Reprinted from PROCEEDINGS OF IEEE-IECEJ-ASJ INTERNATIONAL CONFERENCE
ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, April 7—11, 1986

NEW CLUSTERING ALGORITHMS APPLIED TO SPEAKER INDEPENDENT ISOLATED WORD RECOGNITION

A. Mokeddem* - H. Hugli - F. Pellandini

IMT - Université de Neuchâtel
71, Rue de la Maladière - 2000 NEUCHATEL - Switzerland

ABSTRACT

This paper adresses speaker independent isolated word recognition (SIIWR) performed by an approach which uses multiple template references. In order to find the references we propose to use clustering algorithms based on criterion functions. After the theoretical description of these criterion based clustering algorithms applied to SIIWR, we give the results of tests performed with an isolated word recognizer showing that they perform better than the algorithms previously used in SIIWR.

1. INTRODUCTION

The aim of speaker independent speech recognition is to recognize the speech pronounced by any speaker from a large population. The main problem encountered is the large variation of pronounciation of a same utterance among different speakers. To solve it, the multiple template reference approach [1]-[5] turned out to be sucessful. In this approach one tries to describe all different pronounciations by few and typical pronounciations which, then, represent them all during recognition. They form, once grouped according to the utterance they represent, multiple template references. In selecting the representatives, clustering techniques can be used.

This approach admits implicitly the following hypothesis: among a given population, there exist different typical pronounciations of each word. The question arises whether such groups or clusters of utterances really exist. In order to answer it, we studied in previous work the distribution of the isolated word utterances pronounced by various speakers. Using factor analysis to visualize them, we showed [9],[10] the existence of clusters and justify the hypothesis made above.

In selecting the representatives of a word, automatic clustering can be used. The various known methods can be divided into two categories according to the way they build up clusters. We call cluster-parallel such procedures where all clusters are built up simultaneously and cluster sequential such procedures where clusters are built up sequentially. Among the clustering procedures previously applied to speaker independent speech recognition, k-means [5] (or basic ISODATA [6]) and ISODATA [4] are of the first category, UWA and UFA [2] are of the second. Here we propose two criterion based clustering algorithms. The first proposed, the Criterion based Exchange algorithm (CEx), is a cluster-parallel clustering procedure which improves the partition quality iteratively by transfering elements from cluster to cluster. These element transfers are governed by a criterion function.

The second proposed, the Criterion based Threshold algorithm (CTh), is a cluster-sequential clustering procedure that iteratively removes from the set of elements to be clustered the elements forming the best cluster. At each iteration, the best cluster is chosen among all clusters found by distance thresholding: it is the one which minimizes a criterion function.

2. RECOGNIZER

Here we describe shortly the recognizer used for the tests. The short term energy spectrum is measured every 10 ms by a 14 channel filter bank, covering logarithmicaly the frequency range from 75 Hz to 4800 Hz resulting in a spectrogram $x(k,l)$ where k is the k-th channel and l the l-th instant of sampling. After both amplitude and time normalization [7],[10], we obtain the utterance features as a 14*20 binary matrix X.

The time alignement is done by dynamic time warping (DTW). In the particular form used here, the path range is extended at both the begining and the end of the two words to be compared in such a way that word limit detection errors can be compensated [10].

3. CRITERION BASED CLUSTERING PROCEDURES

Given C, the set of all the elements X_i, i=1...I (different pronounciations of the same word). We look for the few representatives Rk, k=1...m (multiple templates) which describe all elements X_i. Basically the solution has two distinct steps : 1) the clustering itself which divides C into disjoint clusters Ck, k=1...m in such a way that a given criterion be fulfilled, 2) the choice of the representative Rk of each cluster Ck.

CH2243-4/86/0000-2691 $1.00

3.1 Criterion based Exchange procedure CEx)

The CEx is a cluster-parallel iterative clustering procedure producing a fixed number m of clusters. It minimizes a criterion function F by iteratively transfering elements from cluster to cluster in such a way that F decreases.

Algorithm CEx :

1. Choose any initial partition C1, C2,..., Cm

2. a) Find Xi of C for which there exists a cluster Cl such that the transfer of Xi from its cluster Ck to cluster Cl decreases F : $\Delta F(Xi, Ck \rightarrow Cl) < 0$

   b) Stop if no such element Xi of C can be found $\forall Xi \in C$, $l=1,\ldots,m$: $\Delta F(Xi, Ck \rightarrow Cl) \geq 0$

3. Transfer Xi to the cluster Ck' which minimizes ΔF, i.e.
   $\Delta F(Xi, Ck \rightarrow Ck') = \underset{l}{\text{Min}}\ \Delta F(Xi, Ck \rightarrow Cl)$

4. Go to 2.

3.2 Criterion based Threshold procedure (CTh)

The CTh is a cluster-sequential clustering procedure that, as clusters are created, gradually removes the elements from C', the set of elements still to be clustered, until C'is empty. At each iteration a cluster Ck is created which fulfills the threshold condition, i.e., a cluster of elements Xj around a center element Xi whose distances d(Xi,Xj) do not exceed an a priori fixed threshold T.
Now the important point is that, at each iteration, among all possible clusters A(Xi) fulfilling the threshold condition :

$$A(Xi) = \{Xj \in C' \ / \ d(Xi,Xj) < T\}$$

the best is selected, i.e. the one that minimizes the criterion function H(A(Xi)) measuring the homogeneity in A(Xi). Note that, in this case, H applies to a sole cluster.

Algorithmn CTh :

1. Initialization :
   k = 1 (k-th cluster)
   C' = C

2. For each $Xi \in C'$ find the candidate-cluster A(Xi) :
   $A(Xi) = \{Xj \in C' \ / \ d(Xi,Xj) < T\}$

3. Find the candidate-cluster minimizing the criterion function :
   $Ck = A(Xi^*) \ / \ H(A(Xi^*)) \leq H(A(Xi)) \quad \forall Xi \in C'$
4. $C' = C' - Ck$

5. If $C' \neq 0$ then : k = k+1 and Go to 2.
   Else : Stop

With CTh, the number of clusters created is variable and depends on T.

3.3 Criterion function

3.3.1 Definitions

To an element Xi of cluster Ck we associate the following general metrics :

$$L_q(Xi,Ck) = \left(\frac{1}{n_k - 1} \sum_{Xj \in Ck} d(Xi,Xj)^q \right)^{1/q}$$

Note that for q=1, we obtain the mean of distances between all Xj and Xi. For q=2, we obtain the rms value of these distances. For $q=\infty$, we obtain the maximum distance.
From $L_q(Xi,Ck)$ several metrics may be derived which measure the homogeneity of a cluster. These metrics will be used to define criterion functions for CEx and CTh procedures. Let us define the following homogeneity function :

$$M_q(Ck) = \underset{Xi \in Ck}{\text{Min}}\ L_q(Xi,Ck)$$

3.3.2 Criterion function for CEx

Clustering procedures defined above minimize a criterion function which is supposed to measure the quality of a partition. The real world problem is to find which criterion function really measures the partition quality in SIIWR. Several criterion functions were defined and tested. The results published elsewhere [10], have shown the existence of different classes of criterion functions with different behaviour and recognition performance. The following class of criterion functions F_q behaved well :

$$F_q = \sum_k M_q(Ck).(n_k - 1)$$

where n_k is the number of elements in Ck

3.3.3 Criterion function for CTh

The criterion functions for CTh procedures is defined for only one cluster. Here also, various criterion functions based on the homogeneity functions and the number of elements in the candidate-cluster A were defined and tested [10]. The results have shown that the criterion functions based on the pure homogeneity functions behave best. Thus, we use the following criterion function

$$H_q = M_q(A)$$

3.4 Representative of a cluster

The representative Rq(Ck) of a cluster Ck is chosen as follows : Rq(Ck) is the element Xi* of Ck that minimizes the metric Lq (Xi,Ck) in the cluster Ck :

$$Rq(Ck) = Xi^* \text{ such that } Lq(Xi^*,Ck) = \underset{Xi \in Ck}{\text{Min}}\ Lq(Xi,Ck)$$

Note that for $q=\infty$, we obtain the minimax.

All representatives of a given word are used as template to be matched during recognition.

4. EXPERIMENTAL RESULTS

Open tests were performed with the isolated word recognizer already described and using a 13-word french vocabulary (three control words : en-avant, en-arrière, terminer, and the ten digits : zero, un..., neuf). The training data set consisted of one repetition of each word by 25 male and 25 female speakers, while another set of three repetitions of each word by 25 male and 5 female speakers was used for recognition tests.

a) Recognition results with CEx and k-means.

Figure 1 compares the recognition results obtained by cluster-parallel algorithms. The proposed CEx algorithm and the previously used k-means algorithm (basic ISODATA) [4], [6] are compared. For both algorithms, two different ways of selecting the representatives of a cluster were considered : R1(Ck) and R_∞(Ck). Note that for CEx the corresponding criterion function are F1 and F_∞. We can observe, in all cases, the superiority of CEx with respect to k-means in our tests.

b) Recognition results with CTh and UWA.

Figure 2 compares the recognition results obtained by cluster-sequential algorithms. The proposed CTh algorithm and the previously used UWA algorithm [1], [2] are compared. For these algorithms, the threshold, for each word, is chosen iteratively in such a manner that we obtain a number of clusters not exceeding m (m = 6,4,...,1). The representative of a cluster is R_1 (Ck). The criterion-function for CTh is H_1.

We can observe clearly the superiority of CTh algorithm with respect to UWA algorithm in our tests.

c) Variable number of templates per word.

In a given vocabulary, there exist words which are easy to recognize and others which are difficult to recognize. If the total number of reference templates for that vocabulary is fixed, it seems reasonable, a priori, to assign a larger number of templated for words which are difficult to recog - nize than for words which are easy to recognize. Therefore the number of clusters per word m(w) should somehow increase with increasing error rate e(w) of that given word w. As for instance in [9], [10] :

$$m(w) = \frac{e(w)}{\bar{e}}(\bar{m} - 1) + 1$$

where $\bar{e}$ is the global error rate and $\bar{m}$ the average number of reference templates per word.

Figure 3 compares the recognition error rate obtained by the CTh algorithm in the following two cases :
1) CTh algorithm with the same number of clusters for each word, i.e. m(w) = $\bar{m}$;
2) CTh algorithm with a variable number of clusters per word m(w) according to the formula above.

It may be observed that the error rate is significantly reduced when the number of clusters is kept variable as in the second case.

In summary, the three performance figures quantitatively measure :
- the advantage of the CEx algorithm against the k-means algorithm
- the advantage of the CTh algorithm against the UWA algorithm
- the advantage of using a variable number of templates per word.

5. CONCLUSION

In this paper, we presented the two criterion based clustering procedures CEx and CTh and tested their performance. Open tests conducted with an isolated word recognizer have shown the advantage of CEx and CTh algorithms over respectively k-means and UWA algorithms often used in speaker independent speech recognition. Finally, using the CEx algorithm we point out the advantage of the use of a variable number of templates per word.

ACKNOWLEDGEMENTS

This work was supported by the "Commission pour l'Encouragement des Recherches Scientifiques" (CERS no 1158, Bern, Switzerland) and the following companies : ASULAB SA, CEH SA, METTLER SA, HASLER SA, AUTOPHON SA and CIR SA.
We wish to thank the speakers who participated in generating our speech data base.

REFERENCES

[1] L.R. Rabiner, "On Creating References for Speaker Independent Recognition of Isolated Word", IEEE Trans. on ASSP, Vol. ASSP, No 3, pp. 34-42, Feb. 1978

[2] L.R. Rabiner, J.G. Wilpson, "Considerations in Applying Techniques to Speaker-independent Word Recognition", J. Acoust. Soc. Am., Vol. 66, No 3, Sept. 1979.

[3] Niles Les, Harvey F. Silverman, N. Rex Dixon, "A comparison of Three Feature Vector Clustering Procedures in a Speech Recognition Paradigm". Proc. ICASSP 83, pp. 765-768, 1983.

[4] S.E. Levinson, L.R. Rabiner, A.E. Rosenberg and J.G. Wilpson, "Interactive Clustering Techniques for Selecting Speaker-Independent Reference Templates for Isolated Word Recognition", IEEE Trans. on ASSP, Vol. ASSP-27, No 2, April 1979.

[5] J.G. Wilpon, L.R. Rabiner, "A modified k-means clustering Algorithm for Use in Isolated Word Recognition", IEEE Trans. on ASSP, Vol. ASSP-33, No 3, June 1985.

[6] Duda R.O., Hart P.E., "Pattern Classification and Scene Analysis", John Willey, New York, 1973.

[7] Hügli H., Mokeddem A., "Reconnaissance du locuteur et de mots isolés par des systèmes miniaturisés: une comparaison". Proc. Journées d'Electronique 1985, Lausanne, Switzerland, Oct. 1985.

[8] Mokeddem A., Hügli H., Pellandini F., "Evaluation of criterion based clustering procedures for generating multiple template references in speaker independent speech recognition", 7th ICPR, August 1984, Montreal.

[9] Mokeddem A., Hügli H., Pellandini F., "Criterion based clustering techniques applied to speaker independent speech recognition", Proc. of "Digital Processing of Signals on Communications", at University of Loughbrough, England, April 1985.

[10] Mokeddem A., "Analyse de la parole : reconnaissance multilocuteur de mots isolés pour les systèmes miniaturisés", Doctoral thesis, Institut de Microtechnique de l'Université de Neuchâtel, 1985.

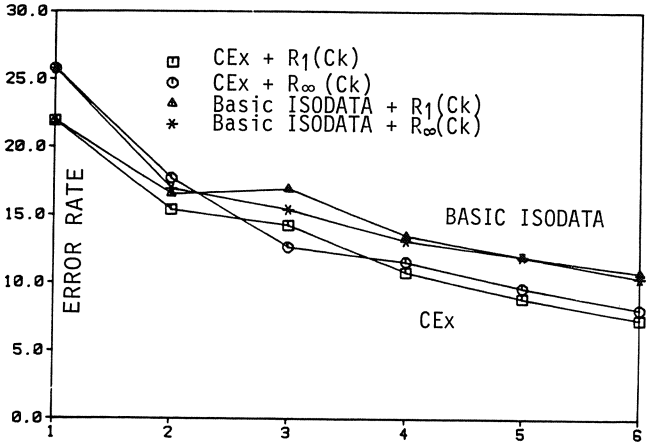

Fig. 1 Recognition error rate using CEx respectively Basic ISODATA clustering procedures

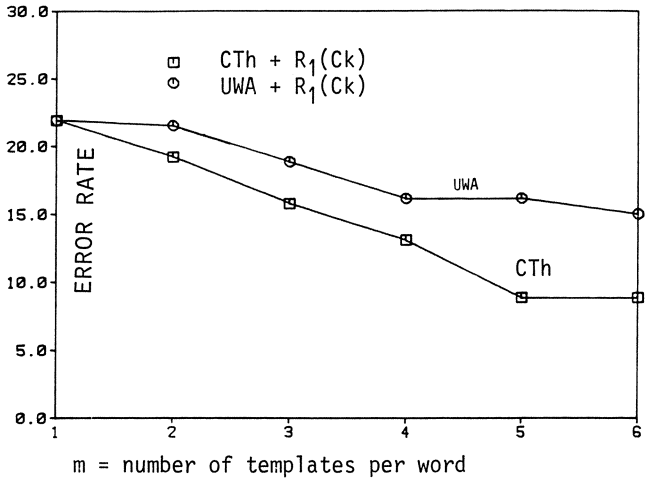

Fig. 2 Recognition error rate using CTh respectively UWA clustering procedures

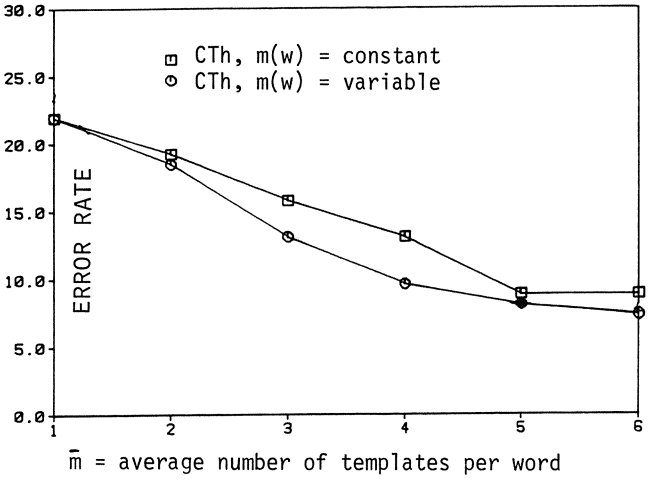

Fig. 3 Recognition error rate for constant respectively variable number of templates per word m(w)

* Institut d'électronique de l'USTO - ORAN - ALGERIA