ISSN:0975-9646

A Combined Rough Sets–K-means Vector Quantization Model for Arabic Speech Recognizer

Elsadig Ahmed Mohamed[1/2], *ACM professional member*;

Hanan Adlan[*/1/3], *ACM professional member,*

Abd. Rahman Ramli[4], *Senior member IEEE*

***Abstract* - Vector quantization (VQ), is considered an efficient data reduction technique, and is used as a preprocessing stage in speech recognition systems. Methods traditionally used for vector quantization are purely numerical methods rather than rule-based methods. Furthermore, most of the experiments performed in previous research work used English data sets, and fewer experiments on Arabic data. In this paper, a vector quantization model that incorporate rough sets attribute reduction and rules generation with a modified version of the K-means clustering algorithm was developed implemented and tested as a part of a speech recognition framework, the learning vector quantization.(LVQ) neural network model was used in the pattern matching stage. Arabic speech data was used in the original experiments, for both speaker dependant and speaker independent tests. The performance was found to be very high in terms of recognition time and recognition rate, which was found to be 98.5% for untrained data, and 99.9% for trained data.**

Keywords: Speech recognition, Rough Sets, Neural Networks, LVQ, Vector quantization.

1.0 Introduction

The main goal for any speech recognition research work is to minimize the amount of data that the speech signal carries in each stage of the speech recognition system (SRS) without loosing any part of the information that the speech signal represents. Information here refers to a higher level than such a pure data; that is information on phonemes and utterances. Researchers usually focus their work on improving the performance in one stage of the recognition process.

Vector Quantization (VQ) is almost considered as a data compression technique by digital signal processing (DSP) researchers. In fact, a lot of work has been done in this field to improve the quality of the speech and image processors. Vector quantization has been applied more and more to reduce the complexity of problems like pattern recognition. In speech recognition, vector quantization used to preprocess speech signal prior to the pattern matching stage. Discrete systems are welcome in real-time implementations since they are less CPU consuming than continuous systems. Better results were obtained in speech recognition due to improvements in vector quantization techniques. Although, some times VQ ignored by people working in speech recognition. It has been shown that VQ methods and their parameters have an important influence on recognition rates. Therefore, some experiments have been done on vector quantization in the framework of a complete speech recognition system. The same recognition algorithms and the same feature vectors have been used to make the comparisons. Many researchers showed the influence of vector quantization on speech recognition system's performance, [16], [17], [2],[3].

Methods used for VQ are mainly numerical methods; they work on continuous real valued attributes, this implies that any small difference in attribute values (such as energy or frequency) for two vectors may result in classifying the two vectors in two classes or clusters, while discrete values methods discard small variations in attribute values. Small variations in attribute values may be a result of external noise in the recording environment or may also be a result of the natural variation in pronouncing an utterance more than once even by the same speaker. Rough sets (RS) approach, which is a relatively new approach to data analysis under uncertainty works mainly on discrete valued attributes. Under rough sets; there are two main training steps in data analysis: first reduction of attributes, and second rule generation. Generated rule can then be used to classify new objects; previous research work on applying rough sets to speech recognition analysis considered complete utterances as information system's objects ignoring the RS data reduction power. That is to apply RS to the pattern matching stage rather than vector quantization stage.

In the other hand, most of the research work in the area of speech recognition was oriented towards English language, a few work was conducted for other languages such as French, Spanish and Mandarin, and very little amount of work was done for Arabic language which has its own special features and importance [18], [19], [20], [21], [22], [23], [24]. Speech signal passes through some preprocessing steps before reaching the pattern matching stage, these steps include signal processing, feature extraction and vector quantization[1],[2],[3].

The main objective of the preprocessing is to reduce the amount of data in the signal and end with a specific pattern representation. Although the traditional k-means algorithm was efficiently used for vector quantization, there are some attempts to use other techniques such as the Kohonen self-organizing neural network, [4],[5],[6].

Rough sets theory represents a new mathematical approach to vagueness and uncertainty, data analysis, data reduction,

The authors: * Corresponding author: mehanan14@gmail.com

1.Dept. of Computer Science, Faculty of Mathematical Sciences, University of Khartoum, P.O.Box321, Sudan

2.Dept. of Computer Science, College of Computer Engineering and Science, Prince Satam Univerity, P.C.11942, P.O.Box151, Elkharj, K.S.A.

3. Dept. of Computer Science, Faculty of Computer and Information Science, Princess Nourah Bint Abdrahman University, P.C, Riyadh. K.S.A.

4. Department of Computer and Communication Systems Engineering, Faculty of Engineering, University Putra Malaysia, Selangor Darul Ehsan,Malaysia

approximate classification, machine learning, and discovery of pattern in data are functions performed by a rough sets analysis. It was one of the first non-statistical methodologies of data analysis. It extends classical set theory by incorporating into the set model the notion of classification as indiscernibility relation, [18], [6], [7].

Using rough sets approach in the field of speech recognition was limited in previous work to the pattern matching stage. That is, to use training speech patterns to generate classification rules that can be used later to classify input words patterns [4],[5].

In this work rough sets approach was used in the preprocessing stages, namely in the vector quantization operation in which feature vectors are quantized or classified to a finite set of codebook classes. Classification rules were generated from training feature vectors set, and a modified form of the standard voter classification algorithm, that use the rough sets generated rules, was applied.

2.0 SPEECH RECOGNITION SYSTEM STRUCTURE

The main components of the speech recognition system can be described as follows:

- Signal processing and feature extraction module that accepts as input words signals and produces feature vectors.
- A vector quantization module that replaces each feature vector by the index of the most suitable code vector in the codebook, a hybrid k-means rough sets method was introduced and implemented here.
- A pattern matching module that match input words patterns with the nearest one, the LVQ NN [10], [11], [12] was used for this work.

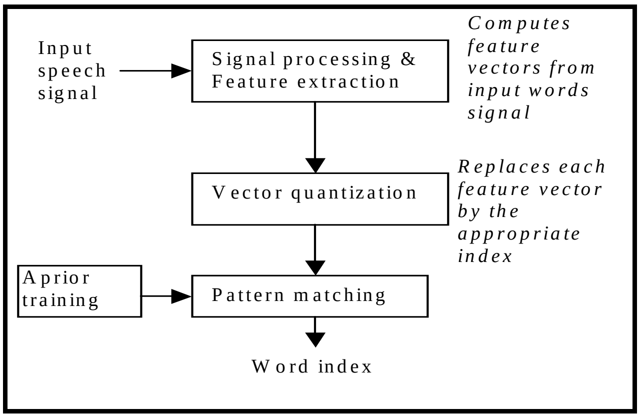
Input speech signal
Signal processing & Feature extraction
Computes feature vectors from input words signal
Vector quantization
Replaces each feature vector by the appropriate index
A prior training
Pattern matching
Word index

Fig. 1: Main components of a speech recognition system

Fig. 1 shows the main components of a speech recognition system, in the figure a speech signal is input to the signal processing and feature extraction module. Quantization is then took place in the vector quantization module. The pattern matching stage performs the appropriate actions by making use of a prior training to produce a word index [1],[2],[3],

2.1 Vector quantization

Vector quantization is an efficient technique of data reduction that still maintains the information needed to characterize different sounds. Vector quantization allows an input vector x = (x_1, x_2, . . ., x_k), to be replaced by a vector y_n drawn from a finite set of N reference vectors {y_i: i = 1, 2, . . . , N} “the code-book”, that minimizes a given distortion measure (Fig. 2).

Each input vector is compared to a codebook vectors. The nearest one according to the distance measure is chosen to represent the input vector. In this way the speech signal can be represented by means of a sequence of “code-words”, namely labels related to the code-book vectors.

The k-means algorithm is being used traditionally to cluster a set of training vectors to build the codebook, and further to classify any input vector of the same size to one of the codebook classes [1].

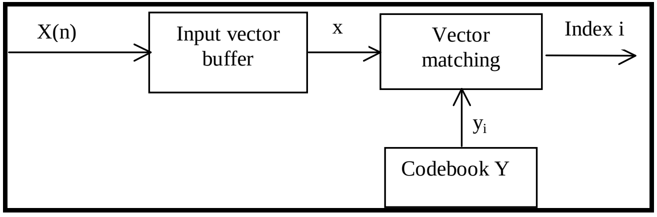

Fig. 2: Vector Quantization

3.0 THE ROUGH SETS THEORY

Rough sets approach is a mathematical approach that can be used for attributes reduction and objects classification based on their attributes. The classification can then be used to access objects or sets of objects where objects are roughly equal or roughly overlap. Since it has introduced by Pawlak in the early 1980's, the rough set theory has been under continuous development, and a fast growing group of researchers and practitioners are interested in this methodology [6].

Rough sets data analysis is a first (and sometimes the sufficient) step in analyzing incomplete or uncertain information. Rough set analysis uses only internal knowledge, and does not rely on prior model assumptions as other models such as fuzzy sets. In other words instead of using external numbers or other additional parameters, rough set analysis utilizes solely the structure of the given data. The numerical value of imprecision is not pre-assumed, as it is in probability theory of fuzzy sets – but is calculated on the basis of approximations which are the fundamental concepts used to express imprecision of knowledge. Consequently, we do not require that an agent assigns precise numerical values to express imprecision of his knowledge, but instead imprecision is expressed by quantitative concepts (approximations) [6], [7].

Rough set theory is based on the assumption that we have some additional information -knowledge, data- about elements of a universe of discourse -or simply universe. Concepts - elementary sets - are elements that exhibit the same information are similar - indiscernible -, and form blocks that can be understood as elementary granules of knowledge about the universe. The concept of rough set has been introduced as a set characterized by its lower and upper approximations.

Knowledge consists of a family of various classification patterns of a domain of interest, which provide explicit facts about reality - together with the reasoning capacity able to deliver implicit fact derivable from explicit knowledge [6], [7].

The lower approximation ($\underline{B}X$) of the set X is the union of elementary sets that are included in X, Whereas the upper approximation ($\overline{B}X$) is the union of all elementary sets that have a none empty intersection with X.. These approximations correspond, respectively to a maximum set including objects that surely belong to X, and a minimal set of objects that possibly belong to X.

$$\underline{B}X = \cup \{E \in \text{U/IND(B)} : E \subseteq X\} \qquad (1)$$

$$\overline{B}X = \cup \{ E \in \text{U/IND(B)}: E \cap X \neq \varnothing\} \qquad (2)$$

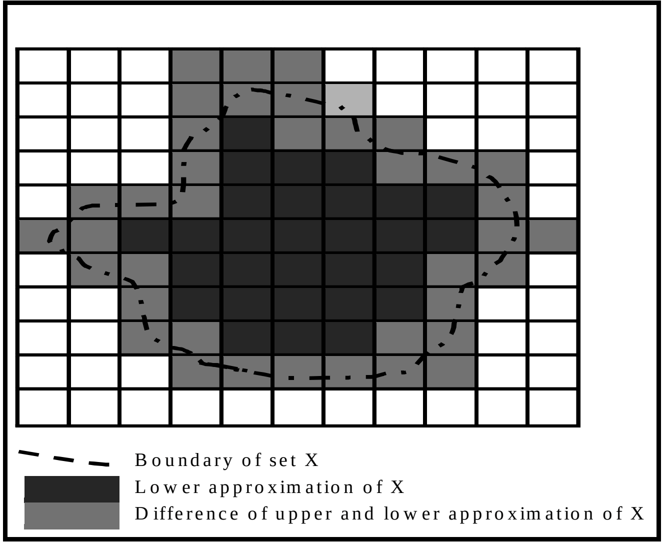

Fig. 3: Rough approximation of sets

The difference between the lower and the upper approximations is a boundary (BNX) set consisting of all objects that cannot be classified with certainty to X or its complement, (Fig. 3).

$$BNX = \overline{B}X - \underline{B}X \qquad (3)$$

4.0 THE SYSTEM

In order to develop the Arabic recognizer, the Combined Rough Sets–K-means vector quantization model was developed according to the following:

4.1 *Signal Processing And Feature Extraction*

The main objective in this stage is to obtain an appropriate representation for the speech signal. First to produce a discrete time representation of the continuous signal s(t), the signal was sampled at a rate of 22050Hz 8bits mono waveform, then the signal is blocked into overlapping frames of 512 samples, each two adjacent frames have 384 common samples, that frame advance is 128 samples [3], [13], [14], [15].

Each frame is then windowed to minimize discontinuities at the beginning and end of each frame. That is to use the windows to taper the signal to zero at the beginning and end of a frame. The window used here is the hamming window (Fig. 4), which has the form of Equation 4

$$w(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right) \qquad (4)$$

For each windowed frame feature vectors were extracted, that is to identify the key components of the speech signal, and to eliminate redundant information in the signal. That is to compute a set of parameters, which capture the transitions in the signal, and are robust enough to represent any phoneme. These parameters are often called the features and are computed at fixed time intervals. Transitions in the signal are important cues that may indicate the encoding of the phonetic information in the speech signal. A feature extractor produces usually a vector every 5-10 ms, an acoustic vector, which represents the salient speech feature of a window of about 20-30 ms. In this work the following parameters were computed for each windowed frame:

- Energy: Average magnitude for the speech amplitude in each analysis frame,
- Zero crossing rate: Rate of where the signal changes from positive to negative or visa versa,
- Correlation coefficient,
- Gain,
- and three LPC parameters.

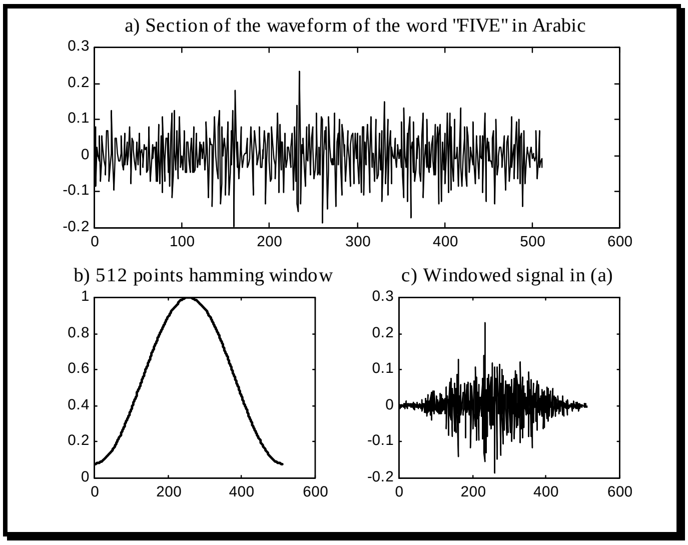

Fig. 4: Section of speech signal, a Hamming window, and a windowed speech signal section

Fig. 4 shows section of speech signal, Hamming window, and a windowed speech signal. In a) section of the waveform of the word “Five” in Arabic is displayed, b) gives the 512 hamming window, while c) shows the windowed signal of a).

4.2 *Rough Sets Based Vector Quantizer*

A hybrid k-means/rough sets approach vector quantizer was proposed, built and tested. The performance of the resulting speech recognizer was found to be very high in terms of recognition rate and time. The following is a description for the developed vector quantization approach.

In the training phase, the k-means clustering algorithm was used to cluster a set of training feature vectors and select the best representing vector for each cluster to build a codebook. One disadvantage of the k-means algorithm is that, the final codebook is highly dependent on the initial codebook. An initialization algorithm was designed and implemented as an attempt to solve this problem.

The codebook was then used to classify the training feature vectors set. The classified feature vectors were then used as an information system to train a rough set engine.

One problem associated with the rough set approach is that it can only deal with discrete data, therefore before performing any processing the information system data should be discretized, that is each condition attribute should have only discrete values. A discretization algorithm was designed for this purpose.

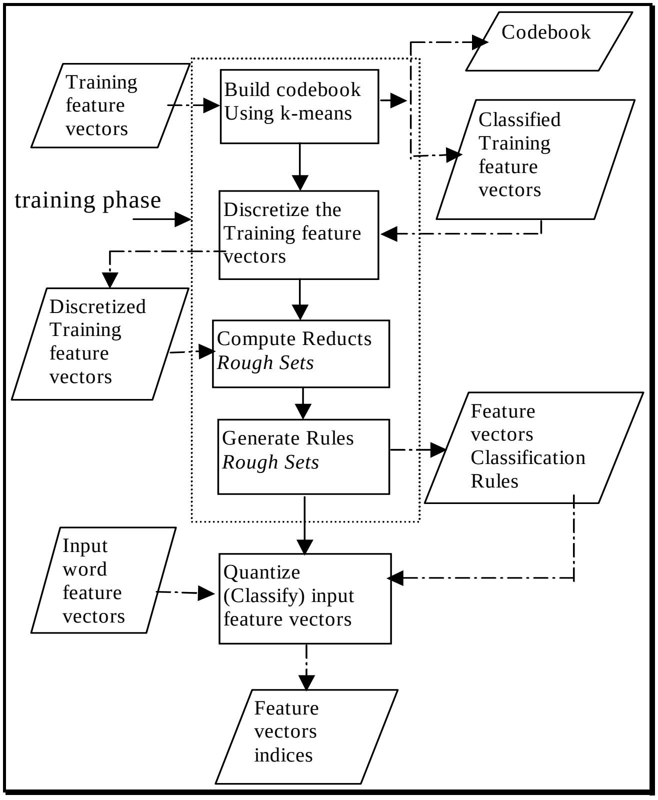
Codebook
Training feature vectors
Build codebook Using k-means
Classified Training feature vectors
training phase
Discretize the Training feature vectors
Discretized Training feature vectors
Compute Reducts *Rough Sets*
Feature vectors Classification Rules
Generate Rules *Rough Sets*
Input word feature vectors
Quantize (Classify) input feature vectors
Feature vectors indices

Fig. 5: Rough sets based quantizer structure.

After discretization the discretized classified training feature vectors are then used as an information system for which reducts were computed using a genetic algorithm, and classification rules were generated to be used to classify input words feature vectors in the testing phase, (fig. 5).

In the testing phase a modified form of the standard voter classification algorithm was used for feature vector classification, The algorithm is described as follows:

- Load and extract rules file;
- Load feature vectors file;
- For each feature vector:
    - Find rules with left hand side vectors match current feature vector;
    - From fired rules right hand side form a set of possible classes;
    - Find the class with a highest number of votes (frequency), this is the required codebook index class.

4.3 ***Pattern Matching:***

In the pattern matching stage, the self-organizing learning vector quantization with different numbers of neurons in input and hidden layers is developed and implemented. Prior to feeding a word pattern to the LVQ neural network we applied a re-sampling procedure that adjusts the pattern data size to match the number of the input layer neurons of the LVQ neural network. This procedure utilizes a signal processing technique that applies an anti-aliasing (low pass) FIR filter.

5.0 EXPERIMENTS AND RESULTS:

The proposed model was tested for small vocabulary, isolated Arabic words, speech recognizer, for both single speaker and multi-speakers.

Attributes computed were the energy, zero-crossing rate, correlation coefficient, gain, Linear Predictive Coding (LPC1, LPC2, LPC3), referred to as condition attributes. Experiments were repeated with different discretization scales from 4 to 20, and for codebook sizes of 32, 50, 64 and 80.

In all experiments, the information system reduction produced only one reduct, which was equal to the condition attributes set in the decision table. Table 1. Displays portion of the decision table.

Table1: Portion of the Decision System

Energy	ZeroCross	Corr	Gain	LPC1	LPC2	LPC3	Class
2	1	3	2	1	3	4	64
2	1	3	2	1	2	4	64
3	1	3	2	2	3	3	64
3	1	3	2	2	3	3	63
3	1	3	2	2	3	3	62
3	1	3	2	2	3	3	58
1	1	2	2	2	2	3	36

The number of rules generated was found to be directly proportional to the discretization scale, (fig. 6). This affect vector quantization time, that is when the number of rules increases quantization time increases and vice versa.

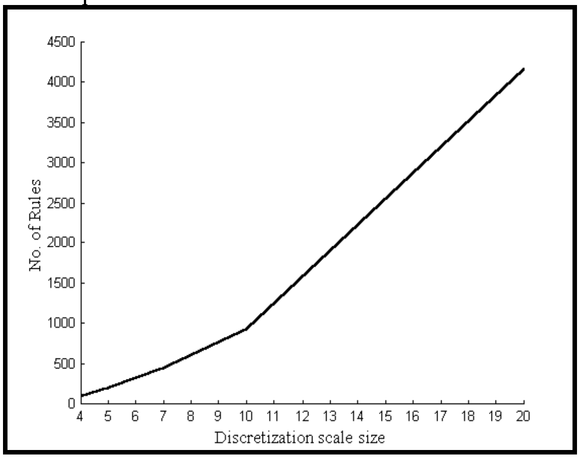

Fig. 6: Number of generated classification rules against discretization scale.

Different codebook sizes were used, quantization time was measured for rough sets and k-means based quantizers, fig. 7 shows that it decreases when discretization scale size decreases, for the rough sets based quantizer, while fig.8

shows that quantization time increases when the codebook size increases, for the k-means quantizer.

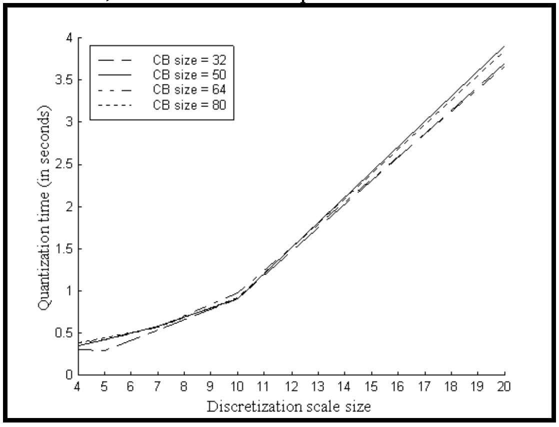

Fig. 7 Vector Quantization time against discretization scale for different codebook sizes, - rough sets based vector quantizer-.

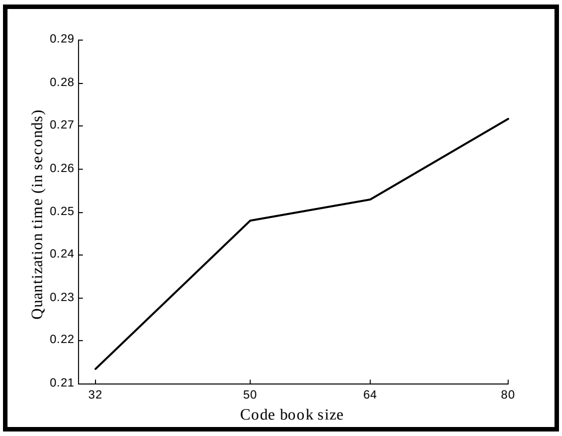

Fig. 8 Vector Quantization time against codebook sizes, - k-means algorithm vector quantizer-.

Using the rough sets based quantizer an average recognition rate of 98.5% was achieved, with codebook size of 64 and discretization scale of 5, Figure 9 shows recognition rates for different codebook sizes. Codebook of 32, 50, 64, and 80 are shown in the figure. Using a testing data drawn from the training data set the recognition rate was found to be 100% in most of the experiments and the average was found to be 99.9%.

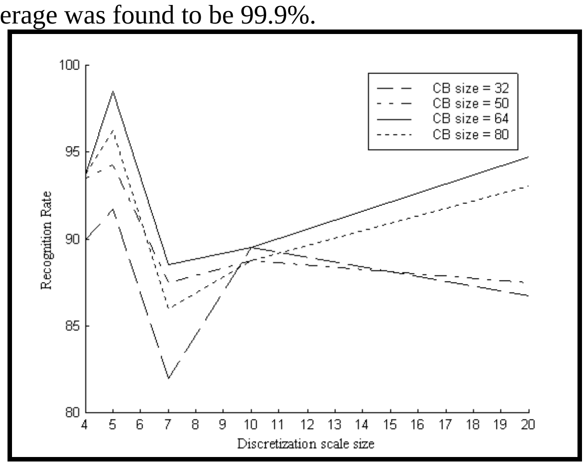

Fig. 9: Recognition rate for different codebook sizes and discretization scales.

6.0 CONCLUSIONS

In this work, a novel quantization model that employs both the traditional k-means algorithm and the rough sets approach was proposed, implemented and test. A codebook initialization algorithm was proposed to enhance the k-means performance. A modified form of the standard voter classification algorithm was used with rough sets generated classification rules.

Experimental results showed recognition rates of 99.9% with trained data, 98.5% with untrained data, and 88% for untrained different speaker data sets. Recognition time, for best recognition rates was found to be of order of 0.5 seconds, which is quite fair, however, recognition time is highly depends on hardware and software environments and implementation.

High recognition rates shows that the rough sets theory can be applied on huge amount of speech data, and moreover promises with good results when applied to similar problems such as speech filtering and image processing data.

REFERENCES

[1] Lawrence Rabiner and Biing-Hawang Juang, "Fundamentals of Speech Recognition", Prentice Hall, 1993.

[2] R. Linggard, D. J. Myers and C. Nightingale Editors, “Neural Networks for Vision, Speech and Natural Language”, BT Telecommunications Series, Chapman & Hall 1992.

[3] S. Nakagawa, K. Shikano, Y. Tohkura, Translated by C. Aschmann “Speech, Hearing and Neural Network Models” IOS Press, 1995.

[4] Andrzej Czyzewski and Andrzej Kaczmarek, “Speeker-independent Recognition of Isolated Words using Rough Sets”, ICS Research Report 34/95,1995.

[5] Bozena Kostek, “Computer Based Recognition of Musical Phrases Using the Rough Set Approach”, ICS Research Report 34/95,1995.

[6] Roman Slowinski, “Intelligent decision support :handbook of applications and advances of the rough sets theory”, Kluwer Academic Publishers,1992.

[7] Zdzislaw Pawlak, “Rough Sets Theoretical Aspects of Reasoning about Data”, Kluwer Academic Publishers 1991.

[8] David P. Morgan, Christopher L. Scofield, “Neural Networks and Speech Processing”, Kluwer Academic Publishers 1991.

[9] Jean Hennebert, Martin Hasler and Herve Dedien, “Neural Networks In Speech Recognition”, Swiss Fedral Institute Of Technology 1998.

[10] James A. Freeman, David M. Skapura,“Neural Networks Algorithms, Applications, and Programming Techniques”, Addison-Wesley Publishing Company 1991.

[11] Laurene Fausett, "Fundamentals of Neural Networks, Architectures, Algorithms, and Applications ", Prentice Hall,1994.

[12] Kohonen, “The Self Organizing Map”, Proceedings of the IEEE, September 1992.

[13] A. M. Kondoz, “Digital Speech Coding for Low Bit Rate Communications Systems”, John Wiley & Sons 1994.

[14] Emmanuel C. Ifeachor and Barrie W. Jervis, "Digital Signal Processing A Practical Approach", Addison-Wesley, 1993.

[15] John R. Deller, Jr., John G. Proakis, and John H. L. Hansen, “Discrete-Time Processing of Speech Signals”, Macmillan Publishing Company 1993.

[16] Abdolreza Hatamloua, Salwani Abdullahb, Hossein and Nezamabadi-pour “A combined approach for clustering based on K-means and gravitational search algorithms”. Swarm and Evolutionary Computation, Volume 6, 47-52, 2012.

[17] Mohammad F. Eltibi, Wesam M. Ashour, “Initializing K-Means Clustering Algorithm using Statistical Information “,International Journal of Computer Applications (0975 – 8887) Volume 29– No.7, September 2011

[18] Jian Zhou1,Guoyin Wang, Yong Yang, Peijun Chen, “Speech Emotion Recognition Based on Rough Set and SVM” Proc. 5th IEEE Int. Conf. on Cognitive Informatics (ICCI'06) 1-4244-0475-4/06/$20.OO @)2006 IEEE.

[19] Ahmad A. M. Abushariah, Teddy S. Gunawan, Othman O. Khalifa, Mohammad Abushariah, “English Digits Speech Recognition System Based on Hidden Markov Models” International Conference on Computer and Communication Engineering (ICCCE 2010), 11-13 May 2010, Kuala Lumpur, Malaysia.

[20] George E. Dahl, Dong Yu, Li Deng, and Alex Acero, “Context-Dependent Pre-Trained Deep Neural Networks for Large-Vocabulary Speech Recognition”, IEEE transactions on Audio, Speech, and Language Processing, VOL. 20, NO. 1, JANUARY 2012.

[21] Basem Ahmed, Tien-Ping Tan “Automatic Speech Recognition of Code Switching Speech using 1-Best Rescoring” International Conference on Asian Language Processing, IEEE 2012.

[22] Vikramjit Mitra, Wen Wang, Andreas Stolcke, Hosung Nam, Colleen Richey, Jiahong Yuan, Mark Liberman “ARTICULATORY TRAJECTORIES FOR LARGE-VOCABULARY SPEECH RECOGNITION” ICASSP 2013, 978-1-4799-0356-6/13/$31.00 ©2013 IEEE.

[23] Vikramjit Mitra, Horacio Franco, Martin Graciarena, Dimitra Vergyri, “MEDIUM-DURATION MODULATION CEPSTRAL FEATURE FOR ROBUST SPEECH RECOGNITION” 2014 IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP).

[24] Nitin Washani, Sandeep Sharma, “Speech Recognition System: A Review”, International Journal of Computer Applications (0975 – 8887) Volume 115 – No. 18, April 2015