Pattern Recognition Vol. 12, pp. 27–34
Pergamon Press Ltd. 1980. Printed in Great Britain

0031-3203/80/0101-0027 $02.00/0

A SELF-SUPERVISED VOWEL RECOGNITION SYSTEM

S. K. PAL, A. K. DATTA and D. DUTTA MAJUMDER
Electronics and Communication Sciences Unit, Indian Statistical Institute,
Calcutta 700 035, India

(*Received* 4 *October* 1978; *in revised form* 14 *June* 1979)

Abstract – The paper describes an adaptive model for computer recognition of vowel sounds with the first three formants as features. The method uses a single pattern training procedure for self-supervised learning and maximum value of fuzzy membership function is the basis of recognition.

The algorithm with selected representative points and a number of guard zones which are ellipsoidal in the three-dimensional feature space around the representative vectors of the classes is taken as a supervisor. An optimum zone for self-supervised learning is found to correspond to half of the class variances beyond which the machine loses its efficiency. A comparison with a nonadaptive recognition system has also been included.

Formants Fuzzy membership function Adaptive classification Vowel recognition
Self-supervisor Optimum guard zone

1. INTRODUCTION

Automatic speech recognition (ASR) involves multilevel multi-class decision processes.[1–10] It covers a wide spectrum ranging from recognition of a vowel, consonant, or isolated word by a single speaker and a limited vocabulary system for a few trained speakers, to a connected speech recognition system with an unlimited vocabulary for a large number of speakers and speech comprehension systems.[8,11] Speech is biological in origin and transmits, over and above the semantic content of the message, information regarding the mood, health, age and sex of the speaker, as well as various other physiological realities and psychological states. The resulting pattern, therefore, manifests a considerable amount of fuzziness. Pattern classes do not exhibit precise boundaries due to inherent vagueness (fuzziness) rather than to randomness in the patterns. Again where the conditional densities of classes are not known and only a small number of design samples are available, a classifier based on similarity or dissimilarity measured within the framework of fuzzy language theory[12] appears to be suitable for their recognition.[13–15]

An adaptive system can be viewed as a learning machine in which the decision of the system gradually approaches the optimal decision by acquiring necessary information from observed patterns. System performance is improved as a result. In a supervised system, the machine requires an extra source of knowledge, usually of a higher order, for correcting the decision taken by the classifier. Bayesian estimation[16] and stochastic approximation[17] can be used for supervised learning to learn unknown parameters successively in a given form of distribution of each class. In a strictly non-supervised adaptive system, the parameters of classification are updated solely on the basis of the decision of the classifier. In some non-supervised machines, the problem of non-supervised learning is often reduced to a process of successive estimation of some unknown parameters in a mixture distribution of all possible pattern classes.[18,19] The convergence of the system to an optimal set of class representative parameters may thus be seriously affected by incorrect decisions of the classifier.[20,21] When an extra source of knowledge on which a supervisory programme could be based is not readily available, the performance of the system becomes completely unpredictable. In the present paper a system of self-supervision based on inherent properties of the class distribution of features is proposed where the convergence property of the system is assured.

The present paper also attempts to demonstrate the adaptive efficiency of the self-supervised recognition system with vowel sounds in CNC (Consonant–Vowel Nucleus–Consonant) context starting with arbitrary representative vectors different from true representatives of the population. The test set contains about 900 utterances by three different informants. A general purpose digital computer (Honeywell 400) was used for analysis.

2. THE MODEL FOR A SELF-SUPERVISED CLASSIFIER

The model for a self-supervisory system may be explained with reference to Fig. 1. The model uses a classifier based on 'maximum fuzzy membership function' which first measures the similarity between the different representative vectors and the input vector and then assigns the input pattern to the class for which the representative vector shows maximum similarity.

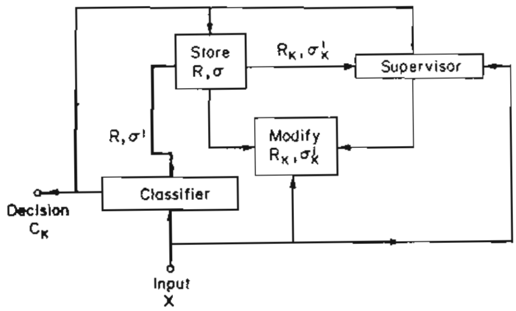

Fig. 1. Block diagram of an adaptive vowel recognition scheme.

In general, a supervisor depends on some extra source of knowledge to check the decision of the classifier for the purpose of updating the classification parameters. In the present model, the supervisor depends upon the basic nature of the class-distribution of the parameters. For all practical problems, the distribution of the members of a class in the feature space has a central tendency and it may be assumed that probability of misclassification near these central tendencies is substantially low. Thus one can construct a region around this central tendency of a class for which the probability of misclassification is so low that an unrestricted updating procedure for the samples falling only in this region would assist the convergence of the system significantly. Such a region, hereinafter referred to as a 'Guard Zone', forms the basis of a supervisory programme. The supervisory system, therefore, needs only to determine if the classified input is within the guard zone for the purpose of inhibition of the updating programme.

As this system uses some inherent properties of the distribution of the same parameters as used by the classifier itself, it may be called a 'self-supervisory' system. A mathematical formulation of such guard zones would require a thorough knowledge of the distribution function of the features for every class. When the distribution functions for the features are not precisely known the size of the guard zones has to be experimentally determined. Figure 2 may help to explain this process. Here the different zonal boundaries as shown by dotted ellipses are described around the selected representative vectors. Various guard zones, the semiaxes of which are respectively the $[1/\lambda]$th $[\lambda^2 = 1, 2, 4, 6, 8]$ part of the corresponding standard deviations $[\sigma]$, are considered for obtaining the optimum guard zone for self-supervised learning.

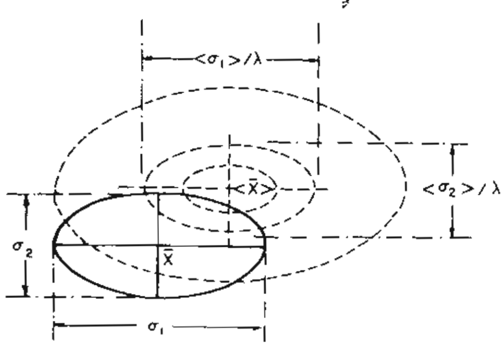

Fig. 2. Selection of representative vectors with the guard zones.

Initially the representative vectors and classification parameters are estimates made from a training sample. Accordingly, we cannot rule out a reasonable discrepancy between these factors and their true values. The model, therefore, should be able to correct a small amount of error in the initial estimates, but the proper convergence of the system should be ensured in order to prevent adverse effects on the classification performance. The system is then tested with a set of initial representative vectors which lie just outside the $\sigma/2$ boundaries of the true representatives.

2.1 *Formulation of the recognition system*

2.1.1 *Fuzzy sets.* A class of events $x_1, x_2, \ldots, x_n$ in the universe of the discourse U is defined as a fuzzy set $[A]$, if the transition of events from membership to non-membership is continuous rather than abrupt. $x_1, x_2, \ldots, x_n$ are called the supports of A at which the value of membership function $\mu_A(x)$ characterizing the grade of membership of x in A is positive ranging between zero and one. A fuzzy set A with its finite number of supports could therefore be viewed as

$$A = \{(\mu_A(x_i), x_i)\}$$

where the support of A which is the subset of U is defined as

$$S(A) = \{x, x \in U, 0 < \mu_A(x) \leqslant 1\}.$$

$\mu_A(x_i)$, being the grade of membership of x_i in A, denotes the degree to which an event x_i may be a member of or belong to A. As it approaches unity, the grade of membership of x_i in A becomes higher.

2.1.2 *Decision rule.* Consider an unknown pattern $X = [x_1, x_2, \ldots, x_n, \ldots, x_N]$ in an N-dimensional vector space Ω_X containing m ill-defined pattern classes, to be recognized with defined set of N-dimensional prototypes $R_1, R_2, \ldots, R_j, \ldots, R_m$ such that

$$R_j^{(l)} \in R_j.$$

$l = 1, 2, \ldots, h_j, h_j$ is the number of reference vectors in set R_j.

The decision of the classifier is based on the magnitude of the fuzzy membership function of the pattern x corresponding to jth $[j = 1, 2, \ldots, m]$ class defined by.[14]

$$\mu_j(X) = \left[1 + \left[\frac{d(X, R_j)}{F_d}\right]^{F_e}\right]^{-1} \tag{1a}$$

where F_e is the exponential fuzzifier
F_d is the denominational fuzzifier

and

$$d(X, R_j) = \min_l \| X - R_j^{(l)} \| \tag{1b}$$

with

$$\|X - R_j^{(l)}\| = \left[\sum_n \left[\frac{x_n - \bar{x}_{jn}^{(l)}}{\sigma_{jn}^{(l)}} \right]^2 \right]^{0.5} \qquad (1c)$$

denotes the weighted Euclidean distance[22,23] between an unknown pattern X and lth reference vector $R_j^{(l)}$ in jth class in which $\bar{x}_{jn}^{(l)}$ and $\sigma_{jn}^{(l)}$ correspond to lth prototype and denote the mean and standard deviation of the features along nth coordinate in jth class.

The fuzzifiers have the effect of altering the ambiguity in a set and hence the overall recognition score.[13-15] The membership function is defined in such a way that it maps N-dimensional feature space into an m-dimensional membership space which is a unit hypercube and should satisfy the following conditions:

(i) $\mu_j(X) \to 0$ as $d(X, R_j) \to \infty$

(ii) $\mu_j(X) \to 1$ $d(X, R_j) \to 0$

(iii) $\mu_j(X)$ increases $d(X, R_j)$ decreases.

Therefore the membership function, $\mu_j(X)$, having a positive value in the interval [0, 1] denotes the degree to which an event X may be a member of or belongs to jth class and the classificatory decision rule would be as follows:

decide: $X \in C_k$ if $\mu_k(X) > \mu_j(X)$

$j = k = 1, 2, \ldots, m, \quad j \neq k$

2.2 *Iterative algorithm for parameter estimation*

The components of reference vector and weight vector for a class used in the decisional algorithm are, respectively, the mean and reciprocal of the standard deviations of the components of the feature vectors of the training sample. The reciprocal of the standard deviation is found to provide appropriate phase weights to patterns for their proper classification.[13-15,24]

The basic idea of the recognition system is to determine the classes of the unknown samples that appear before it one after another. Generally, the input events are in a somewhat randomly mixed sequence. These samples after being classified become members of certain classes and modify the centres and the relative weight vectors of them. The adaptation consists of, in this case, the updating of the components of the mean and the standard deviation vectors.

If $\bar{x}_{n(t)}$ and $\sigma_{n(t)}^2$ represent the mean and variance of a class along nth co-ordinate axis, estimated by first t samples, we note

$$\bar{x}_{n(t)} = \frac{1}{t} \Sigma x_i \qquad (2a)$$

and

$$\sigma_{n(t)}^2 = \frac{1}{t} \Sigma (x_i - \bar{x}_{n(t)})^2$$

$$= \frac{1}{t} \Sigma x_i^2 - \bar{x}_{n(t)}^2, \qquad i = 1, 2, \ldots, t$$

$$= \frac{1}{t} C_{n(t)} - \bar{x}_{n(t)}^2 \qquad (2b)$$

where

$$C_{n(t)} = \Sigma x_i^2, \qquad i = 1, 2, \ldots, t \qquad (2c)$$

Let another sample $x_{(t+1)}$ fall in this class. Then mean and variance are adjusted as follows:

$$\bar{x}_{n(t+1)} = \frac{t}{t+1} \bar{x}_{n(t)} + \frac{1}{t+1} x_{(t+1)} \qquad (3a)$$

$$C_{n(t+1)} = C_{n(t)} + x_{(t+1)}^2 \qquad (3b)$$

$$\sigma_{n(t+1)}^2 = \frac{1}{t+1} C_{n(t+1)} - \bar{x}_{n(t+1)}^2. \qquad (3c)$$

All these equations (3) provide us with an iterative algorithm for the estimation of the mean and variance vectors, given successive samples.

2.3 *Algorithm for a self-supervised learning scheme*

The basic concept of the self-supervised system has been explained earlier. The control of the supervisor is based on the 'decision parameter of the supervisor' [DPS], which, in its turn, is dependent on the size of the guard zone of that class. [DPS] for the jth class is defined by

$$(\mathrm{DPS})_j = \sum_n \left[\frac{x_n - \bar{x}_{jn}}{\sigma'_{jn}} \right]^2 \qquad (4)$$

where

$$\sigma'_{jn} = \sigma_{jn}/\lambda$$

$$n = 1, 2, \ldots, N$$

and λ (a positive constant) is termed as 'zone-controlling parameter' which controls the dimension of the guard zone in N-dimensional vector space $\mathbf{\Omega}_X$.

The supervisor then accepts the decision made by the classifier that $X \in C_k$ only if,

$$(\mathrm{DPS})_k \leqslant 1$$

and the mean and variance vectors of the kth class are correspondingly updated for that input sample X. Otherwise the decision is thought to be doubtful and no other alteration of the class parameters is made. It is to be noted here that the decision parameter would lead to ellipsoidal shapes of the guard zones.

3. EXPERIMENTAL PROCEDURE

3.1 *Data collection*

A vocabulary consisting of Telugu words was selected so as to encompass as many CN and NC combinations as possible with an emphasis on the use of commonly used words. These were recorded by five adult male speakers on an AKAI tape recorder in a large auditorium. On the basis of a listening experiment by 10 listeners, only three speakers, denoted X, Y, Z were selected. We performed a spectrographic analysis of these utterances on a Kay Sonagraph

Model 7029A. The analyses were carried out in the normal mode, using the band 80 Hz–8 KHz with wide band filters having bandwidth 300 Hz.

Formant frequencies F_1, F_2 and F_3 were obtained manually at the steady state of the vowels. Altogether 871 samples were collected. In view of the large amount of data to be handled, the formant frequencies were measured from the base line with a specially constructed scale. Rechecking on 5% of the samples revealed that formant frequencies had been recorded within an accuracy of 10 Hz. Occasionally, steady states were not observed due to the extreme shortness of the vowels. Here measurements were made at the point of congruence of the off-glide and on-glide. The samples which did not depict a prominent third formant were allowed to have an injected average third formant $(F_3)_{av}$, computed over all members of that class of vowels for the particular speaker. The number of samples which fell in this category was 384.

The measured values of the three features F_1, F_2 and F_3 were therefore thought to constitute a three dimensional feature space Ω_X. The significant information available about the event could thus be expressed as a three-dimensional feature vector, $X = [F_1, F_2, F_3]$, $X \in \Omega_X$. The co-ordinates of X would have numerical values indicating the amount of each property of the event.

3.2 *Class definition*

Of the ten Telugu vowels /ə, *a* :, *i*, *i* :, *u*, *u* :, *o*, *o* :, *e*, *e* :/, long and short categories – viz., /*i*, *i* :/, /*u*, *u* :/, /*e*, *e* :/, and /*o*, *o* :/ – were found to differ from one another mainly in duration, while phonetically they are not distinctively different. With this background, the number of pattern classes to be recognized is reduced to six, namely, /ə/, /*a* :/, /*I*/, /*U*/, /*E*/ and /*O*/, which are phonetically different from one another. The feature space so designed could therefore be viewed as constituted by various pattern classes C_1, C_2, C_3, and C_{3l}, C_{4s} and C_{4l}, C_{5s} and C_{5l}, and C_{6s} and C_{6l}, where subscripts s and l stand for shorter and longer categories. Through classification, this three-dimensional feature space is to be divided into six such regions which contain vowels differing only in phonetic features. Some of these regions would contain two subregions, one for the short vowels and another for the longer ones.

3.3 *Representative vector selection*

The next task before classification is, obviously, the selection of reference vectors or 'prototypes' denoting the representative points of each class. It may be noted here that, altogether, ten representative points are to be selected for the ten vowels, though the number of classes for the purpose of classification is only six. Some classes, as we explained above, have two representative points. Now we are interested here in studying the recognition of vowel sounds on the adaptive classification basis with the non-appropriate prototype vectors representing the classes. Similar investigations have already been reported[25] which used non-adaptive, supervised, and non-supervised procedures where the prototype points and corresponding weighting coefficients of a specified class were obtained from five utterances of a single speaker Z selected randomly from each of the categories. In the present experiment, the representative vector of a vowel class was chosen just outside the boundary of an ellipsoid having the three axes equal to the respective standard deviations of the features and mean of the classes as the centre. This is explained in Fig. 2 where <.> denotes estimated value. The standard deviations for defining weighting coefficients corresponding to these representative points were obtained from a specified training set of samples selected randomly from the classes.

4. EXPERIMENTAL RESULTS

The distribution of the samples uttered by three speakers in the F_1–F_2 plane of the vector space is sketched in Fig. 3, where the boundaries among the vowel classes are seen to be ill-defined. Figure 4 represents the position of the mean vectors, together with the standard deviations for all the classes in the three-dimensional feature space. The length of the arrows gives the magnitudes of the respective components of the standard deviation in the same scale. Table 1 shows the components of the selected prototypes and the corresponding weight vectors. These constitute the initial parameter of the recognition system.

Figure 5 shows the flow chart for the complete recognition scheme. The right hand part of the figure contains the algorithms for updating the parameters and the decision-making programmes. Segment switches can convert the scheme into one of the three modes, namely, self-supervised, non-supervised and non-adaptive. In the non-adaptive mode both the updating procedure and the decision-making procedure are shunted. In the non-supervised mode only the decision procedure is shunted. The non-adaptive scheme with fixed mean and weight vectors is presented to demonstrate the efficiency of system adaptation to the new input patterns. The block diagram of the supervisor using self-supervised learning schemes based on guard zones is explained in Fig. 6.

Though the longer and shorter varieties for vowels /*I*/, /*U*/, /*E*/ and /*O*/ are pooled together for the purpose of classification, they were given individual reference vectors and weight vectors computed over the respective set of training samples. Thus in the present experiment, $m = 6$, $N = 3$, $h = 1$ for /ə/ and /*a* :/ and $h = 2$ for /*I*/, /*U*/, /*E*/ and /*O*/. Membership values of an input utterance are computed with respect to all the classes. The input is assigned to kth ($k = 1, 2, \ldots, 6$) class if the membership value of the kth class i.e., μ_k is maximum.

The system started to classify the samples in a random sequence from the sample space Ω_X with the initial set of parameters shown in Table 1. The effect of

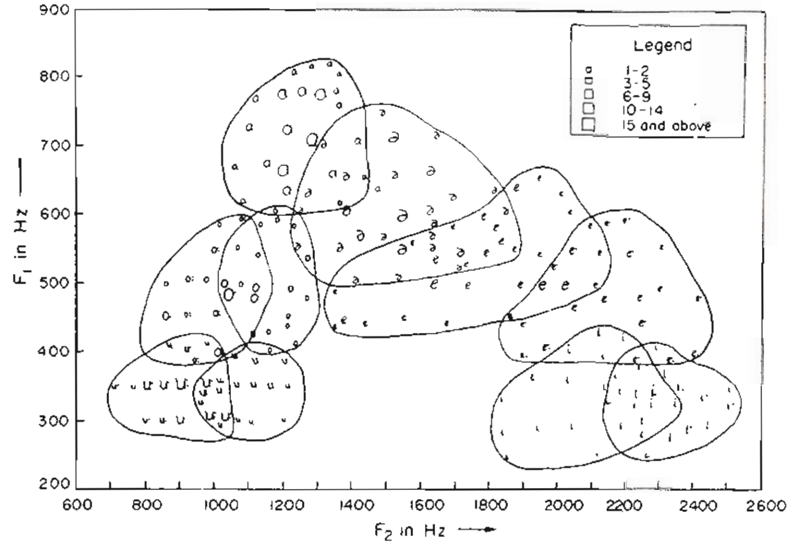

Fig. 3. Distribution of Telugu vowels in F_1–F_2 plane.

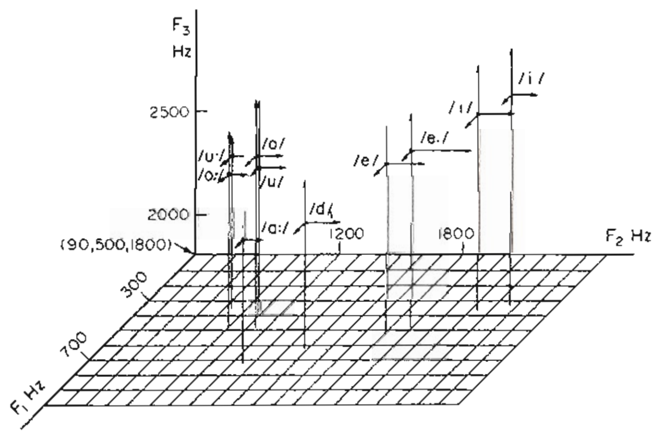

Fig. 4. Location of the mean vectors with standard deviations of the vowel classes in the formant $[F_1, F_2, F_3]$ space.

Table 1. Initial values of mean and standard deviation vectors

	Mean			Standard deviation		
	F_1 Hz	F_2 Hz	F_3 Hz	F_1 Hz	F_2 Hz	F_3 Hz
ə	650	1535	2396	86.483	214.264	154.171
a:	715	1180	2362	79.057	73.951	191.621
i	280	2100	2795	50.000	82.916	129.904
i:	300	2220	2732	40.169	105.185	186.865
u	340	1050	2437	47.404	108.333	203.485
u:	320	865	2532	33.015	72.281	112.164
e	530	2150	2591	64.877	130.466	160.291
e:	550	1965	2725	72.572	397.156	175.076
o	500	1065	2561	72.618	101.358	217.740
o:	525	965	2570	40.000	50.990	227.156

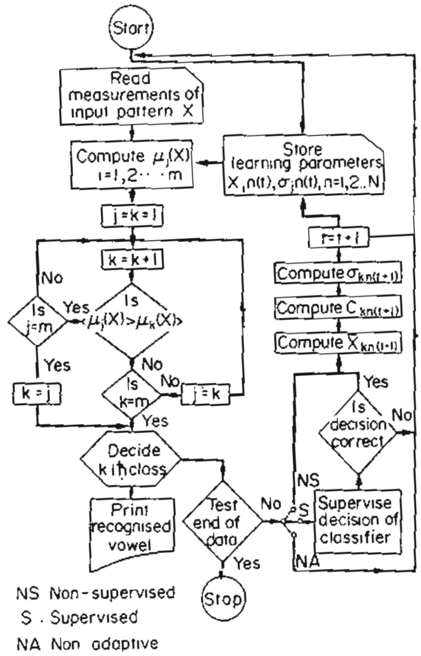

Fig. 5. A flow chart for the recognition system.

the zone controlling parameter λ on classification may be observed in Fig. 7. In this figure the rate of correct classification after every 100 input samples were plotted for each value of λ. As λ increases, the dimension of the guard zones decreases and the corresponding DPS-values increase. Therefore, the chance of correct samples correcting the representative vectors decreases, and the system performance accordingly approaches the non-adaptive recognition case. With the decrease in the value of λ, the zone boundaries, on the other hand, increase, making for a reduction in the DPS-values. The system then behaves more like the non-supervised recognition algorithm, where the chance of wrong samples vitiating the representative vectors increases. It may therefore be expected that for some optimum value of λ, corresponding to a specific size of the guard zone, the performance of the self-supervised system would approach that of a fully supervised system.

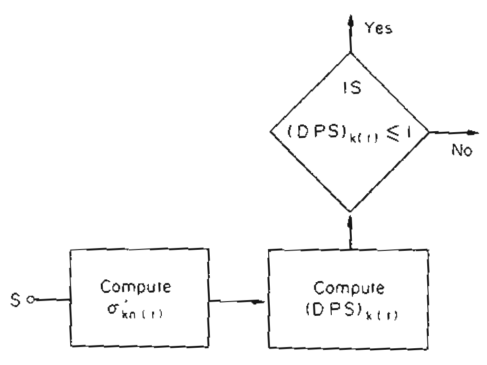

Fig. 6. Block diagram of supervisor for self-supervised learning algorithm.

In Fig. 8, the results obtained using self-supervised learning algorithms are compared with those obtained with non-adaptive and fully-supervised learning algorithms. In fully-supervised learning, the decision of the classifier is verified by an external supervisor and class parameters are altered only if the classification is found to be correct. Otherwise no alteration of the representative and weight vectors was made.

The mean square distance at every instance of the performance curves with different values of λ from that of fully-supervised system is tabulated in Table 2. The curve corresponding to $\lambda=2$ shows the best match to the performance of the fully-supervised system. The classifier with guard zones corresponding to $\lambda=2$ has been found to avail the highest proportion of correct to incorrect samples so that after the several utterances being dealt with by the classifier, the class representative and weight vectors are likely to approach their respective true values.

An average recognition score of about 80% was obtained. The confusion matrix (Table 3) in general, conforms to the vowel distribution (Fig. 3) in the F_1–F_2 plane. It is interesting to note that highest recognition scores of about 89% and 76% were respectively obtained for vowels $/U/$ and $/I/$ which exhibit least overlapping in the F_1–F_2 plane. Further confusion is observed to lie mainly amongst the groups which are adjacent to each other in the vowel diagram.

5. CONCLUSIONS

The model of a self-supervised learning algorithm with a classifier using the fuzzy set theoretical concept has been implemented to the real problems of vowel recognition. The experimental results corroborate the theoretical postulates that such a system would basically approach the supervised learning algorithms in so far as the convergence properties are concerned. The system has been found to approach the performance of a fully supervised system.

System performance for different guard zones selected around the initial representative vectors is studied. With the shrinkage of the zone boundaries, the system behaves like a non-adaptive recognition system, whereas the non-supervised performance is approached by relaxing the boundaries. Optimum results are obtained when the semiaxes of the guard zones defined for the classes correspond to one half the standard deviations along the respective coordinate axes. The confusion in machine performance is seen to be restricted only to neighbouring classes.

6. SUMMARY

The present paper describes an adaptive self-supervised recognition scheme based on fuzzy set theoretic approach. The scheme is tested with a data set of 871 Telugu [an important Indian Language] vowels in CNC [Consonant-Vowel Nucleus-

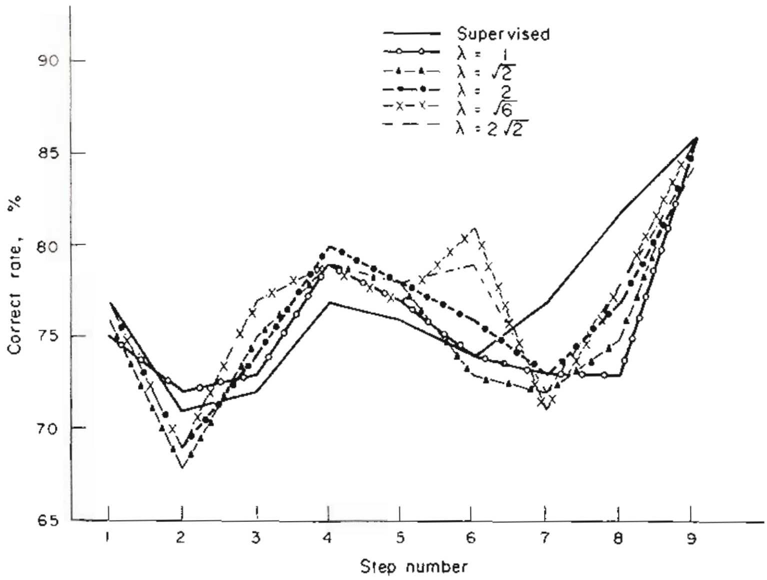

Fig. 7. System performance curve for different values of λ.

Consonant] context spoken by three informants. The basic classification uses fuzzy membership values of the input points with reference to previously chosen representative vectors for each class. The membership function uses weighted Euclidean distance with both exponential and denominational fuzzy generators for creating ambiguity in a set.

The representative vectors are deliberately chosen to be different from the appropriate representative vectors to study the extent of an initial incorrect choice which can be tolerated by the system without any significant loss in ultimate performance. In fact, the system performance has been found to be quite satisfactory even if these vectors have been chosen outside the $\sigma/2$ (σ is the standard deviation) boundary from the mean vectors.

For the purpose of supervision it is assumed here that for an input vector falling within a restricted zone, to be called "guard zone", the probability of its being misclassified is so low that it would not affect the convergence property of the system in any significant way. The supervisory system, therefore, needs only to check whether the classified input is within the guard zone or not for the purpose of inhibition of the updating programme. Various guard zones have been tested to obtain the optimal zone. A comparison of this scheme with the non-adaptive and fully supervised scheme, which uses an extra higher level of knowledge, is also included.

The first three formants of the steady state of the vowels have been taken as the features for classification. These formant frequencies were extracted from the spectragrams done on the spectrum analyser, Kay Sonagraph model 7029A. There are ten sets of Telugu vowels including shorter and longer categories. Through classification this 3-dimensional feature

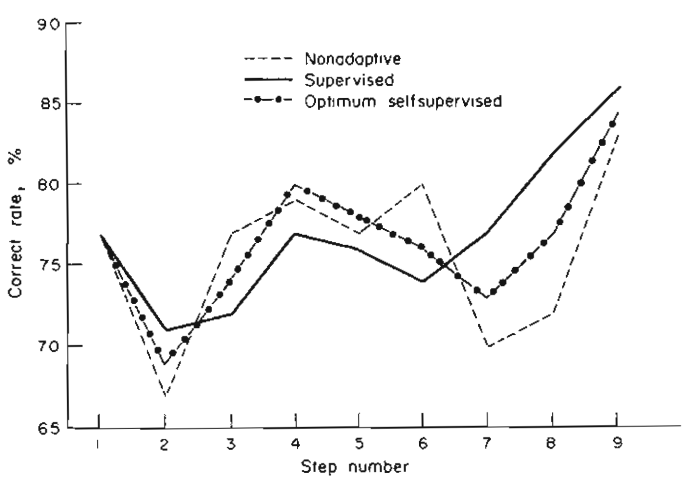

Fig. 8. Curves showing comparative performances of the non-adaptive, self-supervised and fully-supervised system.

Table 2. Mean square distance from the fully-supervised performance curve for different values of λ^2

λ^2	1	2	4	6	8
Mean square distance	12.0	11.33	7.58	11.69	15.0

Table 3. Confusion matrix for vowel recognition with $\lambda = 2$

		Actual class					
		I	*E*	ə	*a*:	*O*	*U*
Recognized as	*I*	131	2				
	E	37	146	9		5	
	ə	4	50	54	22	2	2
	a:			5	63	5	
	O		4	3	4	135	14
	U		5	1		33	135

space is divided into six such regions which contain vowels differing only in phonetic features. Some of these regions would contain two subregions, one of the short and another for long vowels which differ only in duration.

An average recognition score of about 80% has been obtained. As the dimension of the guard zones decreases, the system is found to approach the non-adaptive case. The system approaches a non-supervised adaptive case as the guard zone increases. An optimum value of the guard zone, corresponding to the $\sigma/2$ value of the axes of the ellipsoidal zones, has been found for which the system approaches a fully supervised case.

Acknowledgement – The authors gratefully acknowledge the valuable help rendered by Messrs N. R. Ganguli, B. Mookherjee, S. Ray, S. Chakravarti and Mrs S. DeBhowmik.

REFERENCES

1. D. R. Reddy (ed.), *Speech Recognition: Invited papers of the IEEE Symp.*, Academic Press, New York (1975).
2. D. R. Reddy, Speech Recognition by Machine: A review, *Proc. IEEE*, **64**, (4), 501 (1976).
3. N. Lindgren, Machine recognition of human language, *IEEE Spectrum*, 2 Mar., Apr., May (1965).
4. N. Lindgren, Speech—Man's natural communication, *IEEE Spectrum*, 4 June, p. 75 (1967).
5. D. H. Klatt and K. N. Stevens, On the automatic recognition of continuous speech: Implications from spectrogram reading experiment, *IEEE Trans. Audio Electroacoust.*, **AU-21**, 210 (1973).
6. C. J. Weinstein, S. S. McCandles, L. F. Mondshein and V. W. Zue, A system for acoustic–phonetic analysis of continuous speech, *IEEE Trans. Acoust., Speech. Sig. Process.*, **ASSP-23**, 54 (1975).
7. D. Dutta Majumder and A. K. Datta, An analyser coder for machine recognition of human speech, *J. Inst. Electron. Telecom. Engrs.*, **15**, 233 (1969).
8. V. V. S. Sharma and B. Yegnanarayana, Pattern recognition in speech research, *Proc. All India Interdiscp. symp. Dig. Tech. and Patt. Recog.*, Calcutta, ISI., 15–17 Feb., 1977 (to appear).
9. *Trans. IEEE., on Acoust., Speech and Sig. Process.*, Special Issue on IEEE Symp. on Speech recognition, **ASSP-23**, 1 (1975).
10. *Proc. IEEE.*, Special Issue on Man-Machine communication by voice, **64**, 401 (1976).
11. W. A. Wood and J. Makhoul, Mechanical Inference Problems in continuous speech understanding, *Artificial Intelligence*, **5**, 73 (1974).
12. L. A. Zadeh, K. S. Fu, K. Tanaka and M. Shimura (eds.), *Fuzzy sets and their application to Cognitive and decision processes*, Academic Press, New York (1975).
13. D. Dutta Majumder and S. K. Pal, On Fuzzification, Fuzzy Language and Multicategory Fuzzy Classifier, *Proc. IEEE Int. Conf. on Cybernetics and Society*, Washington, D.C., 19–21 Sept., 1977 (to appear).
14. S. K. Pal and D. Dutta Majumder, Fuzzy Sets and Decisionmaking approaches in vowel and speaker recognition, *IEEE Trans. Syst., Man and Cyberns.*, **SMC-7**, 625 (1977).
15. S. K. Pal and D. Dutta Majumder, On plosive Identification using fuzziness in property sets, *IEEE Trans. Syst., Man and Cynberns.*, **SMC-8**, 302 (1978).
16. N. Abramson and D. Braverman, Learning to Recognise Patterns in a random Environment, *IRE Trans. Inform. Theory*, **8**, 558 (1962).
17. K. S. Fu, Y. T. Chien, Z. J. Nikolic and W. C. Wee, On the Stochastic Approximation and related Learning Techniques, *Tech. Rept. TR-EE* 66-6, Purdue University, Lafayette, Indiana, April 1966.
18. E. A. Patric and J. C. Hancock, Nonsupervised Sequential Classification and Recognition of Patterns, *IEEE Trans. Inform. Theory*, **IT-12**, 362 (1966).
19. Y. T. Chien and K. S. Fu, On Bayesian Learning and Stochastic Approximation, *IEEE Trans. Syst. Man and Cyberns.* **SMC-3**, 28 (1967).
20. K. S. Fu (ed.), *Pattern Recognition and Machine Learning*, Plenum Press, New York (1971).
21. J. M. Mendel and K. S. Fu (eds.), *Adaptive, Learning, and Pattern Recognition Systems – Theory and Applications*, Academic Press, New York (1970).
22. G. S. Sebestyen, *Decision Making Processes in Pattern Recognition*, The Macmillan Co., New York (1962).
23. W. S. Meisel, *Computer-Oriented approaches to pattern recognition*, Academic Press, New York (1972).
24. D. Dutta Majumder, A. K. Datta and S. K. Pal, Computer recognition of Telugu vowel sounds, *J. Comp. Soc. India*, **7**, 14 (1976).
25. S. K. Pal, A. K. Datta and D. Dutta Majumder, Adaptive Learning Algorithm in classification of Fuzzy Patterns: An application to vowels in CNC context, *Int. J. Syst. Sciences*, **9**, 887 (1978).