INT. J. SYSTEMS SCI., 1978, VOL. 9, NO. 8, 873–886

Effect of fuzzification on the plosive cognition system

SANKAR KUMAR PAL†
and DWIJESH DUTTA MAJUMDER†

Fuzzy algorithms provide a simpler and more powerful approach than statistical decision methods for describing non-ideal (fuzzy) environments in which there exists no precise boundary between the categories due to inherent vagueness rather than randomness. This paper attempts to demonstrate the effectiveness of such an algorithm when applied to the computer recognition of patterns of biological origin such as Telugu unaspirated plosives in initial position of large number of utterances in CVC context. A multicategorizer is described in which the fuzzy processor embodies a fuzzy property extractor and a similarity matrix generator. A provision for controlling fuzziness in property sets had been made by keeping two parameters, 'exponential' and 'denominational' fuzzifiers, in the components of property matrices; their effect on recognition score is also studied.

Machines' performances are explained by plotting curves and through confusion matrices when transition, duration and slope of transition from the point of transient release of stop closure to the steady state of only first two formants were used as input features. Voiced stops are differentiated more easily than unvoiced stops, with the maximum overall recognition score ranging from 60% for dentals to 85% for bilabials. The fuzzy hedge 'slightly' when applied to property sets reduces the confusion from that of the hedge 'very' and consecutive utilizations of the operations 'CON', 'DIL' and 'INT' resulted in a wide variation of about 20 to 25% in the recognition score. Such a variation is found to be insignificant beyond an optimum value of the 'exponential fuzzifier'.

1. Introduction

The concept of pattern classification is considered as a mapping from feature space to decision space. Patterns encountered in the real world are either deterministic, probabilistic or fuzzy, and in recognition problems, these are rarely found to be deterministic rather than probabilistic and/or fuzzy. The decision-making system for the classification of probabilistic patterns can be made effective with the knowledge of statistical information about the input patterns. *A priori* information is available from the large number of training sets, where statistical independency is assumed among the components of the pattern. There are certain problems in which the recognition of a pattern is considered essentially fuzzy, because there exist no precise boundaries between categories due to inherent vagueness (or fuzziness) rather than randomness. In such non-ideal environments, particularly when the sample size is small, it is unreasonable to assume the statistical independency of the events, and the fuzzy algorithm (Zadeh 1965, Thomason 1973, Zadeh *et al.* 1975, Pal and Dutta Majumder 1977) then seemed to yield a useful and simple method for the classification of ill-defined patterns. Since speech, a pattern of biological origin, is found to a considerable extent to be fuzzy in nature, its recognition is reasonably expected to be handled within the framework of fuzzy language theory.

Received 15 June 1977.

† Electronics and Communication Sciences Unit, Indian Statistical Institute, Calcutta 700035, India.

Recognition of speech patterns is a very complex problem, involving multi-level decision processes (Dulley and Balashek 1958, Reddy 1966, Dutta Majumder and Datta 1969, Schaffer and Rabiner 1970). In order to achieve clarity of understanding the whole problem of recognition is broken up into various parts, for example, the recognition of vowels and consonants, and recognition of different phonemes in isolation and in connected speech. The steady-state values of the formant frequencies are found to have potential significance in the automatic recognition of vowels (Reddy 1966, Dutta Majumder *et al.* 1977). The transient nature of consonants makes them more difficult than vowels for a machine to recognize. The information regarding the place of articulation of stop consonants is now believed to be present in both the burst spectra and the transition of vowel formants (Halle *et al.* 1957, Lehiste and Peterson 1961, Sharf and Hemeyer 1972). The listening experiment conducted by LaRiviere *et al.* (1975), with segmented and gated speech with ten native undergraduate listeners, revealed that the highest score was obtained when aperiodic and vocalic transition of the CV syllables were presented to the listeners. The aperiodic portion included the burst spectra. The results for recognition of ' p ', ' t ' and ' k ', with target vowels ' i ', ' a ' and ' u ', are a great improvement over the results obtained with vocalic transition alone. Experiments for Telugu unaspirated plosives (Datta *et al.* 1977), using the maximum likelihood ratio as a method of classification, supported the above findings and showed the formant transition to be a characterizing feature of consonant recognition. The role of transition in the perception of stops is also considered to be important both in the ' transition-dependency ' model and the ' integration ' model.

The present study confines itself to the effectiveness of a fuzzy algorithm, and on-glide transition and its rate are employed only as recognition parameters on automatic recognition of initial unaspirated plosives in a CVC (Consonant–Vowel–Consonant) context. Secondly, the effects of fuzzification due to consecutive applications of ' DIL ', ' CON ' and ' INT ' operations on the cognitive system are investigated. It is to be noted that although various acoustic segments such as burst, aperiodic transition, aspiration, etc. seem to contain perceptual cues for stop consonants, the recent trend seems to give emphasis only to the transition. The classification analysis is based on the fuzzy properties extracted from an unknown pattern. A final decision for the purpose of recognition is taken by a machine which compares the magnitudes of similarity vectors for different classes. The amount of ambiguity in property sets is controlled by varying the values of two parameters, F_e and F_d, called ' exponential ' and ' denominational ' fuzzifiers respectively, encountered in an expression for computing the elements of property matrices.

The acoustic features used for the classification of consonants are the on-glide transition ΔF, the duration Δt, and the rate of transition $\Delta F/\Delta t$ from the point of transient release of stop closure to the steady state of the first two formants only. These are extracted from the spectrum analysis of a set of Telugu (an important Indian language) vocabulary containing about 600 commonly used speech units in CVC combination, and uttered by three informants. A Honeywell 400 computer was used for the numerical analysis. Different results are explained through confusion matrices and plotting curves.

2. Fuzzy sets and method of fuzzification

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

$$A = \{\mu_A(x_i), x_i\} \tag{1}$$

where the membership function $\mu_A(x_i)$, which is positive in the interval [0, 1] denotes the degree to which an event x_i may be a member of or belong to A. This characteristic function can be viewed as a weighting coefficient which reflects the ambiguity in a set; as it approaches unity, the grade of membership of an event in A becomes higher.

The operations which effectively create fuzzification on a set A are summarized here.

(i) *Concentration of A :* CON (A)

$$\Rightarrow \mu_{\mathrm{CON}\,(A)}(x) = [\mu_A(x)]^2, \quad \forall x \tag{2 a}$$

(ii) *Dilation of A :* DIL (A)

$$\Rightarrow \mu_{\mathrm{DIL}\,(A)}(x) = [\mu_A(x)]^{1/2}, \quad \forall x \tag{2 b}$$

(iii) *Contrast intensification of A :* INT (A)

$$\Rightarrow \mu_{\mathrm{INT}\,(A)}(x) = \begin{cases} 2[\mu_A(x)]^2, & 0 \leqslant \mu_A(x) \leqslant 0\cdot5 \\ [1 - 2(1-\mu_A(x))^2], & 0\cdot5 \leqslant \mu_A(x) \leqslant 1\cdot0 \end{cases} \tag{2 c}$$

All these operations have the effect of altering the fuzziness of a set. The effect of DIL (A) is opposite to that of CON (A) which reduces the magnitude of $\mu_A(x)$ by a relatively smaller amount for those x with a higher membership value in A compared to those with a low μ_A-value. Contrast intensification, as its name applies, reduces the fuzziness of A by increasing the values of $\mu_A(x)$ which are above 0·5, and decreasing those which are below it.

The method of fuzzification can also be encountered by applying fuzzy hedges on a set. A hedge is an operator which transforms a fuzzy set representing the meaning of a term into another intensified or rarefied fuzzy set. Suppose that the membership value for a pattern to be 'circular' is 0·8, then its membership values for the composite terms '(not (very (circular)))' and '(very (not (circular)))' are 0·36 and 0·04 respectively. Although the operators are same in both cases, the values differ due to non-identical sequence in their application.

3. Fuzzy recognition system

Fuzzy set theory provides a suitable algorithm for the useful classification of imprecisely defined patterns, particularly in problems having a small number of samples, where statistical independency cannot be assumed (non-parametric learning).

Figure 1 shows a state transition diagram of a fuzzy recognition model where $B = (b_1, b_2, \ldots, b_n)$ is the possible output symbol for each input. $\mu_1, \mu_2, \ldots, \mu_n$ are the membership functions corresponding to the outputs associated with each of the outgoing transitions. Null transitions ϕ having no output

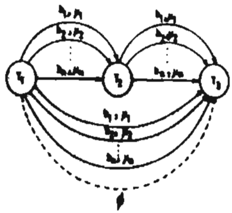

Figure 1. A generalized fuzzy recognition model.

are shown by dotted lines between the initial state T_1 and the final state T_3. These are called deletions. Other transitions described as $T_1 \to T_3$ represent substitutions and, transitions for $T_1 \to T_2 \to T_3$, which produce two output due to wrong segmentation of the input symbol, represent insertions. If the segmentation is perfect (supervised segmentation) deletion and insertion errors will not be present, but substitution error due to misclassification may occur.

Let us now describe a multi-category fuzzy classifier on the basis of properties extracted from a pattern. The property p defined on an event x is a function $p(x)$ which can have values only in the interval [0, 1] (Allen 1974). For example, p_n may denote that the outer boundary of a pattern is circular, or a straight line, or that a lady is beautiful, blonde, or tall. Such a classifier is shown in Fig. 2 where the input pattern and decision (output) of the categorizer are deterministic, but the process of classification is fuzzy.

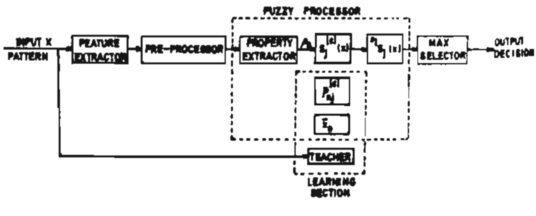

Figure 2. The structure of a fuzzy multi-categorizer.

The preprocessed N-dimensional pattern $X = (x_1, x_2, \ldots, x_n, \ldots, x_N)$ is applied to a fuzzy processor consisting of fuzzy property matrices $F_j^{(l)} = \{p_{nj}^{(l)}\}$ for the lth prototype in category C_j, where $p_{nj}^{(l)}$ denotes the degree to which the property p_n is possessed by the lth prototype in C_j. Since the output of this processor $S_j^{(l)}(X) = \{s_{nj}^{(l)}\}$ represents an N-dimensional fuzzy similarity vector, the non-fuzzy output ${}^{\prime}S_j(X)$ for C_j may be obtained from either of the equations

$$ {}^{\prime}S_j(X) = \max_{n} \min_{l} \{s_{nj}{}^{(l)}\} \tag{3 a}$$

or

$$ {}^{\prime}S_j(X) = \min_{l} |S_j{}^{(l)}(X)| \tag{3 b}$$

Then the pattern X is determined to be from kth class if

$$ {}^{\prime}S_k(X) = \max {}^{\prime}S_j(X), \quad k, j = 1, 2, \ldots, m \text{ and } l = 1, 2, \ldots, h$$

The learning behaviour is expressed by using fuzzy property matrices whose elements are determined from the equations

$$p_n = \left[1 + \left| \frac{\bar{x}_n - x_n}{F_d} \right|^{2.0} \right]^{-F_e} \tag{4}$$

and

$$\bar{x}_n = \max_{j} E\{x\} \tag{5}$$

where $E\{\cdot\}$ denotes the expected value, and F_e and F_d are the 'exponential' and 'denominational' fuzzifiers respectively. For $F_e > 1$, p_n is reduced by a relatively smaller amount for those features with a higher property value compared to those with low p_n. With a further increase in F_e, such reduction of the degree of property increases by a relatively smaller (higher) amount for those events with a higher (lower) property value. The reverse is true for values of $F_e < 1$.

The components of the fuzzy similarity matrices are defined as

$$s_{nj}{}^{(l)} = \left[1 + W \left| \frac{p_n - p_{nj}{}^{(l)}}{p_{nj}{}^{(l)}} \right| \right]^{-1} \tag{6}$$

in which the numerical value of $s_{nj}{}^{(l)}$ denotes the grade of similarity of the nth property of X with the lth prototype in C_j. The positive constant W is the weighting coefficient. The mapping, for the purpose of recognition, from the property plane onto the S-plane, satisfies the conditions

$$s_{nj}{}^{(l)} \begin{cases} \to 1 \\ \\ \to 0 \end{cases} \quad \text{as} \quad |p_n - p_{nj}{}^{(l)}| \begin{cases} \to 0 \\ \\ \to \infty \end{cases}$$

$$\text{increases} \qquad\qquad\qquad \text{decreases}$$

4. Experiment

The test material was prepared from the Telugu vocabulary containing a set of discrete phonetically balanced (PB) speech units in the CVC context. From these PB words, the velars 'k' and 'g', the alveolars 'ṭ' and 'ḍ', the dentals 't' and 'd', and the bilabials 'p' and 'b' in combination with the ten vowels 'ə', 'a:', 'i', 'i:', 'u', 'u:', 'e', 'e:', 'o' and 'o:', including shorter and longer categories, were selected. These speech units were recorded by five male, native educated informants on TDK tape with an AKAI 1710 recorder. By a listening experiment among ten listeners, about

600 samples uttered by three informants in the 30–35 age group were chosen. The entire programme was conducted inside an empty auditorium of approximate dimensions $12 \times 30 \times 6$ m.

4.1. *Feature extraction*

Spectrum analyses of selected words were carried out using a standard Kay Sonagraph 7029A audio-frequency spectrum analyser, which yields a permanent spectrographic display of frequency versus time in the range 5 Hz to 16 kHz. The system was operated in the normal mode in the 80 Hz 8 kHz band with a wide bandpass filter having a resolution of 300 Hz.

The manual extraction of features from spectrograms consisted of the following steps.

(*a*) Extrapolate the transition of the formants to the instant of the release of stop closure, and measure the frequency at that point (beginning of transition) from the base line of the spectrogram.

(*b*) Trace the central line of the formant bands where the formant is parallel to the base line (steady state), and measure the formant value from the base line.

(*c*) Measure the duration of the transition from the point of release of stop closure up to the instant at which the formant reaches a reasonable steady state.

The scale used for the measurement of frequency is derived from the calibrated 500 Hz tone recorded on each of the spectrograms. The accuracy of measurement is within 10 Hz. For every 50 spectrograms, two time-marker recordings, one at the end and one at the beginning, were used. The scale for the measurement of duration was constructed by taking an average of these two recordings. However, throughout the whole recording process no significant difference between these two recordings was observed. The recognition parameters selected for the classification of consonants in a CV context are the amount of on-glide transition of the first two formants ΔF_1 and ΔF_2, their duration Δt, and rate of the transitions $\Delta F_1/\Delta t$ and $\Delta F_2/\Delta t$. The magnitudes of the transitions are obtained by subtracting the values of F at the steady state from those at the beginning of the on-glide transitions. The transition rates are computed by dividing the magnitudes of the transitions by their duration.

In a few cases, for particularly fast informants, it has been noticed that the vowel hardly reaches a stable state. The congruence of on-glide and off-glide in these cases was taken as the steady state. Also, the data was rejected whenever the extrapolation seemed to be confusing. The total number of patterns obtained after processing the spectrograms in this way was only 594.

The respective parameters thus constitute a five-dimensional pattern vector space Ω_X, where each utterance of one of the three speakers may be treated as an event from a population, and each dimension represents an invariant characteristic of that event. Since the longer and shorter categories of a vowel differ more in duration than in phonetic scale, the ten vowels are partitioned into six classes: '∂', 'a:', 'i', 'u', 'e' and 'o'. Therefore each point in the multidimensional vector space associates five measured parameters of a CV context uttered by one of the three informants. The number of such CV occurences for unvoiced/voiced plosive consonants with different target vowels form a vowel–consonant matrix shown in Table 1.

	k	t	t	p	g	ḋ	d	b
ə	19	9	21	22	18	6	24	14
a:	27	5	16	7	13	26	25	11
e	4	4	16	18	9	7	5	5
o	8	8	11	5	6	5	6	6
u	19	6	9	11	26	17	8	16
i	12	12	20	9	5	17	15	6

Table 1. Vowel–consonant matrix showing the number of CV occurences.

4.2. *Method of recognition*

The method consists first of all in recognizing the plosives, irrespective of speakers with *a priori* knowledge of the target vowel. The effect of fuzzification on recognition score is then investigated.

Prototype points chosen for recognition are the average of the coordinate values corresponding to the entire set of samples in a particular class. Properties corresponding to each of the five parameters were computed using eqn. (4) with $F_d = 100$, $m = 4$, $N = 5$ and $h = 1$. Finally, a mapping from the property plane onto the decision plane was carried out, where each component of the similarity vector was obtained via eqn. (6). The inverse of the standard deviation of a recognition feature was used as its weighting coefficient so that W decreased with increasing variance. In one part of the experiment, W was considered to be unity in order to investigate the influence of phase weights associated with the features. Again, in a few cases where standard deviation of the coordinate values in a class was zero, the corresponding W value was set at unity. This is logical, in the sense that a property which has an identical value for all members of a set is an all-important feature of the set, and hence its contribution in the closeness measurement need not be reduced.

The effect of fuzzification on a cognitive system was incorporated by changing only the values of F_e; F_d, being less active compared to F_e in creating ambiguity, it is kept constant at a value of 100. Various values considered for F_e in the experiment are 4, 2, 1, $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$ and $\frac{1}{16}$, such that $F_e = 2$ represents the operation 'CON', $F_e = 4$ represents 'CON (CON)', $F_e = \frac{1}{2}$ represents 'DIL', $F_e = \frac{1}{4}$ represents 'DIL (DIL)', etc. In other words, the fuzzy hedges 'very', 'slightly', and their successive operations are being implemented with these constants. Besides these values of F_e, fuzziness in property sets was also introduced by applying the function 'INT'.

With the above information, the fuzzy similarity matrices denoting the degree of similarity between the pattern and the four classes for a specified value of F_e were formed. To assign a proper class to an unknown pattern, the non-fuzzy decision was adopted by the machine by measuring maximum closeness on the basis of the magnitudes of the similarity vectors.

5. Results

Typical values of the acoustic features for all the unaspirated consonants with the target vowel 'u' are shown in Table 2. Percentage recognition of plosives for different target vowels is given in Table 3 for $F_e = \frac{1}{16}$. Figures 3 and 4 illustrate the variation of recognition score for unvoiced and voiced counterparts respectively with different values of the 'exponential fuzzifier'.

Plosive	ΔF_1 (Hz)	ΔF_2 (Hz)	Δt (ms)
k	0	−150	60
ṭ	50	400	30
t	100	225	40
p	0	100	60
g	−50	−150	45
ḍ	50	200	35
d	0	250	45
b	50	100	30

Table 2. Typical feature values of plosives for target vowel 'u'.

The scores plotted are the average values of the results obtained against all the target vowels. The correct rate of decision rendered by machine in recognizing a consonant is found to increase with decreasing F_e. Recognition varies from about 20 to 25% (except for velars) as F_e changes from $\frac{1}{16}$ to 4. With further reduction of the value of F_e beyond 0·5, the error rate does

Target vowel	k	ṭ	t	p	g	ḍ	d	b
ə	31·58	88·89	38·10	100·00	33·34	50·00	83·34	100·00
a:	48·14	60·00	37·50	100·00	38·46	76·92	40·00	100·00
e	100·00	75·00	75·00	22·23	100·00	85·71	80·00	40·00
o	100·00	62·50	72·73	100·00	100·00	100·00	100·00	66·67
u	100·00	66·67	88·89	90·90	100·00	58·82	25·00	93·75
i	91·67	25·00	70·00	11·12	100·00	64·70	13·34	66·67

Table 3. Percentage of correct classifications of plosives ($F_e = \frac{1}{16}$).

not deteriorate significantly. The s_{ij-} value of an event corresponding to $F_e = 0·5$ indicate its degree of 'slightly belonging' to a class. It could therefore be stated that after an optimum value of the 'exponential fuzzifier is achieved, the ambiguity in property sets is not significantly altered, and the

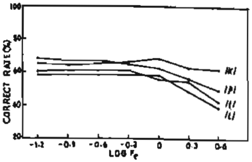

Figure 3. Variation of individual recognition scores (averaged over all the target vowels) for unvoiced plosives with fuzziness in property sets.

variation of machine's performance with fuzzification becomes insignificant. For higher values of F_e, the degree of possessing a property for samples having low property values is reduced by a larger amount compared to those having high property values, and as a result the magnitudes of fuzzy similarity components are decreased for the samples in a common class.

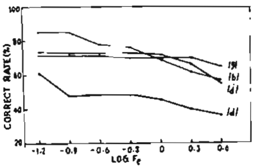

Figure 4. Variation of individual recognition scores (averaged over all the target vowels) for voiced plosives with fuzziness in property sets.

Voiced stops are seen to be differentiated more easily than the unvoiced parts. The increase in the maximum score for voiced over unvoiced stops is about 15% for the ' p '/' b ' and ' t '/' d ' pairs and about 7% for the ' k '/' g ' pair. For the ' t '/' d ' pair, the variation is reversed at higher values of F_e. These results agree well with that of an earlier experiment (Datta *et al.* 1977). The larger formant spreads for voiced stops indicate a greater co-articulation with the following vowels, which is expected to be responsible for better discrimination of the place of articulation for voiced stops once the target vowel is known *a priori*.

The overall percentages of correctness with different target vowels and their variation with fuzziness are shown in Figs. 5 and 6 for unvoiced and voiced plosives respectively. Table 4 shows how the confusion introduced by the machine in making decisions changes for different amount of fuzziness introduced into the property sets. To restrict the size of the paper, only the results of voiced plosives for the target vowel ' u ' are mentioned.

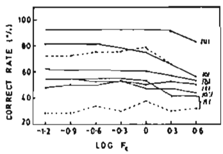

Figure 5. Variation of overall recognition scores for unvoiced plosives (for different target vowels) with fuzziness in property sets.

The figure in a cell represents the number of instances in which the same decision was made by the machine, and the diagonal elements therefore indicate the number of events correctly identified. Confusion tends to be at a minimum as F_c approaches a value of $\frac{1}{8}$. Plosives in the initial position with back target vowels are found to be identified more easily than with other target vowels.

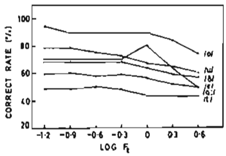

Figure 6. Variation of overall recognition scores for voiced plosives (for different target vowels) with fuzziness in property sets.

The dotted curves in Fig. 5 indicate the scores obtained by computing the similarity components with a weighting coefficient of unity. This was done only for one each of the front and back vowels. It appears, therefore, that the fixation of appropriate phase weights ensures the correct representation of feature importance in classification, leading to a marked improvement in percentage accuracy. Again, it is interesting to note that the discrimination between the scores obtained with and without a weighting coefficient decreases with increasing F_c. When the operation 'INT' is used to significantly reduce the fuzziness of properties, the effect becomes prominent, as can be seen from Figs. 7 and 8, where the correct rates obtained without a weighting coefficient are seen to exceed (except for target vowel 'a:') those

Observed class	Actual class g	$\dot{d}$	d	b
g	26		1	
$\dot{d}$		10	5	1
d		5	2	
b		2		15

(*a*) $F_e = \frac{1}{16}$

Observed class	Actual class g	$\dot{d}$	d	b
g	26			
$\dot{d}$		10	6	1
d		5	2	
b		2		15

(*b*) $F_e = \frac{1}{8}$

Observed class	Actual class g	$\dot{d}$	d	b
g	26	1		
$\dot{d}$		11	6	4
d		5	2	
b				12

(*c*) $F_e = \frac{1}{4}$

Observed class	Actual class g	$\dot{d}$	d	b
g	25			
$\dot{d}$	1	11	6	5
d		5	2	
b		1		11

(*d*) $F_e = \frac{1}{2}$

Observed class	Actual class g	$\dot{d}$	d	b
g	25			
$\dot{d}$	1	11	6	9
d		5	2	
b		1		7

(*e*) $F_e = 1$

Observed class	Actual class g	$\dot{d}$	d	b
g	25	1		
$\dot{d}$	1	11	6	10
d		4	2	
b		1		6

(*f*) $F_e = 2$

Observed class	Actual class g	$\dot{d}$	d	b
g	25			
$\dot{d}$	1	8	5	6
d		8	3	5
b		1		5

(*g*) $F_e = 4$

Observed class	Actual class g	$\dot{d}$	d	b
g	22			
$\dot{d}$	4	14	7	11
d		3	1	
b				5

(*h*) $F_e =$ ' INT '

Table 4. Confusion matrices of voiced plosives with target vowel ' u ' for different values of F_e.

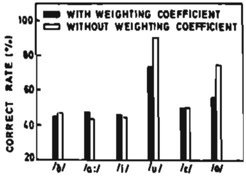

Figure 7. Overall recognition scores of unvoiced plosives for different target vowels when the fuzzifier 'contrast intensification' operates on the property sets.

obtained with a weighting coefficient. These fuzzifiers were found to provide transformations in such a way that the weighting coefficients did not ensure proper representation of the importance of the modified features, which tended to be over-emphasized.

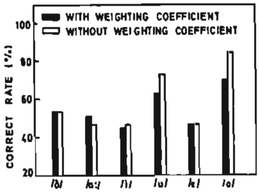

Figure 8. Overall recognition scores of voiced plosives for different target vowels when the fuzzifier 'contrast intensification' operates on the property sets.

6. Discussion and conclusions

Fuzziness in property sets has been implemented in order to study the variation of a machine's performance for the recognition of initial unaspirated plosives in a CVC context. The on-glide transitional data can be used to identify the place of articulation of the voiced and unvoiced stop consonants with *a priori* knowledge of target vowels. It is to be mentioned here that the machine was found to recognize about 82% of vowels using the first three formants (Pal and Dutta Majumder 1977, Dutta Majumder *et al.* 1977). The present results for machine recognition compare well with the human perception as obtained in the listening experiments conducted with segmented and gated speech with ten native undergraduate listeners (LaRiviere *et al.* 1975) and with several bandpass filters and seven postgraduate male listeners (Pal 1974). The results of consonant recognition as conducted by LaRiviere *et al.*

(1975) with aperiodic and vocalic transition for the target vowels 'i', 'a' and 'u', are found to be 0·98, 1·0 and 0·81 for 'p', 1·0, 1·0 and 0·13 for 't' and 0·63, 0·83 and 0·95 for 'k'. In a recent statistical study (Datta *et al.* 1977) using transitional data and the maximum likelihood ratio, the overall recognition scores for these unaspirated plosives were found to be 0·65, 0·69, 0·45, 0·9, 0·95 and 0·72 for unvoiced plosives, and 0·58, 0·65, 0·73, 0·95, 0·85 and 0·53 for voiced plosives, with the target vowels 'ə', 'a:', 'e', 'o', 'u' and 'i'. The corresponding results obtained in our experiment are marginally better. Although the results do not differ much, the classification algorithm has the advantages over the other mentioned that it is simpler, less time-consuming, and does not require so much of the previous information concerning the distribution of the events in a class. In addition, the method is more significant for the following reasons.

(*a*) The burst spectra, an important cue, particularly the antiformants were not included as recognition features.

(*b*) The CV syllables in the experiment were taken from normally spoken words, and therefore the effects of co-articulation from distant vowels and consonants are likely to affect the transitions.

(*c*) The minimum duration of vowels, 250 ms, for arriving at the perfectly steady state could not be achieved in these utterances.

The role played by the 'exponential fuzzifier' is found to be satisfactory for altering the fuzziness within property sets. The fuzzy hedge 'slightly' corresponding to the 'DIL' operator as expected, results in a better classification than does the hedge 'very' ($F_e = 2{\cdot}0$). However, successive application of the 'DIL' operator does not ensure a further increase in the recognition score. Or, in other words, after an optimum value of the 'exponential fuzzifier' is achieved, the fuzziness of property sets is only altered a little, and hence the variation in the score becomes insignificant. A greater variation of about 20 to 25% in the accuracy rate (except for the velars) is achieved with values of F_e ranging from 4 to $\frac{1}{16}$.

The reciprocal of standard deviation is found to provide an appropriate phase weight for measuring the importance of the features, supporting the findings of our previous experiments (Pal and Dutta Majumder 1977, Dutta Majumder *et al.* 1977). This characteristic is found to be significant for property sets with a higher degree of fuzziness. The recognition score is likely to be further improved by the inclusion of the transitional data of the third formant and the data of the burst spectra. The method of feature extraction used in the experiment is by no means perfectly clear and unambiguous. Further investigations are necessary to obtain a better set of reference constants.

Acknowledgment

The authors gratefully acknowledge the valuable help rendered by Shri A. K. Datta, Shri N. R. Ganguli, Shri B. Mukherjee and Shri S. Chakraborty in completing this work.

References

Allen, A. D., 1974, *I.E.E.E. Trans. Syst. Man Cybernet.*, **4**, 66.

Dulley, H., and Balashek, S., 1958, *J. acoust. Soc. Am.*, **30**, 721.

Dutta Majumder, D., and Datta, A. K., 1969, *J. Instn Telecommun. Engrs, New Delhi*, **15**, 233.

Dutta Majumder, D., Datta, A. K., and Pal, S. K., 1977, *J. comput. Soc. India*, **7**, 14.

Datta, A. K., Ganguli, N. R., Mukherjee, B., and Roy, S., 1977, *Proceedings of the All-India Interdisciplinary Symposium on Digital Technique and Pattern Recognition*, 15–17 February (Calcutta : Indian Statistical Institute) (in the press).

Halle, M., Hughes, G. W., and Radley, J. P. A., 1957, *J. acoust. Soc. Am.*, **29**, 107.

LaRiviere, C., Wintz, H., and Herriman, E., 1975, *J. acoust. Soc. Am.*, **57**, 470.

Lehiste, I., and Peterson, G. E., 1961, *J. acoust. Soc. Am.*, **33**, 268.

Pal, S. K., 1974, M.Tech. Thesis, University of Calcutta, India.

Pal, S. K., and Dutta Majumder, D., 1977, *I.E.E.E. Trans. Syst. Man Cybernet.*, **7**, 625.

Reddy, D. R., 1966, *J. acoust. Soc. Am.*, **40**, 307.

Schaffer, R. W., and Rabiner, L. R., 1970, *J. acoust. Soc. Am.*, **47**, 634.

Sharf, D. J., and Hemeyer, T., 1972, *J. acoust. Soc. Am.*, **51**, 652.

Thomason, M. G., 1973, *Pattern Recogn*, **5**, 383.

Zadeh, L. A., 1965, *Inf. Control*, **8**, 338.

Zadeh, L. A., Fu, K. S., Tanaka, K., and Shimura, M., 1975, *Fuzzy Sets and their Applications to Cognitive and Decision Processes* (London : Academic Press).