Fuzzy Set Theoretic Measure for Automatic Feature Evaluation

SANKAR K. PAL, SENIOR MEMBER, IEEE, AND BASABI CHAKRABORTY, STUDENT MEMBER, IEEE

Abstract—The terms *index of fuzziness*, *entropy*, and *π-ness*, which give measures of fuzziness in a set, are used to define an index of feature

Manuscript received April 5, 1984; revised August 27, 1985 and April 21, 1986.

The authors are with the Electronics and Communication Science Unit, Indian Statistical Institute, 203 Barrackpore Trunk Road, Calcutta 700035, India.

IEEE Log Number 8609835.

0018-9472/86/0900-0754$01.00

evaluation in pattern recognition problems in terms of their intraclass and interclass measures. The index value decreases as the reliability of a feature in characterizing and discriminating different classes increases. The algorithm developed has been implemented in cases of vowel and plosive identification problem using formant frequencies and different S and π membership functions.

I. Introduction

The process of selecting the necessary information to present to the decision rule is called *feature selection*. Its main objective is to retain the optimum salient characteristics necessary for the recognition process and to reduce the dimensionality of the measurement space so that effective and easily computable algorithms can be devised for efficient classification.

The criterion of a good feature is that it should be unchanging with any other possible variation within a class, while emphasizing differences that are important in discriminating between patterns of different types. One of the useful techniques to achieve this is clustering transformation [1]–[3], which maximizes/minimizes the interset/intraset distance using a diagonal transformation, such that smaller weights are given to features having larger variance (less reliable). Other separability measures based on information theoretic approach include divergence, Bhattacharyya coefficient, and the Kolmogorov variational distance [1]–[7].

The present work demonstrates an application of the theory of fuzzy sets to the problem of evaluating feature quality. The terms *index of fuzziness* [8], *entropy* [9], and *π-ness* [10] provide measures of fuzziness in a set and are used here to define the measure of separability in terms of their interclass and intraclass measurements. These two types of measurements are found to reflect the concept of interset and intraset distances in classical set theory. An index of feature evaluation is then defined using these measures such that the lower the value of the index for a feature, the greater is the importance (quality) of the feature in recognizing and separating classes in the feature space.

It is also to be mentioned here that the above parameters provide algorithms for automatic segmentation [11] of grey tone image and measuring enhancement quality [12] of an image.

Effectiveness of the algorithm is demonstrated on vowel, plosive consonant, and speaker recognition problems using formant frequencies and their different combinations as feature set and S and π functions [13]–[15] as membership functions.

II. Fuzzy Sets and Measurements of Fuzziness

A. Fuzzy Sets

A fuzzy set A with its finite number of supports $x_1, x_2, \cdots, x_n$ in the universe of discourse U is formally defined as

$$A = \{(\mu_A(x_i), x_i)\}, \quad i = 1, 2, \cdots, n \tag{1}$$

where the characteristic function $\mu_A(x_i)$ known as membership function and having positive value in the interval $[0,1]$ denotes the degree to which an event x_i may be a member of A. A point x_i for which $\mu_A(x_i) = 0.5$ is said to be a crossover point of the fuzzy set A.

Let us now give some measures of fuzziness of a set A. These measures define, on a global sense, the degree of difficulty (ambiguity) in deciding whether an element x_i would be considered as a member of A.

B. Index of Fuzziness

The index of fuzziness γ of a fuzzy set A having n supporting points reflects the degree of ambiguity present in it by measuring the distance between A and its nearest ordinary set $\bar{A}$ and is defined as [8]

$$\gamma(A) = \frac{2}{n^{1/k}} d(A, \bar{A}) \tag{2}$$

where $d(A, \bar{A})$ denotes the distance between A and its nearest ordinary set $\bar{A}$. The set $\bar{A}$ is such that

$$\mu_{\bar{A}}(x_i) = 0, \quad \text{if } \mu_A(x_i) \leq 0.5 \tag{3a}$$

and

$$\mu_{\bar{A}}(x_i) = 1, \quad \text{if } \mu_A(x_i) > 0.5. \tag{3b}$$

The positive constant k appears in order to make $\gamma(A)$ lie between zero and one, and its value depends on the type of distance function used. For example, $k = 1$ for a generalized Hamming distance, whereas $k = 2$ for a Euclidean distance. The corresponding indices of fuzziness are called the linear index of fuzziness $\gamma_l(A)$ and the quadratic index of fuzziness $\gamma_q(A)$. Considering d to be a generalized Hamming distance, we have

$$d_l(A, \bar{A}) = \sum_i |\mu_A(x_i) - \mu_{\bar{A}}(x_i)| = \sum_i \mu_{A \cap \bar{A}}(x_i), \quad i = 1, 2, \cdots, n \tag{4}$$

and

$$\gamma_l(A) = \frac{2}{n} \sum_i \mu_{A \cap \bar{A}}(x_i), \quad i = 1, 2, \cdots, n \tag{5}$$

where $\mu_{A \cap \bar{A}}(x_i)$ denotes the membership of x_i to a set which is the intersection of the fuzzy set A and its complement $\bar{A}$ and is defined as

$$\mu_{A \cap \bar{A}}(x_i) = \min\{\mu_A(x_i), (1 - \mu_A(x_i))\}, \quad i = 1, 2, \cdots, n.$$

Considering d to be an Euclidean distance, we have

$$\gamma_q(A) = \frac{2}{\sqrt{n}} \left[\sum_i (\mu_A(x_i) - \mu_{\bar{A}}(x_i))^2 \right]^{1/2}, \quad i = 1, 2, \cdots, n. \tag{6}$$

C. Entropy

The term entropy of a fuzzy set A is defined according to Deluca and Termini [9] as

$$H(A) = \frac{1}{n \ln 2} \sum_i S_n(\mu_A(x_i)), \quad i = 1, 2, \cdots, n \tag{7}$$

with

$$S_n(\mu_A(x_i)) = -\mu_A(x_i) \ln(\mu_A(x_i)) - (1 - \mu_A(x_i)) \ln(1 - \mu_A(x_i)). \tag{8}$$

In (7) and (8), ln stands for natural logarithm (i.e., base e). However, any other base would serve the purpose because of the normalization factor $\ln 2$ in (7).

$\gamma(A)$ and $H(A)$ are such that (from (5)–(7))

$$\gamma_{\min} = H_{\min} = 0 (\min), \text{ for } \mu_A(x_i) = 0 \text{ or } 1 \quad \text{for all } i \tag{9a}$$

$$\gamma_{\max} = H_{\max} = 1 (\max), \text{ for } \mu_A(x_i) = 0.5, \quad \text{for all } i. \tag{9b}$$

Suppose $\mu_A(x_i) = 0.5$, for all i. Then $\mu_{\bar{A}}(x_i) = 0$, for all i, and

$$\gamma_l(A) = \frac{2}{n} \sum_i \left(\frac{1}{2}\right) = \frac{2}{n} \cdot \frac{n}{2} = 1$$

$$\gamma_q(A) = \frac{2}{\sqrt{n}} \left(\sum_i \left(\frac{1}{2}\right)^2 \right)^{1/2} = \frac{2}{\sqrt{n}} \cdot \frac{\sqrt{n}}{2} = 1$$

and

$$H(A) = \frac{1}{n \ln 2} \sum_i \left(-\ln \frac{1}{2} \right) = \frac{1}{n \ln 2} \cdot n \ln 2 = 1.$$

Therefore, γ and H increase monotonically in the interval $[0, 0.5]$

and decrease monotonically in $[0.5, 1]$ with a maximum of one at $\mu = 0.5$.

D. π-ness

The π-ness of A is defined as [10]

$$\pi(A) = \frac{1}{n} \sum_i G_\pi(x_i), \qquad i = 1, 2, \cdots, n \tag{10}$$

where G_π is any π function as explained in the Section III.

G_π $(0 \leq G_\pi \leq 1)$ increases monotonically in $[x_i = 0$ to $x_i = x_{max}/2$, say$]$ and then decreases monotonically in $[x_{max}/2, x_{max}]$ with a maximum of unity at $x_{max}/2$, where x_{max} denotes the maximum value of x_i.

III. Membership Functions

Let us now consider different S and π functions to obtain $\mu_A(x_i)$ from x_i. The standard S function as defined by Zadeh [13] has the form

$$\begin{aligned} \mu_{AS}(x_i; a, b, c) &= 0, & x_i \leq a \quad & (11a)\\ &= 2[(x_i - a)/(c - a)]^2, & a \leq x_i \leq b \quad & (11b)\\ &= 1 - 2[(x_i - c)/(c - a)]^2, & b \leq x_i \leq c \quad & (11c)\\ &= 1, & x_i \geq c \quad & (11d) \end{aligned}$$

in the interval $[a, c]$ with $b = (a + c)/2$. The parameter b is known as the crossover point for which $\mu_{AS}(b) = S(b; a, b, c) = 0.5$.

Similarly, the standard π function has the form

$$\begin{aligned} \mu_{A\pi}(x_i; a, c, a') &= \mu_{AS}(x_i; a, b, c), & x_i \leq c \quad & (12a)\\ &= 1 - \mu_{AS}(x_i; c, b', a'), & x_i \geq c \quad & (12b) \end{aligned}$$

in the interval $[a, a']$ with $c = (a + a')/2$, $b = (a + c)/2$, and $b' = (a' + c)/2$. b and b' are the crossover points, i.e., $\mu_{A\pi}(b) = \mu_{A\pi}(b') = 0.5$, and c is the central point at which $\mu_{A\pi} = 1$.

Instead of using the standard S and π functions one can also consider the following equation as defined by Pal and Dutta Majumder [14], [15]

$$\mu_A(x_i) = G(x_i) = \left[1 + \left(\frac{|\hat{x}_n - x_i|}{F_d}\right)\right]^{-F_e} \tag{13}$$

which approximates the standard membership functions.

F_e and F_d (two positive constants) are known respectively as exponential and denominational fuzzy generators and control the crossover point, bandwidth, and hence the symmetry of the curve about the crossover point. $\hat{x}_n$ is the reference constant such that the function represents an S-type function G_S for $\hat{x}_n = x_{max}$ and a π-type function G_π for $\hat{x}_n = x_l$, $0 < x_l < x_{max}$, where x_{max} represents the maximum value of x_i.

IV. Feature Evaluation Index

Let $C_1, C_2, \cdots, C_j, \cdots, C_m$ be the m-pattern classes in an N-dimensional $(X_1, X_2, \cdots, X_q, \cdots, X_N)$ feature space Q_X. Also, let n_j $(j = 1, 2, \cdots, m)$ be the number of samples available from class C_j. The algorithms for computing γ, H, and π-ness values of the classes in order to provide a quantitative index for feature evaluation are described in this section.

A. Computation of γ and H Using Standard π Function

Let us consider the standard π function (12) for computing γ and H of C_j along the qth component and take the parameters of the function as

$$c = (x_{qj})_{av} \tag{14a}$$

$$b' = c + \max\left\{\left|(x_{qj})_{av} - (x_{qj})_{max}\right|, \left|(x_{qj})_{av} - (x_{qj})_{min}\right|\right\} \tag{14b}$$

with

$$b = 2c - b' \tag{14c}$$

$$a = 2b - c \tag{14d}$$

$$a' = 2b' - c \tag{14e}$$

where $(x_{qj})_{av}$, $(x_{qj})_{max}$, and $(x_{qj})_{min}$ denote the mean, maximum, and minimum values respectively, computed along the qth coordinate axis, over all the n_j samples in C_j.

Since $\mu(c) = \mu((x_{qj})_{av}) = 1$, the values of γ and H are zero at $c = (x_{qj})_{av}$ and would tend to unity (9) as we move away from c towards either b or b' of the π function (i.e., from mean towards boundary of C_j). The lower the value of γ or H along the qth component in C_j, the greater would be the number of samples having $\mu(x) = 1$ (or, the less would be the difficulty in deciding whether an element x can be considered, on the basis of its qth measurement, a member of C_j or not) and hence the greater would be the tendency of the samples to cluster around its mean value, resulting in less internal scatter or less intraset distance or more compactness of the samples along the qth axis within C_j. Therefore, the reliability (goodness) of a feature in characterizing a class increases as its corresponding γ or H value within the class (computed with π function) decreases.

The value of γ or H thus obtained along the qth coordinate axis in C_j may be denoted by γ_{qj}^π or H_{qj}^π.

Let us now pool together the classes C_j and C_k $(j, k = 1, 2, \cdots, m, j \neq k)$ and compute the mean $(x_{qjk})_{av}$, maximum $(x_{qjk})_{max}$ and minimum $(x_{qjk})_{min}$ values of the qth component over all the samples (numbering $n_j + n_k$). The value of γ or H so computed with (14) would therefore increase as the goodness of the qth feature in discriminating pattern classes C_j and C_k increases, because there would be fewer samples around the mean $(x_{qjk})_{av}$ of the combined class, resulting in γ or $H = 0$, and more samples far from the $(x_{qjk})_{av}$, giving γ or $H = 1$. Let us denote the γ and H value so computed by γ_{qjk}^π and H_{qjk}^π, which increase as the separation between C_j and C_k (i.e., separation between b and b') along the qth dimension increases or, in other words, as the steepness of π function decreases.

It is to be mentioned here that one can also replace $(x_{qj})_{av}$, $(x_{qj})_{max}$ and $(x_{qj})_{min}$ of (14) by $(x_{qjk})_{av}$, $(x_{qj})_{av}$, and $(x_{qk})_{av}$, respectively, to compute γ_{qjk} or H_{qjk}. In this case, only their absolute values but not their behavior, as described previously, would be affected.

B. Computation of γ and H Using Standard S Function

For computing γ and H of C_j along the qth component let us now take the parameters of S function (11) as

$$b = (x_{qj})_{av} \tag{15a}$$

$$c = b + \max\left\{\left|(x_{qj})_{av} - (x_{qj})_{max}\right|, \left|(x_{qj})_{av} - (x_{qj})_{min}\right|\right\} \tag{15b}$$

where

$$a = 2b - c. \tag{15c}$$

Since $\mu(b) = \mu((x_{qj})_{av}) = 0.5$, the values of γ and H are 1 at $b = (x_{qj})_{av}$ and would tend to zero (9) as we move away from b towards either c or a of the S function. The higher the value of γ or H, the greater would be the number of samples having $\mu(x) \simeq 0.5$ and hence the greater would be the tendency of the samples to cluster around its mean value, resulting in less internal scatter within the class. Therefore, unlike the case with π function, the reliability (goodness) of a feature in characterizing a class C_j increases as its corresponding γ_{qj}^s or H_{qj}^s value within the class increases.

Similarly, if we now pool together the classes C_j and C_k $(j, k = 1, 2, \cdots, m, j \neq k)$ and compute the mean, maximum and minimum values of the qth component over all the samples $(n_j + n_k)$, then the value of γ_{qjk}^s or H_{qjk}^s so computed with (15) would therefore decrease as the goodness of the qth feature in

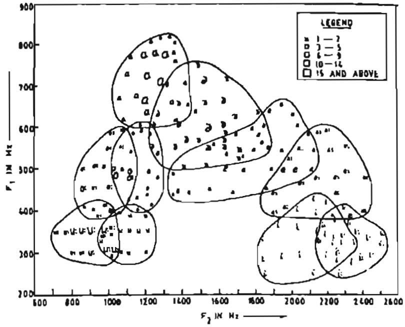

Fig. 1. Vowel diagram in F_1–F_2 plane.

discriminating pattern classes C_j and C_k increases; because there would be fewer samples around the mean of the classes C_j and C_k, resulting in γ or $H \simeq 1$, and more samples far from the mean, giving γ, or $H \simeq 0$.

It therefore appears that the γ_{qj} (or H_{qj}) and γ_{qjk} (or H_{qjk}) reflect the concept of intraset and interset distances, respectively, in a classical feature-selection problem. With decrease in intraset and interset distances along qth component in C_j the values of γ_{qj} (or H_{qj}) and γ_{qjk} (or H_{qjk}) are seen to decrease, or increase, when computed using the π, or S function.

C. Computation of π-ness

Similarly, for computing π_{qj} along the qth dimension in C_j, the parameters of the π function are set as follows:

$$c = (x_{qj})_{av} \tag{16a}$$

$$a' = c + \max\left\{\left|(x_{qj})_{av} - (x_{qj})_{max}\right|, \left|(x_{qj})_{av} - (x_{qj})_{min}\right|\right\} \tag{16b}$$

with

$$a = 2c - a', \qquad b = (a + c)/2, \qquad b' = (a' + c)/2. \tag{16c}$$

For computing π_{qjk}, the classes C_j and C_k are pooled together and these parameters are obtained from $(n_j + n_k)$ samples. Like the γ' (or H') value obtained with S function, π_{qj} and π_{qjk} decrease as intraset and interset distances in C_j decrease.

Considering these intraclass and interclass measures in each case, the problem of evaluating feature quality in Q_X therefore reduces to minimizing/maximizing the values of

$$\gamma_{qj}^{\pi} \text{ or } H_{qj}^{\pi}/\gamma_{qj}^{s} \text{ or } H_{qj}^{2} \text{ or } \pi_{qj}$$

while maximizing/minimizing the values of

$$\gamma_{qjk}^{\pi} \text{ or } H_{qjk}^{\pi}/\gamma_{qjk}^{s} \text{ or } H_{qjk}^{s} \text{ or } \pi_{qjk}.$$

The feature-evaluation index for the qth feature is accordingly defined as

$$(\mathrm{FEI})_q = \frac{d_{qj} + d_{qk}}{d_{qjk}},$$

$$j, k = 1, 2, \cdots, m, \; j \neq k, \; q = 1, 2, \cdots, N \tag{17a}$$

where d stands for γ^{π} or H^{π} and

$$(\mathrm{FEI})_q = \frac{d_{qjk}}{d_{qj} + d_{qk}} \tag{17b}$$

where d stands for γ' or H' or π-ness. The lower the value of $(\mathrm{FEI})_q$, the higher is, therefore, the quality (importance) of the qth feature in characterizing and discriminating different classes in Q_X.

V. Implementation and Results

For implementation of the above algorithm, the test material is prepared from a set of nearly 600 discrete phonetically balanced speech units in consonant-vowel-consonant Telugu (a major Indian Language) vocabulary uttered by three male speakers in the age group of 30–35 years.

For vowel sounds of ten classes (δ, a, i, i:, u, u:, e, e:, o and o:) including shorter and longer categories, the first three formant frequencies at the steady state (F_1, F_2, and F_3) are obtained through spectrum analysis.

For consonants, eight unaspirated plosive sounds namely the velars /k, g/, the alveolars /t, d/, the dentals /t, d/, and the bilabials /p, b/ in combination with six vowel groups (δ, a, E, I, O, U) are selected. The formant frequencies are measured at the initial and the final state of the plosives. The details of processing and formant extraction are available in [14]–[16].

A. Vowel Recognition

A set of 496 vowel sounds of ten different classes are used here as the data set with F_1, F_2, and F_3 as the features. Fig. 1 shows the feature space of vowels corresponding to F_1 and F_2 when longer and shorter categories are treated separately.

Fig. 2 shows the order of importance of formants in recognizing and discriminating different vowels as obtained with intraclass measures (diagonal cells) and FEI values (off-diagonal cells). Results using only S function in computing γ_i and H values are shown here. Lower triangular part of the matrix corresponds to the results obtained with standard S and π functions ((11) and (12)) whereas, the upper triangular portion gives the results corresponding to their approximated versions (13). While using (13) we selected the parameters as follows.

For S-type Function:

$$\hat{x}_s = (x_{qj})_{max}, \qquad \text{for computing } \gamma_{qj} \text{ or } H_{qj} \tag{18a}$$

$$= (x_{qjk})_{max}, \qquad \text{for computing } \gamma_{qjk} \text{ or } H_{qjk}. \tag{18b}$$

F_e and F_d were selected in such a way that $(x_q)_{av}$ corresponds to the crossover point, i.e., $G_S((x_q)_{av}) = 0.5$. To keep the crossover point fixed at $(x_q)_{av}$, different values of F_e and F_d may be used to result in various slopes of S function.

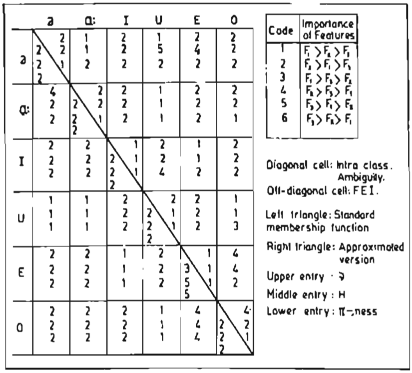

Fig. 2. Order of importance of features.

For π-type Function:

$$\hat{x}_n = (x_{qj})_{av}, \quad \text{for computing } \pi_{qj} \tag{19a}$$

$$= (x_{qjk}), \quad \text{for computing } \pi_{qjk}. \tag{19b}$$

As crossover points have no importance here in measuring π-ness, the selection of fuzzifiers is not crucial.

The results using (13) (as shown in the lower trianglular part of Fig. 2) were obtained for $F_e = 1/16, 1/8, 1/4, 1/2$, and 1 with the crossover point at $(x_q)_{av}$. These values of F_e were also found to yield optimum recognition score in earlier investigations on vowel and plosive identification [14], [15]. For computing π_{qj} and π_{qjk}, F_d was selected to be 50 for $F_e = 1/16$.

Again, the order of importance as shown in Fig. 2 was obtained after pooling together the shorter and longer counterparts (differing mainly in duration) of a vowel. In a part of the experiment the shorter and longer categories were treated separately, and the order of importance of formants for the corresponding γ_{qj}, H_{qj}, and π_{qj} values (intraclass measures) is listed in Table I. This is included for comparison with the diagonal entries of Fig. 2.

For vowel recognition (except for /E/, as shown from Fig. 2) the first two formants are found to be much more important than F_3 (which is mainly responsible for speaker identification). Furthermore, better result has been obtained for the cases when the shorter and longer categories are pooled together than the cases when they are treated separately. The result agrees well with previous investigation [14]. From the FEI measures of different pair of classes (off-diagonal cells of Fig. 2), F_1 is seen to be more important than F_2 in discriminating the class combinations /U, O/, /I, E/, /a, U/, and /δ, U/, i.e., between /front and front/ or /back and back/vowels. For the other combinations, i.e., discriminating between /front and back/ vowels, F_2 is found to be the strongest feature. The above findings can readily be verified from Fig. 1.

Typical FEI values for F_1, F_2, and F_3 are shown in Table II to illustrate the relative difference in importance among the formants in characterizing a class.

Similar investigations have also been made in case of speaker identification problem using the same data set (Fig. 1) and $\{F_1 F_2, F_3, F_3 - F_2, F_3 - F_1, F_3/F_2, F_3/F_1\}$ as the feature set. FEI values have been computed for each of the three speakers individually for all the vowel classes. Contrary to the vowel recognition problem, F_3 and its combinations were found here to yield lower FEI values, i.e., more important than F_1 and F_2—resembling well the earlier report [14].

TABLE I
INTRACLASS AMBIGUITIES FOR SHORT AND LONG VOWEL CLASSES

Membership Function	Vowel Class							
	i	*i:*	*u*	*u:*	*e*	*e:*	*o*	*o:*
	2	1	2	3	4	1	2	2
Standard	2	1	2	6	6	1	2	2
	1	1	2	6	6	1	2	2
Approximated	3	1	2	1	5	1	2	1
version	3	1	2	1	5	1	2	1
	1	1	2	1	1	1	2	1

1: $F_1F_2F_3$ 2: $F_2F_1F_3$
3: $F_1F_3F_2$ 4: $F_2F_3F_1$
5: $F_3F_1F_2$ 6: $F_3F_2F_1$

B. *Plosive Recognition*

A set of 588 unaspirated plosive consonants are used as the data set with ΔF_1, ΔF_2 (the difference of the initial and final values of the first and second formants), ΔT (duration), $\Delta F_1/\Delta T$, $\Delta F_2/\Delta T$ (the rates of transition) as the feature set.

The order of importance of the features for plosive recognition according to FEI values does not seem to be very regular as has been obtained in case of vowel recognition problem. Here all five features have more or less importance in determining the plosive classes, contrary to the case of vowel recognition, where F_3 has much less importance than F_1 and F_2 in defining the vowel classes. However, a qualitative assessment has been adopted here to formulate an idea about the quality of the features based on the measure of FEI.

Table III shows the number of times each feature has occupied a particular position of importance on the basis of FEI measure using γ, H, and π-ness values and different target vowels. Results corresponding to both standard membership functions and their approximated versions are included for comparison.

TABLE II
TYPICAL FEI VALUES OF THE FORMANTS USING STANDARD MEMBERSHIP FUNCTIONS

Vowel Classes	FEI Values According to the Parameter: Index of Fuzziness F_1	F_2	F_3	Entropy F_1	F_2	F_3	π-ness F_1	F_2	F_3
δ, I	0.4521	0.3378	0.7142	0.4829	0.3874	0.6226	0.4828	0.3778	0.6402
a, U	0.2093	0.3649	0.5970	0.3038	0.4300	0.5860	0.2755	0.4253	0.5963
I, E	0.4503	0.4591	0.5894	0.4822	0.4841	0.5581	0.4795	0.4846	0.5649
I, O	0.3133	0.1018	0.6329	0.3994	0.1811	0.5586	0.3826	0.1483	0.5645
U, E	0.4538	0.2696	0.5635	0.4839	0.3289	0.5282	0.4822	0.3158	0.5286
U, O	0.3126	0.5196	0.5720	0.3948	0.4999	0.5547	0.3859	0.4994	0.5630

TABLE III
IMPORTANCE OF PLOSIVE FEATURES ACCORDING TO FEI

Membership Function	Position of Importance: First	Second	Third
Standard	$I_{36}, III_{34}, II_{20}$	V_{42}, III_{25}, I_{17}	IV_{36}, I_{22}, II_{21}
Approximate	IV_{61}, V_{28}, I_{13}	III_{46}, I_{24}, V_{20}	$III_{46}, I_{25}, II_{22}$
Standard	$I_{43}, III_{30}, II_{22}$	$III_{31}, II_{26}, I_{24}$	$II_{39}, III_{30}, V_{26}$
Approximate	IV_{62}, I_{17}, II_{12}	$III_{41}, I_{25}, IV_{17}$	$III4_{4}, I_{24}, V_{21}$

1) Suffixes indicate the number of times the feature has occurred in the position.

2)I, II, III, IV, V represent the features ΔF_1, ΔF_2, ΔT, $\Delta F_1/\Delta T$, $\Delta F_2/\Delta T$, respectively

Let us now consider the case of unvoiced plosive sounds with the standard membership function. The features ΔF_1, ΔT, and ΔF_2 were first in order of importance 36, 34, and 20 times, respectively. They occupied the second position 42, 25, and 17 times, respectively, and the third 36, 22, and 21 times, respectively. Considering the first three number of occurrences in first two positions and the first two number of occurrences in first two positions, it is seen that the set $(\Delta F_1, \Delta T)$ is more effective than $(\Delta F_2/\Delta T, \Delta F_2)$, which is again more important than $\Delta F_1/\Delta T$ in discriminating unvoiced plosive sounds. Similarly, the features $(\Delta F_1, \Delta T$, and $\Delta F_2)$ (particularly, $\Delta F_1, \Delta T)$ are seen to be more reliable than the others in characterizing voiced plosives, using the standard membership function.

Let us now consider the cases of using approximate version of the membership functions (13). To discriminate unvoiced plosives the set $(\Delta F_1/\Delta T, \Delta F_2/\Delta T)$ gives better characterizing feature than ΔF_1 and ΔT; whereas for the voiced counterparts $(\Delta F_1/\Delta T, \Delta F_1, \Delta T)$ came out to be the best feature set.

From these discussions, the features ΔF_1 and ΔT are overall found to be the most important in characterizing and discriminating different plosive sounds. The result conforms to the earlier findings [15], [16] obtained from the point of automatic plosive sound recognition.

VI. DISCUSSION AND CONCLUSION

An algorithm for automatic evaluation of feature quality in pattern recognition has been described using the terms *index of fuzziness*, *entropy*, and *π-ness* of a fuzzy set. These terms are used to define measures of separability between classes and compactness within a class when they are implemented with S and π membership functions. For example, when these measures are implemented with π function, d^{π}_{qj} (d stands for γ or H) then decreases as compactness within jth class along qth direction increases and d^{π}_{qjk} increases as separability between C_j and C_k increases in the qth direction. If the classes C_j and C_k do not differ in mean value but differ in second order moment, i.e., variances are different, then

$$\left[(x_{qj})_{\max} - (x_{qj})_{\min}\right] > \left[(x_{qk})_{\max} - (x_{qk})_{\min}\right]$$

(assuming C_j with larger variance than C_k). From (12) and (14) we have

$$d^{\pi}_{qj} > d^{\pi}_{qk},$$

i.e., the qth feature is more important in recognizing the kth class than the jth class. Value of the interset ambiguity d^{π}_{qjk} as expected, then decreases showing the deterioration in reliability (goodness) of the qth feature in discriminating C_j from C_k. Similar behavior would also be reflected for the third (representing skewness of a class) and higher order moments when they are different for C_j and C_k with the same mean value. The algorithm is found to provide satisfactory order of importance of the features in characterizing speech sounds, in discriminating different classes of speeches and also in identifying a speaker.

Since F_3 and its higher formants $(F_4, F_5, \cdots)$ are mostly responsible for identifying a speaker, we have considered in our experiment only F_3 in addition to F_1 and F_2 for evaluating feature quality in vowel recognition problem.

It is to be mentioned here that the well-known statistical measures of feature evaluation such as Bhattacharyya coefficient, divergence, Kolmogorov variational distance, etc., theoretically take into account the interdependence of feature variables. Their computation involves multivariate numerical integration and estimation of probability density functions [4]. In practice in their computation, the features are usually treated individually to avoid computational difficulty [17]. Our proposed measure also treats the features individually. In fact, the present algorithm attempts to rank individual features according to their importance in characterizing and discriminating classes. Combination of features in doing so is not of interest. Furthermore, even in the case of independent feature, the algorithm is computationally more efficient than the aforesaid statistical measures.

ACKNOWLEDGMENT

The authors gratefully acknowledge Prof. D. Dutta Majumder for his interest in this work and Mrs. S. De Bhowmik and Mr. S. Chakraborty for preparing the manuscript.

REFERENCES

[1] K. Fukunaga, *Introduction to Statistical Pattern Recognition and Machine Learning*. New York: Academic, 1972.
[2] J. T. Tou and R. C. Gonzalez, *Pattern Recognition Principles*. Reading, MA: Addison-Wesley, 1974.
[3] G. S. Sebestyen, *Decision Making Process in Pattern Recognition*. New York: Macmillan, 1962.
[4] R. A. Devijver and J. Kittler, *Pattern Recognition—A Statistical Approach*. London: Prentice-Hall, 1982.
[5] S. K. Pal and D. Dutta Majumder, *Fuzzy Mathematical Approach to Pattern Recognition*. New York: Wiley (Halsted Press), 1986.
[6] A. Bhattacharyya, "On a measure of divergence between two multinomial populations," *Sankhya*, vol. 6, pp. 401–406, 1946.
[7] T. Kailath, "The divergence and Bhattacharyya distance measures in signal detection," *IEEE Trans. Commun. Tech.*, vol. CT-15, pp. 52–60, 1967.
[8] A. Kaufmann, *Introduction to the Theory of Fuzzy Subsets—Fundamental Theoretical Elements*, vol. 1. New York: Academic, 1975.
[9] A. Deluca and S. Termini, "A definition of a nonprobabilistic entropy in the setting of fuzzy set theory," *Inform. Contr.*, vol. 20, pp. 301–302, 1972.
[10] S. K. Pal, "Fuzzy set theoretic approach-A tool for speech and image recognition," in *Pattern Recognition Theory and Applications*. Amsterdam: D. Reidel, 1982, pp. 103–117.

[11] S. K. Pal, R. A. King, and A. A. Hashim, "Automatic grey level thresholding through index of fuzziness and entropy," *Patt. Recognition Lett.*, vol. 1, pp. 141–146, Mar. 1983

[12] —, "A note on the quantitative measure of image enhancement through fuzziness," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol PAMI-4, pp. 204–208, 1982.

[13] L. A. Zadeh, "Calculus of fuzzy restrictions," in *Fuzzy Sets and Their Application to Cognitive and Decision Process*. London: Academic, pp. 1–39, 1975.

[14] S. K. Pal and D. Dutta Majumder, "Fuzzy sets and decision making approaches in vowel and speaker recognition," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-7, pp. 625–629, 1977.

[15] —, "On automatic plosive identification using fuzziness in property sets," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-8, pp. 302–308, Apr 1978.

[16] A. K. Datta, N. R. Ganguli, and S. Ray, "Recognition of unaspirated plosives—A statistical approach," *IEEE Trans Acoustics, Speech, Signal Processing*, vol. ASSP-28, pp. 85–91, Feb. 1980.

[17] S. Ray, "The effectiveness of features in pattern recognition," Ph.D. thesis, Imperial College, Univ. of London, 1984.