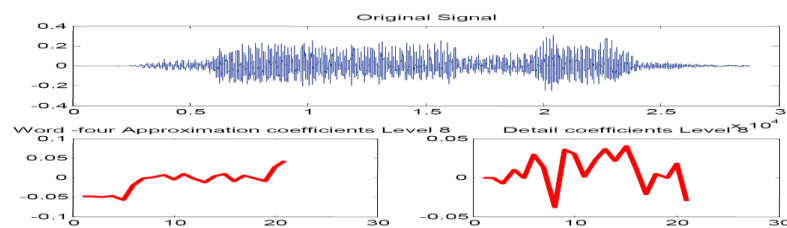


Figure 1: Decomposition of digit four with DWT

The feature vectors obtained from level 8 are then classified using ANN. In both DWT and WPD, we have divided the spoken digits database into three sets. The training set, validation set and the testing set. 70% of the data is used for training the Neural Networks. 15% of the data set is used for validation which is used to optimize the model and the remaining 15% for testing. After classification, the confusion matrix is generated where the diagonal elements in the matrix contain the instances that are correctly classified for each spoken digit. The confusion matrix obtained is illustrated in Table 2 given below.

Class	0	1	2	3	4	5	6	7	8	9
0	331	7	9	12	6	7	5	10	8	5
1	8	362	2	6	9	1	4	5	2	1
2	4	5	346	8	7	5	9	6	3	7
3	3	7	10	322	8	11	9	5	12	13
4	9	8	7	6	330	5	12	8	9	6
5	4	3	5	2	6	368	2	3	6	1

6	5	5	2	6	7	5	344	9	7	10
7	4	1	5	3	2	6	4	363	6	6
8	1	0	2	3	2	1	4	1	386	0
9	9	6	4	8	5	7	3	2	5	351

Table 2: Confusion Matrix obtained using DWT & ANN

6.2 Experiments using WPD

Like DWT, the db4 type of mother wavelet is used for extracting features in WPD also. In this experiment also the spoken digits are consecutively decomposed into approximation and exhaustive coefficients up to level 8. The feature vectors obtained from decomposing both the high frequency and low frequency components are given to Neural Networks for classification.

The original signal and the 8th level decomposition coefficients of spoken digit seven obtained using WPD are shown in figure 2.

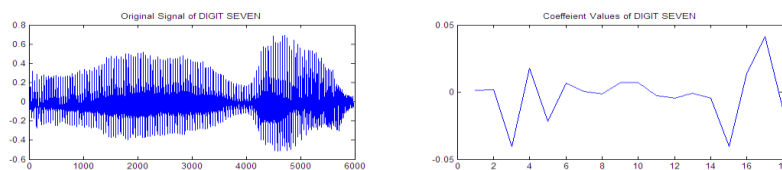


Figure 2: Decomposition of digit seven using WPD

The confusion matrix obtained using ANN and WPD feature vectors is given in Table 3 given below.

Class	0	1	2	3	4	5	6	7	8	9
0	321	8	10	7	2	11	6	9	13	13
1	4	350	8	11	3	1	9	2	5	7
2	9	2	341	5	8	9	6	1	8	11
3	13	6	13	312	14	9	7	8	5	13
4	8	7	10	9	324	8	11	13	1	9
5	1	4	3	8	5	361	3	4	4	7
6	8	9	6	6	4	8	339	9	4	7
7	4	5	3	1	8	6	3	359	4	7
8	1	1	3	1	4	3	0	5	377	5
9	5	10	9	9	7	8	7	5	4	336

Table 3: Confusion Matrix obtained using WPD & ANN

7 Comparison of Results

The classification results indicate that, both feature extraction methods produced reliable and good recognition accuracies when used along with ANN. The overall recognition accuracies attained using DWT and WPD are presented in table 4.

Feature Extraction method	No of speakers	Total samples	Correctly classified	Recognition accuracy %
DWT	400	4000	3503	87.575
WPD	400	4000	3420	85.50

Table 4: Comparison of Results

It is clear from the table that the tools employed yield reasonable recognition accuracies when used along with ANN. Among the two methods, DWT outperforms WPD in recognition accuracy.

8 Conclusions and Future Work

ATM transactions have now become an indispensable part of our lives. So, a speech recognition system is designed for ATM transactions using two methods DWT and WPD and the efficiency of these are tested and evaluated using Neural Networks as classifier. Good recognition rates are obtained for both the methods. The accuracy rate achieved by means of DWT and ANN combination is proved to be more than WPD and ANN method combined. The experiment results show that wavelet transforms are suitable for the analysis of non-stationary signals like speech and along with Neural Networks they can effectively extract the features from the speech signal and classify. As a future work, the number of speakers can be increased to have better generalization. Considering the future scope of the work, other pattern classifiers like Genetic algorithms, Support Vector Machines, Fuzzy set approaches etc. can also be used. Since ATM transactions are prone to security issues, speech recognition technologies along with biometric recognition methods give better security during transactions.

References

- [1] Lawrence R. Rabiner, Ronald W. Schafer, *Introduction to Digital Speech Processing*, Foundations and Trends in Signal Processing **1(2)** (2007), 1-194.
- [2] Samudravijaya K, *Speech and Speaker recognition: a tutorial.*, Proc. International Workshop on Technology Development in Indian Languages,

Kolkata (2003).

- [3] Kuldeep Kumar, R. K. Aggarwal, *Hindi Speech Recognition System Using Htk*, International Journal of Computing and Business Research **2(2)** (2011), 2229-6166.
- [4] <http://learnartificialneuralnetworks.com/speechrecognition.html>
- [5] Benzeghiba, M., De Mori, R., Deroo, O., Dupont, S., Erbes, T., Jouvett, D., Rose, R., *Automatic speech recognition and speech variability: A review*, Speech communication **49(10)** (2007), 763-786.
- [6] Picone J.W., *Signal Modelling Technique in Speech Recognition*, Proc. of the IEEE **81(9)** (1993), 1215-1247.
- [7] Stolcke A., Shriberg E., Ferrer L., Kajarekar S., Sonmez K., TurG., *Speech Recognition As Feature Extraction For Speaker Recognition*, SAFE (2007), 11-13.
- [8] Hu Dingyin, Li Wei, Chen Xi, *Feature Extraction of Motor Imagery EEG Signals based on Wavelet Packet Decomposition*, Proc. of the 2011 IEEE International Conference on Complex Medical Engineering **15(2)** (2011), 694-697.
- [9] S. Mallat, *A wavelet Tour of Signal Processing*, Academic Press (1999).
- [10] S. Chan Woo, C.Peng Lin, R. Osman, *Development of a Speaker Recognition System using Wavelets and Artificial Neural networks*, Proc. of 2001 Int. Symposium on Intelligent Multimedia, Video and Speech processing, Hong Kong (2001), 413-416.
- [11] S. Kadambe, P. Srinivasan, *Application of Adaptive Wavelets for Speech*, Optical Engineering **33(7)** (1994), 2204-2211.
- [12] S.G Mallat, *A Theory for Multi-resolution Signal Decomposition: The Wavelet Representation*, IEEE Transactions on Pattern Analysis and Machine Intelligence **11(1)** (1989), 674-693.
- [13] http://en.wikipedia.org/wiki/Wavelet_packet_decomposition
- [14] Y. Hao, X. Zhu, *A New Feature in Speech Recognition based on Wavelet Transform*, IEEE 5th Inter. Conf. on Signal Processing **3(1)** (2000), 1526-1529.
- [15] Economou GK, Lymberopoulos D, Spiropoulos K, Goumas PD., *A New Perspective in Learning Pattern Generation for Teaching Neural Networks*, Neural Networks **12(4-5)** (1999), 767-775.
- [16] Anil K. Jain, Robert P.W. Duin, Jianchang Mao, *Statistical Pattern Recognition: A Review*, IEEE Transactions on Pattern Analysis and Machine Intelligence **22(1)** (2000), 4-37.
- [17] Sonia Sunny, David Peter S., K. Poullose Jacob, *Optimal Daubechies Wavelets for Recognizing Isolated Spoken Words with Artificial Neural Networks Classifier*, International Journal of Wisdom Based Computing, **2(1)** (2012), 35-41.

