ARCHIVES OF ACOUSTICS
13, 3–4, 205–218 (1988)

AUTOMATIC VOICE RECOGNITION IN OPEN SETS

CZESŁAW BASZTURA, WOJCIECH MAJEWSKI, JERZY JURKIEWICZ

Institute of Telecommunication and Acoustics, Wrocław Technical University
(53/55 B. Prusa St., 50–317 Wrocław)

Every system of automatic voice recognition can be divided into three parts: the voice source, the measuring system (the system of parameter extraction) and the classifier. The object of interest of the present paper is the classifier, in which emphasis is laid upon the procedure capable of recognizing voices in open sets. The methodology of investigations, analysis of the problem and the recognizing algorithm has been presented. Also the experimental results accounting for choice optimization and extraneous voice discrimination as well as the problem of choosing the threshold values for a given recognition strategy has been discussed.

Każdy system automatycznego rozpoznawania obrazów można podzielić na trzy części: źródło sygnału, układ pomiarowy (układ ekstrakcji parametrów) oraz klasyfikator. Przedmiotem zainteresowania niniejszej pracy jest klasyfikator z punktu widzenia procedury przydatnej do rozpoznawania głosów w zbiorach otwartych. Przedstawiono metodykę badań, analizę problemu i algorytm rozpoznawania oraz omówiono wyniki eksperymentów uwzględniających optymalizację wyboru, dyskryminujących obce głosy, oraz wybór wartości progowych dla określonej strategii rozpoznawania systemu.

1. Introduction

In the problem of automatic recognition of the acoustic (and not only acoustic) patterns the cases appear when it can not be assumed a priori that the currently recognized object or it's representing pattern belongs to a fixed set of object or pattern classes. In the process of speech recognition the number of recognized linguistic units is limited in general to a given set of classes (closed set), whereas because of practical reasons, it is not always possible to assume the analogical limitation in the task of recognizing the speaker's voice [1, 6, 7]. Then the problem arises of developing a recognition algorithm capable of dealing with the open set of patterns, i.e. an algorithm which wouldn't need the assumption that an input voice pattern of an unknown speaker must belong to the given set. The concept of such approach to the voice recognition problem has been presented in the paper [6]. Here the analysis of this problem and the complete recognition algorithm together with

the experimental results is described and discussed. One of the main targets of the present work was to carry out the probability analysis of errors and risk concerned with making a decision for an algorithm of recognition in open sets as a function of the approximation method of extraneous voice patterns distribution and the discrimination threshold.

2. Methodological assumptions and recognition procedure

Analogically to the majority of automatic recognition systems in automatic voice recognition (AVR) as the description of utterances the patterns are used. A pattern is an ordered set of numerical parameters belonging to a specified parameter space (observation space) χ^K (K denotes the space dimension). These parameter sets form some specified distributions characterised by the probability densities Q $(\mathbf{x}|m)$ where $\mathbf{x}$ is a parameter vector (voice pattern) in a multi – dimensional space, and m is an index of speaker's voice or, generally, a class index.

In the recognition process the classical Bayesian decision criterion accounts for the probability with which the recognised pattern y represents the class m, i.e. $P(m|\mathbf{y})$ [4, 5].

This probability is connected with conditional probability densities by the Bayes relation

$$P(m|\mathbf{y}) = \frac{Q(\mathbf{y}|m)P_m}{\sum_{l=1}^{M} Q(\mathbf{y}|l)P_l}; \quad m = 1, 2, \ldots, M \tag{1}$$

where M – number of classes, P_m – probability of occurrence for a pattern belonging to the m–th class. The problem of recognition in the classical approach resolves itself into finding the minimum of risk $R_m(\mathbf{y})$ concerned with assigning the pattern $\mathbf{y}$ to the class m:

$$R_m(\mathbf{y}) = \sum_{l=1}^{M} C_{m,l} P(l|\mathbf{y}); \quad m = 1, 2, \ldots M, \tag{2}$$

where $C_{m,l}$ – loss matrix element which denotes the value of loss resulting from assigning the pattern from the class l to the class m [4].

Taking into account the fact that for a given vector pattern $\mathbf{y}$ the denominator in (1) is constant, the expression (2) can be rewritten in the form(1)

$$R_m(\mathbf{y}) = \sum_{l=1}^{M} C_{m,l} Q(\mathbf{y}|l)P_l; \quad m = 1, 2, \ldots M. \tag{3}$$

(1) by factoring the constant out and neglecting it.

2.1. *The Bayesian decision criterion for open sets*

In the case of recognition in open sets the set of voice classes of the speakers consists of a subset M of the known recognized classes (closed set) and one multiobject class corresponding to the complement of the subset of known speakers' voices, which is named the ground or the class of extraneous voices.

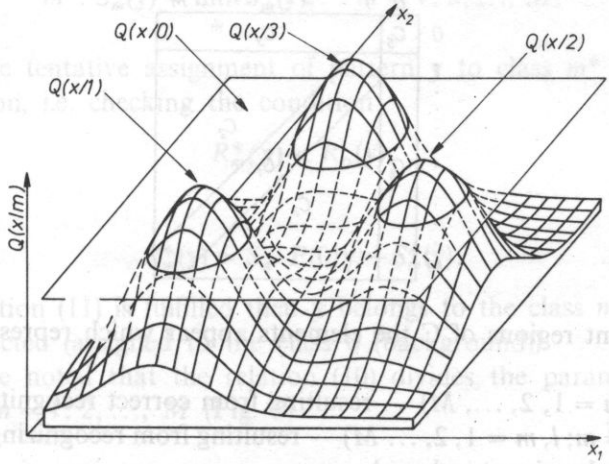


Fig. 1. Distributions of conditional probability densities in case of a two-dimensional space ($K = 2$)

For the subset of known pattern classes $m = 1, 2, \ldots M$ it can be assumed that the conditional distributions are normal distributions with the conditional probability density $Q(\mathbf{x}|m)$ expressed by the relation (4) (cf. Fig. 1):

$$Q(\mathbf{x}|m) = (2\pi)^{\frac{-K}{2}} |\mathbf{B}_m|^{\frac{-1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mathbf{W}_m)^{T_r} \mathbf{B}_m^{-1}(\mathbf{x}-\mathbf{W}_m)\right\} \tag{4}$$

where

$$B_m = \frac{1}{I_m - 1} \sum_{i=1}^{I_m} (\mathbf{W}_m - \mathbf{x}_{m,i})(\mathbf{W}_m - \mathbf{x}_{m,i})^{T_r} \tag{5}$$

$\mathbf{B}_m$ – covariance matrix of the intra-class dissipations and

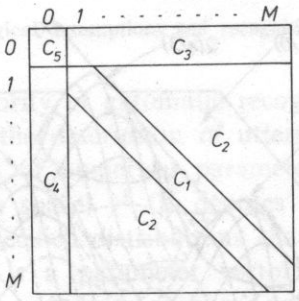
$$\mathbf{W}_m = \frac{1}{I_m} \sum_{i=1}^{I_m} \mathbf{x}_{m,i} \tag{6}$$

is the mean vector for a class, $m = 1, 2, \ldots M$, M – number of classes (speakers' voice), $i = 1, 2, \ldots I$, I – number of repetitions of an utterance for the m-th class in the training sequence TS, K – dimension of the parameter vector, T_r – denotation of the vector transposition. The conditional distribution of the background ($m = 0$) ([2])

$Q(\mathbf{x}|0)$ is in general case a multimodal distribution with a significant number of modes, tending to infinity. In general its measurement or assumption of an analytical form is impossible.

Before analyzing some possibilities of approximation of the distribution $Q(\mathbf{x}|0)$ the structure of the loss matrix should be examined.

In the loss matrix $\mathbf{C}$ the following regions can be distinguished:



In the subsequent regions of C the elements appear which represent the following values:

C_1 $(C_{m,m};\ m = 1, 2, \ldots, M)$ – resulting from correct recognition of pattern m,

$C_2(C_{l,m};\ l \neq m;\ l, m = 1, 2, \ldots M)$ – resulting from recognizing a representative of class m as belonging to class l,

$C_3(C_{0,m};\ m = 1, 2, M)$ – resulting from rejecting a representative of class m considering it as not belonging to the closed set,

$C_4\, C_{l,0};(l = 1, 2, \ldots, M)$ – resulting from assigning on extraneous patterns class representative to class l from the closed set,

C_5 [3] $(C_{0,0})$ – resulting from the correct rejection (assigning to the background) of a voice representative from ouside the closed set.

Taking into account the above described structure of the matrix $\mathbf{C}$ the relation (2) can be transformed as follows:

$$R_0(\mathbf{y}) = C_{00}\, P(O|\mathbf{y}) + \sum_{i=1}^{M} C_{0l}\, P(l|\mathbf{y}); \qquad m = 0, \tag{7}$$

$$R_m(\mathbf{y}) = C_{m,0} P(O|\mathbf{y}) + S_m(\mathbf{y}); \qquad m = 1, 2, \ldots, M, \tag{8}$$

where

$$S_m(\mathbf{y}) = \sum_{i=1}^{M} C_{m,j}\, P(l|\mathbf{y}); \qquad m = 1, 2, \ldots, M, \tag{9}$$

---

[2] Introduction of an additional class $m = 0$ necessitates for an appropriate modification of the summation range in Eqs. (1), (2), (3).

[3] The regions C_1 and C_5 represent the losses resulting from the correct decision, therefore usually the zero values are assumed in these regions, or if there are some special reasons the gains can appear here (with the opposite sign).

is the risk of a decision that **y** represents the class m from the closed set. Numerically this value is proportional to the value of risk (cf. Eq. (2)).

If in the region C_4 all the values are equal to a constant ($C_{m,0} = S_4$; $m = 1, 2, \ldots, M$; it will be assumed that this condition is valid from now on), then the recognition process in open sets can be divided into two stages

1) recognition in the closed set, i.e. finding

$$m^*: S_m^*(\mathbf{y}) = \min_{\hat{m}} S_m(\mathbf{y}); \quad m = 1, 2, \ldots, M, \tag{10}$$

what denotes the tentative assignment of pattern **y** to class m^* and

2) verification, i.e. checking the condition

$$R_m^*(\mathbf{y}) < R_0(\mathbf{y}), \tag{11}$$

where

$$R_m^*(\mathbf{y}) = S_4 \cdot P(0|\mathbf{y}) + S_m^*(\mathbf{y}). \tag{12}$$

If the condition (11) is fulfilled then **y** belongs to the class m^* otherwise the pattern **y** is rejected (assigned to the class 0 (background)).

It should be noted that the relation (10) divides the parameter space into M regions χ_m^K; $m = 1, 2, \ldots, M$ (Fig. 2)

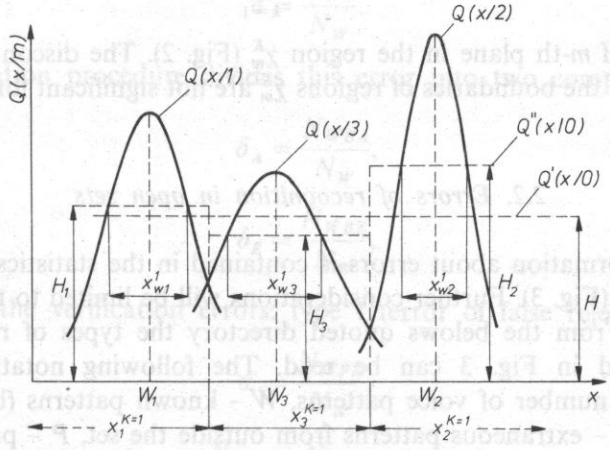


Fig. 2. Illustrations for the method of approximation of $Q(x|0)$ and determination of threshold values in a one-dimensional space $K = 1$, $Q'(x|0)$ – Case 1 (Eq. (39)), $Q''(x|0)$ – Case 2 (Eq. (40))

$$\mathbf{x} \in \chi_m^K \text{ if } S_m(\mathbf{x}) < S_l(\mathbf{x}); \; l \neq m, \; l = 1, 2, \ldots M, \tag{13}$$

and the relation

$$S_m(\mathbf{x}) = \min_{l \neq m} S_l(\mathbf{x}) \tag{14}$$

is the equation of the boundary of the region χ_m^K with respect to the vector $\mathbf{x}$. Therefore, recognition in the closed subset is equivalent to finding

$$m^*: \ \mathbf{y} \in \chi_m^K. \tag{15}$$

Similarly, the verification process (Eq. (11)) in every region χ_m^K specifies the subregion $\chi_{w_m}^K$ (Fig. 2). From the above it results that the exact knowledge of the whole distribution $Q(\mathbf{x}|0)$ is not needed. It is only necessary to know the boundaries of the regions χ_m^K; $m = 1, 2, \ldots M$ or the distribution $Q(\mathbf{x}|0)$ in some neighbourhood of these boundaries. If the values in the regions of matrix $\mathbf{C}$ do not differ significantly, then most frequently the following relation takes place:

$$\mathbf{W}_m \in \chi_{w_m}^K; \ m = 1, 2, \ldots, M, \tag{16}$$

where $\mathbf{W}_m$ as in Eq. (6).

Hence, the region $\chi_{w_m}^K$ includes some neighbourhood of the vertex of distribution $Q(x|m)$; $m = 1, 2, \ldots M$ (Fig. 2). An approximation of $Q(\mathbf{x}|0)$ by M planes, one per each region χ_m^K, can be proposed then

$$Q(\mathbf{x}|0) = H_m(\mathbf{x}), \quad m{:}\mathbf{x} \in \chi_m^K, \tag{17}$$

where

$$H_m(\mathbf{x}) = \cdot h_{m,0} + \sum_{k=1}^{K} h_{m,k} \cdot x_k \tag{18}$$

is the equation of m-th plane in the region χ_m^K (Fig. 2). The discontinuities of such approximation at the boundaries of regions χ_m^K are not significant for the verification process.

2.2. Errors of recognition in open sets

The full information about errors is contained in the statistics of recognitions and verifications (Fig. 3). Further considerations will be limited to the global errors of recognition. From the belows quoted directory the types of recognitions and eliminations used in Fig. 3 can be read. The following notations have been introduced: N – number of voice patterns, W – known patterns (belonging to the closed set M), O – extraneous patterns from outside the set, P – patterns correctly initially recognized, B – patterns erroneously initially recognized, A – patterns accepted by the verifier, E – patterns eliminated by the verifier.

EXAMPLE: N_{WPE} denotes the number of patterns recognized by the classifier and then rejected in the verification process.

If in the previously distinguished regions in the loss matrix its elements are equal

$$P_m = \frac{1}{M} P_W; \quad m = 1, 2, \ldots, M, \tag{19}$$

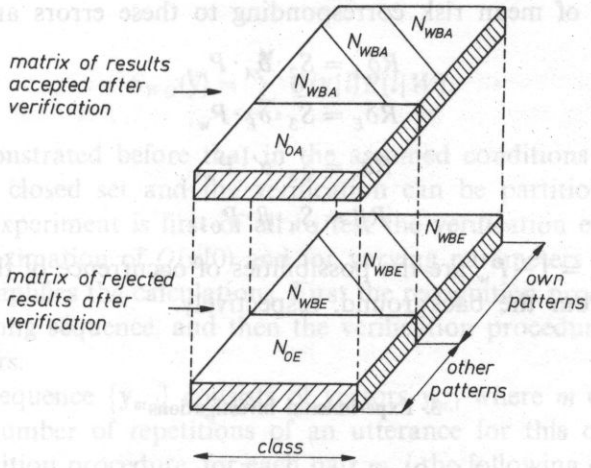


Fig. 3. Recognition and verification statistics in an open test set

where P_W – probability of occurrence of any representative of a closed set, then the following errors can be found (Eqs (20) to (26)). Inside the closed set the statistics of erroneous recognitions are represented by

$$\delta = \frac{N_{WB}}{N_W}. \tag{20}$$

The verification procedure divides this error into two components:

$$\delta_A = \frac{N_{WBA}}{N_W}, \tag{21}$$

$$\delta_E = \frac{N_{WBE}}{N_W}, \tag{22}$$

and introduces the verification errors: type I (error of false rejection)

$$\alpha' = \frac{N_{WPE}}{N_W}, \tag{23}$$

$$\alpha'' = \frac{N_{WPE}+N_{WB}}{N_W} = \alpha' + \delta, \tag{24}$$

$$\alpha''' = \frac{N_{WPE}+N_{WBE}}{N_W} = \alpha' + \delta_E, \tag{25}$$

and type II (error of false acceptation):

$$\beta = \frac{N_{OA}}{N_O}. \tag{26}$$

The components of mean risk corresponding to these errors are

$$R\delta_A = S_2 \cdot \delta_A \cdot P_w, \tag{27}$$

$$R\delta_E = S_3 \cdot \delta_E \cdot P_w, \tag{28}$$

$$R\alpha' = S_3 \cdot \alpha' \cdot P_w, \tag{29}$$

$$R\beta = S_4 \cdot \beta \cdot P_0, \tag{30}$$

where P_w and $P_0 = 1 - P_w$ are the possibilities of occurrence of the voice from the closed set and from the background, respectively.

3. Experimental investigations

3.1. Organisation of experiment

Values of the loss matrix **C** elements can be fixed or can be the parameters of an experiment. There are no reasons to differentiate them in the distinguished regions and it has been assumed that in the regions C_1 and C_5 the values are zero, whereas in the regions C_2, C_3 and C_4 the values are 1, S_3 and S_4, respectively, where S_3 and S_4 are the parameters of the experiment. Since recognition and verification consists in the choice of minimum risk, the addition of an arbitrary constant to the matrix **C** does not change the results of recognition. Further, the scaled matrix **C** will be applied, with the values in regions C_1, C_2, C_3, C_4 and C_5 equal to $-1, 0, S_3 - 1, S_4 - 1, -1$, respectively. Under these assumptions the relations (7), (8) and (9) can be transformed as follows

$$R_0(\mathbf{y}) = -Q(\mathbf{y}|0)P_0 + (S_3 - 1)S_0(\mathbf{y});\ m = 0, \tag{31}$$

$$R_m(\mathbf{y}) = (S_4 - 1)Q(\mathbf{y}|0)P_0 + S_m(\mathbf{y});\ m = 1, 2, \ldots, M, \tag{32}$$

where

$$S_m(\mathbf{y}) = -Q(\mathbf{y}|m)P_m;\ m = 1, 2, \ldots, M, \tag{33}$$

$$S_0(\mathbf{y}) = \sum_{l=1}^{M} Q(\mathbf{y}|l)P_l;\ m = 0, \tag{34}$$

and the verifying relation can be rewritten in the form

$$S_m^*(\mathbf{y}) < -S_4 Q(\mathbf{y}|0)P_0 - (1 - S_3) \cdot S_0(\mathbf{y}). \tag{35}$$

Having introduced $P_l = P(l|W)P_W$, where $P(l|W)$ denotes the conditional probability inside the closed set, the relation (35) can be transformed as follows:

$$S_m^*(\mathbf{y}) < -S_4 Q(\mathbf{y}|0)P_0 - (1 - S_3)S_{WO}(\mathbf{y}), \tag{36}$$

where

$$S_{WO}(\mathbf{y}) = \sum_{l=1}^{M} Q(\mathbf{y}|l)P(l|W). \tag{37}$$

It has been demonstrated before that in the assumed conditions the recognitions procedure in the closed set and the verification can be partitioned.

The aim of experiment is first of all to test the verification errors for various methods of approximation of $Q(\mathbf{x}|0)$ and for varying parameters S_3 and S_4. Such decomposition simplifies the calculations. First the recognition procedure is realised for the whole testing sequence, and then the verification procedure is repeated for various parameters.

The testing sequence $\{\mathbf{y}_{m,i}\}$ consists of vectors $\mathbf{y}_{m,i}$ where m denotes the class index and i – number of repetitions of an utterance for this class.

In the recognition procedure, for each pair m, i the following quantities can be found

$$m^*_{m,i},\ S^*_m(\mathbf{y}_{m,i}),\ S_0(\mathbf{y}_{m,i}), \tag{38}$$

i.e. the three values which allow for to repeat quickly the verification procedure many times according to Eq. (36).

The choice of possible types of planes Eq. (18) has been limited to two simple cases (cf Fig. 2)

$$Q(\mathbf{x}|0) = \mathrm{H}, \quad \mathrm{H} - \text{constant}, \tag{39}$$

$$Q(\mathbf{x}|0) = \mathrm{H}_m;\ m{:}\mathbf{x} \in \chi^K_m. \tag{40}$$

For the first case the threshold value H can be described as

$$\mathrm{H} = \gamma \cdot Q_{śr}, \tag{41}$$

where

$$Q_{śr} = \frac{1}{M} \sum_{m=1}^{M} Q(W_m|m), \tag{42}$$

and γ is a coefficient (a parameter of the experiment).

For the second case Eq. (40) two versions of the method of introducing the thresholds H_m for each class can be distinguished

a) $$\mathrm{H}_m = \gamma \cdot Q(W_m|m);\ m = 1, 2, \ldots M, \tag{43}$$

with the same coefficient for all the classes and

b) $$\mathrm{H}_m = \gamma_m \cdot Q(W_m|m);\ m = 1, 2, \ldots M. \tag{44}$$

with the coefficient γ_m chosen individually for each class on the basis of the testing sequence, so that the verification risk component in the given class would be minimum, whilst the values N_{WPE} and N_{OA} are taken into account for each class.

3.2. *The experiment*

The experimental investigation aimed at the verification of the proposed algorithm and the method of analysis for open sets has been realized with the following agreements taken into account:

a) As the parameters in the observation space, the values extracted from a fixed signal (a key phrase) will be used, the efficiency of which for voice recognition has been well established. It has been assumed that the parameter vector will be formed from the distribution of time intervals between the speech signal zero-crossings in the sentence „Jutro będzie ładny dzień" (pol: "It will be a fine day tomorrow") [1, 2, 3]. Value of the dimension of the space K had been initially assumed equal to 7; it has been reduced then to $K = 4$, because parameters with the greatest discrimination force have been chosen.

b) The mean object of interest will be the verification procedure, for the fixed parameters of the signals and of the parameter extraction system.

c) The training sequence TS is the set of vectors $\{\mathbf{x}_{m,i}\}$ $m = 1, 2, \ldots, M$ (M – number of voices in the closed set) $i = 1, 2, \ldots, I_m$ (it has been assumed $I_m = I$ – number of repetitions for each class).

d) It has been assumed that the training sequence will consist of the utterance patterns of $M = 10$ speakers per $I = 10$ repetitions. The test sequence will contain the patterns of 10 speakers × 10 repetitions from the closed set and of 10 speakers × 10 repetitions as extraneous utterances. Together, the open test set consisted then of 200 utterances comming from 20 speakers.

The analysis and recognition has been carried out with the use of programs written on ZX Spectrum microcomputer.

With these programs the following calculations have been realized:

a) The mean vector $\mathbf{W}_m$ for 10 classes from the training set were calculated (Table 1).

b) The estimators of conditional probability densities $Q(\mathbf{y}|m)$ and risk $R_m(\mathbf{y})$ were found. In the Table 2 the conditional densities have been confronted.

c) The values of errors α and β and the value of verification risk as a function of γ were calculated and plotted for the cases given by Eqs (39) and (40). The diagrams of these functions are presented in the Fig. 4a, b, c.

d) The matrices of recognitions and verifications were calculated for various

Table 1. Vectors of class standards

Parameter k	Pattern n										Pattern medium
	1	2	3	4	5	6	7	8	9	10	
1	−1.094	−0.037	−0.583	−0.655	0.629	−0.649	−0.196	−0.003	1.017	−0.825	−0.240
2	−0.840	−0.224	−0.478	−0.575	−0.158	−0.690	−0.370	−0.288	0.151	−0.6064	−0.408
3	−1.379	−0.518	−0.652	−0.952	−0.978	−1.165	−0.764	−0.772	−0.100	−0.726	−0.801
4	0.298	−0.493	1.066	−0.749	1.359	−1.238	−0.939	−0.539	−0.588	−0.158	−0.198

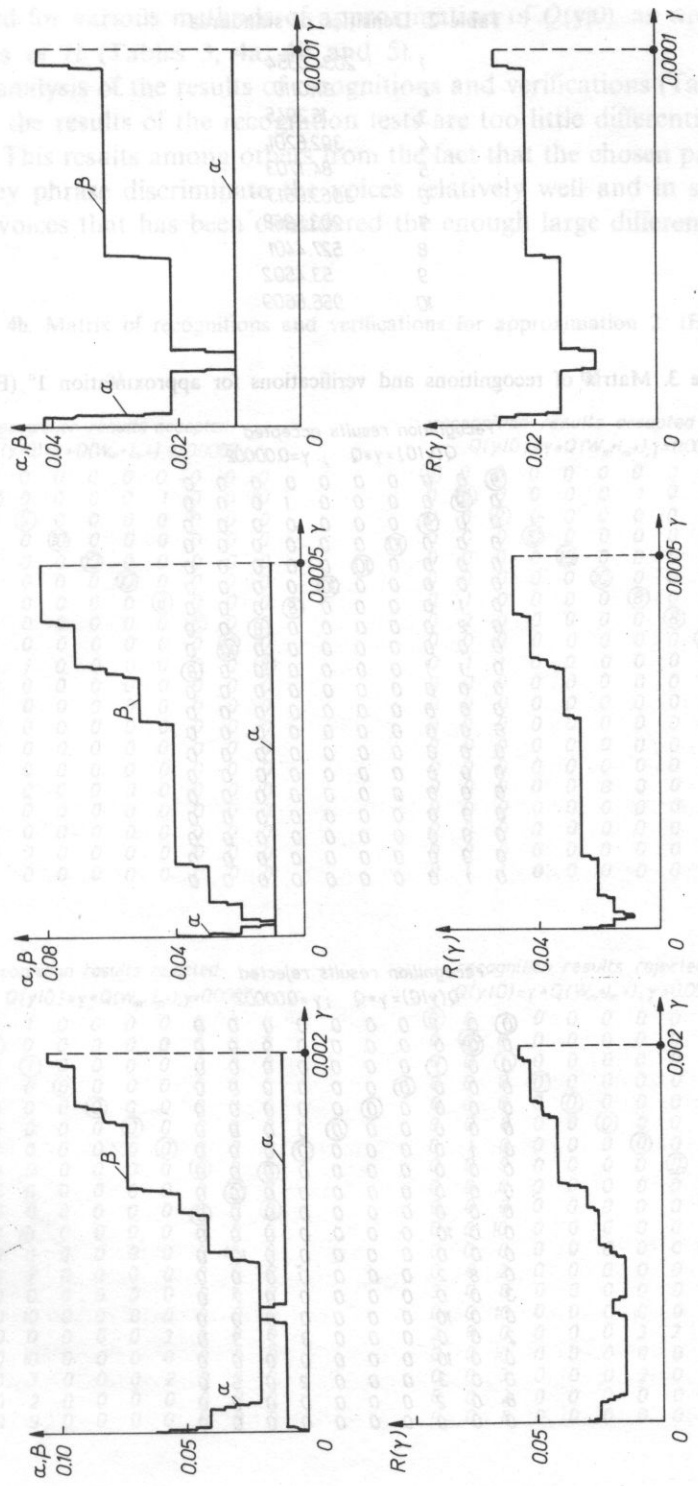


Fig. 4. Graphs of errors α and β and risk function R depending on coefficient γ

Table 2. Densities of standards

1	2034.0354
2	37.0110
3	16.2945
4	392.6201
5	84.1703
6	3063.0513
7	203.5968
8	527.4401
9	53.4502
10	956.6609

Table 3. Matrix of recognitions and verifications for approximation 1° (Eq. (39))

recognition results accepted

$Q(y/0) = \gamma * Q \quad ; \ \gamma = 0.00002$

(9)	0	0	0	0	0	0	0	0	0
0	(9)	0	0	0	0	1	0	0	0
0	0	(9)	0	0	0	0	0	0	0
0	0	0	(10)	0	0	0	0	0	0
0	0	0	0	(10)	0	0	0	0	0
0	0	0	0	0	(10)	0	0	0	0
0	1	0	0	0	0	(8)	0	0	0
0	2	0	0	0	0	0	(8)	0	0
0	0	0	0	0	0	0	0	(10)	0
0	1	1	0	0	0	0	0	0	(8)
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0

recognition results rejected

$Q(y/0) = \gamma * Q \quad ; \ \gamma = 0.00002$

(0)	0	1	0	0	0	0	0	0	0
0	(0)	0	0	0	0	0	0	0	0
0	0	(1)	0	0	0	0	0	0	0
0	0	0	(0)	0	0	0	0	0	0
0	0	0	0	(0)	0	0	0	0	0
0	0	0	0	0	(0)	0	0	0	0
0	1	0	0	0	0	(0)	0	0	0
0	0	0	0	0	0	0	(0)	0	0
0	0	0	0	0	0	0	0	(0)	0
0	0	0	0	0	0	0	0	0	(0)
0	0	10	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	10	0
0	8	2	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	7	0
0	0	10	0	0	0	0	0	0	0
2	0	0	0	0	0	3	2	3	0
0	0	10	0	0	0	0	0	0	0
0	0	3	0	0	0	2	0	5	0
8	0	2	0	0	0	0	0	0	0
0	0	9	0	0	0	0	0	0	0

values of γ and for various methods of approximation of $Q(y|0)$, as well as for the optimal values of H (Tables 3, 4a, 4b and 5).

After an analysis of the results of recognitions and verifications (Table 6) it can be stated that the results of the recognition tests are too little differentiated for the studied cases. This results among others from the fact that the chosen parameters of the present key phrase discriminate the voices relatively well and in such a small testing set of voices that has been considered the enough large differences couldn't appear. What's more, it occurred not profitable to choose the individual optimal threshold (Eq. (44)). It should be clearly underlined here that the methodological correctness of the method of choice of the discriminating quantities has been confirmed, what was the main target of the study, although the small number of the background patterns did not allow for significant differentiation of the cases of approximation of $Q(y|0)$ (cf. Eqs. (39), (40), (41) and (42)).

Table 4a and **4b**. Matrix of recognitions and verifications for approximation 2° (Eq. (40), (43))

a) recognition results accepted, $Q(y|0)=\gamma * Q(W_m+l_m+); \gamma=0.0007$

⑥	0	0	0	0	0	0	0	0	0
0	⑨	0	0	0	0	1	0	0	0
0	0	⑨	0	0	0	0	0	0	0
0	0	0	⑩	0	0	0	0	0	0
0	0	0	0	⑩	0	0	0	0	0
0	0	0	0	0	⑩	0	0	0	0
0	1	0	0	0	0	⑧	0	0	0
0	2	0	0	0	0	0	⑧	0	0
0	0	0	0	0	0	0	0	⑩	0
0	1	1	0	0	0	0	0	0	⑧
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0

b) recognition results accepted, $Q(y|0)=\gamma * Q(W_m+l_m+); \gamma=0.0003$

⑨	0	0	0	0	0	0	0	0	0
0	⑨	0	0	0	0	1	0	0	0
0	0	⑨	0	0	0	0	0	0	0
0	0	0	⑩	0	0	0	0	0	0
0	0	0	0	⑩	0	0	0	0	0
0	0	0	0	0	⑩	0	0	0	0
0	1	0	0	0	0	⑧	0	0	0
0	2	0	0	0	0	0	⑧	0	0
0	0	0	0	0	0	0	0	⑩	0
0	1	1	0	0	0	0	0	0	⑧
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	1	1	0	0	0	0	0	0	0

a) recognition results rejected, $Q(y|0)=\gamma * Q(W_m+l_m+); \gamma=0.0007$

①	0	1	0	0	0	0	0	0	0
0	⓪	0	0	0	0	0	0	0	0
0	0	①	0	0	0	0	0	0	0
0	0	0	⓪	0	0	0	0	0	0
0	0	0	0	⓪	0	0	0	0	0
0	0	0	0	0	⓪	0	0	0	0
0	1	0	0	0	0	⓪	0	0	0
0	0	0	0	0	0	0	⓪	0	0
0	0	0	0	0	0	0	0	⓪	0
0	0	0	0	0	0	0	0	0	⓪
0	0	10	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	10	0
0	8	2	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	7	0
0	0	10	0	0	0	0	0	0	0
2	0	0	0	0	0	3	2	3	0
0	0	10	0	0	0	0	0	0	0
0	0	3	0	0	0	2	0	5	0
8	0	2	0	0	0	0	0	0	0
0	0	9	0	0	0	0	0	0	0

b) recognition results rejected, $Q(y|0)=\gamma * Q(W_m+l_m+); \gamma=0.0003$

⓪	0	1	0	0	0	0	0	0	0
0	⓪	0	0	0	0	0	0	0	0
0	0	①	0	0	0	0	0	0	0
0	0	0	⓪	0	0	0	0	0	0
0	0	0	0	⓪	0	0	0	0	0
0	0	0	0	0	⓪	0	0	0	0
0	1	0	0	0	0	⓪	0	0	0
0	0	0	0	0	0	0	⓪	0	0
0	0	0	0	0	0	0	0	⓪	0
0	0	0	0	0	0	0	0	0	⓪
0	0	10	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	10	0
0	8	2	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	7	0
0	0	10	0	0	0	0	0	0	0
2	0	0	0	0	0	3	2	3	0
0	0	10	0	0	0	0	0	0	0
0	0	3	0	0	0	2	0	5	0
8	0	2	0	0	0	0	0	0	0
0	0	8	0	0	0	0	0	0	0

Table 5. Matrix of recognitions and verifications for approximation 2° including optimized coefficient γ_m

recognition results accepted

$Q(y|0) = H_{m*opt}$

⑨	0	0	0	0	0	0	0	0	0
0	⑨	0	0	0	0	1	0	0	0
0	0	⑨	0	0	0	0	0	0	0
0	0	0	⑩	0	0	0	0	0	0
0	0	0	0	⑩	0	0	0	0	0
0	0	0	0	0	⑩	0	0	0	0
0	1	0	0	0	0	⑧	0	0	0
0	2	0	0	0	0	0	⑧	0	0
0	0	0	0	0	0	0	0	⑩	0
0	1	1	0	0	0	0	0	0	⑧
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0

recognition results rejected

$Q(y|0) = H_{m*opt}$

⓪	0	1	0	0	0	0	0	0	0
0	⓪	0	0	0	0	0	0	0	0
0	0	①	0	0	0	0	0	0	0
0	0	0	⓪	0	0	0	0	0	0
0	0	0	0	⓪	0	0	0	0	0
0	0	0	0	0	⓪	0	0	0	0
0	1	0	0	0	0	⓪	0	0	0
0	0	0	0	0	0	0	⓪	0	0
0	0	0	0	0	0	0	0	⓪	0
0	0	0	0	0	0	0	0	0	⓪
0	0	10	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	10	0
0	8	2	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	7	0
0	0	10	0	0	0	0	0	0	0
2	0	0	0	0	0	3	2	3	0
0	0	10	0	0	0	0	0	0	0
0	0	3	0	0	0	2	0	5	0
8	0	2	0	0	0	0	0	0	0
0	0	9	0	0	0	0	0	0	0

appear. What's more, it occurred not profitable to choose the individual optimal threshold (Eq. (44)). It should be clearly underlined here that the methodological correctness of the method of choice of the discriminating quantities has been confirmed, what was the main target of the study; whereas the small number of the background patterns did not allow for significant differentiation of the cases of approximation of $Q(\mathbf{x}|0)$ (cf Eqs. (39), (40), (43) and (44)).

Table 6 Confrontation of experimental results (for $S_3 = S_4 = 1$)

N	Case 1° (Eq. (39))	Case 2° (Eq. (43))		Case 2° (γ_m individual) (Eq. (44))
	$\gamma = 0.00002$	$\gamma = 0.0007$	$\gamma = 0.0003$	
N_{WPA}	91	90	91	91
N_{WBA}	6	6	6	6
N_{OA}	1	1	2	1
N_{WPE}	1	2	1	1
N_{WBE}	2	2	2	2
N_{OE}	99	99	98	99
Errors	%	%	%	%
δ	8	8	8	8
δ_A	6	6	6	6
δ_E	2	2	2	2
α'	1	2	1	1
α''	9	10	9	9
α'''	3	4	3	3
β	1	1	2	1

4. Conclusion

The above presented methodological considerations justify the statement that the proposed algorithm of voice recognition in open sets is a flexible procedure that allows for fitting the global recognition characteristics, which are the errors α and β, to a certain operating strategy of the system.

For the chosen parameters patterns describing voices there always exists a possibility of carrying out an optimization of the recognition process by an appropriate choice of background approximation, i.e. by selecting the optimal threshold values H_m.

This is the basic advantage of the method. It has been experimentally verified for the population of 20 voice classes including (10 extraneous ones).

If the tests had been carried out for a larger population of voices especially extraneous, then finding a more distinct optimal threshold H_m could have been expected with regard to smaller granularity. In Fig. 4a the minimum of risk is not univocally determined because of large granularity of data e.g., two flat minima appear what has been already explained above by a too small size of the testing sequence.

Irrespective of the above, in every experiment on automatic voice recognition the following factors will always influence the numerical values: a) choice of the key phrase, b) observation space parameters, c) method of forming the voice patterns.

In a fixed, chosen pattern preparation system these factors can be treated as the values which have a parametric influence. The further studies on automatic voice recognition in open sets wall concentrate on the larger number of classes M, mainly on the background voices, and on other parameters describing voices of speakers.

References

[1] Cz. BASZTURA, *Automatic speaker recognition by zero-crossing analysis of the speech signal*, In: Speech Analysis and Synthesis (W. Jassem[ed.]), 5 PWN, p. 9–40, Warszawa 1980.

[2] Cz. BASZTURA, W. MAJEWSKI, *The application of long-term analysis of the zero-crossing of a speech signal in automatic speaker identification*, Arch. of Acoustics **3**, 1, 3–15 (1978).

[3] Cz. BASZTURA, J. JURKIEWICZ, *The zero-crossing analysis of a speech signal in the short-term method of automatic speaker identification*, Arch. of Acoustics **3**, 3, 185–196 (1978).

[4] J. Z. CYPKIN, *Basics of the learning systems* (in Polish) WNT, Warszawa 1973.

[5] А. Л. ГОРЕЛИК, В. А. СКРИПКИН, *Методы розпознавания*, Высшая Школа, Москва 1977.

[6] W. MAJEWSKI, Cz. BASZTURA, *Speaker recognition in open sets*, In: Proceedings of the Tenth International Congress of Phonetic Sciences M.P.R. Van den Broecke and A. Cohen [eds] Foris Publications, p. 322–325, Dordrecht 1984.

[7] A. E. ROSENBERG, *Automatic speaker verification: a review*, Proc. IEEE, **64**, 4, 475–487 (1976).

Received on May 15, 1986; revised version on September 22, 1986.