23783
04.
A8

A NEURO-FUZZY APPROACH TO ISOLATED WORD RECOGNITION

submitted by
SAROJ KUMAR DAS

under the supervision of
PROF. KUMAR SANKAR RAY
ELECTRONICS & COMMUNICATION SCIENCES UNIT

A dissertation
submitted in the partial fulfillment for the degree of
Master of Technology
in
Computer Science

INDIAN STATISTICAL INSTITUTE LIBRARY
C 23783
SEP 1997
CALCUTTA

INDIAN STATISTICAL INSTITUTE
203, B.T ROAD.
CALCUTTA - 700035
1997

Indian Statistical Institute
Calcutta - 700 035

To whom it may concern

This is to certify that the thesis **A Neuro Fuzzy Approach to Isolated Word recognition** *is a bona fide piece of work of Saroj Kumar Das, done under my supervision and is to be taken as a partial fulfilment of the requirements for the degree of Master of Technology in Computer Science of Indian Statistical Institute, Calcutta.*

Date : 28/07/1997

(K. S. Ray)
Professor
E.C.S.U., I.S.I.
Calcutta.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to Dr K. S. Ray, Professor, E.C.S.U, Indian Statistical Institute,Calcutta for his constant help,support and guidance during the successful completion of the dissertation work.It was on account of his persistent encouragement at the moment of stress that ultimately could lead to fruition of the endeavor.

I also owe a lot to many others, especially the staff of E.C.S.U., for their direct and productive involvement and for creating an amicable atmosphere at the laboratories.

Abstract

The dissertation deals with the applicatioin of fusion technology of artificial neural network and fuzzy logic to the contradistinction of Bengali words by considering the first two formant frequencies of the embedded vowel patterns in them.To facilitate the said work we also give a new model of interpretation to the multidimensional fuzzy reasoning and realize the same through a backpropagation type neural network,where the learning stage of the network uses fuzzy linguistic statements as input.Once learned, the nonfuzzy features can be classified by fuzzification of the desired features using fuzzy singleton and then feeding the ensuing fuzzy set to the recognition system followed by defuzzification of the result.The proposed model has been successfully tested for vowel recognition problem of some selected Bengali words.

Contents

STATISTICAL
23783

1 INTRODUCTION

The area of Fuzzy Logic has received a tremendous amount of attention after the presentation of the Compositional Rule of Inference by Zadeh.Since then a large number of literatures have focused their attention on fuzzy reasoning and its applications.However, a similar reasoning for the multidimensional case has received few attention. So,in this paper to tackle the pattern recognition problems for isolated word recognition, we give a new interpretation to multidimensional fuzzy reasoning. The new interpretation of multidimensional fuzzy reasoning (**MFR**) can be hopefully realized through a Back Propagation type Neural Network.Recently, fuzzy reasoning and neural network have independently received considerable attention in the field of engineering and applied sciences.They share the common characteristics that each one promises to provide satisfactory solutions for flexible information processing.To compensate for the inherent demerits of each field by the merits of the other and vice versa, an attempt has been to generate an effective and successful fusion technology to coelesce the two fields.Recent results on fusion technology are reported in Takagi and Hayashi in 1991 and Takagi et al in 1992.In the course of developing such fusion technology the researchers claim that the existing fuzzy reasoning suffers from the following defficiencies.

1. The lack of a definite method to determine the membership function.
2. The lack of a learning function or adaptability which can be overcome by neural network driven fuzzy reasoning.

Here, our present work deals with generation of an alternative fusion technology for the problem of isolated word recognition by realizing the new interpretation of MFR through a Back Propagation type Neural Network. We do so because of the following problems of the existing MFR.

1. There is no definite method for finding out the appropriate translating rule for the law of implication.
2. There is practical difficulty in handling a multidimensional relation.
3. There is lack of a definite method to determine appropriate compositional rule.
4. And finally, the lack of a learning function or adaptability.

But here we don't question the lack of a definite method for determining membership function of a fuzzy set.This is because fuzzy membership can be successfully taken to be the quantification of human perception.It may so happen that the level of perception varies according to the individuality.But then, we can set a tolerance level and rule out certain absurd type of perception, such as assigning membership function as

$$1/1 + .3/2 + .1/3 + .2/4$$

for the fuzzy set **Big** in the universe of discourse $\{1,2,3,4\}$.In certain cases heuristic measures are taken to tune and fit the membership function more accurately.

2 STATEMENT OF THE PROBLEM

The problem undertaken here is "How to recognize isolated words by the fusion technology of Neural network and Fuzzy reasoning". The subject matter inherently falls under the class of pattern recognition problems applied to speech signals. Before getting into the problem itself, we would like to present the conventional approach to Multidimensional Fuzzy Reasoning (MFR) and our new approach to suit the pattern recognition problem better.

Multidimensional fuzzy implications(MFI) like

$$\text{if}(X \text{ is } A, Y \text{ is } B) \text{ then } Z \text{ is } C$$

where A,B and C are fuzzy sets, generally have the following two types of conventional interpretations taken in the multidimensional case in the compositional rule of inference.

1. if X is A and Y is B then Z is C
2. if X is A then if Y is B then Z is C

According to Tsukamoto, the above multidimensional fuzzy implication can be represented as

$$\begin{gathered}\text{if } X \text{ is } A \text{ then } Z \text{ is } C\\ \text{and}\\ \text{if } Y \text{ is } B \text{ then } Z \text{ is } C.\end{gathered}$$

and we take the fuzzy set $C' \cap C''$ for the consequence of reasoning, where C' and C" are the inferred values from the first and second implications respectively.

To tackle the pattern recognition problem using Multidimensional fuzzy reasoning (MFR), we modify the above interpretation as follows to suit this special case.

$$\begin{gathered}\text{if } X \text{ is } A \text{ then } Z \text{ is } C_1\\ \text{and}\\ \text{if } Y \text{ is } B \text{ then } Z \text{ is } C_2.\end{gathered}$$

where we take $C = C_1 \cap C_2$ and the intersection $C' \cap C''$, where C' and C'' are the inferred values from the first and second implications respectively, is taken to be the consequence of reasoning.

The new interpretation differs from the Tsukamoto's model only at the interpretation of the consequent part of each of the decomposed fuzzy implication(DFI) of the MFI.According to Tsukamoto, the consequent parts of the DFI's of an MFI are the same as the consequent part of the said MFI.Whereas, in our newly proposed model,the consequent part of MFI is the intersection of the consequent parts of DFI's.Thus, the linguistic connective 'and' has a more meaningful logical interpretation $\cap$ in the latter case than in the former.

We introduced here the notion of what we call the Fuzzy Feature Vector $\mathbf{F_f}$ (fig 1), whose elements are the antecedent parts of the DFI's which are basically linguistic features like F_1 is big and F_2 is small, where small, big etc are represented by fuzzy sets.The tip of the fuzzy feature vector is not necessarily a crisp point;it rather represents a population of pattern(**ABCD**).In the context of pattern recognition the consequent part of MFI represents the possibility of occurrence of each class on the pattern space.According to a particular MFI, a particular class $\tilde{\mathbf{C}}_\mathbf{i}$ may have higher possibility of occurrence in pattern space R^2(for simplicity we take R^2 instead of R^n) than a class $\tilde{\mathbf{C}}_\mathbf{j}$.This is indicated by the fact that the portion of tip of $\mathbf{F_f}$ in class $\tilde{\mathbf{C}}_\mathbf{i}$ is more than that in class $\tilde{\mathbf{C}}_\mathbf{j}$. Again according to the same MFI, the class $\tilde{\mathbf{C}}_\mathbf{k}$ is having zero possibility of occurrence.

There are generally two basic approaches to the problems of pattern recognition, namely

1. Decision Theoretic Approach

2. Syntactic Approach .

Since the MFR approach to pattern recognition is similar to the decision theoretic approach, we have followed the latter religiously in our present work. In this approach each pattern is represented by a vector of features.The feature space is divided into a number of regions, each of which represents either a prototype pattern or a cluster of patterns. A decision function maps the given pattern to previously determined regions.

In the MFR approach to pattern recognition, each element of the fuzzy fea-

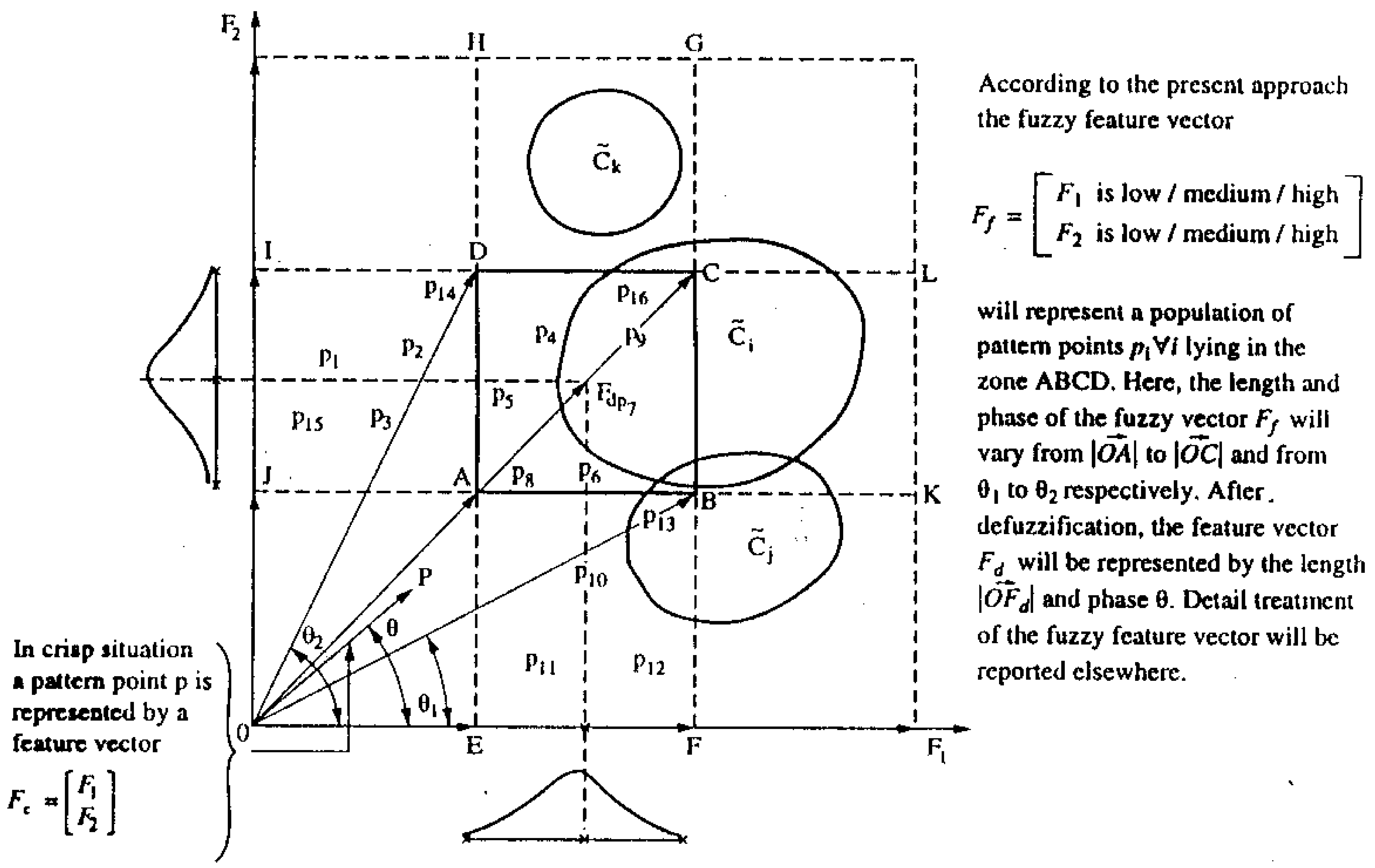

FIGURE 1. Representation of feature vectors F_c F_f and F_d.

ture vector $\mathbf{F_f}$, as mentioned earlier, is represented by a fuzzy linguistic variable instead of a real number.To cite an example, let $\mathbf{F} = (\mathbf{F_1}, \mathbf{F_2})^\mathbf{T}$, where ^{T} denotes transpose,F_1 represents first formant frequency and F_2 is the second formant frequency of a speech signal.In the decision theoretic approach, F_1 and F_2 are two features and are represented by crisp values like 500 Hz and 950 Hz, whereas in the MFR approach they are represented by fuzzy linguistic variables like F_1 is small and F_2 is medium.The elements of the feature vector($\mathbf{F_f}$) which are represented by fuzzy linguistic variables are characterized by their membership functions which fuzzily partition the pattern space as shown in fig 2. The antecedent parts of the DFI's are constituted by these elements of the feature vector $\mathbf{F_f}$. The consequent part of the MFI represents the possibility of occurrence of each class in the fuzzily partitioned pattern space. The multidimensional fuzzy **if then** rules are decomposed according to the new model, which then maps the given patterns to previously determined regions.

The process of recognition is usually divided into two steps, learning and then classification.The learning process involves feature extraction, feature selection, clustering and determination of the appropriate fuzzy **if then** rules which will constitute a decision function.The main stages in classification are extraction of a selected set of features application of the **if then** rules and decision making based on the result. Here, a set of predetermined features is extracted from an unknown pattern presented to the system. Then the **if then** rules are applied on the above extracted features to determine the possibility of occurrence of the pattern to various classes.

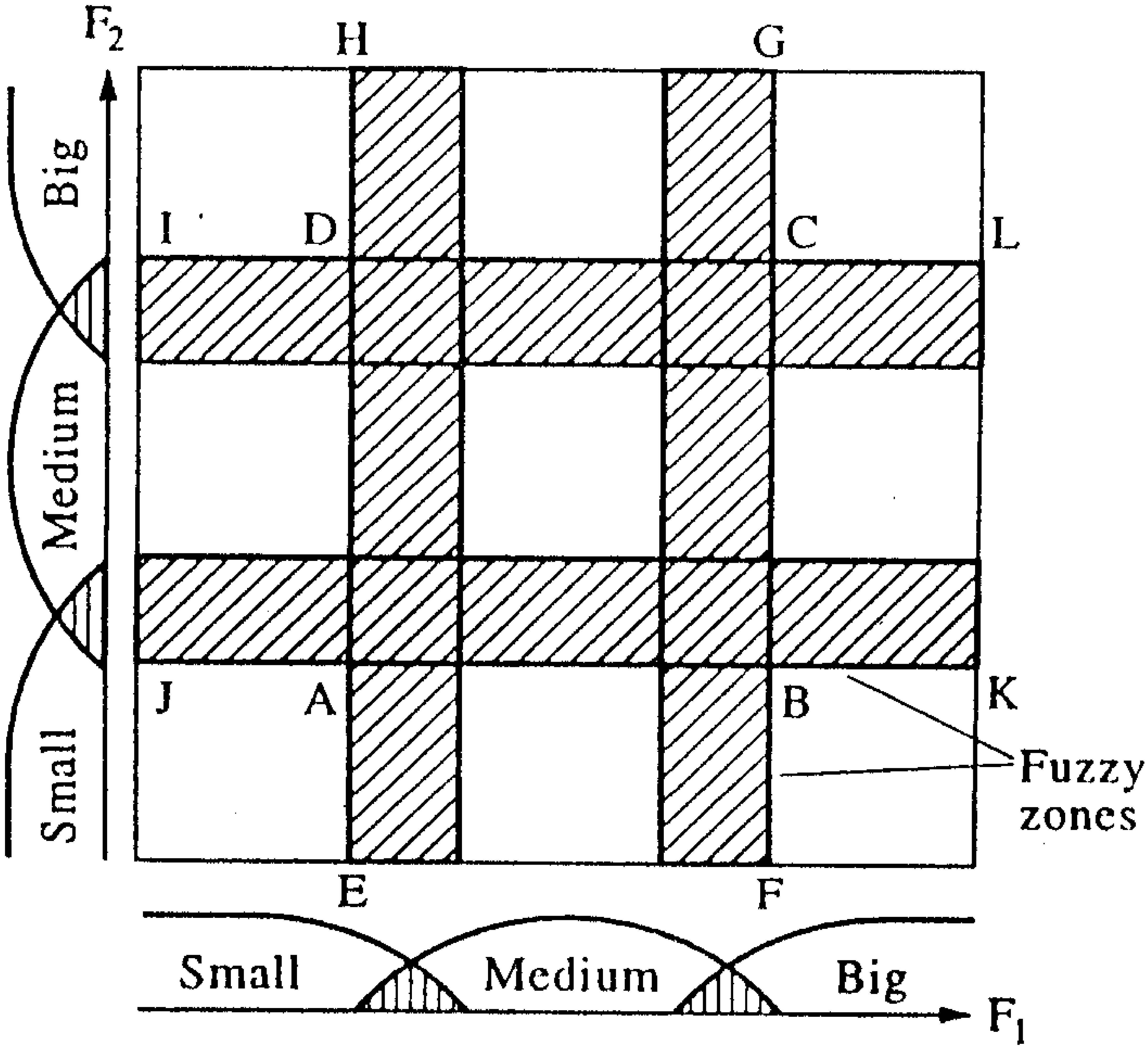

FIGURE 2. Fuzzy partitioning of pattern space.

3 BACK PROPAGATION TYPE NEURAL NETWORK AND IMPLEMENTATION OF THE NEW INTERPRETATION OF MFR

The invention of the backpropagation algorithm has played a large part in the resurgence of interest in artificial neural networks.Backpropagation is a systematic method for training artificial neural networks.It has a mathematical foundation that is strong if not highly practical.Despite its limitations, backpropagation has dramatically expanded the range of problems to which artificial neural networks can be successfully applied.

The fundamental building block of a back propagation neural networks is the neuron.A set of inputs is applied to it, either from the outside or from a previous layer.Each of these is multiplied by a weight, and the products are summed.The summation of the products is called **NET**, and must be calculated for each neuron in the network.After **NET** is calculated, a suitable activation function **F** is applied to modify it,thereby producing the signal **OUT**.

The objective of training the neural network is to adjust the weights so that application of a set of inputs produces the desired set of outputs (this combination of an input-output pair is called input vector).Before starting the training process, all the weights are initialized to small random numbers to ensure that the network is not saturated.Then the following 5-step feedback algorithm is applied.

1. Select the next training pair from the training set ; apply the input vectors to the network input.
2. Calculate the output of the network.
3. Calculate the error between the network output and the desired output.
4. Adjust the weights of the network in a way that minimizes the error.
5. Repeat steps 1 through 4 for each vector in the training set until the error for the entire set is acceptably low.

In the case of our newly proposed model of MFR, the first expression represents one law of implication while the second expression represents another law of implication.Both of these laws of implication can independently be realized through two separate backpropagation type neural networks, called the three layered perceptrons (consisting of input layer, hidden layer and output layer).

The input of each neural network is the antecedent part of the **if then** clause and is represented by a fuzzy membership function.The reference output of the network is the consequent part of the clause,which also is a fuzzy set and is represented by another fuzzy membership function.This indicates the possibility of occurrence of each class.

Each network is independently trained by a set of fuzzy **if then** statements using the previously mentioned algorithm.Once the training part for each individual network is over, we combine the outputs of both the networks by an intersection operator(fig 3).This finally results in another fuzzy set from which final inference is drawn as to the occurrence of each pattern in the pattern space.

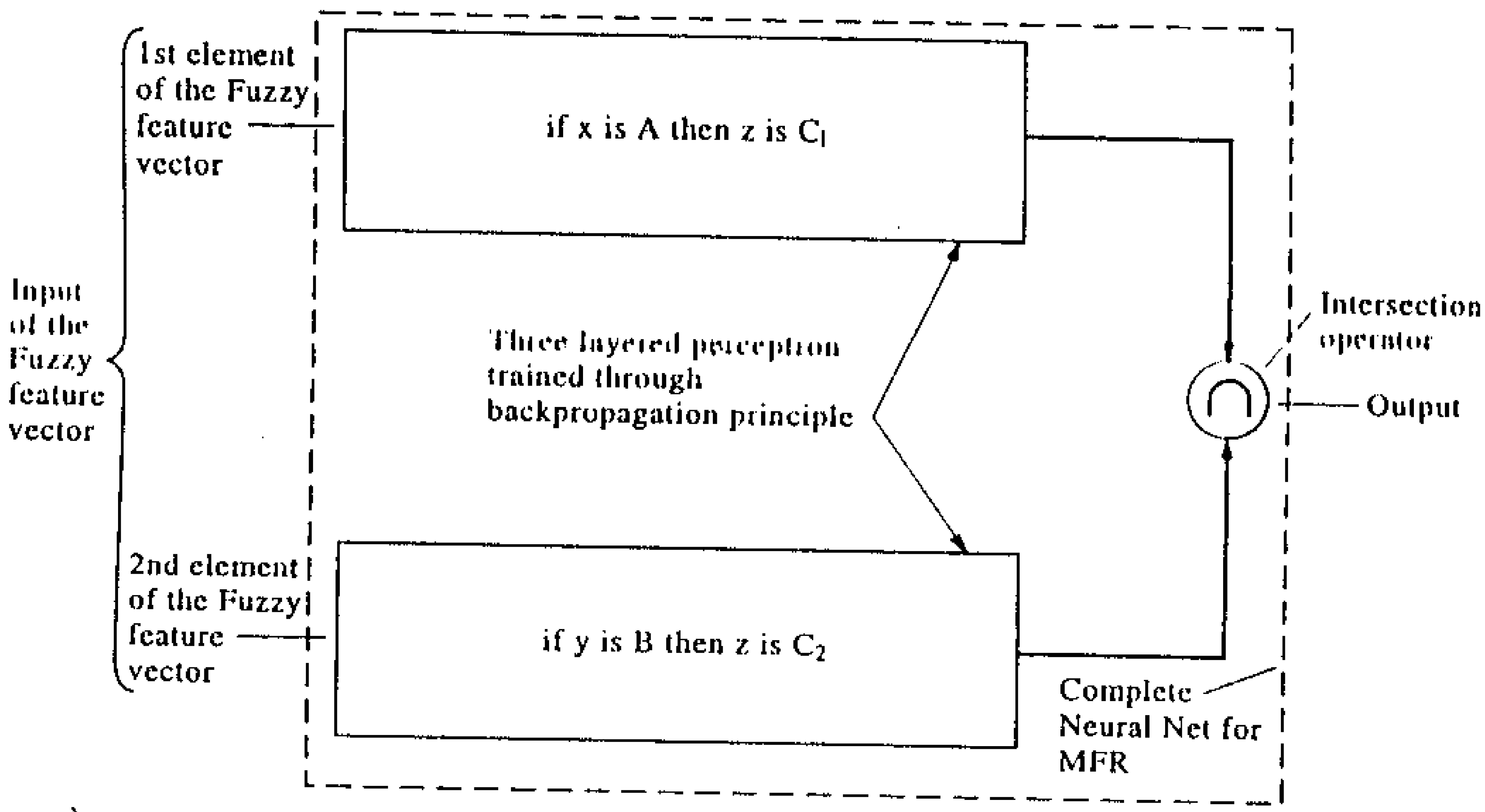

FIGURE 3. Realization of two-dimensional MFR through MLP type neural network.

4 FORMULATION OF THE PROBLEM

At the learning stage, first of all we discretize the universe of discourses of the features $\mathbf{F_1}$ and $\mathbf{F_2}$. This is known as quantization which quantizes a universe into a certain number of segments (known as quantization levels).Each segment is labelled as a generic element and forms a discrete universe.A fuzzy set is then defined for each of these generic elements by assigning appropriate grade of membership values. As already mentioned,the membership function of the consequent part of the if-then rule represents the possibility of occurrence of each class on the fuzzily partitioned feature space.At the learning stage of the classifier, we depend on an expert's knowledge which is captured through fuzzy **if then** rules.

Depending on the nature of the membership function,we fuzzily partition the feature space and generate the **if then** rules to classify the patterns.Then we test the validity of the rules by classifying some known patterns.If the classification results are satisfactory, we proceed further; otherwise we tune the rules by changing the grades of the membership function of the antecedent part and the consequent part.

It is worth mentioning here that we also need to map the entire feature space. Hence we can consider two additional classes g and **h** to represent respectively the empty portions in the bottom right corner and top left corner of the feature space(fig 4).

At the classification stage, the selected features are fuzzified using the concept of fuzzy singleton.The classification results produce the possibility of occurrence of each pattern at different fuzzily partitioned regions.At the time of taking a nonfuzzy decision out of this fuzzy classification, known as defuzzification, we can go by selecting the class having highest possibility value.In case of a controversial case, which happens normally in the fuzzy zones we have to state the equal possibility of a pattern belonging to more than one class.Such a conclusion is quite natural which normally does not exist in the conventional approach of classification.In some cases, patterns in the fuzzy zones are classified with "**almost equal**" possibility of occurrence for more than one class.

STATISTICAL I
C 23783
SEP
CALCUTTA

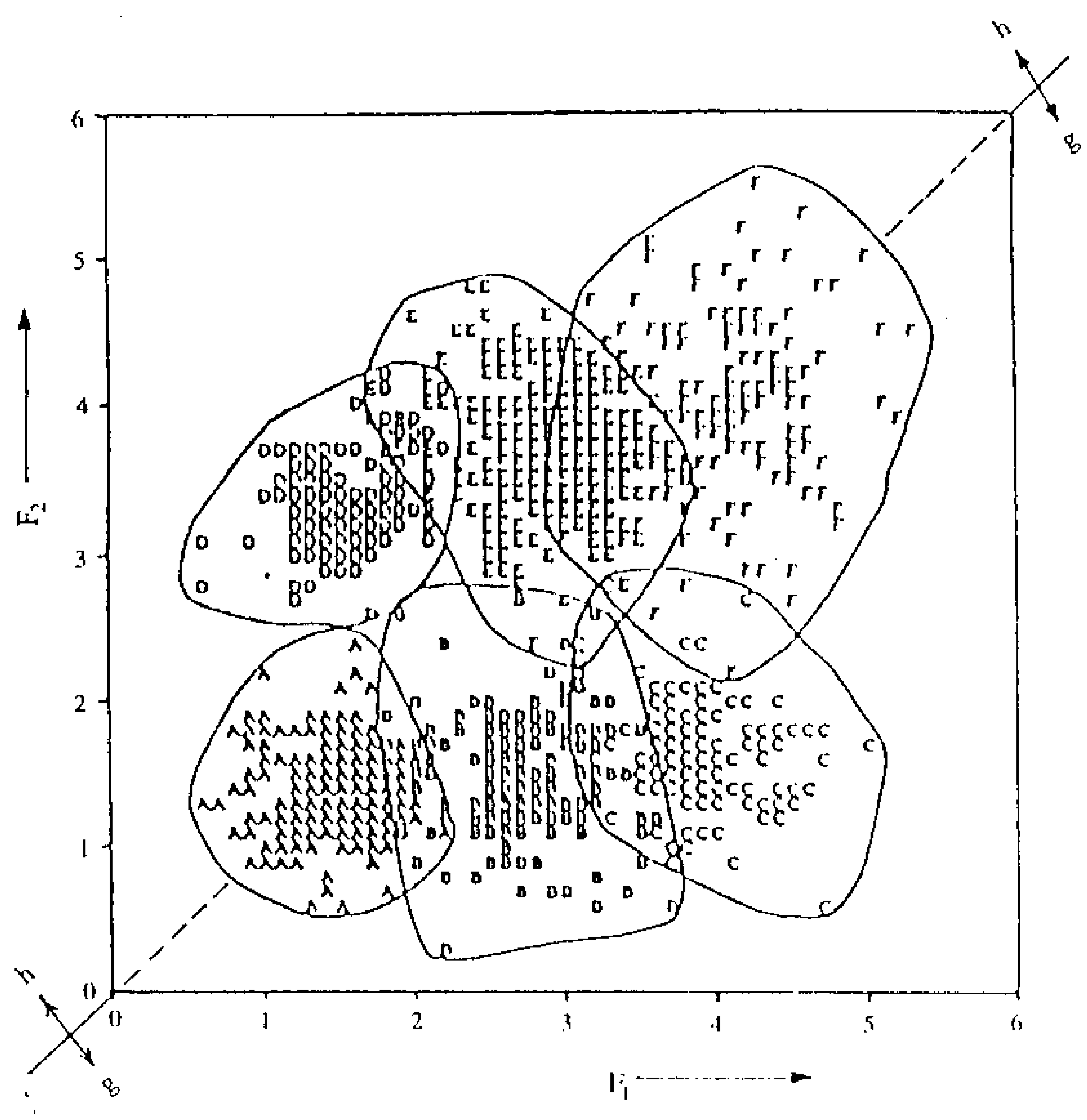

FIGURE 4 The first synthetic data in the F_1–F_2 plane.

5 EXPERIMENT WITH CERTAIN BENGALI WORDS

The experiment of isolated word recognition was conducted with the vowel patterns of certain carefully selected Bengali words taken from the utterances of 10 male and 10 female speakers who are consciously phonetically informed beings.Of all these records, data for three male informants have been considered on the basis of good and clear spectrographs.The age of the informants varies between 30 to 55 years.The analysis was carried out on Kay Sonograph model 7029-A .The bandwidth range was 80 Hz to 8 kHz;the resolution used was 300Hz. The acoustic data which were the first four formant frequencies and the duration was derived from the spectrographs.Since it was impossible to isolate the vowels from the embedded consonants, appropriate segmentation procedure was adopted to fix transitions as well as steady state of vowels.

We considered the distribution of $F_1 - F_2$ plane(fig 5)for recognition of Bengali vowel pattern.Quantization of Bengali features are done as shown in the following tables:1,2,3.The rules and the distribution patterns are shown in table-4.The result of contradistinction was quite satisfactory for these selected words.

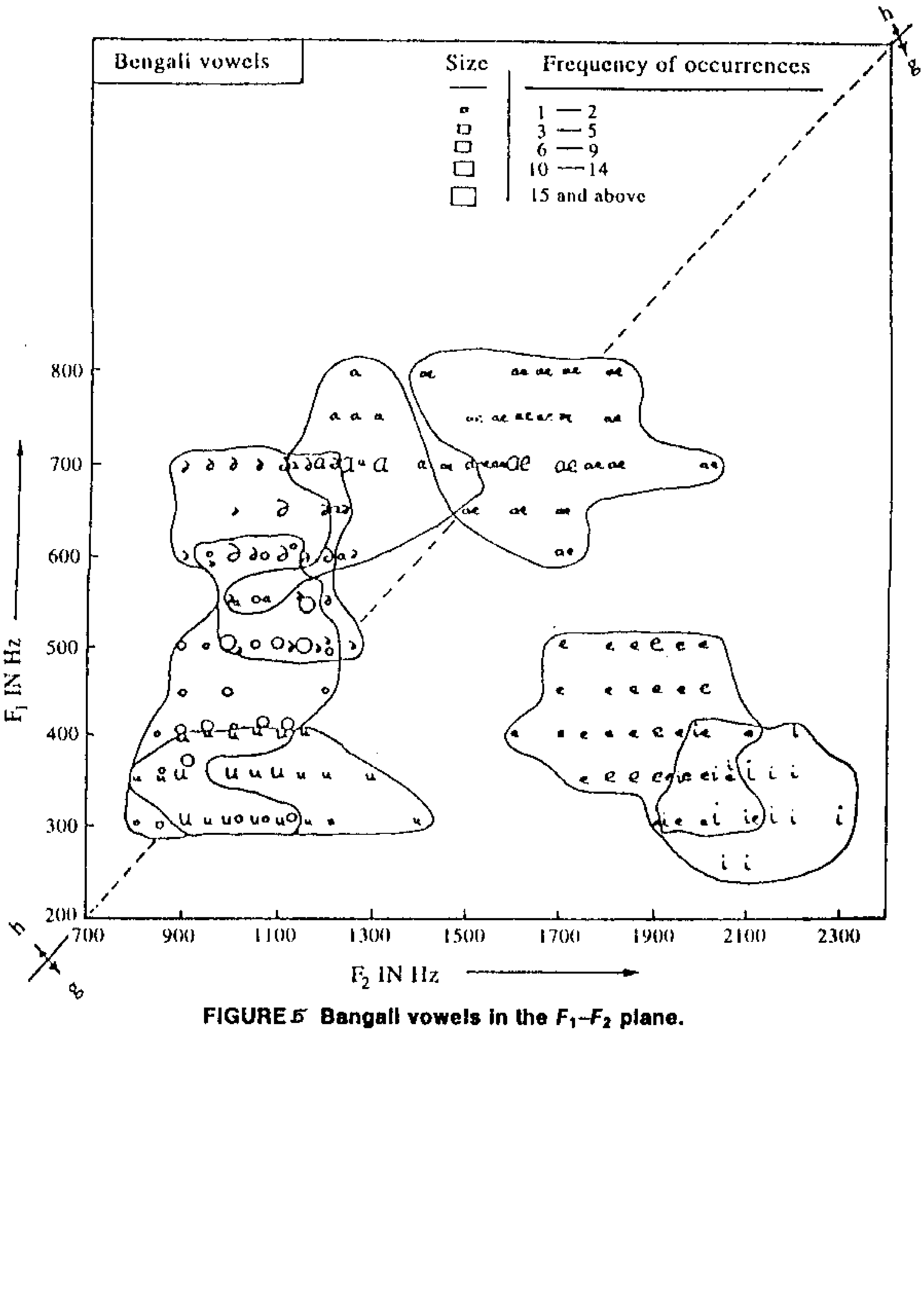

FIGURE 5 Bangali vowels in the F_1–F_2 plane.

TABLE 1 Average Formant Frequencies of Bengali Vowels

Phonetic symbol	F_1 Hz	F_2 Hz	F_3 Hz
/ *u* /	327	935	2198
/ *o* /	438	1015	2308
/ ə /	626	1095	2391
/ ə˞ /	695	1326	2424
/ *ae* /	681	1663	2320
/ *e* /	374	1935	2410
/ *i* /	304	2095	2565

TABLE 2 Quantization of the Feature Space for Feature 2 of Bengali Vowels

	Small	More or less Medium	Medium	Big
$700 \leqslant F_2 \leqslant 800$	1	0	0	0
$800 < F_2 \leqslant 900$	0.7	0	0	0
$900 < F_2 \leqslant 1000$	0.5	0.1	0	0
$1000 < F_2 \leqslant 1100$	0.3	0.5	0	0
$1100 < F_2 \leqslant 1200$	0.1	1	0	0
$1200 < F_2 \leqslant 1300$	0	0.5	0	0
$1300 < F_2 \leqslant 1400$	0	0.1	0.1	0
$1400 < F_2 \leqslant 1500$	0	0	0.3	0
$1500 < F_2 \leqslant 1600$	0	0	0.5	0
$1600 < F_2 \leqslant 1700$	0	0	0.7	0
$1700 < F_2 \leqslant 1800$	0	0	1	0
$1800 < F_2 \leqslant 1900$	0	0	0.5	0.1
$1900 < F_2 \leqslant 2000$	0	0	0.3	0.3
$2000 < F_2 \leqslant 2100$	0	0	0.1	0.5
$2100 < F_2 \leqslant 2200$	0	0	0	0.7
$2200 < F_2 \leqslant 2300$	0	0	0	1

TABLE 3 Quantization of the Feature Space for Feature 1 of Bengali Vowels

	Small	Medium	Big
$200 \leqslant F_1 \leqslant 250$	1	0	0
$250 < F_1 \leqslant 300$	0.7	0	0
$300 < F_1 \leqslant 350$	0.5	0	0
$350 < F_1 \leqslant 400$	0.3	0	0
$400 < F_1 \leqslant 450$	0.1	0.1	0
$450 < F_1 \leqslant 500$	0	0.5	0
$500 < F_1 \leqslant 550$	0	1	0
$550 < F_1 \leqslant 600$	0	0.5	0.1
$600 < F_1 \leqslant 650$	0	0.1	0.3
$650 < F_1 \leqslant 700$	0	0	0.5
$700 < F_1 \leqslant 750$	0	0	0.7
$750 < F_1 \leqslant 800$	0	0	1

TABLE 4 Results of Bengali Vowels of Figure 5

Number of Hidden nodes = 6 (for each 3 – layered neural network)
Rules used for learning the net

	Antedecent	Consequent								
		u	o	ə	a	ae	e	i	g	h
NN_1^1	If F_1 is small then	Very high	Medium	Nil	Nil	Nil	Medium	Very high	Very low	Very low
	If F_1 is medium then	Nil	Very high	Medium	Very low	Very low	More or less high	Nil	Very low	Very low
	If F_1 is big then	Nil	Very low	High	Very high	Very high	Nil	Nil	Very low	Very low
NN_2^2	If F_2 is small then	High	Very high	Very high	Very low	Nil	Nil	Nil	Very low	Very low
	If F_2 is more or less medium then	Very high	Very high	Very high	High	Very low	Nil	Nil	Very low	Very low
	If F_2 is medium then	Very low	Nil	Nil	Very low	Very high	High	Nil	Very low	Very low
	If F_2 is big then	Nil	Nil	Nil	Nil	Very low	More or less high	Very high	Very low	Very low

1 learned in 168 iterations

2 learned in 164 iterations

Recognition score

Hard partitioning

Class u	*Class o*	*Class ə*	*Class a*	*Class ae*	*Class e*	*Class i*	*Overall score*
74.6%	92.9%	32.8%	42.8%	76%	96.3%	63.9%	70.2%

Fuzzy partitioning

Class u	*Class o*	*Class ə*	*Class a*	*Class ae*	*Class e*	*Class i*	*Overall score*
100%	100%	94.3%	85.7%	76%	100%	91.8%	91.8%

6 CONCLUSION

The new interpretation of MFR to pattern recognition problems is found to be quite promising. Although presented for two dimensional case(R^2), the approach can easily be extended to problems in R^n, using similar technique for learning stage (through backpropagation type neural network) and with a similar fuzzy reasoning.

We also compare the performance of the newly proposed model with those of the existing classifiers like Bayesian classifier etc and in every case the comparative results are satisfactory.

7 REFERENCES

1. Franzini, M. A. (1987). Speech Recognition with backpropagation. Proceedings of 9th Annual Conference of the Engineering in Medicine and Biology Society.
2. Fukami, S., Mizumoto, M. & Tanaka, K. (1980). Some Considerations on Fuzzy Conditional Inference. Fuzzy Sets and Systems, 4, 243-273.
3. Zadeh, L. A. (1970). Theory of Approximate Reasoning.In J. E. Hayes, Donald Michie & L. I. Mikulich (Eds), Machine Intelligence(pp. 149-194). Chichester: Horwood.
4. Tsukamoto, Y. (1979). An Approach to fuzzy reasoning method. In M. M. Gupta, R. K. Ragade & R. R. Yager(Eds),Advances in fuzzy set theory and applications.
5. Takagi, H., & Hayashi, I (1991). NN-driven fuzzy reasoning. International Journal of Approximate Reasoning.

INDIAN STATISTICAL INSTITUTE
C 23783
4 SEP 1997
CALCUTTA