TABLE I
Results of the test on the method proposed in this paper

Speakers	V→V	V→UV	%	UV→UV	UV→V	%
F1	383	1	0.3	91	4	4.2
F2	363	4	1.1	106	7	6.2
M1	413	0	0	59	5	7.8

TABLE II
Results of the test on the method of Atal - Rabiner

Speakers	V→V	V→UV	%	UV→UV	UV→V	%
F1	372	12	3.1	94	1	1.1
F2	361	6	1.6	106	7	6.2
M1	410	3	0.7	62	2	3.1

V→V denotes the number of Voiced segments classified as Voiced; V→UV denotes the number of Voiced segments classified as Unvoiced; and so on. % denotes the misclassification rate.

On the whole the method here proposed appears offering attractive performances for further evaluation; computer performances are indeed comparable with those obtained by using the method of Atal-Rabiner. Then, in particular, the examination of the sequence of the classifier decisions has shown the uselessness of a smoothing algorithm after the decisor,useful in other methods(e.g. [1]).

4.CONCLUSIONS

In this paper we have presented a method of Voiced/Unvoiced classification based on the maximum a posteriori probability criterion , which has given good experimental results.

This method is a semplified version of more sophisticated Voiced /Unvoiced /Silence classification criterion, in which either the a priori probabilities of the classes either the p.d.f. of the measurements vector $\underline{x}_k$ are updated.We are actually testing this more complex method of classification.

- REFERENCES

[1] B.S. Atal - L.R. Rabiner,"A pattern recognition approach to Voiced /Unvoiced/Silence Classification with applications to Speech recognition" . IEEE-Trans.Acoust.Speech Signal Processing. Vol.ASSP-24 No.3, June 1976.

[2] L.J. Siegel,"A procedure for using pattern claffication technique to obtain a Voiced/Unvoiced classifier". IEEE-Trans.Acoust.Speech Signal Processing. Vol.ASSP-27 No.1, Febr. 1979.

G.BRUNO(*),M.D.DI BENEDETTO(**),M.G.DI BENEDETTO(***),A.GILIO(*)
(*) Ist.Comunicazioni Elettriche, Fac.Ingegneria, Univ. Roma
(**) IBM France, Centre Scientifique de Paris
(***) Fac.Ingegneria, Univ. Roma (student)

A BAYESIAN APPROACH FOR VOICED/UNVOICED CLASSIFICATION OF SEGMENTS OF A SPEECH SIGNAL

ABSTRACT - In this paper a method for Voiced/Unvoiced classification of segments of a speech signal is presented. For this purpose, on each segment four different measurements are made. On the basis of such measurements the a posteriori probabilities of the two classes are determined and the decision is made by means of the maximum a posteriori probability criterion. The a priori probabilities of the classes are updated for each segment . In this context, it is assumed that the sequence of the classes constitutes a 1st order time-omogeneous Markov chain. The experimental results of the method,included in the paper, have been satisfactory.

1. INTRODUCTION

In the area of analysis and synthesis of the speech signal and , in particular, in that of continuous speech recognition, a relevant factor consists in discriminating, among the sounds produced by the vocal tract, those which are voiced from those which are unvoiced. For this purpose, the signal is subdivided into segments, typically of 10 to 20 msec duration . On each segment 4 types of measurements are made, on the basis of which the Voiced/Unvoiced class is chosen. The methods proposed by previous authors are of two types: probabilistics (see e.g. [1]) and deterministics(see e.g. [2]).

The method here presented (sec.2) is probabilistic. In particular , for each segment, on the basis of the measurements,the a posteriori probabilities of the two classes are determined and the class, for which the above probability is maximum, is chosen. In addition, using the a posteriori probabilities that served for choosing the segment class and assuming that the sequence of segment classes constitutes a 1st order time-omogeneous Markov chain, the a priori probabilities of the classes,which are necessary for the classification of the successive segment, are updated. Finally,it is assumed that each vector composed by the 4 different measurements has a multidimensional gaussian p.d.f., with parameters which depend upon the class to which the segment is supposed to belong.

In this paper we give results obtained by preliminary experimentation of the method proposed (sec.3), carried out by using 4 different measurements, appropriately chosen among those which are commonly employed [1], [2]. The training has been made by employing two female speakers, while the testing has been made on two female speakers, one of which was used also in the training, and one male speaker.

2. CLASSIFICATION METHOD

Let C_1 and C_2 indicate the classes of the voiced and unvoiced sounds respectively. Let S_k, $S_k \in \{C_1, C_2\}$, represent the unknown class to which

the kth segment to be classified belongs and $\underline{x}_k$ the corresponding measurements vector. By denoting with $f(\underline{x}_k/C_j)$ the p.d.f. of $\underline{x}_k$ conditioned to C_j , it is assumed for sake of simplicity, but with no loss of generality , that:

$$f(\underline{x}_k/C_j) = (2\pi)^{-2}\left|\underline{R}_j\right|^{-1/2}\exp\left[-1/2\ (\underline{x}_k-\underline{m}_j)^t\underline{R}_j^{-1}(\underline{x}_k-\underline{m}_j)\right] = g(\underline{x}_k;\ \underline{m}_j,\underline{R}_j)\ ,\quad j=1,2\ ,\ \forall\ k \tag{1}$$

The classification of the kth segment and the procedure for the preliminary updating of the successive classification are made as follows:

a) On the basis of the a priori probabilities $P_k(C_j)$, j=1,2, and of the observation $\underline{x}_k$, the a posteriori probabilities $P_k(C_j/\underline{x}_k)$, j=1,2, are computed;

b) The decision is made for the class for which the a posteriori probability is maximum;

c) Using the a posteriori probabilities $P_k(C_j/\underline{x}_k)$, j=1,2, and the transition probabilities $p_{ij} = P(C_j/C_i)$, i = 1,2, j = 1,2, the a priori probabilities $P_{k+1}(C_j)$, j=1,2, which are necessary for the classification of the (k+1)th segment are computed. The a priori probabilities $P_1(C_j)$, j = 1,2, are estimated on the basis of the training data.

A schematic diagram of the algorithm is illustrated in fig.1 .

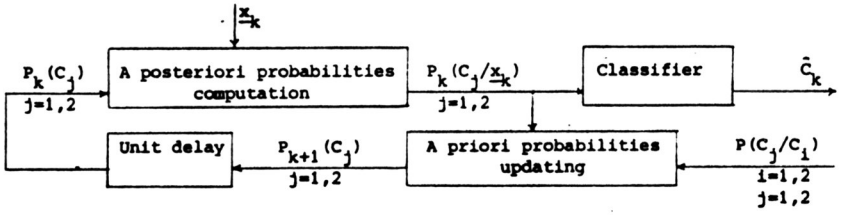

fig.1

Let us now consider the various steps of the algorithm:

a) Since $P_k(C_j) = P(S_k=C_j/\underline{x}_1,\ldots,\underline{x}_{k-1})$ and $P_k(C_j/\underline{x}_k) = P(S_k=C_j/\underline{x}_1,\ldots,\underline{x}_k)$, from Bayes formula we obtain:

$$P_k(C_j/\underline{x}_k) = \frac{P_k(C_j)f(\underline{x}_k/S_k=C_j,\underline{x}_1,\ldots,\underline{x}_{k-1})}{\sum_{h=1}^{2}P_k(C_h)f(\underline{x}_k/S_k=C_h,\underline{x}_1,\ldots,\underline{x}_{k-1})}\ ,\quad j=1,2,\ \forall\ k \tag{2}$$

In addition, supposing that, $\forall$ k, it holds:

$$f(\underline{x}_k/S_k=C_j,\underline{x}_1,\ldots,\underline{x}_{k-1}) = f(\underline{x}_k/S_k=C_j)\ ,\quad j=1,2 \tag{3}$$

from eqs (1), (2) and (3) we have:

$$P_k(C_j/\underline{x}_k) = \frac{P_k(C_j)g(\underline{x}_k;\ \underline{m}_j,\underline{R}_j)}{\sum_{h=1}^{2}P_k(C_h)g(\underline{x}_k;\ \underline{m}_h,\underline{R}_h)}\ ,\quad j=1,2,\ \forall\ k \tag{4}$$

b) The classification is made by putting $\hat{C}_k = C_1$ when $P_k(C_1/\underline{x}_k) > 1/2$,

$\hat{C}_k = C_2$ when $P_k(C_1/\underline{x}_k) < 1/2$, and choosing at random if $P_k(C_1/\underline{x}_k) = 1/2$.

c) For the purpose of classifying the (k+1)th segment we observe at first that: $P_{k+1}(C_j) = P(S_{k+1}=C_j/\underline{x}_1,\ldots,\underline{x}_k)$. Moreover, since it is reasonable to believe that the probabilistic structure of the language is independent of the speech production mechanism, we can assume that:

$$P(S_{k+1}=C_j/S_1,\ldots,S_k,\underline{x}_1,\ldots,\underline{x}_k) = P(S_{k+1}=C_j/S_1,\ldots,S_k),\ j=1,2,\ \forall\ k, \tag{5}$$

In addition,by introducing the hypothesis that the sequence of the classes constitutes a 1st order time-omogeneous Markov chain, we have:

$$P(S_{k+1}=C_j/S_1,\ldots,S_k) = P(S_{k+1}=C_j/S_k)\ ,\quad j=1,2,\ \forall\ k \tag{6}$$

Then, from eqs (5) and (6) it follows:

$$P_{k+1}(C_j) = \sum_{S_1}\ldots\sum_{S_k}P(S_{k+1}=C_j,S_1,\ldots,S_k/\underline{x}_1,\ldots,\underline{x}_k) = \sum_{i=1}^{2}p_{ij}P_k(C_i/\underline{x}_k)\ ,\quad j=1,2,\ \forall\ k\ . \tag{7}$$

3.EXPERIMENTAL RESULTS

In this paragraph preliminary results concerning the application of the method previously illustrated are given. The method of classification has been experimented at the "Centre Scientifique de Paris, IBM France", in the framework of the "Deaf Children" project. The laboratory in which the test of the algorithm has been carried out disposed of an IBM Series/1 computer. A dynamic microphone has been used. The speech signal has been filtered at 4.8 kHz by means of a Butterworth lowpass-band and 8 pole filter , which permits the sampling of the signal at a frequency of 10 kHz. The signal so sampled has been stocked on a disk and then analyzed, by subdividing it into segments of 12.8 msec duration, corresponding to 128 samples of the signal. From each segment we have extracted, with standard procedures , the following features: 1) Energy of the signal, 2) Zero-crossings, 3) Autocorrelation coefficient at unit sample delay, 4) First predictor coefficient .

During the training, for each of the two classes, the expected value and the covariance matrix have been estimated. For this purpose , the segments of five sentences(for a total of 1946 segments), pronounced by two female speakers have been analyzed. In addition, on the basis of the analysis of 4936 segments, the a priori probabilities of the classes for the first segment and the transition probabilities have been evaluated.The various initial entities have been estimated by using classic methods.The segments have been classified manually, by using standard procedures (examination of the spectrogramm and waveform).

The algorithm has been tested on three sentences:the first pronounced by one of the two female speakers of the training(F1),the second by another external female speaker(F2)and the third by a male speaker(M1).On the basis of the same training data the algorithm proposed by Atal-Rabiner 1 has been tested (without smoothing),using the same sentences and speakers . The results obtained with the two methods are illustrated in table I and Table II. As regards the two female speakers, both methods have given satisfactory results. In the case of male speaker,for both methods the results can be considered still satisfactory,though male speakers were not used in the tra ining. On 1446 segments the misclassification rate was 1.4% for the method here proposed and 2.1% for the Atal-Rabiner method.

TABLE I

Results of the test on the method proposed in this paper

Speakers	V→V	V→UV	%	UV→UV	UV→V	%
F1	383	1	0.3	91	4	4.2
F2	363	4	1.1	106	7	6.2
M1	413	0	0	59	5	7.8

TABLE II

Results of the test on the method of Atal - Rabiner

Speakers	V→V	V→UV	%	UV→UV	UV→V	%
F1	372	12	3.1	94	1	1.1
F2	361	6	1.6	106	7	6.2
M1	410	3	0.7	62	2	3.1

V→V denotes the number of Voiced segments classified as Voiced; V→UV denotes the number of Voiced segments classified as Unvoiced; and so on. % denotes the misclassification rate.

On the whole the method here proposed appears offering attractive performances for further evaluation; computer performances are indeed comparable with those obtained by using the method of Atal-Rabiner. Then, in particular, the examination of the sequence of the classifier decisions has shown the uselessness of a smoothing algorithm after the decisor,useful in other methods(e.g. [1]).

4.CONCLUSIONS

In this paper we have presented a method of Voiced/Unvoiced classification based on the maximum a posteriori probability criterion , which has given good experimental results.

This method is a semplified version of more sophisticated Voiced /Unvoiced /Silence classification criterion, in which either the a priori probabilities of the classes either the p.d.f. of the measurements vector $\underline{x}_k$ are updated.We are actually testing this more complex method of classification.

- REFERENCES

[1] B.S. Atal - L.R. Rabiner,"A pattern recognition approach to Voiced /Unvoiced/Silence Classification with applications to Speech recognition" . IEEE-Trans.Acoust.Speech Signal Processing. Vol.ASSP-24 No.3, June 1976.

[2] L.J. Siegel,"A procedure for using pattern claffication technique to obtain a Voiced/Unvoiced classifier". IEEE-Trans.Acoust.Speech Signal Processing. Vol.ASSP-27 No.1, Febr. 1979.