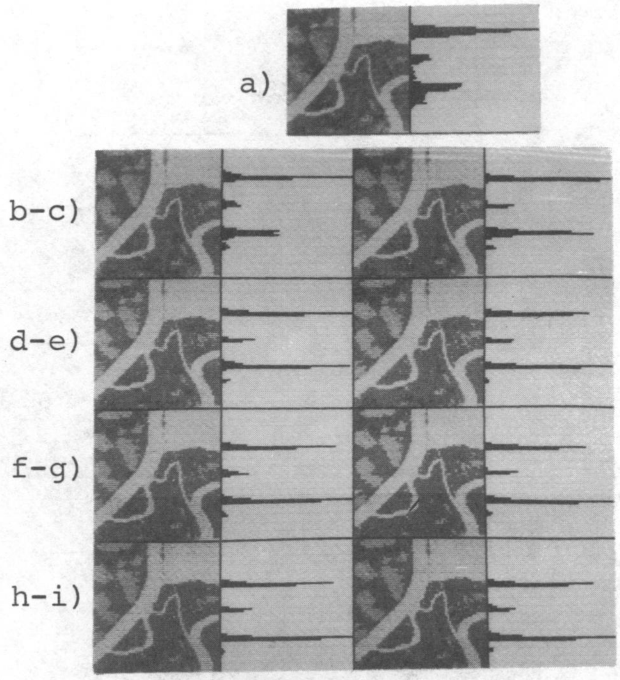

Fig. 7. Result of limiting fraction of points that can change gray level in any bin from 3/5 to 4/5 in Fig. 1.

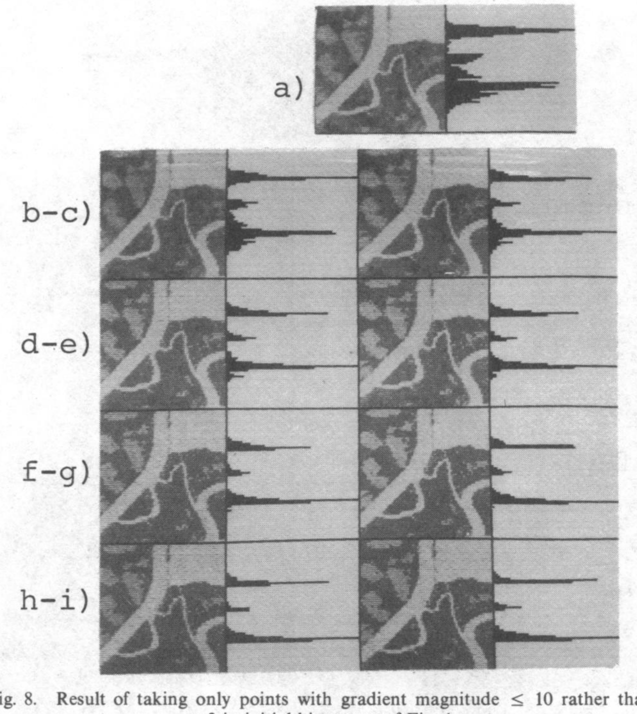

Fig. 8. Result of taking only points with gradient magnitude ≤ 10 rather than ≤ 3 in initial histogram of Fig. 1.

We then normalize the set of three fractions by dividing each by their sum. The result tells us what fraction (randomly selected) of the points of level $i+j$ will be changed to level i (or vice versa, if the fraction is negative). The entire process is then iterated.

One restriction was imposed on this process, namely that no more than 3/5 of the points in any bin have their gray levels changed at any one iteration. The process is not very sensitive to choice of this fraction. For example, in Fig. 7, 4/5 was used instead of 3/5, and it seems to make little difference. The process is also not very sensitive to the cutoff on gradient magnitude; in Fig. 8, 10 was used instead of 3, with little effect on the results.

In conclusion, the process works well for the images that we have used. It is computationally inexpensive, since only a few iterations are required. It therefore deserves consideration as an alternative to other types of peak detection or one-dimensional clustering schemes. [Of course, like all such schemes, it must be used with caution, since it cannot take into account the possibility that small regions in an image may have special significance even though they do not give rise to significant peaks on the image's histogram, or that the histogram peaks do not correspond to homogeneous image regions. It is also difficult to predict what results such schemes will produce if applied to histograms that do not have well-defined peak structures.]

Acknowledgment

The help of Ms. Shelly Rowe in preparing this paper is gratefully acknowledged.

References

[1] A. Rosenfeld, "Iterative methods in image analysis," in *Proc. IEEE Conf. Pattern Recognition and Image Processing*, 1977, 14–18.
[2] R. B. Eberlein, "An iterative gradient edge detection technique," *Computer Graphics and Image Processing* 5, 1976, 245–253.

On Automatic Plosive Identification Using Fuzziness in Property Sets

SANKAR K. PAL AND DWIJESH DUTTA MAJUMDER

Abstract—**The present paper is a continuation of our previous work [1] in vowel and speaker recognition and is an attempt to demonstrate the effectiveness of fuzzy algorithms developed for computer recognition of unaspirated plosives in CVC context. The maximum recognition score for voiced ranges from 60 percent for dentals to 85 percent for bilabials when the on-glide formant transitional data alone were considered as input. Variation in recognition score due to multiple applications of the operators CON, DIL, and INT is about 20–25 percent and becomes insignificant after optimum fuzziness is achieved.**

I. Introduction

The present study is a continuation of that in our previous correspondence [1] on fuzzy sets and decisionmaking approaches in vowel and speaker recognition where the decisional algorithms were based on the maximum values of membership function and magnitude of similarity vectors among the fuzzy property sets of a pattern. Two constant parameters which appeared in the exponent and denominator of the expressions for membership-value and property-value played the role of fuzzy generators. The object of the present correspondence is to demonstrate the effectiveness of using fuzziness in property sets as a classificatory method and

Manuscript received July 26, 1977; revised October 27, 1977.
The authors are with the Electronics and Communication Sciences Unit, Indian Statistical Institute, Calcutta 700 035, India.

0018-9472/78/0400-0302$00.75

also the formant transitional data alone as characteristic features to automatic recognition of initial unaspirated plosives in CVC context. The function of the fuzzy operators CON, DIL, and INT [2], [3] was implemented by changing the value of the exponential fuzzy generators. Their effect on plosive accuracy rate is also investigated.

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 [4]–[6]. The listening experiment conducted by LaRiviere *et al.* [7] with segmented and gated speech and 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 spectra portion included the burst spectra. The result for recognition of /p/, /t/, and /k/ with target vowels /i/, /a/, and /u/ using aperiodic plus vocalic transition is much improved from that obtained with vocalic transition alone. A recent experiment for Telugu unaspirated plosives [8], using a maximum likelihood ratio as a test of classification, supported the above findings and showed the characterizing behavior of formant transitional data for consonant recognition. To examine the above claim, the acoustic features in the present experiment for classification of consonants were considered to be the on-glide transition (ΔF), the duration (Δt), and rate of transition ($\Delta F/\Delta t$) from the point of transient release of stop closure to the steady state of formants. In the present experiment only the first two formants were considered for selection of the transitional features. These had been extracted from spectrum analysis of a set of Telugu vocabulary containing about 600 commonly used speech units in CVC combination and uttered by three informants. Results are furnished through confusion matrices and plotting curves.

II. Fuzzy Sets and Operators

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

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

where the membership function $\mu_A(x_i)$ having positive value 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 generate a fuzzy set A are summarized here.

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

$$\Rightarrow \mu_{\mathrm{CON}\,(A)}(x) = [\mu_A(x)]^2, \qquad \text{for all } x. \tag{2a}$$

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

$$\Rightarrow \mu_{\mathrm{DIL}\,(A)}(x) = [\mu_A(x)]^{0.5}, \qquad \text{for all } x. \tag{2b}$$

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

$$\Rightarrow \mu_{\mathrm{INT}\,(A)}(x) = \begin{cases} 2[\mu_A(x)]^2, & 0 \le \mu_A(x) \le 0.5 \quad \text{(2c)} \\ [1 - 2(1 - \mu_A(x))^2], & 0.5 \le \mu_A(x) \le 1.0. \quad \text{(2d)} \end{cases}$$

All these operations have the effect of altering the fuzziness of a set. The effect of DIL (A) is opposite to that of concentration, which reduces the magnitude of $\mu_A(x)$ by a relatively smaller amount for those x having higher membership value in A compared to those with low μ_A values. 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.

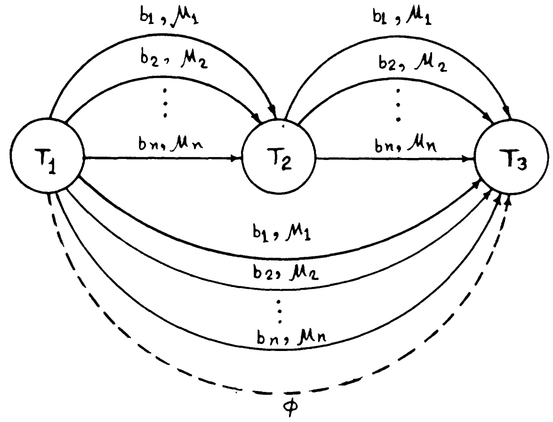

Fig. 1. Generalized fuzzy recognition model.

III. Fuzzy Recognition System

The theory of fuzzy sets provides a suitable algorithm for classification of imprecisely defined patterns, particularly, in problems having a small number of samples where statistical independency can not be assumed (nonparametric learning).

Fig. 1 shows a state transition diagram of a fuzzy recognition model where $\boldsymbol{B} = (b_1, b_2, \cdots, b_n)$ is the possible output symbol for each input. $\mu_1, \mu_2, \cdots, \mu_n$ are the membership functions corresponding to the output associated with each of the outgoing transitions. Null transitions $\varnothing$ having no output are shown by dotted lines between initial state T_1 and final state T_3. These are called deletions. Other transitions described as $T_1 \rightarrow T_3$ represent substitutions, and transitions for $T_1 \rightarrow T_2 \rightarrow T_3$, which produce two outputs due to wrong segmentation of the input symbol, denote insertions. If the segmentations are perfect (supervized) deletion and insertion errors will not be present, but error due to misclassification may occur. Let us now describe in brief the multicategory fuzzy classifier on the basis of properties extracted from a pattern where the error is present only due to misrecognition of patterns. The property p defined on an event x is a function $p(x)$ which can have values only in the interval [0,1]. A set of these functions $p(x)$, which assigns the degree of possession of some property p by an event x, constitutes what is called a property set. For example, p_n may denote the property that the outer boundary of a pattern is a circular or straight line or that a person is blonde, tall, etc. Such a classifier is shown in Fig. 2 where the input pattern and decision (output) of the categorizer is deterministic, but the process of classification is fuzzy.

The problem of feature extraction is to select a map of the form $\boldsymbol{X} = f(\boldsymbol{Y})$, where $\boldsymbol{Y} = (y_1, y_2, \cdots, y_p, \cdots, y_P)$ is a P-dimensional measurement vector derived from a pattern and $\boldsymbol{X} = (x_1, x_2, \cdots, x_n, \cdots, x_N)$ is an N-dimensional feature vector. The number of features are usually less than the measurements, and the feature vector $\boldsymbol{X}$ should adequately characterize the particular class into which the pattern is to be allocated.

The preprocessed N-dimensional pattern $\boldsymbol{X}$ is then applied to a fuzzy processor consisting of fuzzy property matrices $\boldsymbol{F}_j^{(l)} = \{p_{nj}^{(l)}\}$ for the lth prototype in a category C_j, where $p_{nj}^{(l)}$ denotes the degree to which property p_n is possessed by the lth prototype in C_j. Since the output of the processor $\boldsymbol{S}_j^{(l)}(\boldsymbol{X}) = \{s_{nj}^{(l)}\}$ represents an N-dimensional fuzzy similarity vector, the nonfuzzy output $\eta S_j(\boldsymbol{X})$ may be obtained from either of the following two equations:

$$\eta S_j(\boldsymbol{X}) = \max_n \min_l \{s_{nj}^{(l)}\} \tag{3a}$$

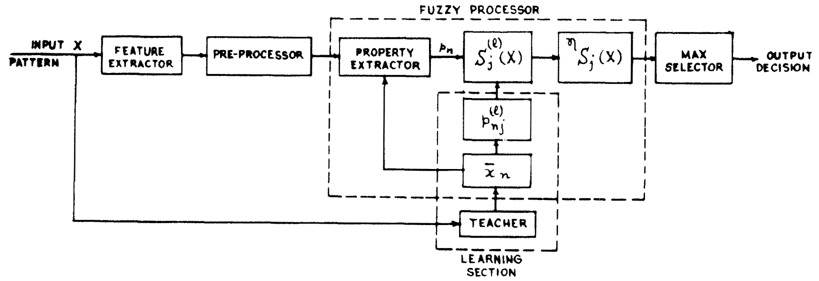

Fig. 2. Structure of fuzzy multicategorizer.

and

$$\eta S_j(X) = \min_l |S_j^{(l)}(X)|. \tag{3b}$$

Then it is decided that the pattern X is from the kth class if

$$\eta S_k(X) = \max_j \eta S_j(X), \qquad k, j = 1, 2, \cdots, m, \quad l = 1, 2, \cdots, h.$$

The numerical value of $s_{nj}^{(l)}$ in the foregoing represents the grade of similarity between the nth property of X and that of the lth prototype in C_j. m and h denote the number of pattern classes and number of prototype vectors in each class, respectively.

The learning behavior is reflected by using the effect of a fuzzy generator in property matrices whose elements are redefined by the following equations [1]:

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

and

$$\bar{x}_n = \max_j E\{x\} = \max_j \frac{1}{M_j} \sum_i x_{ijn}, \qquad i = 1, 2, \cdots, M_j \tag{5}$$

where M_j is the total number of features x_{ijn} representing the nth components of training samples associated with class C_j, $E\{x\}$ denotes the expected value of x, and F_e and F_d are, respectively, the exponential and denominational fuzzy generators. For $F_e > 1$, the p_n value reduces by a relatively smaller amount for those features having higher property values compared to those with low p_n value. With further increase of F_e, such reduction of the degree of property increases by a relatively smaller (higher) amount for those events with higher (lower) property values. The reverse is the true for values of $F_e < 1$.

IV. Implementation to Plosive Recognition

The test material was prepared from the Telugu vocabulary containing a set of discrete phonetically balanced (PB) speech units in CVC context. Out of these words, the velars /k, g/, the alveolar /ṫ, ḋ/, the dentals /t, d/, and the bilabials /p, b/ in combination with ten vowels /ə/, /a:/, /iˀ/, /i:/, /uˀ/, /u:/, /eˀ/, /e:/, /oˀ/, and /o:/ including shorter and longer categories had been selected. The experimental setup for recording and spectrographic display of these words, including nature of speakers and measurement process, is described in our previous communication [1].

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 base line (steady state), and measure the formant value from the base line.

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

The recognition parameters selected for classification of consonants in CV context are the amount of on-glide transitions of the first two formants (ΔF_1,ΔF_2), their duration (Δt), and rate of transitions ($\Delta F_1/\Delta t$,$\Delta F_2/\Delta t$). The number of transitions is obtained by subtracting the F values at the steady state from those at the beginning of on-glide transitions, and transition rates are computed by dividing the number of transitions by their duration. The total number of samples obtained after processing the spectrograms is only 594. The respective parameters thus constitute a five-dimensional pattern vector space Ω_X, where each point associates five measured parameters of a CV context uttered by one of the three informants.

Prototype points chosen for recognition are the averages of the coordinate values corresponding to entire set of samples in a particular class. Properties corresponding to each of the five parameters were computed with $F_d = 100$, m (number of pattern classes) $= 4$, $N = 5$, and h (number of prototypes in each class) $= 1$. The effect of fuzzy generation on the cognitive system was incorporated by changing only the values of F_e. F_d is kept constant at a value of 100 since it creates less ambiguity than F_e. 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 operator 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 multiple 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 were formed, which denote the degree of similarity of the pattern with the four classes for a specified value of F_e. The method of evaluating the similarity matrices has already been reported [1], where the reciprocal of the standard deviation of a component was used as its weighting coefficient so that the features with increasing variance have a decreasing weighting coefficient. Again, in a few cases, where the standard deviation of the coordinate values in a class was zero, the corresponding coefficient was set at unity. This is logical since an attribute occurring in identical magnitudes in all members of a set is an important feature of the set. Hence, its contribution to the similarity measurement need not be reduced. To assign a proper class to an unknown pattern, the nonfuzzy decision was adopted by the machine by measuring the maximum closeness on the basis of the magnitudes of the similarity vectors. For instance, the property values of an event corresponding to $F_e = 0.5$ indicate its degree of "slightly having" the properties, and the components in the fuzzy similarity vectors denote its corresponding grades of membership of "slightly belonging" to the respective classes, etc. The class possessing maximum nonfuzzy

TABLE I
TYPICAL FEATURE VALUES OF PLOSIVES FOR TARGET VOWEL /u/

Plosives	ΔF_1 (Hz)	ΔF_2 (Hz)	Δt (msec)
/k /	0	-150	60
/ṫ /	50	400	30
/t /	100	225	40
/p /	0	100	60
/g /	- 50	-150	45
/ḋ /	50	200	33
/d /	0	250	45
/b /	50	100	30

output as measured with (3b) is then decided to be the proper class of that event for the hedge "slightly."

V. RESULTS

Typical values of the acoustic features for all unaspirated plosives with target nucleus /u/ are shown in Table I. The percentages of recognition of plosives for different target vowels when F_e is set at a value of $\frac{1}{16}$ are given in Table II. Fig. 3 illustrates the variation of the recognition score for unvoiced and voiced counterparts with different values of the exponential fuzzifier. Scores plotted are the average values of the results obtained for all target vowels. The rate of correct decisions rendered by machine in the recognition of a consonant is found to increase with a decrease in the magnitude of F_e. Variation in recognition is about 20–25 percent (except for velars) as F_e changes from $\frac{1}{16}$ to 8. With further reduction of the value of F_e beyond 0.5, the error rate does not deteriorate significantly. The s_{nj} values of an event corresponding to $F_e = 0.5$ indicate its degree of "slightly belonging" to the respective classes. It could therefore be stated that after an optimum value of the exponential fuzzifier is achieved, the ambiguity in property sets is not much altered, and the variation of machine's performance with fuzzification becomes insignificant. For higher values of F_e, the property degree for samples having low property value is reduced by a larger amount than the samples having high property value. As a result, the magnitudes of fuzzy similarity components are decreased for the samples in a common class.

Voiced stops are seen to be differentiated better than the unvoiced parts. The increase in maximum score for voiced over unvoiced is about 15 percent for /p, b/ and /ṫ, ḋ/ pairs and about 7 percent 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 those

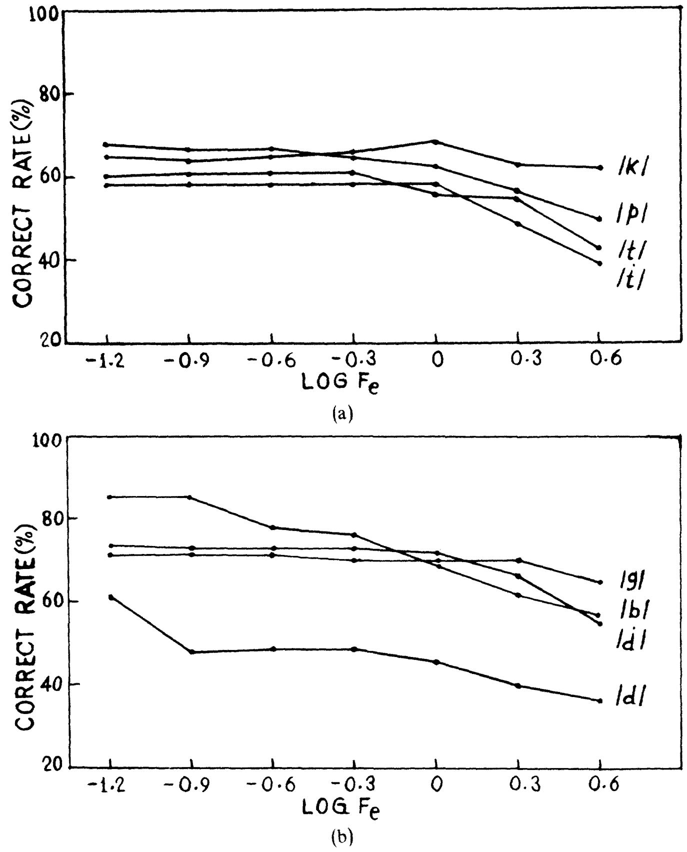

Fig. 3. Variation of overall recognition score (averaged over all target vowels) for plosives with fuzziness in property sets. (a) Unvoiced. (b) Voiced.

of an earlier experiment [8]. The larger formant spreads for voiced stops indicate a larger extent of coarticulation with the vowels that follow them. This larger coarticulation 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 sketched in Fig. 4 for unvoiced and voiced plosives. Table III explains how the confusion made by machine in taking correct decision changes for different amount of fuzziness introduced in property sets. For brevity, only the results of voiced plosives for the target vowel /u/ are mentioned. The figures represent the number of instances in which the same decision was made by machine, and diagonal elements therefore indicate the number of events correctly identified. Confusion tends to be minimum as F_e approaches a

TABLE II
PERCENTAGE OF CORRECT CLASSIFICATION OF PLOSIVES ($F_e = \frac{1}{16}$)

Target Vowel	/k/	/ṫ/	/t/	/p/	/g/	/ḋ/	/d/	/b/
/ɔ /	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 III
CONFUSION MATRICES OF VOICED PLOSIVES WITH TARGET VOWEL /u/ FOR DIFFERENT VALUES OF F_e

(a) F_e = 1/16

Observed Class \ Actual Class	/g/	/ḋ/	/d/	/b/
/g/	26		1	
/ḋ/		10	5	1
/d/		5	2	
/b/		2		15

(b) F_e = 1/8

Observed Class \ Actual Class	/g/	/ḋ/	/d/	/b/
/g/	26			
/ḋ/		10	6	1
/d/		5	2	
/b/		2		15

(c) F_e = 1/4

Observed Class \ Actual Class	/g/	/ḋ/	/d/	/b/
/g/	26	1		
/ḋ/		11	6	4
/d/		5	2	
/b/				12

(d) F_e = 1/2

Observed Class \ Actual Class	/g/	/ḋ/	/d/	/b/
/g/	25			
/ḋ/	1	11	6	5
/d/		5	2	
/b/		1		11

(e) F_e = 1

Observed Class \ Actual Class	/g/	/ḋ/	/d/	/b/
/g/	25			
/ḋ/	1	11	6	9
/d/		5	2	
/b/		1		7

(f) F_e = 2

Observed Class \ Actual Class	/g/	/ḋ/	/d/	/b/
/g/	25	1		
/ḋ/	1	11	6	10
/d/		4	2	
/b/		1		6

(g) F_e = 4

Observed Class \ Actual Class	/g/	/ḋ/	/d/	/b/
/g/	25			
/ḋ/	1	8	5	6
/d/		8	3	5
/b/		1		5

(h) F_e = 'INT'

Observed Class \ Actual Class	/g/	/ḋ/	/d/	/b/
/g/	22			
/ḋ/	4	14	7	11
/d/		3	1	
/b/				5

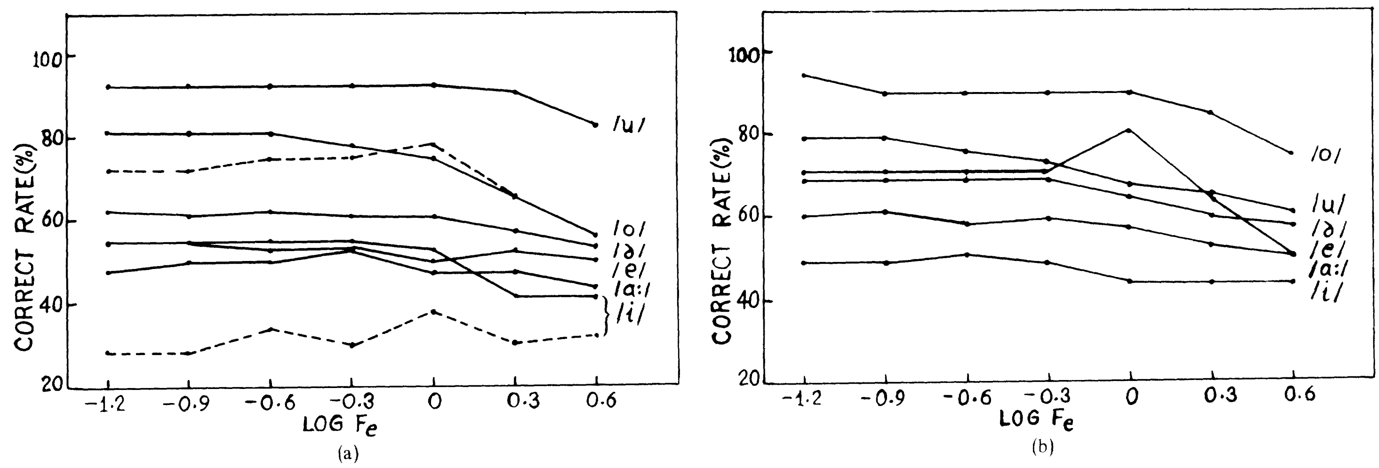

Fig. 4. Variation of overall recognition score of plosives (for different target vowels) with fuzziness in property sets. (a) Unvoiced. (b) Voiced.

value of $\frac{1}{8}$. Plosives in initial positions with back target vowels are better identified than other target vowels.

The dotted curves in Fig. 4(a) indicate the scores when similarity components were computed with a unity weighting coefficient. This was done for only one each of the front and back vowels. It therefore appears that fixation of appropriate phase weights ensures correct representation of feature importance in classification, and the percentage of accuracy is much improved. Again it is interesting to note that the discrimination between the scores obtained with and without weighting coefficients decreases with an increase in the value of F_e. When the operation INT is used to provide significant reduction in fuzziness among properties, the effect becomes prominant, as seen in Fig. 5, where the correct rates without weighting coefficients are seen to exceed (except for the target vowel /a:/) those with weighting coefficients. It was revealed on investigation that these fuzzy generators provide transformation in such a way that the weighting coefficients do not ensure proper representation of the importance to the modified features, rather, over-importances are given to them. In other words, the property components corresponding to those

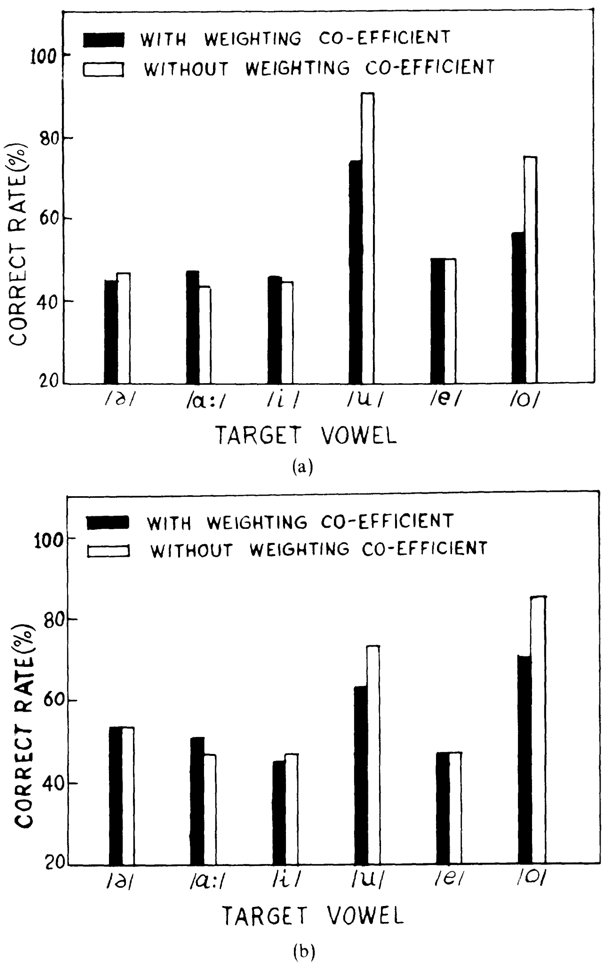

Fig. 5. Overall recognition score of plosives for different target vowels when contrast intensification is operated on property sets. (a) Unvoiced. (b) Voiced.

fuzzifiers themselves represent the proper importances to the characterizing features, and no additional factor is required for appropriate classification of patterns.

VI. Discussion and Conclusions

Fuzziness in property sets has been implemented in order to study the variation of the machine's performance for recognition of initial unaspirated plosives in CVC context. The on-glide transitional data can be used to identify the place of articulation of the voiced and unvoiced stop consonants with the *a priori* knowledge of target vowels. It should be mentioned here that the machine was found to recognize vowels with an 82 percent [1], [9] accuracy rate using the first three formants. Present results of machine recognition compares well with human perception as obtained in the listening experiments conducted with segmented and gated speech with ten native undergraduate listeners [7] and with several band pass filters and seven postgraduate male listeners [10]. The results of consonant recognition as conducted by LaRiviere *et al.* [7] with aperiodic and vocalic transition for the target vowels /i/, /a/, and /u/ are found to be 0.98, 1.0, 0.81 for /p/, 1.0, 1.0, 0.13 for /t/, and 0.63, 0.83, 0.95 for /u/. In a recent statistical study [8] using transitional data and the maximum likelihood ratio, the overall recognition scores for the unaspirated plosives were 0.65, 0.69, 0.45, 0.9, 0.95, and 0.72 for unvoiced and 0.58, 0.65, 0.73, 0.95, 0.85, and 0.53 for voiced with target vowels /ə/, /a:/, /e/, /o/, /u/, and /i/.

The corresponding figures obtained in our experiment are comparable. Although the results do not differ much, the classification algorithm as compared to the reported methods is simpler, than available ones, less time consuming, and does not require much previous information over the distribution of the events in a class. The results do agree well with those of other recent studies [11]–[13] showing the conceptual and practical advantages of fuzzy algorithms over the probabilistic approach in demonstrating ill-defined patterns where pattern indeterminacy is due to inherent vagueness rather than randomness of events and especially when the sample size is small. In addition the method is more significant from the following viewpoints.

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, coarticulation from distant vowels and consonants is likely to affect the transitions.

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

The role played by the exponential fuzzy generator is found to be satisfactory in altering the fuzziness within property sets. The fuzzy hedge "slightly," corresponding to the DIL operation as expected, results in better classification than that of the hedge "very" ($F_e = 2.0$). But successive application of the DIL operation does not ensure an increase in recognition score. In other words, after an optimum value of the exponential fuzzy generator is achieved, the fuzziness in property sets is not much altered. Hence, the variation of the score becomes insignificant. A wide variation of about 20–25 percent (excepting the velars) in accuracy rate is achieved with different values of F_e ranging from 8 to $\frac{1}{16}$.

The reciprocal of the standard deviation provides appropriate phase weights in measuring the importance of the features. This supports the findings in our previous communications [1], [9]. This characteristic is also found to be significant for the property sets having higher degrees of fuzziness. The recognition score is likely to be further improved by the inclusion of the transitional data of a third formant and the data of the burst spectra. Since this work is purposely restricted to show only the effectiveness of formant transitional data as a recognition parameter and the fuzzy-set theoretic approach as a decisional algorithm for automatic consonant recognition, these parameters were excluded.

Acknowledgment

The authors gratefully acknowledge the valuable assistance of A. K. Datta, N. R. Ganguli, B. Mookherjee, S. Chakraborty, and J. Gupta in completing this work.

References

[1] S. K. Pal and D. Dutta Majumder, "Fuzzy sets and decisionmaking approaches in vowel and speaker recognition," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-7, pp. 625–629, Aug. 1977.
[2] L. A. Zadeh, "Outline of a new approach to the analysis of complex systems and decision processes," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, pp. 28–44, Jan. 1973.
[3] L. A. Zadeh, K. S. Fu, K. Tanaka, and M. Shimura, *Fuzzy Sets and Their Application to Cognitive and Decision Processes.* London: Academic, 1975.
[4] M. Halle, G. W. Hughes, and J. P. A. Radley, "Acoustic properties of stop consonants," *J. Acoust. Soc. Am.*, vol. 29, pp. 107–116, 1957.
[5] L. Lehiste and G. E. Peterson, "Transitions, glides and dipthongs," *J. Acoust. Soc. Am.*, vol. 33, pp. 268–277, 1961.
[6] D. J. Sharf and T. Hemeyer, "Identification of place of consonant articulation from vowel formant transition," *J. Acoust. Soc. Am.*, vol. 51, pp. 652–658, 1972.
[7] C. LaRiviere, H. Wintz, and E. Herriman, "Vocalic transitions in the perception of voiceless initial stops," *J. Acoust. Soc. Am.*, vol. 57, pp. 470–475, 1975.
[8] A. K. Datta, N. R. Ganguli, B. Mukherjee, and S. Roy, "Formant transition: A cue for automatic recognition of plosives," in *Proc. All India Interdisciplinary Symp. Digital Tech. and Pattern Recog.*, Feb. 15–17, I.S.I., Calcutta (to appear).
[9] D. Dutta Majumder, A. K. Datta, and S. K. Pal, "Computer recognition of Telugu vowel sounds," *J. Comp. Soc. India*, vol. 7, pp. 14–20, Dec. 1976.
[10] S. K. Pal, "Study of articulation score affected by spectral space and amplitude range limitation," M. Tech. thesis, Inst. Radiophysics and Electronics, Univ. Calcutta, India, 1974.

[11] M. G. Thomason, "Finite fuzzy automata, regular fuzzy language and pattern recognition," *Pattern Recog.*, vol. 5, pp. 383–390, Dec. 1973.
[12] W. J. M. Kickert and H. Koppelaar, "Application of fuzzy set theory to syntactic pattern recognition of handwritten capitals," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-6, pp. 148–151, Feb. 1976.
[13] S. K. Pal, D. Dutta Majumder, and B. B. Chaudhury, "Fuzzy sets in handwritten character recognition," in *Proc. All India Interdisciplinary Symp. Digital Tech. and Pattern Recog.*, Feb. 15–17, I.S.I., Calcutta (to appear).

A Notation for Computer Animation

S. P. MUDUR AND J. H. SINGH

***Abstract*—Graphics primitives are usually designed with the aim of efficient representation and generation of complex static pictures. They are therefore highly unsuitable for the description of the dynamics of animation. Animators express a single animation sequence in just a few English sentences. Most animation systems described in the literature are extensions of existing graphics systems. Expressing even a simple animation sequence can be quite cumbersome in these systems. The essential concepts of animation have been delineated. A few computational primitives for animation are presented. A notation is also introduced in which an animation sequence may be described in a few "computational sentences."**

I. Concepts

An animation sequence essentially involves the dynamic generation of a series of images (picture frames) of a set of objects (picture-cels), in which each image is a slight alteration of the previous image. This slight alteration may involve

1) transformation of the earlier picture frame without change in the total picture information content of the set of picture-cels, e.g., successive positions of an object when falling under gravity;
2) addition or deletion of a little picture information from the earlier picture frame, thus altering the individual picture-cel description itself, e.g., one object dissolving into another.

An animation sequence extends over a number of frames. The slight alterations of the image from frame to frame are not necessarily identical; rather, these alterations take place according to a particular pattern. A complete animated film consists of a series of animation sequences. Computer animation implies the ability to describe picture-cels in a computer language and the ability to computationally express the alterations in picture frames of an animation sequence. The term animation essentially implies a changing situation. This dynamic aspect of animation is not intrinsic to graphics system design, which aims at efficient representation and generation of static pictures of high complexity.

An animator is usually able to express the animation in a sequence in just a few English sentences. For example, the line grows from a dot in the left end of the screen and in three seconds forms a line across the screen, and this growth starts slowly, picks up speed, and then slows down again.

Most computer animation systems today are designed as extensions of general purpose graphics systems [2]. As a result, expressing (i.e., programming for) even a simple animation sequence may involve a few pages of program! Unless the animation in a sequence can be expressed in just a few "computational sentences," it is very unlikely that computer animation will acquire any popularity with animators.

It is important, therefore, that computer animation be reworked from the viewpoint of the fundamentals of animation, i.e., at the basic conceptual level, and not at the level of software, implementation, techniques, etc. Of course, ultimately these things determine the efficiency of a system when actually used by animators. But these can only follow if a strong formal foundation exists. We attempt to lay this foundation in the rest of this correspondence.

If one reviews existing animation systems [2], one will see that what is presented in the remaining sections is available in some system or another. Our aim, however, is to unify and generalize the concepts for which these systems have particular implementations.

Manuscript received May 27, 1977; revised August 3, 1977.
The authors are with the National Center for Software Development and Computing Techniques, Tata Institute of Fundamental Research, Bombay 400 005, India.

II. Computer Animation Primitives

A. Picture Alteration

The image alterations in successive frames must be expressed as computable functions (procedures), which henceforth will be referred to as alter-image functions.

Notation: If f is an alter-image function and p is a picture-cel, then we shall adopt the notation $f \cdot p$ to represent the fact that f is successively applied to p to obtain the images of p.

Application of an alter-image function to a picture-cel results in another picture cell. Therefore, any number of alter-image functions may be applied in sequence before the final image of this picture-cel in a picture frame is obtained. Thus $f2 \cdot f1 \cdot p$ represents the fact that $f1$ is applied to p to obtain, let us say, p', and then $f2$ is applied to p' to obtain p'', whose image then occurs in the picture frame.

The general form is

$$fn \cdot fn-1 \cdots f2 \cdot f2 \cdot p = fn \cdot (fn-1 \cdot (\cdots (f2 \cdot (f1 \cdot p)) \cdots))$$

where the parentheses have been used to define the order of precedence of application of the alter-image functions.

B. Pattern of Alteration

The pattern of alteration also must be defined as a computable function. This function will henceforth be referred to as the alter-pattern function. The alter-pattern function determines the number of times (≥ 0) the associated alter-image function must be applied to a picture-cel in order to obtain its image in the current picture frame.

Notation: If a is the alter-pattern function, then $f\ @\ a \cdot p$ is used to represent the fact that f according to a is applied to p.

The general form of the alteration of a picture-cel is

$$fn\ @\ an \cdot fn-1\ @\ an-1 \cdots f2\ @\ a2 \cdot f1\ @\ a1 \cdot p.$$

To end the discussion on alter-functions, we shall briefly consider a small example. The sequence involves making a square into a circle. We define an alter-function as follows. Move the mid-point of every vector in the picture-cel along the radius of the circumscribing circle until the circumference is hit. The point is moved according to an alter-pattern function, let us say, a constant velocity pattern. These functions will make the square into an octagon, the octagon into a 16-sided polygon, and so on.

0018-9472/78/0400-0308$00.75