Maximum Likelihood Methods in Vowel Recognition: A Comparative Study

A. K. DATTA, N. R. GANGULI, AND S. RAY

***Abstract*—Vowel classification accuracy is studied using a generalized maximum likelihood ratio method. It is shown that two simplifying assumptions can reduce computation times by as much as a factor of five while producing practically no change in recognition accuracy. The two simplifying assumptions remove cross correlation terms and produce an Euclidean distance discriminant function. The vowels are taken from 350 multisyllabic isolated words spoken by five male speakers. The vowels occur in a variety of pre- and postconsonantal contexts. The recognition scores obtained for vowels are 83 percent. The effect of grouping of similar vowels on recognition scores is found to be marginal. The high back and high front vowels show better recognition scores (92-94 percent). In general, recognition performance for individual vowels follows a definite trend with respect to the vowel diagram. A reasonable similarity is observed between confusion matrix and the distribution of vowels in first and second formant frequency ($F_1 - F_2$) plane.**

***Index Terms*—Automatic speech recognition (ASR), discriminant score, feature extraction, intergroup, intragroup, maximum likelihood ratio, phoneme, vowel recognition.**

INTRODUCTION

Automatic speech recognition (ASR) systems attempt to translate human speech to control machines, enter data into intelligent terminals, and transform speech into scripts. Such systems often employ phonemic recognition schemes and the classification of vowels plays a key role in such schemes. Extensive investigations on the problems of automatic recognition of vowels have produced recognition scores varying from about 70 to 90 percent according to various degrees of restrictions regarding the number of speakers, number of syllables in a word, continuity of speech, variety of contexts, etc. [5]-[13]. Various classification algorithms using different features have been employed. The variability associated with the features used in ASR has prompted the use of statistical and fuzzy set theoretic approaches in this field [9], [14]-[20]. Maximum likelihood method has been shown to be a potentially effective tool [5], [17], [19], [20].

The present paper endeavors mainly to evaluate simplifying assumptions those can reduce computation time of various statistical methods based on maximum likelihood ratios in the recognition of vowels in different consonantal contexts for different speakers. The simplified version of the maximum likelihood method used here is based on the assumptions of intergroup independence of the variation of the features and intragroup independence of the features together with constant volume for the classes. A comparison of these methods with respect to accuracy of classification and computational time is given. The most notable feature is that about 85 percent of the computational time may be saved by using the simplest versions of the maximum likelihood methods with only negligible loss in recognition score. The vowel data have been collected from about 350 commonly used multisyllabic Telugu (a major Indian language) words spoken by five adult male informants. The major vowels classified are /u/, /u:/, /o/, /o:/, /ɑ/, /ə/, /e/, /e:/, /i/, and /i:/. The grouping of the similar vowels, differing mainly in duration reduces the number of classes to six, e.g., /u/, /o/, /ɑ/, /ə/, /e/, and /i/.

The paper also presents a comparison of the recognition behavior of the vowels with respect to its phonetic qualities like articulatory positions, positions in the ($F_1 - F_2$) plane and also with respect to a grouping based on the duration of vowels having same phonetic qualities. A comparison of the results with some other relevant published results is included.

Manuscript received May 12, 1980; revised March 22, 1982.

The authors are with the Electronics and Communication Sciences Unit, Indian Statistical Institute, Calcutta 700035, India.

EXPERIMENTAL PROCEDURE

The test material is prepared from a set of about 350 commonly used Telugu words recorded by five male educated native informants in the age group of 25-30 years. The nonsense words are totally avoided. The recordings are made inside an empty auditorium with an AKAI 1710 recorder on TDK tapes. The spectrographic analysis is done on a KAY Sonagraph Model 7029-A. The system bandwidth has been 80 Hz to 8 kHz with a resolution of 300 Hz.

The two main purposes of the experiment were 1) to evaluate the suitability of maximum likelihood methods for vowel identification from the measurement of formant frequencies at the steady state when these are influenced by the effect of coarticulation as obtainable in natural speech and 2) to assess the effect of various simplifying assumptions which reduce computational time on the recognition score. The recognition of vowels is known to be significantly influenced by the consonantal contexts [21]-[24]. The vowels have been shown to be better identified in a particular context, e.g., h-t and h-d [24]. A better evaluation of the recognition scheme can be made if the vowels are subjected to the perturbing effects of various consonantal contexts as prevail in a natural language. The words are therefore so selected that the vowels occur in as many pre- and postconsonantal contexts as possible. The data for all the ten major vowels have been taken in CNC (consonant-nucleous vowel-consonant) contexts with 26 nonnasal consonants (Table I). However nasal consonants like /m/, /n/, /ɳ/ do occur in the words, but not in the immediate preceding position to the selected vowels. On an average every vowel has 20 and 17 varieties of pre- and postconsonantal contexts, respectively. The longest word has four syllables and the shortest one two. The average number of syllables in the vocabulary is 2.4.

The formant frequencies are considered to be the most important features for vowel identification [4], [25], [26]. In the present experiment the frequencies of the first three formants F_1, F_2, and F_3 measured at the steady state of the vowels [1]-[3] have been taken as the features for classification. The frequencies are measured manually from the base line (i.e., zero line) of the spectrogram to the central line of formant bands where these are parallel to the time axis. The accuracy of measurement of the formant frequencies has been ±50 Hz.

The ten vowels considered for classification are /u/, /u:/, /o/, /o:/, /ɑ/, /ə/, /e/, /e:/, /i/, /i:/. Two different sets of experiments are performed. In the first set A, the similar vowels, /u/-/u:/, /o/-/o:/, /e/-/e:/ and /i/-/i:/ which differ mainly in duration (Table II), have been paired together. Here parameters for classification for all the ten vowels have been taken separately for the purpose of computing the discriminant scores. But for the purpose of assigning the class, the two vowels in each of the four vowel pairs mentioned above are

TABLE I
CONSONANTAL CONTEXT OF CLASSIFIED VOWELS

consonant	vowel /u/		/u:/		/o/		/o:/		/ɑ/		/ə/		/i/		/i:/		/e/		/e:/	
	I	F	I	F	I	F	I	F	I	F	I	F	I	F	I	F	I	F	I	F
/k/	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+	+
/kh/	+	+	-	-	-	-	-	-	+	+	+	+	+	+	-	-	-	-	+	+
/g/	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
/gh/	+	-	+	-	-	-	+	-	+	-	+	+	+	+	-	-	-	-	-	-
/ṭ/	-	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	-	+	-
/ṭh/	-	+	-	-	-	-	+	-	+	+	+	+	+	-	+	+	-	+	+	-
/ḍ/	+	+	-	+	+	-	+	+	+	+	+	+	+	+	+	-	+	+	+	-
/ḍh/	-	-	-	+	-	-	-	+	+	+	+	-	+	-	-	+	-	-	-	-
/t/	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
/th/	-	-	-	+	-	-	-	-	+	+	+	+	-	-	-	+	-	-	-	-
/d/	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-
/dh/	+	+	+	+	-	-	-	+	+	+	+	+	+	+	-	+	-	-	-	+
/p/	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+
/ph/	+	-	-	-	-	-	-	-	+	-	+	-	+	-	-	-	-	-	+	-
/b/	+	+	+	+	+	+	+	-	+	+	+	+	+	-	+	-	+	-	+	+
/bh/	+	+	+	-	-	-	+	+	+	-	+	-	+	-	+	-	-	-	+	-
/s/	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+
/ʃ/	+	+	-	+	-	-	+	+	+	+	+	+	+	+	+	-	-	-	+	+
/h/	+	+	+	-	-	+	+	+	+	+	+	-	+	-	+	-	+	-	+	+
/tʃ/	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
/tʃh/	+	+	+	-	-	-	-	-	+	-	+	-	+	-	-	-	-	-	+	-
/dz/	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
/dzh/	+	-	+	-	-	-	-	-	+	-	+	-	+	-	-	-	-	-	-	-
/l/	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
/ɭ/	-	+	-	+	-	+	-	+	+	+	+	+	+	-	-	-	-	+	-	+
/r/	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

I = Initial position

F = Final Position

\+ = Occurrence of the consonant in the specified context

\- = Absence of the consonant in the specified context

TABLE II
DURATIONS OF TELUGU VOWELS

Telugu Vowel	/u/	/u:/	/o/	/o:/	/ɑ/	/ə/	/e/	/e:/	/i/	/i:/
Duration ms	100	275	112	252	283	100	113	284	96	237

considered to belong to the same class. In the second set (set B) of experiments the similar vowels have been mixed together and considered as a single class both for the purpose of computation of parameters and for classification. Each of these two sets consists of three experiments using three different methods. Thus, altogether six experiments have been done on the same data set of about 3000 samples for all the five speakers.

The training set consisted of 20 percent of the samples drawn randomly from the whole data set. This size has been shown to be adequate for obtaining a fairly good estimate of the parameters for the representation of a class [13].

The vowel recognition schemes may be understood in the context of a general ASR system based on phonemes identification as shown in Fig. 1. The segmentation of the Fourier transform of continuous speech into phonemes and partnemes (i.e., the parts of phonemes having completely different acoustical properties, like burst and friction in affricates) is necessary for such a system. Detection of segment categories on the basis of a somewhat coarser analysis of the Fourier map helps selection of optimum feature sets and classification techniques for different phonemes. Thus there is a branching off after determination of segment categories specifically for the purpose of classification of phonemes in that category. The results of such classification go sequentially from different branches into a syntax operated reconstruction facility using linguistic rules. The classifiers used in the present experiments operate in two phases. As it is a nonadaptive system the samples are switched to the training phase initially. Here the estimates of the mean vector μ_k and the dispersion matrix

TABLE III
RECOGNITION SCORE VERSUS COMPUTATIONAL EFFICIENCY

Set	Method	Score (%)	Computational time (%)	Effect of simplifying assumption			Effect of Grouping		
				Method	Loss in score (%)	Gain in time (%)	Set	Loss in score (%)	Gain in time (%)
A	1	82.9	100	(1-2)	+ 1.5	60	(A_1-B_1)	+ 2.2	35
	2	81.6	40	(2-3)	- 1.2	50	(A_2-B_2)	0	37
	3	82.6	20	(1-3)	+ 0.4	80	(A_3-B_3)	+ 0.6	25
B	1	81.1	65	(1-2)	- 0.6	61			
	2	81.6	25	(2-3)	- 0.6	40			
	3	82.1	15	(1-3)	- 1.2	77			

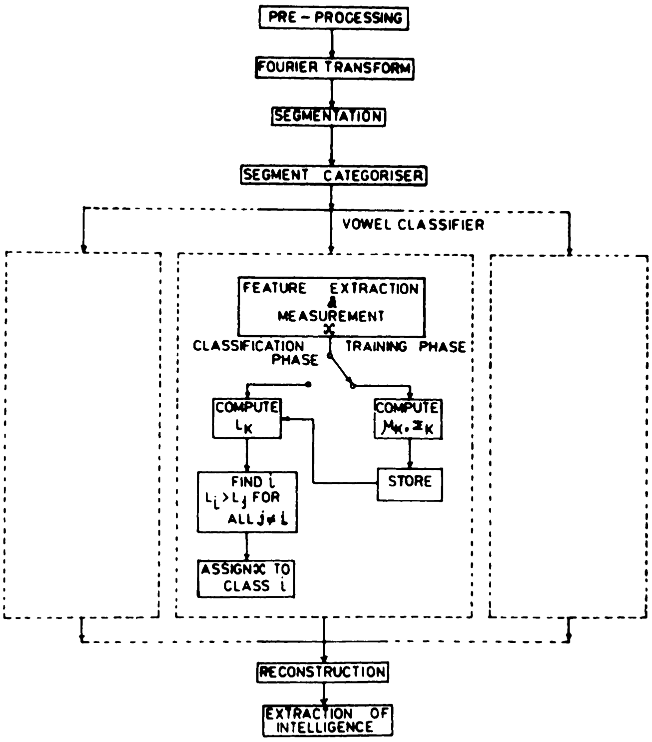

Fig. 1. Block diagram of general ASR.

Σ_k for each class are computed from the samples in the training set. These values are stored for future use in the classification phase. In the classification phase the discriminant score L_k for each input vector for all the classes are computed first. The class number i for which the score is maximum is obtained through usual sorting procedure.

METHODS OF CLASSIFICATION

Statistical classificatory methods using parametric representation need the distribution function to be known *a priori.* The formant frequencies have been reported to follow normal distribution [2]. The feature vector for classification of vowels, therefore, may be taken to be multivariate normal.

Let $x' = (x_1, x_2, \cdots, x_n)$ be an n-dimensional feature vector. Let x be multivariate normal with parameters μ_k and Σ_k, $k = 1, 2, \cdots, m$, where m is the number of groups. In the general maximum likelihood methods of classification, under the assumption of equal *a priori* probability of the groups, x is assigned to the group k for which discriminant score L_k defined by (1) is maximum [27]:

$$L_k = -\tfrac{1}{2} \log_e |\Sigma_k| - \tfrac{1}{2}(x - \mu_k)'\Sigma_k^{-1}(x - \mu_k) \tag{1}$$

where Σ_k is the dispersion matrix whose ijth component is the covariance between ith and jth features in relation to the kth class.

The assumption of group independence of the variability of the features equalizes all the dispersion matrices and consequently the expression to be maximized reduces to

$$L_k' = \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda^{ij} x_i \mu_{kj} - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda^{ij} \mu_{ki} \mu_{kj} \tag{2}$$

where

$$\Sigma_1 = \Sigma_2 = \cdots = \Sigma_m = \Sigma, \text{ say}$$

and

$$\Sigma^{-1} = (\lambda^{ij}).$$

If the features are assumed to be independent and the regions occupied by various classes are taken to be equal in volume then 1) Σ_k's are diagonal and 2) $|\Sigma_k|$'s are equal. With assumption 2), expression (1) reduces to

$$-\tfrac{1}{2}[C + (x - \mu_k)'\Sigma_k^{-1}(x - \mu_k)] \tag{3}$$

where C is a constant.

Assumption 1) reduces expression (3), after simplification, to

$$-\frac{1}{2}\left[C + \sum_{i=1}^{n} \frac{(x_i - \mu_{ki})^2}{\sigma_{ki}^2}\right]. \tag{3a}$$

Thus the classification problem reduces to minimization of the simple expression

$$D_k = \sum_{i=1}^{n} \frac{(x_i - \mu_{ki})^2}{\sigma_{ki}^2}. \tag{3b}$$

It may be noted that D_k is the square of the Euclidean distance of the sample weighted by the inverse of variance from the mean of the kth pattern class.

Altogether three classification methods as has been explained above were used. Method 1, 2, and 3 referred in Tables III and IV used expressions (1), (2), and (3b), respectively.

TABLE IV
VOWELWISE RECOGNITION SCORE WITH SUBGROUPING AND WITHOUT SUBGROUPING FOR DIFFERENT METHODS

	Method					
	1		2		3	
	set		set		set	
Vowel	A	B	A	B	A	B
/u/	96.0	94.1	95.4	94.1	96.7	90.1
/o/	79.0	78.1	76.4	78.7	75.3	77.6
/ɑ/	86.9	88.8	88.0	90.8	84.0	84.0
/ə/	60.8	62.3	54.0	59.6	59.7	69.3
/e/	75.0	69.2	76.4	64.4	82.2	81.3
/i/	92.1	92.1	90.4	92.7	87.5	87.5

TABLE V
CONFUSION MATRIX FOR EXPERIMENT I SET A

		Recognised as										
		/u/	/u:/	/o/	/o:/	/ɑ/	/ə/	/e/	/e:/	/i/	/i:/	Total No. of Observations.
Actual class	/u/	157	56	7	12	0	1	0	0	0	0	233
	/u:/	20	253	0	0	0	0	0	0	0	0	273
	/o/	57	6	131	66	13	4	0	0	0	0	277
	/o:/	36	10	24	253	0	0	0	0	0	0	323
	/ɑ/	0	0	11	0	259	28	0	0	0	0	298
	/ə/	1	0	13	0	50	146	30	0	0	0	240
	/e/	5	0	3	0	0	32	254	74	0	0	368
	/e:/	3	0	24	0	0	6	70	121	76	24	324
	/i/	0	0	0	0	0	0	6	26	139	114	285
	/i:/	0	0	0	0	0	0	0	14	22	259	295
		Total										2916

EXPERIMENTAL RESULTS AND DISCUSSION

Effect of Simplifying Assumptions and Grouping

Table III presents the average recognition scores of the two sets of experiments and the corresponding average computation time for classification of a sample in each experiment expressed in percent of average computation time for method I in set A. It also summarizes the differences between the sets A and B and between the three methods of classification with respect to the behavior of recognition score and the computation time. It is immediately apparent from the table that the various degrees of simplification used in the classification technique, while affecting the computation time significantly, affect the recognition score only marginally. Similar experiments with respect to recognition of consonants have shown larger differences [19]. It may be of interest to observe that while method 3 of set A requires only 20 percent of the computation time required by the most generalized approach, corresponding effective loss in the recognition performance is only 0.3 percent. The result is somewhat unexpected. The correlation coefficients computed from the covariance matrices used in the classification show negligible correlation of the first formant with the other two formants. In some cases, however, the second formant has been found to be well correlated with the third formant. But in general one could observe that the correlations between the features are low and hence the result.

In the present experiment the difference between the highest and lowest scores, respectively for set A and set B with method 1, is only 2.2 percent of the actual performance, i.e., of 81.1 percent. The effect of grouping has been observed by several researchers in the classification of vowels [5]. In the

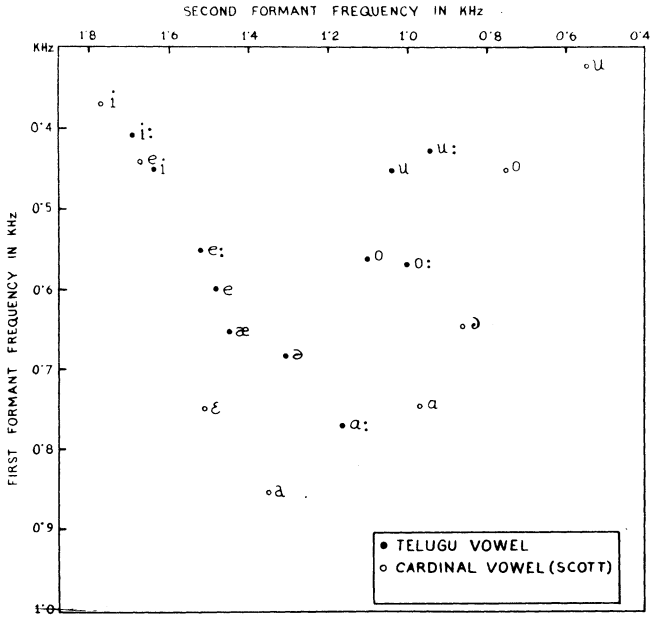

Fig. 2. Telugu vowel diagram.

experiments of set A the vowels having short and long variants were subgrouped. The resulting improvement in the performance of classification, though consistent, is really small (Table III). A reference to the confusion matrix (Table V) at once reveals that the misclassification of these subgroups primarily lie within themselves. In the recognition of vowels, therefore, these vowels can be taken together for primary classification. The longer variants can be easily distinguished from the shorter counterparts by examining the duration of vowels (Table II). In fact, Table II shows that the durations of longer are more than double of those of the shorter.

Recognition Behavior of Vowels

The vowels have been recognized with the score in the range of 70–92 percent depending on the number of speakers, size of vocabulary, variety of context, classificatory techniques, and features selected [5]–[13]. However, results reported from experiments in environments similar to the present exhibit comparable performances. Pal *et al.* [9], Plom *et al.*, [11], and Pols *et al.*, [5] reported recognition scores of 82, 85, and 87 percent, respectively. Particular mention may be made of the work of Pols *et al.*, who used an approach very similar to the present one. They used a larger number of speakers though the vowels were used in monosyllabic h–t contexts. Thus the variance due to the variability of context and effect of coarticulation of distant phonemes was absent. The results obtained in the present experiment 83 percent are comparable to the above results.

Table IV represents the recognition score of each vowel corresponding to all six experiments. High vowels /i/ and /u/ show consistently better recognition scores in all the six experiments. Midvowels /ə/, /e/, and /o/, particularly /ə/, show comparatively poor recognition scores. The average position of different Telugu vowels (in mel scale) with respect to the cardinal vowels (Scott, set 4) [28] is shown in Fig. 2. The vowel /æ/ in vowel diagram is excluded from present classification. The number of samples were so small that an effective statistical classification for this vowel could not be undertaken. The diagram reveals that the vowels /i/ and /u/ occupy two end positions and have interfering vowels only on one side. Furthermore, the area of confusion of these two vowels in proportion to the total respective areas is smallest (Fig. 3). Vowel /ɑ/ also has only one significant confusion area, that is with /ə/. The recognition score for /ɑ/ is also good. The mid central vowel /ə/ has the largest area of confusion with the three adjoining vowels /ɑ/, /e/, and /o/. Consequently, error of classification for this vowel is also high. Referring to the vowel diagram for Telugu vowels (Fig. 2) one may observe a general trend for the performance of individual vowels. The recognition scores for vowels occupying the end positions are high. The recognition score for vowel /ə/ which is most centrally placed in the vowel diagram is lowest. It may be observed that whenever a vowel is between two other vowels in the vowel diagram the recognition score becomes low.

The confusion of vowels in classification as revealed in the confusion matrix conforms with the distribution of vowels in the $(F_1 - F_2)$ plane. It may be seen from the confusion matrix that the errors are mainly confined to the neighboring squares (Table V) which correlate with the neighboring region in Fig. 3. Moreover, the amount of errors can also be seen to follow the area of confusion in proportion to the total area.

Conclusion

The use of the special case of maximum likelihood method, with the assumptions of independence of features and equal volume of class regions, cuts down the effective computational time to a great extent without significant loss in performance. The computation time may be further reduced by reducing the number of classes through the grouping of similar vowels, differing mainly in durations. As a general rule the vowels in the free ends of the vowel diagram show better recognition scores.

Encouraging results for classification of unaspirated plosives using same methods and same basic acoustic features have been reported earlier [19]. One can reasonably visualize the application of the same technique for aspirated plosives and laterals/semivowels which respectively resemble unaspirated plosives and vowels. A primary ASR system based on phoneme recognition at the primary level may be developed along this line provided other phonemes particularly the fricatives could be identified in a similar way. Some investigations in this area

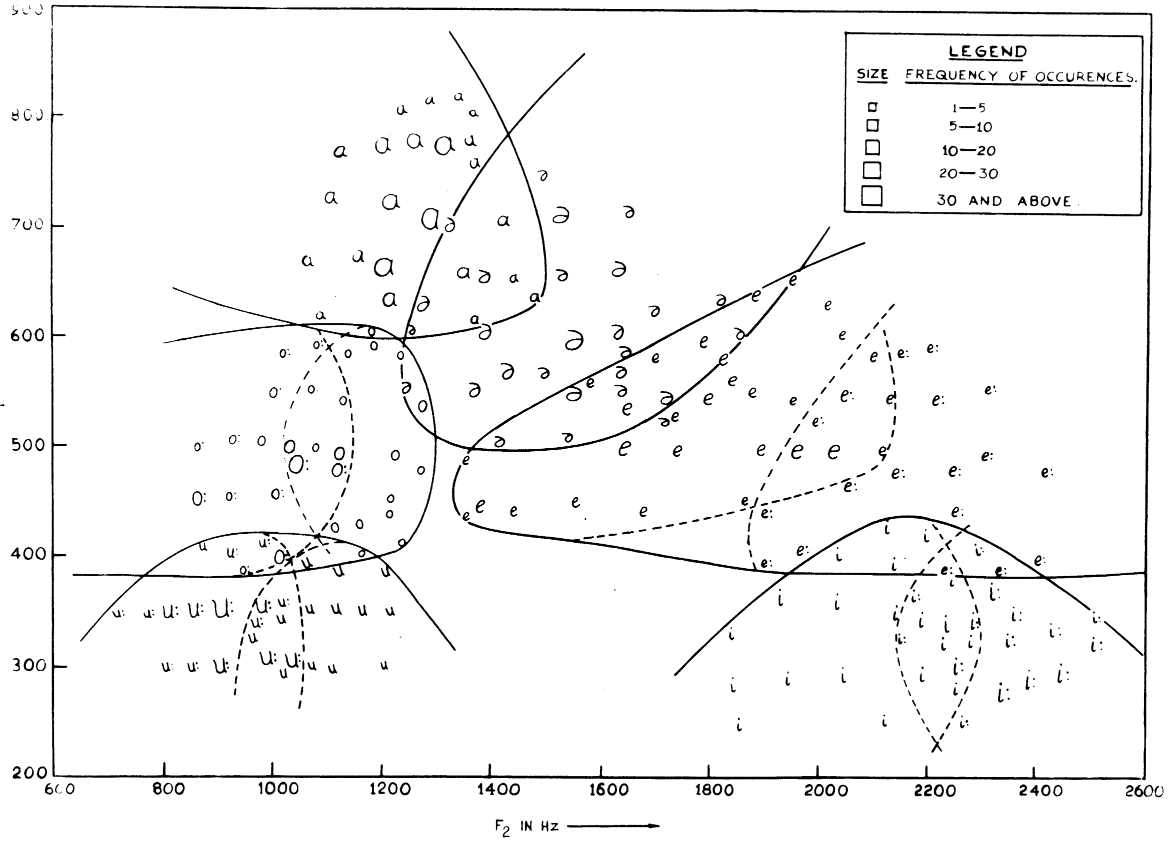

Fig. 3. Distribution of vowels in $(F_1 - F_2)$ plane.

are, therefore, necessary before an ASR system based on maximum likelihood method of classification can be realized.

Acknowledgment

The authors express their thanks to Prof. D. Dutta Majumder, Head, Electronics and Communication Sciences Unit, Indian Statistical Institute, for his valuable suggestions, to the Andhra Association of Calcutta for voice recording and S. De Bhowmick for typing work.

References

[1] G. Fant, *Acoustic Theory of Speech Production*. 's-Gravenhage: Mouton, 1960.

[2] D. Dutta Majumder, A. K. Datta, and N. R. Ganguli, "Acoustic feature of Telugu vowels," *Acustica*, vol. 41, pp. 55-64, 1978.

[3] D. Dutta Majumder, A. K. Datta, and N. R. Ganguli, "Some studies on acoustic features of human speech in relation to Hindi speech sounds," *Indian J. Phys.*, vol. 47, pp. 598-613, 1973.

[4] H. Fujasiki, N. Nakamura, and K. Yoshimune, "Analysis, normalization and recognition of sustained Japanese vowels," *J. Acoust. Soc. Japan*, vol. 26, pp. 152-154, 1970.

[5] L. C. W. Pols, H. R. C. Tromp, and R. Plomp, "Frequency analysis of Dutch vowels from male speakers," *J. Acoust. Soc. Amer.*, vol. 53, pp. 1093-1101, 1973.

[6] M. W. Cannon, "A method of analysis and recognition for voiced vowels," *IEEE Trans. Audio Electroacoust.*, vol. AU-16, pp. 154-158, 1968.

[7] T. Sakai and S. Doshita, "The automatic speech recognition system for conversational sound," *IEEE Trans. Electron. Comput.*, vol. EC-12, pp. 835-846, 1963.

[8] J. W. Forgie and C. D. Forgie, "Results obtained from a vowel recognition computer program," *J. Acoust. Soc. Amer.*, vol. 31, pp. 1480-1489, 1959.

[9] 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.

[10] Bezdel and Chandler, "Results of analysis for recognition of vowels by computer using zero crossing data," *Proc. IEEE*, vol. 112, p. 2000, 1963.

[11] R. Plomp, L. C. W. Pols, and J. P. Van De Geer, "Dimensional analysis of vowel spectra," *J. Acoust. Soc. Amer.*, vol. 41, pp. 707-712, 1967.

[12] W. Klein, R. Plomp, and L. C. W. Pols, "Vowel spectra, vowel spaces, and vowel identification," *J. Acoust. Soc. Amer.*, vol. 40, pp. 999-1009, 1970.

[13] D. Dutta Majumder, A. K. Datta, and S. K. Pal, "Computer recognition of Telugu vowel sounds," *J. Comput. Soc. India*, vol. 7, pp. 14-28, 1976.

[14] F. Itakura, "Minimum prediction residual principal applied to speech recognition," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-23, pp. 67-72, 1975.

[15] M. R. Sambur and L. R. Rabiner, "A statistical decision approach to the recognition of connected digits," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-24, pp. 550-558, 1976.

[16] P. V. de Souza, "Statistical tests and distance measures for LPC coefficients," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-25, pp. 554-559, 1977.

[17] F. Itakura and S. Saite, "Analysis synthesis telephony based on the maximum likelihood method," in *Proc. 6th Int. Congr. Acoust.*, 1968, paper C-5-5.

[18] F. Jelinek, "Continuous speech recognition by statistical methods," *Proc. IEEE*, vol. 64, no. 4, 1976.

[19] A. K. Datta, N. R. Ganguli and S. Ray, "Recognition of unaspirated plosives: A statistical approach," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, no. 1, pp. 85-91, 1980.

[20] A. K. Datta, N. R. Ganguli, and S. Ray, "Computer recognition of consonantal speech sound," in *Proc. 4th Int. Joint Conf. Pat-*

tern Recognition, Kyoto, Japan, Session-B8, Nov. 1978, pp. 1047-1049.

[21] D. B. Fry, A. B. Abramson, P. D. Eimas, and A. M. Liberman, "The identification and discrimination of synthetic vowels," *Language and Speech*, vol. 5, pp. 171-189, 1962.

[22] K. N. Stevens and A. S. House, "Perturbation of vowel articulation by consonantal context: An acoustical study," *J. Speech Hearing Res.*, vol. 6, pp. 111-128, 1963.

[23] B. E. F. Lindblom and M. Studdert-Kenedy, "On the role of formant transitions in vowel recognition," *J. Acoust. Soc. Amer.*, vol. 42, p. 830, 1967.

[24] J. B. Miller and W. A. Ainsworth, "Identification of synthetic isolated vowels and vowels in h–d context," *Acustica*, vol. 27, pp. 278-282, 1972.

[25] P. C. Delattre, A. M. Liberman, F. S. Cooper, and L. J. Gerstman, "An experimental study of the acoustic determination of vowel color: Observations on one-and-two formant vowels synthesized from spectrographic patterns," *World*, vol. 8, pp. 195-210, 1952.

[26] L. J. Gerstman, "Classification of self normalized vowels," *IEEE Trans. Audio Electroacoust.*, vol. AU-16, 1968.

[27] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*. New York, Wiley, 1958, p. 147.

[28] P. Ladgefoged, *Three Areas of Experimental Phonetics*. Oxford: Oxford Univ. Press, 1967.

Correction to "Segmentation of Images Having Unimodal Distributions"

BIR BHANU AND OLIVIER D. FAUGERAS

In the above paper,[1] three figures have been inadvertently switched. The correct figures are as follows.

Fig. 7(a) should be Fig. 7(e).
Fig. 7(e) should be Fig. 7(a).
Fig. 8(c) should be Fig. 8(e).
Fig. 8(e) should be Fig. 8(c).
Fig. 11(j) should be Fig. 11(l).
Fig. 11(l) should be Fig. 11(j).

Manuscript received July 12, 1982.

B. Bhanu is with the Aeronutronic Division, Ford Aerospace and Communications Corporation, Newport Beach, CA 92660.

O. D. Faugeras is with INRIA, Rocquencourt, France, and the University of Paris XI, Paris, France.

[1] B. Bhanu and O. D. Faugeras, *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-4, pp. 408-419, July 1982.