2014 International Joint Conference on Neural Networks (IJCNN)
July 6-11, 2014, Beijing, China

A Novel Intelligent System for Speech Recognition

Washington Luis Santos Silva* and Ginalber Luiz de Oliveira Serra*
*Federal Institute of Education, Science and Technology
Department of Electroelectronics, Laboratory of Computational Intelligence Applied to Technology
Av. Getulio Vargas, 04, Monte Castelo, CEP: 65030-005, Sao Luis, Maranhao, Brazil
Email: washington.wlss@ifma.edu.br;ginalber@ifma.edu.br

***Abstract*—The concept of fuzzy sets and fuzzy logic is widely used to propose several methods applied to modeling, classification and pattern recognition problem. This paper proposes an Intelligent Methodology for Speech Recognition (IMSR). In addition to pre-processing, with mel-cepstral coefficients, the Discrete Cosine Transform (DCT) is used to generate a two-dimensional time matrix 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 with minimum number of parameters for final recognition. 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*—Fuzzy Systems; Automatic Speech Recognition; Genetic Algorithms; Discrete Cosine Transform; Instelligent System.**

I. INTRODUCTION

The goal of an Automatic Speech Recognition System (ASR) is to convert a speech signal to mathematic coding of the spoken words, accurately and efficiently, independent of the device used to record the speech (i.e., the transducer or microphone), the speakers accent, or the acoustic environment in which the speaker is located (e.g., quiet office, noisy room, outdoors). That is, the ultimate goal, which has not yet been achieved, is to perform as well as a human listener. The parametrization of a speech signal is the first step in speech recognition process. Several popular signal analysis standards techniques have emerged in the literature. These algorithms are intended to produce a perceptually meaningful parametric representation of the speech signal: parameters that emulate some behavior observed in human auditory and perceptual systems[1].

The selection of the better parametric representation for speech signal is a very important task of developing any speech recognition system. The goal of selecting the best way to encode the signal is to compress the speech data information, eliminating non-phonetic analysis of the signal and improving the aspects of the signal which contributes significantly to detect phonetic differences of speech sounds [2].

This Paper follows the formalism developed by Andrews [3]. The problem of pattern recognition might be formulated as follows: Let S_k classes, indexed by k = 1,2,3...K, and $S_k \subset \mathbb{R}^n$, if any pattern space is taken with dimension $\mathbb{R}^x$, where $x \leq n$, it should transform this space into a new pattern space with dimension $\mathbb{R}^a$, where $a < x \leq n$. Therefore 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:

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

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:

$$\Phi_{\mathbf{A}} = \mathbf{A}^T [\Phi_x] \mathbf{A} \tag{2}$$

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. This probability distribution does not model the speech temporal structure. A second weakness of the HMMs is that the observation vectors within each state are assumed uncorrelated, and these vectors are correlated. Artificial Neural Networks are widely used to improve the recognition performance for speech recognition.

To overcome this problem, a variety of different hybrid techniques with HMM and ANN were proposed retaining its their ability to treat the temporal variability of the speech signal. However, analysis with ANN requires lots input data, which can be a limiting factor for computing performance. The above considerations has motivated us to propose a hybrid intelligent architecture based on liguistic genarilization of the Mamdani Fuzzy Inference System, the optmization of the genetic algorithm (GA) and the capacity reduction feature of the Discrete Cosine Transform (DCT).

The main contribution of the proposed methodology can be addressed as follow:

1) It is not based directly on the modeling of the state/word, but on the global changes in the spectral characteristics of each word and their correlation at the time, in order to justify the dynamic structure of

the observation vectors, including global and local variations;

2) It utilizes a minimal number of parameter in the speech signal codification ;
3) It utilizes an intelligent system to classify and recognize the speech patterns; A speech signal is encoded in two-dimensional time matrix through of the MFCC and DCT coefficients, the mean and variance of the parameters of each speech pattern are used to develop the rule base of a linguistic fuzzy inference system, which is optimized by genetic algorithm so resulting recognition system with best performance.

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): $'0','1','2','3','4','5','6','7','8','9'$. This paper demonstrates the potential of DCT and fuzzy inference system in speech recognition [5].

II. A Hybrid-Intelligent Methodology for Speech Recognition

The proposed Recognition System Block Diagram is depicted in Fig.1.

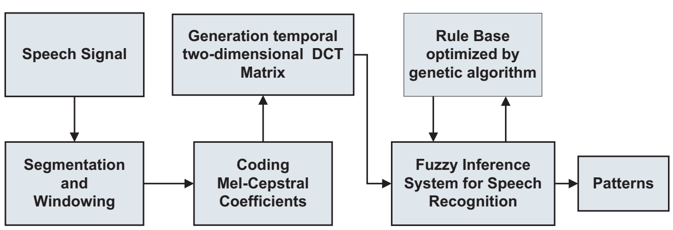

Fig. 1: Block diagram of the proposed recongnition system.

There is no standard set of features for speech recognition. Instead, various combinations of acoustic, articulatory, and auditory features have been utilized in a range of speech recognition systems. The most popular acoustic features have been the (LPC-derived) mel-frequency cepstrum coefficients and their derivatives. Initially, the speech signal is digitizing, so it is divided in segments they which are windowed and encoded in a set of parameters defined by the order of mel-frenquency cepstrum coefficients (MFCC). The DCT coefficients are computed and the two-dimensional time DCT matrix is generated, based on each speech signal to be recognized.

A. Two-Dimensional Time Matrix DCT Coding

At the stage encoding is applying a DCT feature extractor to remove unwanted additional data from the speech samples, this way, the speech frames are transformed into a DCT space. DCT-II-E coefficients, used in this paper, are calculated by following relation

$$X(k) = \sum_{n=0}^{N-1} \alpha(n) x(n) cos \frac{(2k+1)n\pi}{2N} \quad (3)$$

$k = 0, 1, 2, ..., N-1$, and

$$\alpha(n) = \begin{cases} \sqrt{1/N}, & \text{if } n = 0 \\ \sqrt{2/N}, & \text{otherwise} \end{cases}$$

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:

$$C_k(n,T) = \frac{1}{N} \sum_{t=1}^{T} mfcc_k(t) cos \frac{(2t-1)n\pi}{2T} \quad (4)$$

where $k, 1 \leq k \leq K$, 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, 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. Thus, there is a two-dimensional time matrix $C_k(n,T)$ for each input speech signal. The elements of the matrix are obtained as follows:

1) 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 $m \in \{0, 1, 2, ..., 9\}$.
2) 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 (4). 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 $(K = 2) \times (N = 2)$.
3) Finally, the matrices of mean CM_{kn}^j (5) and variances CV_{kn}^j(6) 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 be optimized by genetic algorithm.

$$CM_{kn}^j = \frac{1}{M} \sum_{m=0}^{M-1} C_{kn}^{jm} \quad (5)$$

$$CV_{kn}^j(var) = \frac{1}{M-1} \sum_{m=0}^{M-1} \left[C_{kn}^{jm} - \left(\frac{1}{M} \sum_{m=0}^{M-1} C_{kn}^{jm} \right) \right]^2 \quad (6)$$

Given the fuzzy set A' input, the fuzzy set B' output, should be obtained by the relational max-t composition. This relationship is given by.

$$B' = A' \circ Ru \quad (7)$$

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. In this paper the compositional inference is used. Generally, a fuzzy rule base is given by:

$$Ru^l : IF\ \ x_1\ \ is\ \ A_1^l\ \ and...and\ \ x_n\ \ is\ \ A_n^l\ \ THEN\ \ y\ \ is\ \ B^l \quad (8)$$

where A_i^l and B^l are fuzzy set in $U_i \subset \mathbb{R}$ and $V \subset \mathbb{R}$, and $x = (x_1, x_2, ..., x_n)^T \in U$ 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, $l = 1, 2, ...M$.

B. Rule base used for speech recognition

From the coefficients of the matrices $\mathbf{C}_{kn}^j$ with $j = 0, 1, 2, ..., 9$, $k = 1, 2$ and $n = 1, 2$ 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:

$$Ru^j : IF\ \ C_{kn}^j\ \ THEN\ \ y^j \quad (9)$$

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:

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

where the operator I should be any t-norm.

C. 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 $j = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9$, 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:

$$\mu_{Ru} = \mu_{c_{kn}^j} \circ \mu_{y^j} \quad (11)$$

This relational surface is the fuzzy system rule base for recognition optimized by genetic algorithm to maximize the speech recognition.

D. 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 $c_{kn}^j\big|^{k=1,2;n=1,2}$ of the matrix. The set of rules of the fuzzy relation is given by:

Rule Bases

$$\boldsymbol{IF}\ \ c_{kn}^j\big|^{k=1,2;n=1,2}\ \ \boldsymbol{THEN}\ \ y^j \quad (12)$$

Modus Ponens

$$\boldsymbol{IF}\ \ c_{kn}^{'j}\big|^{k=1,2;n=1,2}\ \ \boldsymbol{THEN}\ \ y^{'j} \quad (13)$$

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 presented in section 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:

$$\mu_{y'^j} = \mu_{c_{kn}^{'j}} \circ \mu_{(Ru)} \quad (14)$$

$$y^{'} = mom(\mu_{y'^j}) = mean\{y | \mu_{y'^j} = max_{y \in Y}(\mu_{y'^j})\} \quad (15)$$

E. Optimization of Relational Surface with Genetic Algorithm

The continuous genetic algorithm is configured with a population size of 100, generations of 300, with mutations probability of 15% and two individuals (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.

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_{kn}^j, its fuzzifieds components $\mu_{c_{kn}^j}$ had been optimized by the GA that maximize the total of successful recognition. The optimization process was performed with 50 realizations of the genetic algorithm.

The best result of the recognition processing is shown in Fig.2. The total number of hits using GA was 94 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 is shown in Fig.3. In this case the total number of correct answers was 46 digits correctly identified. The ralational surface of the best individual in the first generation is shows in Fig.4.

The optimum individual, presents the features in Fig.5 and Fig.6.

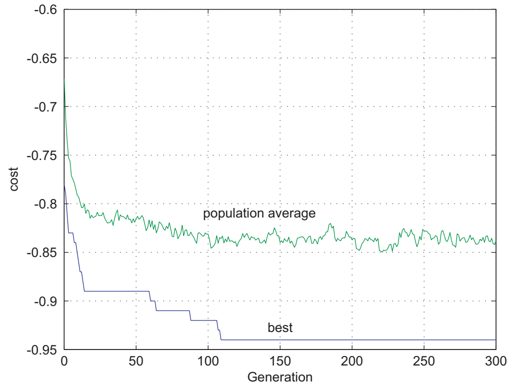

Fig. 2: Plot of the best results obtained in the training process.

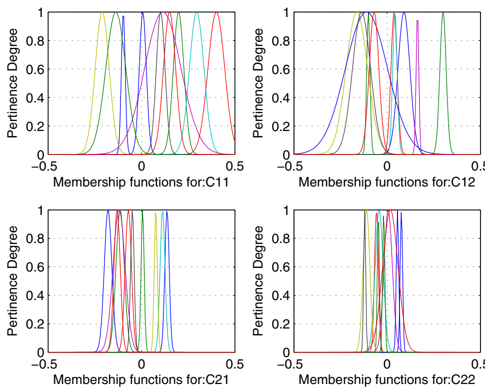

Fig. 3: Membership functions for c_{kn}^{j} in the 1st generation.

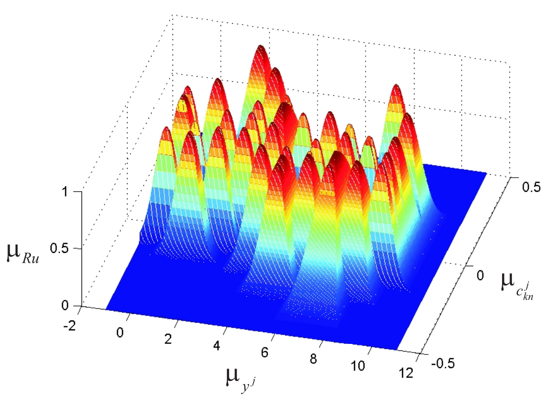

Fig. 4: Relational surface (μ_{Ru}) in the 1st generation.

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, two-dimensional time matrix cepstral coefficients $\mathbf{C}_{kn}^{j}$ was generated and they were used in the test procedure in which six types of tests were performed:

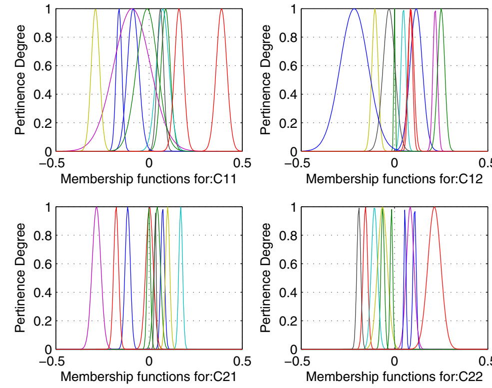

Fig. 5: Membership functions for c_{kn}^{j} optmized by GA.

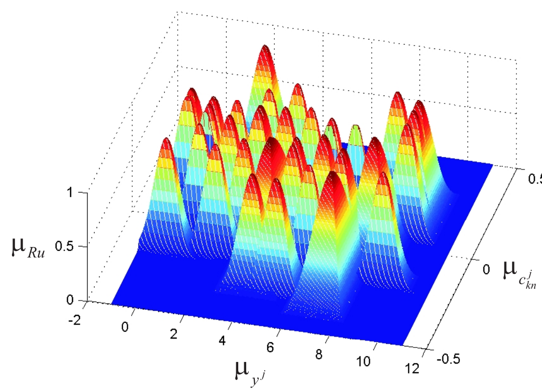

Fig. 6: Relational surface (μ_{Ru}) optmized by GA.

Training: Recognition Optimized by IMSR (5 Female and 5 Male Speakers);
TEST 1: Validation - Strictly speaker dependent recognition, in which 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);

In the Fig. 7 to 12 presents the comparative analysis of the HMM with two, three and four states, two, three and four gaussians mixtures by state and order analysis equal 12, i.e., the number of mel-cepstral parameter equal 12 to HMM and the Intelligent Methodology for Speech Recognition(**IMSR**) with two, three and four mel-cepstral parameter for speech recognition. With the data points obtained experimentally, a fit curve for all tests was mapped, and the amount of parameters needed to obtain 100% accuracy is estimated by these curves with two tested recognizers. In the Fig.7 it is observed that with two mel-cepstral parameters could be a number of hits equal to 94% with three parameters to 98% and 99% with a total of four mel-cepstral parameters. Thus, through the fit curve, a total of 100 % accuracy can be reached, since that tuned properly the parameters of the genetic algorithm. In the Fig.8 and Fig.9 is shown an estimate of 100% accuracy with approximately 5 mel-cepstral parameters. In the Fig.10 is shown an estimate of 100% accuracy with a total of 7 mel-cepstral parameters. It is noteworthy that these results are strictly speaker dependent (Fig.8) and with partial speaker dependence, respectively (Fig.9 and Fig.10). In the Fig.11 and Fig.12, where tests are independent of the speaker gets a higher estimate of the number of mel-cepstral parameters 7 and 16, respectively, to reach the 100% accuracy.

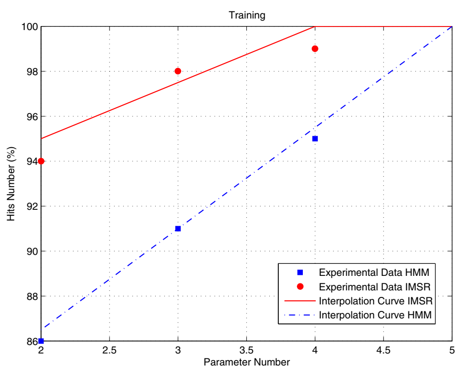

Fig. 7: Results in the training.

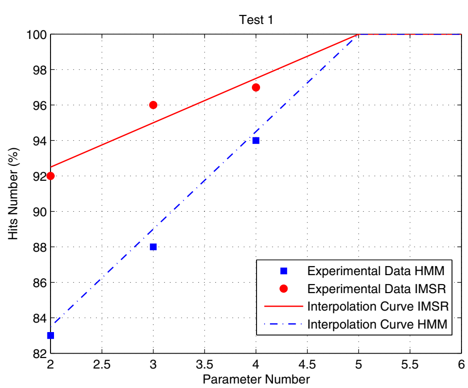

Fig. 8: Validation Test 1.

Experimental results and mel-cepstral parameter were presented in Table I for tests performed in this paper.From data obtained, it is observed that even with a lower number of parameters, resulting from the encoding of the speech signal, similar results are obtained and are compared with methodologies more complex with a larger number of parameter.

C. Comparison with other intelligent methods

The goal with this comparison is to show that the proposed method, even with a minimal number of input parameters, produces similar results comparable with other intelligent techniques with a substantial amount of input data. For this

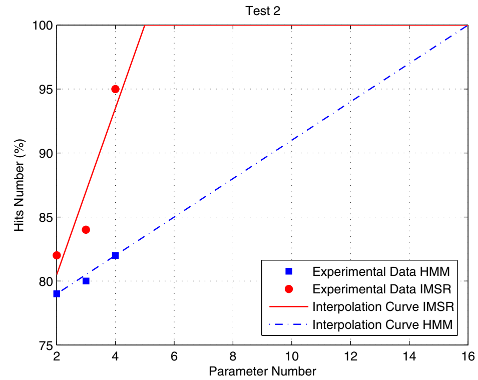

Fig. 9: Validation Test 2.

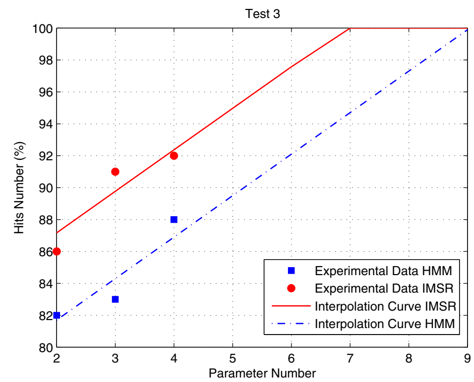

Fig. 10: Validation Test 3.

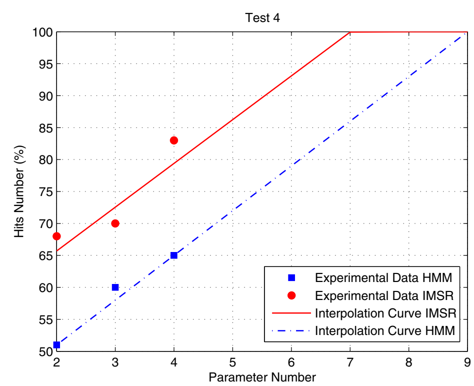

Fig. 11: Validation Test 4.

TABLE I: Results for Proposed Methodology

IMSR Parameters	IMSR Traning	IMSR Test-01	IMSR Test-02	IMSR Test-03	IMSR Test-04	IMSR Test-05
MFCC=2, DCT=2	94	92	82	86	68	65
MFCC=3, DCT=3	98	96	84	91	70	68
MFCC=4, DCT=4	99	97	95	92	84	70
(Estimate)MFCC=4, DCT=4	100	-	-	-	-	-
(Estimate)MFCC=5, DCT=5	-	100	-	-	-	-
(Estimate)MFCC=5, DCT=5	-	-	100	-	-	-
(Estimate)MFCC=7, DCT=7	-	-	-	100	-	-
(Estimate)MFCC=7, DCT=7	-	-	-	-	100	-
(Estimate)MFCC=16, DCT=16	-	-	-	-	-	100

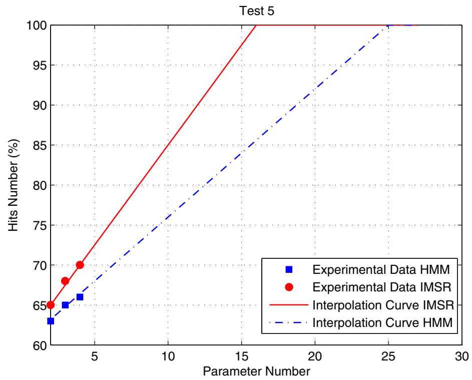

Fig. 12: Validation Test 5.

comparison the results presented in the article [8] were used. The work cited [8] use the same patterns used in proposed work, i.e., brazilian locutions (digits). The following describes the database used in [8]: Ťhe database consists of spoken digits in Portuguese collected during a period of three months, from eighty-two men aged between 18 and 42 years-old. The sampling rate of the recording is 22050Hz. Altogether, the database has 216 sequences of 10 digits (0 - 9) each, totalling 10 classes and 2.160 examples. Thus, it is a balanced dataset considering the class distribution.2 That it uses a MFCC with 13 coefficients and Line Spectral Frequencies with orders 24 and 48. In the proposed methodology, it is use A IMSR com order 2, 3 and 4, with a amount of 4, 9 and 16 input data for the recognition. In the Table II there are describ scenarios used by [8] and in the Table III there are described the obtained results in the work perfomed by [8] for comparation.

TABLE II: Experiment scenarios description performed by [8]

Scenario	Inducer/Settings
1-NN	Nearest Neibor
5-NN	5- Nearest Neighbor weighted by inverse distance (one)
7-NN	7- Nearest Neighbor weighted by inverse distance
9-NN	9- Nearest Neighbor weighted by inverse distance
SVM-Poly1	Suport Vector Machine with Polynomial Kernel with Degree 1
SVM-Poly2	Suport Vector Machine with Polynomial Kernel with Degree 2
SVM-Poly3	Suport Vector Machine with Polynomial Kernel with Degree 3
SVM-RBF0.01	Suport Vector Machine with RBF Kernel with Gamma =0.01
SVM-RBF0.05	Suport Vector Machine with RBF Kernel with Gamma =0.05
SVM-RBF0.1	Suport Vector Machine with RBF Kernel with Gamma =0.1
NB	Naive Bayes
RF	Random Forest

TABLE III: Mean accuracy for three analyzed methods on 12 scenario performed by [8]

Scenario	MFCC	24LSF	48 LSF
1-NN	86.33	92.92	93.03
5-NN	89.52	95.57	95.66
7-NN	89.61	95.82	95.98
9-NN	90.20	96.13	95.67
SVM-Poly1	97.96	98.85	99.30
SVM-Poly2	97.88	98.77	99.31
SVM-Poly3	97.91	98.75	99.17
SVM-RBF0.01	93.62	97.93	98.64
SVM-RBF0.05	96.88	98.54	98.70
SVM-RBF0.1	97.19	98.32	98.02
NB	90.63	94.86	94.72
RF	91.83	96.36	95.89

IV. CONCLUSION

Evaluating the results, it is observed that the proposed speech recognizer, even with a minimal number of parameters in the generated patterns, was reliably able to extract 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. Other intelligent techniques, such as neural networks, svm, or hybrid intelligent techniques, can achieve good results in speech recognition, however, they usually suffer from the curse of dimensionality, with a high number of parameters and a heavy computational load. The proposed methodology can work with a small number of parameters, maintaining a reasonable number of hits, which indicates its ability to discard redundant information not necessary to the process of recognition. It is believed that with proper treatment of the signal to noise ratio in the process of training and testing, the proposed speech recognizer may improve its performance:Increase the speech bank with different accents; Improve the performance of genetic algorithm to 100% recognition in the training process; Use Nonlinear Predicitve Coding for feature extraction in speech recognition and Use Digital Filter in the speech signal to be recognized.

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 Maranhao by its infrastructure for this research and experimental results, and Master and PhD program in Eletrical Engineering at the Federal University of Maranhao (UFMA).

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.

[4] A. A. M. Abushariah, T. S. Gunawan, O. O. Khalifa, 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.Y. Azar, 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.

[6] G. Gosztolya, J. Dombi, 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.

[7] X. Zhang, X. Wang, S. Zhang and F. Yu. “Approximating the True Domain of Fuzzy Inference Sentence with Genetic Algorithm”. Seventh International Conference on Fuzzy Systems and Knowledge Discovery-FSKD 2010, pp.114-118, doi:10.1109/FSKD.2010.5569735.

[8] D.F. Silva, V.M.A. de Souza, G.E. Batista and R. Giusti.“Spoken Digit Recognition in Portuguese Using Line Spectral Frequencies”. Lectures Notes in Artificial Intelligence-LNAI 7637. Springer-Verlag Berlin Heidelberg. pp.241-250. ISBN 978-3-642-34653-8.