A Hybrid Method for Extraction of Low-Order Features for Speech Recognition Application

Washington Luis Santos Silva*
*Laboratory of Eletronics Instruments
Federal Institute of Education, Science and Technology
São Luis, Maranhão, Brazil
e-mail:washington.wlss@ifma.edu.br

Ginalber Luiz de Oliveira Serra†
†Laboratory of Computational Intelligence Applied to Technology
Federal Institute of Education, Science and Technology
São Luis, Maranhão, Brazil
e-mail:ginalber@ifma.edu.br

***Abstract*—The concept of fuzzy sets and fuzzy logic is widely used in the proposal of several methods applied to systems modeling, classification and pattern recognition problem. This paper proposes a genetic-fuzzy system for extraction of low-order features for speech recognition application. In addition to pre-processing, with mel-cepstral coefficients, the Discrete Cosine Transform (DCT) is used to generate a two-dimensional time matrix with the features of low-order for each pattern to be recognized. A genetic algorithm is used to optimize a Mamdani fuzzy inference system in order to obtain the best model for final recognition. The proposed method used in this paper was named Hibrid Method for Extraction of Low-Order Features for Speech Recognition Application (HMFE). Experimental results for speech recognition applied to Brazilian language show the efficiency of the proposed methodology compared to methodologies widely used and cited in the literature.**

***Keywords*–Discrete Cosine Transform; Speech Recognition; Fuzzy Systems; Genetic Algorithm.**

I. INTRODUCTION

Parameterization of an analog speech signal is the first step in speech recognition process. Several popular signal analysis techniques have emerged as standards in the literature. These algorithms are intended to produce a perceptually meaningful parametric representation of the speech signal: parameters that can emulate some behavior observed in human auditory and perceptual systems. Actually, these algorithms are also designed to maximize recognition performance [1][2]. The selection of best representation for parametric speech signal is a very important task of developing any speech recognition system. The problem of pattern recognition might be formulated as follows: Let S_k classes, where $k \in \{1, 2, 3...K\}$, and $S_k \subset \Re^n$. If any pattern space is taken with dimension $\Re^x$, where $x \leq n$, it should transform this space into a new pattern space with dimension $\Re^a$, where $a < x \leq n$. Then assuming a statistical measure or second order model for each S_k, through a covariance function represented by $\left[\Phi_x^{(k)}\right]$, the covariance matrix of the general pattern recognition problem becomes:

$$[\Phi_x] = \sum_{k=1}^{K} P(S_k) \left[\Phi_x^{(k)}\right] \quad (1)$$

where $P(S_k)$ is a distribution function of the class S_k, *a priori*, with $0 \leq P(S_k) \leq 1$. A linear transformation operator through the matrix $\mathbf{A}$ maps the pattern space in a transformed space where the columns are orthogonal basis vectors of this matrix $\mathbf{A}$. The patterns of the new space are linear combinations of the original axes as structure of the matrix $\mathbf{A}$. The statistics of second order in the transformed space are given by:

$$\Phi_{\mathbf{A}} = \mathbf{A}^T[\Phi_x]\mathbf{A} \quad (2)$$

where $\Phi_{\mathbf{A}}$ is the covariance matrix which corresponds to the space generated by the matrix $\mathbf{A}$ and the operator $[\cdot]^T$ corresponds to the transpose of a matrix. Thus, it can extract features that provide greater discriminatory power for classification from the dimension of the space generated [3]. One of the most widespread techniques for pattern speech recognition is the "Hidden Markov Model" (HMM)[4]. A well known deficiency of the classical HMMs is the poor modeling of the acoustic events related to each state. Since the probability of recursion to the same state is constant, the probability of the acoustic event related to the state is exponentially decreasing. A second weakness of the HMMs is that the observation vectors within each state are assumed uncorrelated, and these vectors are correlated [5]. To overcome these drawbacks, robust recognizer has been proposed, since it has been experimentally shown that spectral variations are discriminant features for similar sounds. Several errors occur because an observation sequence is decoded by a few states typically absorbing low-energy frames [6]. The other states, instead, are quickly crossed because their distribution do not adapt well to the rest of the observation. Therefore, these errors do not depend on the intrinsic confusion of the words with similar sound, but on the poor modeling of the acoustic event which produces hypothesis weakly related to the acoustics of the correct word [7]. In order to justify the dynamic structure of the observation vectors, including global and local variations, this paper proposes a speech recognition system for isolated digits that are not based directly on the modeling of the state/word, but based on the global changes in the spectral characteristics of each word and their correlation in time, two important features partially explored by classical HMM [8][9].

Recently several works on digit recognition has been presented using MFCC classifiers and Neural Networks [10][11][12], Hybrid HMM-Suport Vector Machine (HMM-SVM) [13], Sparse Systems for Speech Recognition [14],

 ISBN: 978-1-61208-237-0

Hybrid Robust Voice Activity Detection System [15], Wolof Speech Recognition with Limited vocabulary Based HMM and Toolkit [16], Real-Time Robust Speech Recognition using Compact Support Vector Machines [17], Digit Recognition with Confidence [18], and others.

A. Proposed Methodology

In this proposal, a speech signal is encoded and parameterized in a two-dimensional time matrix with four parameters of the speech signal. After coding, the mean and variance of each pattern are used to generate the rule base of Mamdani fuzzy inference system. The mean and variance are optimized using genetic algorithm in order to have the best performance of the recognition system. This paper consider as patterns the Brazilian locutions (digits): $'0', '1', '2', '3', '4', '5', '6', '7', '8', '9'$. The Discrete Cosine Transform (DCT) [19][20] is used to encoding the speech patterns. The use of DCT in data compression and pattern classification has been increased in recent years, mainly due to the fact its performance is much closer to the results obtained by the Karhunen-Loève transform which is considered optimal for a variety of criteria such as mean square error of truncation and entropy [21]. This paper demonstrates the potential of DCT and fuzzy inference system in speech recognition [22]. These two tools have shown good results in the temporal modeling of speech signal [23].

II. Speech Recognition System

The proposed recognition system HMFE block diagram is depicted in Fig. 1.

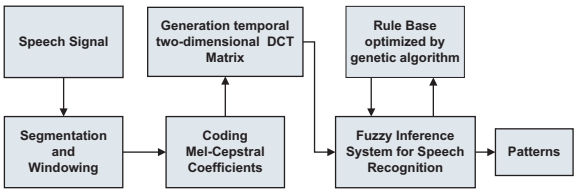

Fig. 1. Block diagram of the proposed recognition system HMFE.

A. Pre-processing Speech Signal

Initially, the speech signal is digitizing, so it is divided in segments which are windowed and encoded in a set of parameters defined by the order of mel-cepstral coefficients (MFCC). The DCT coefficients are computed and the two-dimensinal time DCT matrix is generated, based on each speech signal to be recognized.

1) Segmentation and windowing of the speech signal: When a window is applied to a given signal, it selects a small portion of this signal, named frame, to be analyzed. The duration of the frame T is defined as the total of time over which a set of parameters is considered valid. The duration of the frame is used to determine the total of time from successive calculations of parameters [1]. It is necessary to use a process called overlap to control how quickly the signal parameters may change from frame to frame because the windows at the ends of the analyzed signal have an excessive smoothing in their samples. In speech processing, the Hamming window is widely. This paper uses the hamming window with duration time (frames) of 10ms with 50% overlap between frames, thus, only a fraction of the signal is changed for each new frame .

2) Mel-Cepstrais Coefficients Coding: Experiments on human perception have shown that complex sound frequencies within a certain bandwidth of a nominal frequency should not be individually identified. When one of the components of this sound is out of bandwidth, this component can not be distinguished. Normally, it is considered a critical bandwidth for speech from 10 % to 20 % of the center frequency of the sound. One of the most popular way to map the frequency of a given sound signal for perceptual frequencies values, i.e., to be capable of exciting the human hearing range is the Mel-Scale [2].

B. Two-Dimensional Time Matrix DCT Coding

The two-dimensional time matrix as the result of DCT in a sequence of T mel-cepstral coefficients observation vectors on the time axis, is given by:

$$C_k(n,T) = \frac{1}{N}\sum_{t=1}^{T} mfcc_k(t) cos\frac{(2t-1)n\pi}{2T} \qquad (3)$$

where $mfcc$ are the mel-cepstral coefficients, and $k, 1 \leq k \leq K$, is the k-th (line) component of t-th frame of the matrix and $n, 1 \leq n \leq N$ (column) is the order of DCT. Thus, the two-dimensional time matrix [24], where the interesting low-order coefficients k and n that encode the long-term variations of the spectral envelope of the speech signal is obtained [7].

For a given spoken word *P* (digit), ten examples of utterances of *P* are gotten. This way it has itself $P_0^0, P_1^0, ..., P_9^0, P_0^1, P_1^1 ..., P_9^1, P_0^2, P_1^2, ..., P_9^2, ..., P_m^j$, where $j \in \{0, 1, 2, ..., 9\}$ and $m \in \{0, 1, 2, ..., 9\}$. Each frame of a given example of the word ***P*** generates a total of K mel-cepstral coefficients and the significant features are taken for each frame along time. The *N-th* order DCT is computed for each mel-cepstral coefficient of same order within the frames distributed along the time axis, i.e., c_1 of the frame t_1, c_1 of the frame $t_2, ...$, c_1 of the frame t_T, c_2 of the frame t_1, c_2 of the frame $t_2, ...,$, c_2 of the frame t_T, and so on, generating elements $\{c_{11}, c_{12}, c_{13}, ..., c_{1N}\}$, $\{c_{21}, c_{22}, c_{23}, ..., c_{2N}\}$, $\{c_{K1}, c_{K2}, c_{K3}, ..., c_{KN}\}$ of the matrix given in equation (3). Therefore, a two-dimensional time matrix DCT is generated for each example of the word ***P***. In this paper, the two-dimensional time matrices generated has order $(K = 2) \times (N = 2)$.

Finally, the matrices of mean CM_{kn}^j (4) and variances CV_{kn}^j(5) are generated. The parameters of CM_{kn}^j and CV_{kn}^j are used to produce Gaussians matrices C_{kn}^j which will be used as fundamental information for implementation of the fuzzy recognition system. The parameters of this matrix will

 ISBN: 978-1-61208-237-0

be optimized by genetic algorithm.

$$CM_{kn}^{j} = \frac{1}{M}\sum_{m=0}^{M-1} C_{kn}^{jm} \quad (4)$$

$$CV_{kn}^{j}(var) = \frac{1}{M-1}\sum_{m=0}^{M-1}\left[C_{kn}^{jm} - \left(\frac{1}{M}\sum_{m=0}^{M-1} C_{kn}^{jm}\right)\right]^{2} \quad (5)$$

where M=10.

C. Rule Base Used for Speech Recognition

Given the fuzzy set A input, the fuzzy set B output, should be obtained by the relational max-t composition [25]. This relationship is given by.

$$B = A \circ Ru \quad (6)$$

where Ru is a fuzzy relational rules base.

The fuzzy rule base of practical systems usually consists of more than one rule. There are two ways to infer a set of rules: Inference based on composition and inference based on individual rules [26][27]. In this paper the compositional inference is used. Generally, a fuzzy rule base is given by:

$$Ru^{l} : IF\ \ x_1\ \ is\ \ A_1^{l}\ \ and...and\ \ x_n\ \ is\ \ A_n^{l}\ \ THEN\ \ y\ \ is\ \ B^{l} \quad (7)$$

where A_i^l and B^l are fuzzy set in $U_i \subset \Re$ and $V \subset \Re$, and $x \in \{x_1, x_2, ..., x_n\}^T \in U$ and $y \in V$ are input and output variables of fuzzy system, respectively. Let M be the number of rules in the fuzzy rule base; that is, $l \in \{1, 2, ...M\}$.

From the coefficients of the matrices $\mathbf{C}_{kn}^{j}$ with $j \in \{0, 1, 2, ..., 9\}$, $k \in \{1, 2\}$ and $n \in \{1, 2\}$ generated during the training process, representing the mean and variance of each pattern j a rule base with $M = 40$ individual rules is obtained and given by:

$$Ru^{j} : IF\ \ C_{kn}^{j}\ \ THEN\ \ y^{j} \quad (8)$$

In this paper, the training process is based on the fuzzy relation Ru^j using the Mamdani implication. The rule base Ru^j should be considered a relation $R(X \times Y) \rightarrow [0, 1]$, computed by:

$$\mu_{Ru}(x, y) = I(\mu_A(x), \mu_B(y)) \quad (9)$$

where the operator I should be any t-norm [28][29][30]. Given the fuzzy set A' input, the fuzzy set B' output might be obtained by ***max-min*** composition, [26]. For a minimum t-norm and max-min composition it yields:

$$\mu_{(Ru)}(x, y) = I(\mu_A(x), \mu_B(y)) = min(\mu_A(x), \mu_B(y)) \quad (10)$$

$$\mu_{(B')} = max_x min_{x,y}(\mu_{A'}(x), \mu_{(Ru)}(x, y)) \quad (11)$$

D. Generation of Fuzzy Patterns

The elements of the matrix C_{kn}^{j} were used to generate Gaussians membership functions in the process of fuzzification. For each trained model j the Gaussians memberships functions $\mu_{c_{kn}^{j}}$ are generated, corresponding to the elements c_{kn}^{j} of the two-dimensional time matrix $\mathbf{C}_{kn}^{j}$ with $j \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$, where j is the model used in training. The training system for generation of fuzzy patterns is based on the encoding of the speech signal ***s(t)***, generating the parameters of the matrix C_{kn}^{j}. Then, these parameters are fuzzified, and they are related to properly fuzzified output y^j by the relational implications, generating a relational surface $\mu_{(Ru)}$, given by:

$$\mu_{Ru} = \mu_{c_{kn}^{j}} \circ \mu_{y^j} \quad (12)$$

This relational surface is the fuzzy system rule base for recognition optimized by genetic algorithm to maximize the speech recognition. The training system is shown in Fig. 2.

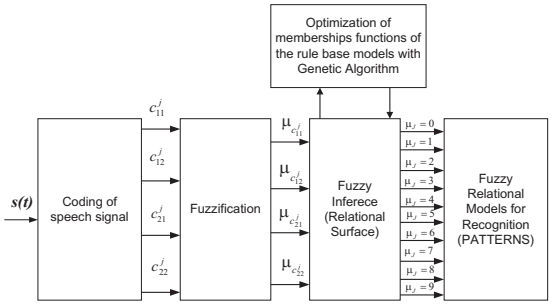

Fig. 2. Generation Systems Fuzzifieds Models.

E. Fuzzy Inference System for Speech Recognition Decision

The decision phase is performed by a fuzzy inference system based on the set of rules obtained from the mean and variance matrices of two dimensions time of each spoken digit. In this paper, a matrix with minimum number of parameters (2×2) in order to allow a satisfactory performance compared to pattern recognizers available in the literature. The elements of the matrices $\mathbf{C}_{kn}^{j}$ are used by the fuzzy inference system to generate four Gaussian membership functions corresponding to each element $c_{kn}^{j}\Big|^{k\in\{1,2\};n\in\{1,2\}}$ of the matrix. The set of rules of the fuzzy relation is given by:

Rule Bases

$$\textbf{\textit{IF}}\ \ c_{kn}^{j}\Big|^{k\in\{1,2\};n\in\{1,2\}}\ \ \textbf{\textit{THEN}}\ \ y^{j} \quad (13)$$

Modus Ponens

$$\textbf{\textit{IF}}\ \ c_{kn}^{'j}\Big|^{k\in\{1,2\};n\in\{1,2\}}\ \ \textbf{\textit{THEN}}\ \ y^{'j} \quad (14)$$

From the set of rules of the fuzzy relation between antecedent and consequent, a data matrix for the given implication is obtained. After the training process, the relational surfaces is generated based on the rule base and implication method

 ISBN: 978-1-61208-237-0

presented in Section II.D. The speech signal is encoded to be recognized and their parameters are evaluated in relation to the functions of each patterns on the surfaces and the degree of membership is obtained. The final decision for the pattern is taken according to the $max - min$ composition between the input parameters and the data contained in the relational surfaces. The process of defuzzification for the pattern recognition is based on the ***mean of maxima (mom)*** method given by:

$$\mu_{y'^j} = \mu_{c'^j_{kn}} \circ \mu_{(Ru)} \quad (15)$$

$$y^{'} = mom(\mu_{y'^j}) = mean\{y|\mu_{y'^j} = max_{y \in Y}(\mu_{y'^j})\} \quad (16)$$

F. Optimization of Relational Surface with Genetic Algorithm

The continuous genetic algorithm [31][32] is configured with a population size of 100, generations of 300, with mutations probability of 15% and two chromosomes, with 40 genes each, to optimize a cost function with 80 variables, which are the mean and variances of the patterns to be recognized by the proposed fuzzy recognition system. The genetic algorithm was used to optimize the variations of mean and variances of each pattern in order to maximize the successful recognition process. For example, for the pattern of the spoken word "*zero*" is generated ten two-dimensional time matrix. For each element of the matrix C^j_{kn} coefficients are determined with variations minimum and maximum , and the coefficient $c_{11} \in [c_{11(minimum)}\ c_{11(maximum)}]$, $c_{12} \in [c_{12(minimum)}\ c_{12(maximum)}]$, $c_{21} \in [c_{21(minimum)}\ c_{21(maximum)}]$, $c_{22} \in [c_{22(minimum)}\ c_{22(maximum)}]$. Thus, it has eight time varying parameters for each pattern which correspond to eighty parameters to be optimized by genetic algorithm [33].

III. Experimental Results

A. System Training

The patterns to be used in the recognition process were obtained from ten speakers who are speaking the digits 0 until 9. After pre-processing of the speech signal and fuzzification of the matrix C^j_{kn}, its fuzzifieds components $\mu_{c^j_{kn}}$ had been optimized by the GA that maximize the total of successful recognition. The optimization process was performed with 16 realizations of the genetic algorithm.The best result of the recognition processing by HMFE is shown in Fig. 3. The total number of hits using GA was 92 digits correctly identified in the training process. The relational surface generated for this result was used for validation process. The best individual in the first generation of the GA is shown in Fig. 4. In this case the total number of correct answers was 46 digits correctly identified. The relational surface of the best individual in the first generation of the GA is shown in Fig. 5.

In Fig. 6 are shown the features of the Gaussians membership functions of the optimum individual after the training process. This figure also shows a better distribution of the Gaussians membership functions organized by the GA during the training process. Fig. 7 presents the relational surface generated by Gaussians membership functions of the optimum individual after the training process, already organized by the GA. The better distribution of the Gaussians membership functions, made by the GA shown in Fig. 6 and Fig. 7 improved results due to reduction intrinsic confusion of the Gaussians membership functions shown in Fig. 4 and Fig. 5.

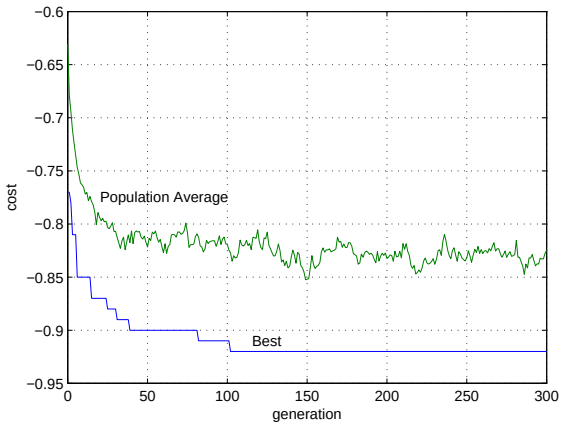

Fig. 3. Plot of the best results obtained in the training process.

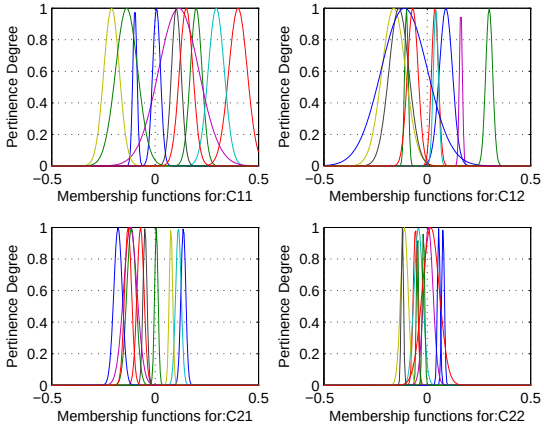

Fig. 4. Membership functions for c^j_{kn} in the 1st generation.

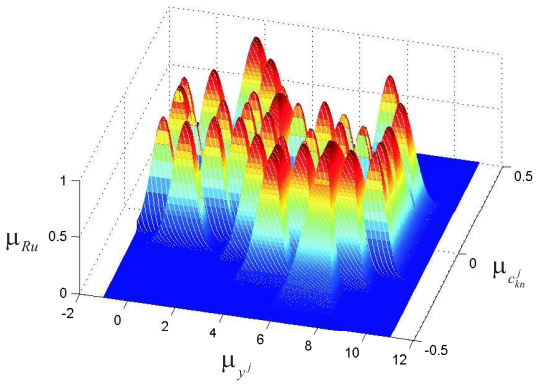

Fig. 5. Relational surface (μ_{Ru}) in the 1st generation.

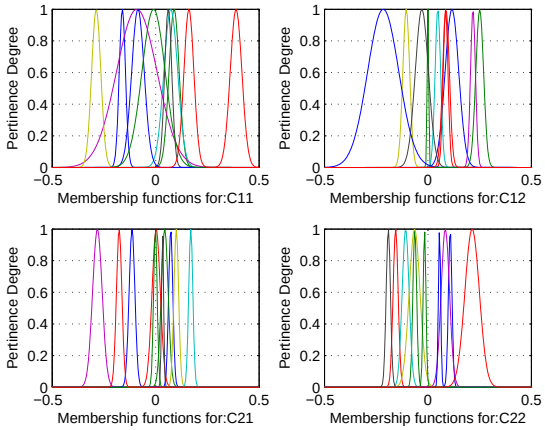

Fig. 6. Membership functions for c_{kn}^{j} optmized by GA.

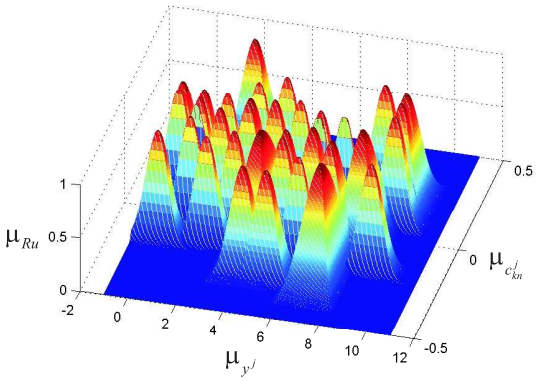

Fig. 7. Relational surface (μ_{Ru}) optmized by GA.

B. System Test - Validation

In this step, 100 locutions uttered in a room with controlled noise level and 500 locutions uttered in an environment without any kind of noise control were used. For every ten examples of each spoken digit, was generated two-dimensional time matrix cepstral coefficients $\mathbf{C}_{kn}^{j}$ and they were used in the test procedure. Six types of tests where performed:

Training: Recognition Optimized by HMFE (5 Female and 5 Male Speakers)

TEST 1: Validation - Strictly speaker dependent recognition, where the words used for training and testing were spoken by a same group of 10 speakers(5 Female and 5 Male Speakers).

TEST 2: Validation test- Recognition based on the partial dependence of the speaker with two examples for each ten examples of each digit(Female Speaker).

TEST 3: Validation test- Recognition based on the partial dependence of the speaker with two examples for each ten examples of each digit(Male Speaker).

TEST 4: Validation test- Recognition independent of the Speaker, where the speaker does not have influence in the training process(Female Speaker).

TEST 5: Validation test- Recognition independent of the Speaker, where the speaker does not have influence in the training process(Male Speaker).

Figures 8 - 13 present the comparative analysis of the HMM with two, three and four states, two, three and four Gaussians mixtures by state and order analysis, i.e., the number of mel-cepstral parameter equal 12 and HMFE with two, three and four parameteres for speech recognition.

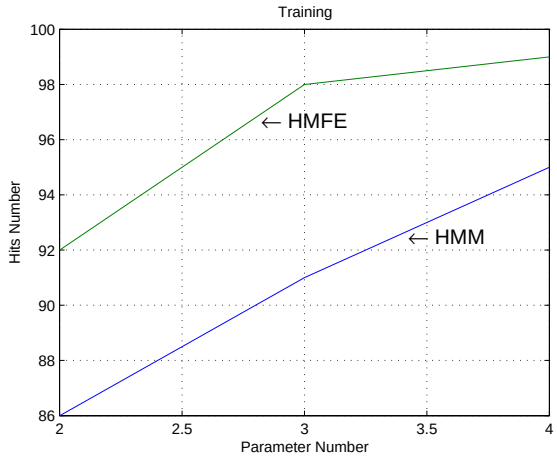

Fig. 8. Results for the digits used in the training.

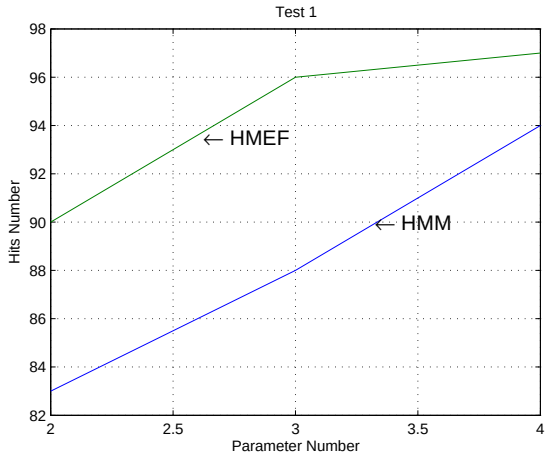

Fig. 9. Validation Test 1.

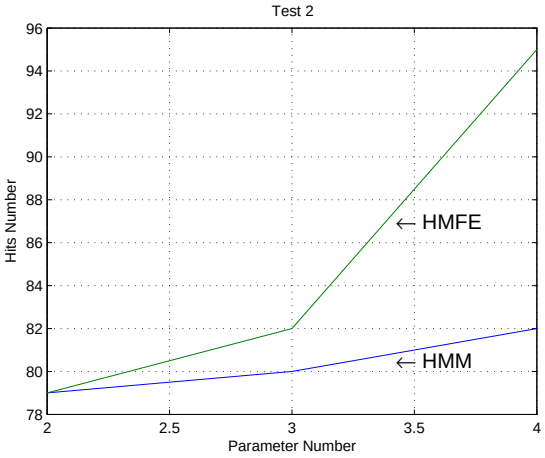

Fig. 10. Validation Test 2.

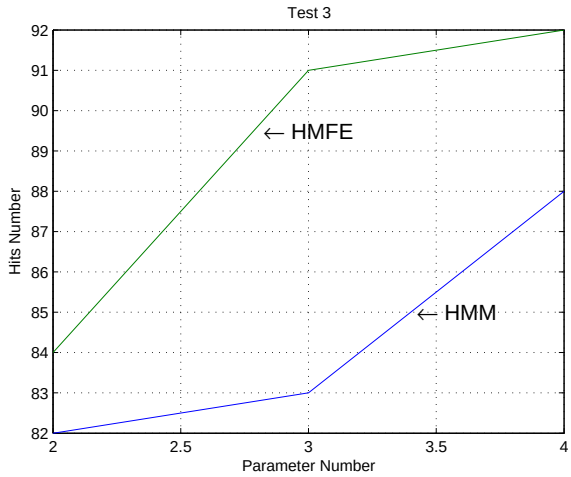

Fig. 11. Validation Test 3.

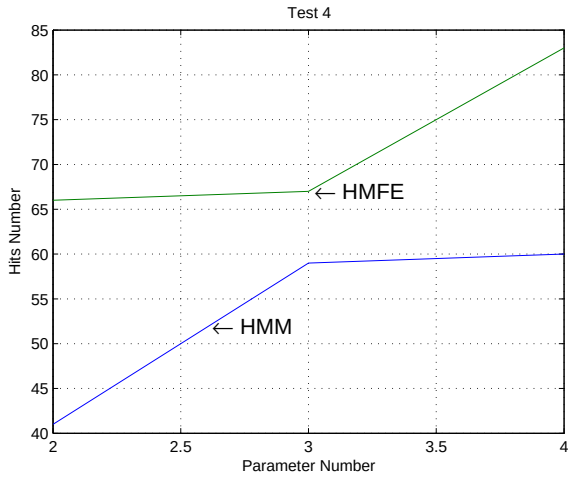

Fig. 12. Validation Test 4.

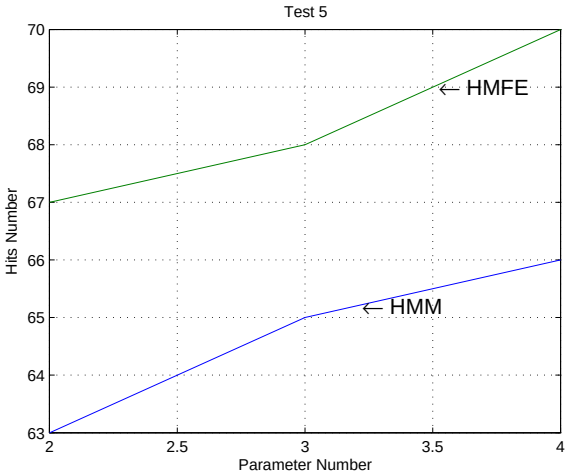

Fig. 13. Validation Test 5.

IV. Conclusion and Future Work

Evaluating the results, it is observed that the proposed method for extraction of low-order features for speech recognition application (HMFE), even with a minimal parameters number in the generated patterns was able to extract more reliably the temporal characteristics of the speech signal and produce good recognition results compared with the traditional HMM. To obtain equivalent results with HMM is necessary to increase the state number and/or mixture number. An increase in the order of the analysis above 12 does not improve significantly the performance of HMM. Any particular technique of noise reduction, such as those commonly used in HMM-based recognizers, was not used during the development of this paper. It is believed that with proper treatment of the signal to noise ratio in the process of training and testing, the HMFE Recognizer may improve its performance:

1) Increase the speech bank with different accents;
2) Use Nonlinear Predictive Coding for feature extraction in speech recognition;
3) Use Digital Filter in the speech signal to be recognized.
4) Increase the parameters number used.

Acknowledgment

The authors would like to thank FAPEMA for financial support, research group of computational intelligence applied to technology at the Federal Institute of Education, Science and Technology of the Maranhão by its infrastructure for this research and experimental results.

References

[1] J.W. Picone, “Signal Modeling Techniques in Speech Recognition”, IEEE Transactions on Computer, vol. 81, 9th edition, Apr. 1993, pp. 1215-1247, doi: 10.1109/5.237532.

[2] L. Rabiner and J. Biing-Hwang, “Fundamentals of Speech Recognition”, Prentice Hall, New Jersey, 1993.

[3] H. C. Andrews, “Multidimensional Rotations in Feature Selection”, IEEE Transaction on Computers, Sep. 1971, pp. 1045-1051, doi: 10.1109/T-C.1971.223400.

 ISBN: 978-1-61208-237-0

[4] A. A. M. Abushariah, T. S. Gunawan, O. O. Khalifa and M. A. M. Abushariah, “English Digits Speech Recognition System Based on Hidden Markov Models”, International Conference on Computer and Communication Engineer (ICCCE 2010), Kuala Lumpur, Malaysia, May 2010, pp. 1-5, doi:10.1109/ICCCE.2010.5556819.

[5] M.Wachter, M. Matton, K. Demuynck, P.K. Wambacq, R. Cools and D. Compernolle, “Template-Based Continuous Speech Recognition”, IEEE Transactions on Audio, Speech, and Language Processing, vol. 15, no. 4, May 2007, pp. 1377-1390, doi: 10.1109/TASL.2007.894524.

[6] Y. Ariki, S. Mizuta, M. Nagata and T. Sakai, “Spoken-Word Recognition Using Dynamic Features Analysed by Two-Dimensional Cepstrum”, IEEE Proceedings, vol. 136, no. 2, Apr. 1989, pp. 133-140.

[7] P. L. L. Fissore and E. Rivera, “Using Word Temporal Structure in HMM Speech Recognition”, ICASSP 97, vol. 2, Munich, Germany, Apr. 1997, pp. 975-978, doi: 10.1109/ICASSP.1997.596101.

[8] A. Revathi and Y. Venkataramani, “Speaker Independent Continuous Speech and Isolated Digit Recognition using VQ and HMM”, International Conference on Communications and Signal Processing (ICCSP), Calicut, India, Feb. 2011, pp. 198-202, doi: 10.1109/ICCSP.2011.5739300.

[9] J. Deng, M. Bouchard and T. H. Yeap, “Feature Enhancement for Noisy Speech Recognition with a Time-Variant Linear Predictive HMM Structure”, IEEE Transactions on Audio, Speech, and Language Processing, vol. 16, no. 5, Jul. 2008, pp. 891-899, doi: 10.1109/TASL.2004.924593.

[10] D. B. Hanchate, M. Nalawade, M. Pawar, V. Pohale and P. K. Maurya, “Vocal Digit Recognition Using Artificial Neural Network”, 2nd International Conference on Coumputer Engineering and Technology, vol. 6, Chendu, China, Apr. 2010, pp. 88-91, doi: 10.1109/ICCET.2010.5486314.

[11] R. K. Aggarwal and M. Dave, “Application of Genetically Optimized Neural Networks for Hindi Speech Recognition System”, World Congress on Information and Communication Technologies (WICT), Mumbai, India, Dec. 2011, pp. 512-517, doi: 10.1109/WICT.2011.6141298.

[12] S. M. Azam, Z. A. Mansor, M. S. Mughal and S. Moshin, “Urdu Spoken Digits Recognition Using Classifield MFCC and Backpropagation Neural Network”, 4th International Conference on Computer Graphics, Imaging and Visualization (CGIV), Bangkok, Thailand, Aug. 2007, pp. 414-418, doi: 10.1109/CGIV.2007.85.

[13] S. A. Hejazi, R. Kazemi and S. Ghaemmaghami, “Isolated Persian Digit Recognition Using a Hybrid HMM-SVM”, International Symposium on Intelligent Signal Processing and Communications Systems (ISPACS), Bangkok, Thailand, Dec. 2008, pp. 1-4, doi: 10.1109/ISPACS.2009.4806757.

[14] M. Mohammed, E. Bijov, C. Xavier, A. K. Yasif and V. Supriya, “Robust Automatic Speech Recognition Systems:HMM Vesus Sparse”, Third International Conference on Intelligent Systems modelling and Simulation, Kinabalu, Malaysia, Feb. 2012, pp. 339-342, doi: 10.1109/ISMS.2012.66.

[15] C. Ganesh, H. Kumar and P. T. Vanathi, “Performance Analysis of Hybrid Robust Automatic Speech Recognition System”, IEEE International Conference on Signal Processing, Computing and Control (ISPCC), Solan, India, Mar 2012, pp. 1-4, doi: 10.1109/ISMS.2012.66.

[16] J. K. Tamgo, E. Barnard, C. Lishou and M. Richome, “Wolof Speech Recognition Model of Digits and Limited-Vocabulary Based on HMM and ToolKit”, 14th International Conference on Computer Modelling and Simulation (UKSim), Cambridge, United Kingdom, Mar. 2012, pp. 389-395, doi: 10.1109/UKSim.2012.118.

[17] R. Solera, A. Moral, C. Moreno, M. Ramon and F. Maria, “Real-Time Robust Automatic Speech Recognition Using Compact Support Vector Machines”, IEEE Transactions on Audio, Speech, and Language Processing, vol.20, no. 4, May 2012, pp. 1347-1361, doi: 10.1109/TASL.2011.2178597.

[18] G. E. Sakr and I. H. Elhajj, “Digit Recognition with Confidence”, IEEE Workshop on Signal Processing Systems (SiPS), Beirut, Lebanon, Oct. 2011, pp. 299-304, doi: 10.1109/SiPS.2011.6088993.

[19] T. N. N. Ahmed and K. Rao, “Discrete Cosine Transform”, IEEE Transaction on Computers, vol.c-24, 2th edition, Jan. 1974, pp. 90-93, doi: 10.1109/T-C.1974.223784.

[20] P. C. J. Zhou, “Generalized Discrete Cosine Transform”, Pacific-Asia Conference on Circuits, Communications and System, Chegdu, China, May 2009, pp. 449-452, doi: 10.1109/PACCS.2009.62.

[21] M. Effros, H.Feng and K. Zeger, “Suboptimality of the KarhunenLove Transform for Transform Coding”, IEEE Transactions on Information Theory, vol. 50, no. 8, Aug. 2004, pp. 293-302, doi: 10.1109/DCC.2003.1194020.

[22] J. Zeng and Z. Q. Liu, “Type-2 Fuzzy Hidden Markov Models and their Application to Speech Recognition”, IEEE Transactions on Fuzzy Systems, vol. 14, no. 3, Jun. 2006, pp. 454-467, doi: 10.1109/TFUZZ.2006.876366.

[23] W. L. S. Silva and G. L. O. Serra, “Proposta de Metodologia TCD-Fuzzy para Reconhecimentos de Voz”, X SBAI Simposio Brasileiro de Automacao Inteligente, Sao Joao del-Rei, Brasil, Sep. 2011, pp. 1054-1059.

[24] M.Y. Azar and F. Razzazi, “A DCT Based Nonlinear Predictive Coding for Feature Extraction in Speech Recognition Systems”, IEEE International Conference on Computational Intelligente for Measurement Systems and Applications, Istanbul, Turkey, Jul. 2008, pp. 19-22, doi: 10.1109/CIMSA.2008.4595825.

[25] M. Mas, M. Monserrat, J. Torrens and E. Trillas, “A Survey on Fuzzy Implication Functions”, IEEE Transactions on Fuzzy Systems, vol.15, no.6, Dec. 2007. pp. 1107-1121, doi: 10.1109/TFUZZ.2007.896304.

[26] L.-X. Wang, “A Course in Fuzzy Systems and Control”, Prentice Hall, 1994.

[27] C. Gang, “Discussion of Approximation Properties of Minimum Inference Fuzzy System”, Proceedings of the 29th Chinese Control Conference, Beijing, China, Jul. 2010, pp. 2540-2546.

[28] R. Babuska, “Fuzzy Modeling for Control”, Kluwer Academic Publishers, 1998.

[29] H. Seki, H. Ishii and M. Mizumoto, “On the Monotonicity of Fuzzy-Inference Methods Related to TS Inference Method”, IEEE Transactions on Fuzzy Systems, vol. 18, no. 3, Jun. 2010, pp. 629-634, doi: 10.1109/TFUZZ.2010.2046668.

[30] G. Gosztolya, J. Dombi and A. Kocsor, “Applying the Generalized Dombi Operator Family to the Speech Recognition Task”, Journal of Computing and Information Technology - CIT, vol. 17, no. 9, 2009, pp. 285-293, doi: :10.2498/cit.1001284.

[31] R. L. Haupt and S. E. Haupt, “Pratical Genetic Algorithms”, John Wiley & Sons, Inc, 2004.

[32] C. Tang, E. Lai and Y. C. Wang, “Distributed Fuzzy Rules for Preprocessing of Speech Segmentation with Genetic Algorithm”, Fuzzy-IEEE Conference 1997, vol. 1, Barcelona, Spain, Jul. 1997, pp. 427-431, doi: 10.1109/FUZZY.1997.616406.

[33] K. Tang, K. Man, Z. Liu and S. Kwong, “Minimal Fuzzy Memberships and Rules Using Hierarchical Genetic Algorithms”, IEEE Transactions on Industrial Eletronics, vol. 45, no. 1, Feb. 1998. pp. 427-431, doi: 10.1109/FUZZY.1997.616406.

 ISBN: 978-1-61208-237-0