ARCHIVES OF ACOUSTICS
16, 2, 237–247 (1991)

IMAGE SIMILARITY FUNCTIONS IN NON-PARAMETRIC ALGORITHMS OF VOICE IDENTIFICATION

CZ. BASZTURA AND J. ZUK

Institute of Telecommunication and Acoustics of the Wrocław Technical University
(50-317 Wrocław, ul. B. Prusa 53/55)

This paper is dedicated to the question of the choice of a function of similarity between images in non-parametric alogorithms of voice recognition. The usefulness of 10 similarity functions (8 distances and 2 nearness'es) in three non-parametric identification algorithms — *NN* (nearest neighbour), *k–NN* (*k*–nearest neighbours) and *NM* (nearest mean) — was investigated for three sets of parameters (1 natural and 2 normalized). Results obtained for a population of speakers from a closed set with size $M = 20$ (after 10 repetitions of the learning and test sequences) have proved that the Camberr distance function prevails in all types of parameters and algorithms. Other functions ensure a differentiated discrimination force strongly dependent on the algorithm and form of parameters. Limited usefulness of the square of Mahalonobis distance in comparison to other similarity functions was proved, as well as generally worse results for the *NM* algorithm.

Praca jest poświęcona problemowi doboru funkcji podobieństwa pomiędzy obrazami w nieparametrycznych algorytmach rozpoznawania głosów. Dla trzech zespołów parametrów (1 naturalnego i 2 znormalizowanych), pochodzących z ekstrakcji sygnału mowy hasła kluczowego, zbadano przydatność 10 funkcji podobieństwa (8 odległości i 2 bliskości) w trzech nieparametrycznych algorytmach identyfikacji: *NN* (najbliższy sąsiad), *k-NN* (k–najbliższych sąsiadów) oraz *NM* (najbliższa średnia). Uzyskane wyniki dla populacji mówców zbioru zamkniętego o liczebności $M = 20$ (po 10 powtórzeń ciągu uczącego się i testowego), wykazały zdecydowaną przewagę funkcji odległości Camberra we wszystkich rodzajach parametrów i algorytmów. Pozostałe funkcje zapewniają zróżnicowaną siłę dyskryminacyjną zależną mocno od algorytmu i postaci parametrów. Wykazano słabą przydatność kwadratu odległości Mahalanobisa w porównaniu z innymi funkcjami podobieństwa oraz ogólnie gorsze wyniki dla algorytmu *NM*.

1. Introduction

Computer recognition of voices includes several partial procedures which can be divided into three basic blocks:

a) source
b) measurement block
c) classification block

The sender of the signal (speaker) and the set of phenomena and conditions related with sending and registration of the speech signal is the source. The measurement block includes processing and analysis procedures of the input signal $u(t)$. This signal is the speaker's voice representation in acoustic images $\mathbf{x}$. Quantity $\mathbf{x}$ usually denotes vectors from the space of parameters R^P (P – dimension of the space).

The classification block is a set of procedures or a procedure converting the input vector information $\mathbf{x}$ into a scalar m from the space of classes and decisions. Quantity m is the indicator of voices among which the system included the recognised signal. In parametric recognition algorithms with full (or estimated) probabilistic information the problem of finding the function of similarity of recognised voice's image and the standard is included in the classification algorithm [1]. "Voice standards" in the classical Bayes algorithm [1] are contained in a multidimensional distribution of conditional probability $p(x|m)$ (x–vector of individual's parameters, m–class speaker's number). The decision criterion is based on the minimum of average risk, which includes the loss matrix and the probability of appearance of images from the given class and of course the $p(x|m)$ distribution [1].

There is always a definite correlation between the space of parameters and accepted functions of similarity of the recognised image and standards in non-parametric recognition algorithms [1, 4, 5]. Frequently the simplification of the classification procedure in non-parametric agorithms leads to worse recognition results, because a similarity functions inadequate to the space of parameters is applied. The significance of this problem with regard to automatic speech recognition is among others confirned by TADEUSIEWICZ's paper [5] which present the usability evaluation of the similarity function (in the form of distance measures) in the recognition of vowel in the Polish language; and by the paper by Zalewski [6] who analysed the effectiveness of distance measures in the recognition of speakers with the application of linear predictive coding (LPC). Also BASZTURA [3] tired to check 8 chosen similarity functions as indices for speech transmission quality estimation. Because of frequent use of computer voice recognition in classification procedures it seems advisable to investigate a group of chosen similarity functions with regards to their effectiveness, using homogeneous experimental material.

2. Methods

To achieve a clear evaluation of the influence of investigated similarity functions on results of voice identification, the comparative procedure has to be free of all types of variability which influence the evaluation. At the same time it is advisable and necessary to check the "behaviour" of the similarity function in definite non-parametric classification algorithms. Considering this, the folowing assumption concerning methods made to systemize further experiments:

a) Voices of 20 speakers (men) aged 20–35 in a so-called closed set (i.e. recognised speakers will be included among the set of speakers in the learning sequences) were accepted as phonetic material.

b) In order to eliminate the influence of information which is not individual (linguistic and sociolinguistic) on identification results, a short-term analysis model with a fixed key-word for all statements was chosen. The maxim „Jutro będzie ładny dzień" ("tomorrow will be a fine day") was chosen as the key-word. It was used previously in paper [3] among others. The test series wās *TS* recorded 7 days after the recoding of the learning sequences.

c) Vectors with components x_p (V_p) which are numbers corresponding with the number of time intervals between zero-crossings of the speech signal were applied as individual parameters forming images of statements x and standards V [2]. Components x_p are calculated from:

$$x_p = x(t_{p-1}, t_p) = \begin{cases} x(t_{p-1}, t_p)+1 & \text{for} \quad t_j \in (t_{p-1}, t_p\rangle \\ x(t_{p-1}, t_p) & \text{for} \quad t_j \notin (t_{p-1}, t_p\rangle \end{cases} \tag{2.1}$$

where: t_p — boundary values of so-called time channes; $p = 1, 2, 3, \ldots, P$, P – number of time channels. It was accepted that $P = 7$, while t_p was chosen in accordance with the exponential division [2] from range $t \in (0.2 \text{ ms}–6.2 \text{ ms})$.

d) It was accepted that the usability of the similarity function will be evaluated for three most frequently applied heuristic classification algorithms, i.e. *NN* (Nearest Neighbour) *k–NN* (*k*–Nearest Neighbours) and *NM* (Nearest Mean). These algorithms have the following form:

NN algorithm

Image x belongs to class (voice) m, i.e. $x \to m$ if

$$FP(x, V_{m,i}) < FP(x, V_{l,i}) \tag{2.2}$$

where: $m = 1, 2, \ldots, M$; M – number of classes (voices), $\text{-} = 1, 2, \ldots, m-1, m+1, \ldots, M$, $V_{m,i} = X_{m,i}$–image of speaker's voice $i = 1, 2, \ldots, I_m$, I_m – number of repetitions of the statement in the learning sequences.

k–NN Algorithm

Functions of similarity (let us accept these as distances) between image x and all images in the learning series $x_{m,i}$ are calculated and ordered according to increasing order (decreasing order for nearness functions). Then first k distance values are considered and it is determined how many of them correspond with individual classes. If among k minimal distances there is $k_1, k_2, \ldots, k_m$ which belong to first, second, ... etc. class respectively, then values k_m are accepted as new similarity functions. Image x belongs to class m, i.e. $x \to m$ if

$$k_m > k_i$$
$$l = 1, 2, 3, \ldots, m-1, m+1, \ldots, M \tag{2.3}$$

Value k is chosen in suitable proportion to the length of the learning sequences LS.

NM Algorithm

Most frequently the mean vector the voice (class) standard in the NM algorithm. The decision rule of the algorithm is as follows:

$$x \to m \quad \text{if} \quad FP(x, V_m) < FP(x, V_l) \tag{2.4}$$

where

$$V_m = x_m = \frac{1}{I_m} \sum_{i=1}^{I_m} x_{m,i} \tag{2.5}$$

m, l — as in expressions (2.2) and (2.3).

e) A set of ten similarity functions was chosen for investigation from among known similarity functions. Eight of them are distance functions, also called distance measures, while two are nearness functions. These functions corresponded with given below relationships (between x as the recognised image and V as the standard image). The first group of similarity functions can be noted with Minkowski's dependence:

$$d^{\mathrm{MIN}}(x, V) = \Big[\sum_{p=1}^{P} |x_p - V_p|^r \Big]^{1/r}; \; r \geq 1 \tag{2.6}$$

where: $p = 1, 2, \ldots, P$, x_p — p-th — element of vector x, V_p — p-th — element of vector V. For $r = 1$ d^{MIN} is known as Hamming's distance or street distance [6]. For $r = 2$ it is the Euclides metric. Two Minkowski's distances were additionally accepted for investigation, mamely d^{MIN3} $(r = 3)$ and d^{MIN5} $(r = 5)$.

Other similarity functions are as follows:

CHI–square distance

$$d^{\mathrm{CHI}}(x, V) = \sum_{p=1}^{P} \frac{1}{x_p + V_p} \left[\frac{X_p}{\sum_{p=1}^{P} x_p} - \frac{V_p}{\sum_{p=1}^{P} V_p} \right] \tag{2.7}$$

Czebyszew's distance

$$d^{\mathrm{CZE}}(x, V) = \max_p (x_p - V_p) \tag{2.8}$$

Camberr's distance

$$d^{\mathrm{CAM}}(x, V) = \sum_{p=1}^{P} \frac{x_p - V_p}{x_p + V_p} \tag{2.9}$$

square of Mahalanobis'es distance

$$d^{\mathrm{MAH}}(x, V) = (x-V)^{Tr} C^{-1}(x-V) \tag{2.10}$$

where C–mean covariance matrix (intraclass scatterings [1])

directional cos nearness function

$$b^{\cos}(x, V) = \frac{x \cdot V^{\mathrm{Tr}}}{|x||V|} \tag{2.11}$$

and

Tanimoto's nearness function

$$b^{\mathrm{TAN}}(x, V) = \frac{x \cdot V^{Tr}}{x \cdot x^{Tr} + V \cdot V^{Tr} - x \cdot V^{Tr}} \tag{2.12}$$

3. Identification experiment

An experiment of voice identification was carried out in compliance with paragraph 2. It was aimed at the determination of numerical relations and dependences for 3 identification algorithms and 10 similarity functions.

It was accepted that the learning series LS will consist of 200 statements of 20 speakers (20×10 repetitions), while the test series will also consist of 10 statements of every speaker. Numerical methods were used for parameter extraction ($P = 7$). Statements were recorded on professional equipment in a quiet room. The band of the signal was limited to the 75 – 4500 Hz range. The sampling frequency of the a/d converter was equal to $f_{pr} = 10\,000$ samples/s and the dynamics were described by a 10 bit word.

All experiments were repeated for all three forms in order to analyse the influence of the form of sets of parameters on the effectiveness of identification. The first set ($ZP1$) is a set of measurement parameters (Table 1). The second ($ZP2$) is a set of parameters with components normalized with respect to the value of their variability range (Table 2) (expression 2.1)).

Let

$$\Delta x_p^{sr} = \frac{1}{M} \sum_{m=1}^{M} (x_{m,p}^{\max} - x_{m,p}^{\min}) \tag{3.1}$$

be the mean variability range of the p – element, where

$$x_{m,p}^{\max} = \max_i \{x_{m,i,p}\} \tag{3.2}$$

and

$$x_{m,p}^{\min} = \min_i \{x_{m,i,p}\} \tag{3.3}$$

Table 1. Set of parameters–$ZP1$ (not normalized). An example of 20 repetitions for speaker 1.

Series	Repetition	Parameter no 1	2	3	4	5	6	7
LS	1	68	98	77	104	133	261	59
	2	75	110	76	92	127	213	63
	3	105	127	101	103	175	236	68
	4	65	87	75	97	127	267	73
	5	117	113	95	78	136	161	27
	6	121	216	144	88	120	200	42
	7	139	187	153	114	116	180	36
	8	127	195	146	105	117	175	39
	9	79	106	80	97	128	225	62
	10	126	182	137	95	154	245	46
TS	1	120	117	83	81	126	183	32
	2	109	140	116	130	187	226	101
	3	99	130	126	117	178	225	96
	4	87	125	91	97	174	250	56
	5	108	129	88	109	156	199	54
	6	96	136	94	110	154	201	53
	7	116	170	134	102	52	259	41
	8	122	201	137	77	106	185	37
	9	78	117	84	97	93	158	37
	10	84	120	69	103	95	155	45

We determine the maximal mean variation range

$$\Delta x^{sr\max} = \max_p \{\Delta x_p^{sr}\} \tag{3.4}$$

The regraduated p element is calculated from

$$x_p^{(ZP2)} = x_p \frac{\Delta x^{sr\max}}{\Delta x^{sr}} \tag{3.5}$$

The third set ($ZP3$) is a set of parameter with components normalized with respect to the variations range of their variances (Table 2)

$$x_p^{(ZP3)} = x_p \frac{\delta^{\max}}{\delta_p^{\max}} \tag{3.6}$$

where

$$\delta^{\max} = \max_p \{\delta_p\} \tag{3.7}$$

Table 2. Set of parameters–$ZP2$ normalized with respect to maximal range of parameter's variability. An example of 20 repetitions for speaker 2

Series	Repetition	Parameter no						
		1	2	3	4	5	6	7
LS	1	135	98	101	297	791	1178	630
	2	149	110	100	263	755	961	673
	3	209	127	133	294	1041	1065	727
	4	129	87	99	277	755	1205	289
	5	233	113	125	223	809	727	289
	6	241	216	189	251	714	902	449
	7	277	187	201	325	690	812	385
	8	253	195	192	300	696	790	417
	9	157	106	105	277	761	1015	663
	10	251	182	180	271	916	1106	492
TS	1	239	117	109	231	749	826	342
	2	217	140	153	371	1112	1020	1079
	3	197	130	166	334	1059	1015	1026
	4	173	125	120	277	1035	1128	598
	5	215	129	116	311	928	898	577
	6	191	136	124	314	916	907	566
	7	231	170	176	291	309	1169	438
	8	243	201	180	220	630	835	395
	9	155	117	110	277	553	713	395
	10	167	120	91	294	565	699	481

and

$$\delta_p = \frac{1}{M} \sum_{m=1}^{M} \delta_{m,p} \tag{3.8}$$

while $\delta_{m,p}$ – variance of p parameter of m speaker calculated on the basis of learning series *LS*.

4. Analysis of results and conclusions

The series of carried out identification experiments led to definite comparisons and analysis aimed at the usability evaluation of individual similarity functions in investigated non-parametric identification algorithms. Two additional sets of parameters ($ZP2$ and $ZP3$) resulting from normalizing transformations improved the results in terms of static likelihood. The following conclusions can be drawn from the set of results presented in Tables 4, 5, 6 and 7:

Table 3. Set of parameters–$ZP3$ normalized with respect to the variability range of parameters' variances. An example of 20 repetitions for speaker 3

Series	Repetition	Parameter no						
		1	2	3	4	5	6	7
LS	1	122	98	97	274	732	1094	597
	2	134	110	96	242	699	892	637
	3	188	127	127	271	963	989	688
	4	116	87	95	255	699	1119	738
	5	209	113	120	205	748	675	273
	6	216	216	182	232	660	838	425
	7	249	187	193	300	638	754	364
	8	227	195	184	276	644	733	395
	9	141	106	101	255	704	943	627
	10	225	182	173	250	847	1027	465
TS	1	215	117	105	213	693	767	324
	2	195	140	146	342	1029	947	1022
	3	177	130	159	308	979	943	971
	4	156	125	115	255	957	1047	566
	5	193	129	111	287	858	834	546
	6	172	136	119	290	847	842	536
	7	208	170	169	269	286	1085	415
	8	218	201	173	203	583	775	374
	9	140	117	106	255	512	662	374
	10	150	120	87	271	523	649	455

Table 4. Results of voice identification in % for sets of parameters–$ZP1$.

Algo-rithm	Measure	1	2	3	4	5	6	7	8	9	10
		d^{HAM}	d^{EUK}	d^{MIN3}	d^{MIN5}	d^{CHI}	d^{CZE}	d^{CAM}	d^{MAH}	d^{COS}	d^{TAN}
NN	S_r	91.5	90.5	90.5	88.5	92.0	87.0	95.5	91.0	93.0	91.5
	δ	12.3	12.8	12.8	15.0	15.8	17.5	12.8	14.1	10.8	12.7
	$S_{r\max}$	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	$S_{r\min}$	60.0	60.0	60.0	50.0	40.0	40.0	50.0	60.0	60.0	60.0
k-NN	S_r	90.5	89.5	87.0	86.5	89.5	83.5	94.5	87.5	89.0	89.0
	δ	14.3	14.7	15.9	18.4	23.3	21.6	11.8	14.5	12.1	14.8
	$S_{r\max}$	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	$S_{r\min}$	50.0	50.0	50.0	30.0	10.0	20.0	50.0	50.0	60.0	50.0
NM	S_r	88.0	84.5	81.0	81.0	89.5	81.0	96.0	80.5	87.0	85.0
	δ	19.0	21.4	22.2	24.0	23.1	24.3	12.7	24.8	15.9	21.4
	$S_{r\max}$	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	$S_{r\min}$	30.0	30.0	30.0	30.0	20.0	30.0	50.0	10.0	40.0	30.0

Table 5. Results of voice identification in % for sets of parameters–$ZP2$

Algorithm	Measure	1 d^{HAM}	2 d^{EUK}	3 d^{MIN3}	4 d^{MIN5}	5 d^{CHI}	6 d^{CZE}	7 d^{CAM}	8 d^{MAH}	9 d^{COS}	10 d^{TAN}
NN	s_r	95.0	97.0	97.0	95.5	94.0	93.5	95.5	91.0	93.0	97.0
	δ	14.0	9.2	9.2	10.0	14.3	13.1	12.8	14.1	14.5	9.8
	$s_{r\max}$	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	$s_{r\min}$	50.0	70.0	70.0	70.0	40.0	50.0	50.0	60.0	40.0	60.0
k-NN	s_r	95.0	93.5	92.5	92.0	90.5	90.5	94.5	87.5	95.0	95.0
	δ	12.4	15.7	15.5	15.8	21.6	17.6	11.9	14.5	14.0	11.9
	$s_{r\max}$	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	$s_{r\min}$	60.0	40.0	40.0	40.0	20.0	30.0	50.0	50.0	40.0	50.0
NM	s_r	94.5	95.0	95.0	94.5	90.5	93.0	96.0	80.5	93.0	95.5
	δ	12.3	11.5	11.0	11.0	23.1	10.6	12.7	24.8	15.9	8.9
	$s_{r\max}$	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	$s_{r\min}$	60.0	60.0	60.0	60.0	20.0	60.0	50.0	10.0	40.0	70.0

Table 6. Results of voice identification in % for sets of parameters–$ZP3$

Algorithm	Measure	1 d^{HAM}	2 d^{EUK}	3 d^{MIN3}	4 d^{MIN5}	5 d^{CHI}	6 d^{CZE}	7 d^{CAM}	8 d^{MAH}	9 d^{COS}	10 d^{TAN}
NN	s_r	95.5	97.0	97.0	95.5	93.5	94.5	95.5	91.0	93.0	96.0
	δ	12.3	9.2	9.2	9.4	15.0	11.9	12.8	14.1	14.2	11.9
	$s_{r\max}$	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	$s_{r\min}$	60.0	70.0	70.0	70.0	40.0	50.0	50.0	60.0	40.0	50.0
k-NN	s_r	95.0	94.5	94.5	92.5	90.5	91.5	94.5	87.5	95.0	95.0
	δ	12.4	12.8	12.8	14.8	21.6	17.6	11.9	14.5	14.0	11.9
	$s_{r\max}$	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	$s_{r\min}$	60.0	60.0	70.0	60.0	20.0	60.0	50.0	10.0	60.0	70.0
NM	s_r	94.5	95.0	95.0	94.5	90.5	93.0	96.0	80.5	93.0	95.5
	δ	12.3	11.5	11.0	11.0	23.1	10.6	12.7	24.8	15.9	8.9
	$s_{r\max}$	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	$s_{r\min}$	60.0	60.0	60.0	60.0	20.0	60.0	50.0	10.0	40.0	70.0

1) Positively best average results of correct identification for sets of natural parameters (not normalized) $ZP1$ come from Camberr's distance function (enc. 9). This is due to a somewhat normalizing form of this function. Differences with respect to other similarity functions are smallest for the NN algorithm (2.5%), and greatest for NM (6.5%). Tests of significance performed for differences of results between d^{CAM} and $b^{\cos}$, d^{CHI}, which give closest average results of correct identification (see Table 4), have indicated the significance of these differences on significance lovel $\alpha = 0.05$.

Tabele 7. Parameters of the voice identification experiment arranged according to decreasing values of correct decisions

N°	Set of parameters	FP	Algorithm	s_r [%]	σ [%]	s_{rmax} [%]	s_{rmax} [%]
1	$ZP2$	Euk	NN	97.0	9.2	100	70
2	$ZP3$	Euk	NN	97.0	9.2	100	70
3	$ZP2$	Min3	NN	97.0	9.2	100	70
4	$ZP3$	Min3	NN	97.0	9.2	100	70
5	$ZP2$	Tan	NN	97.0	9.8	100	60
6	$ZP3$	Tan	NN	96.0	11.9	100	50
7	$ZP1$	Cam	NM	96.0	12.7	100	50
8	$ZP2$	Cam	NM	96.0	12.7	100	50
9	$ZP3$	Cam	NM	96.0	12.7	100	50
10	$ZP2$	Tan	NM	95.5	8.9	100	70
11	$ZP3$	Tan	NM	95.5	8.9	100	70
12	$ZP3$	Min3	NM	95.5	9.4	100	70
13	$ZP3$	Min5	NN	95.5	9.4	100	70
14	$ZP2$	Min5	NN	95.5	10.0	100	70
15	$ZP1$	Cam	NN	95.5	10.0	100	50
16	$ZP3$	Ham	NN	95.5	12.3	100	60
17	$ZP2$	Cam	NN	95.5	12.8	100	50
18	$ZP3$	Cam	NN	95.5	12.8	100	50

2) For sets of parameters $ZP1$ all other similarity functions gave best identification results for the NN algorithm and worst for the NM algorithm. This is also confirmed by so-called minimal probabilities of correct identification s_{rmin} for individual speakers. Their values decreased to 30, 20 and even 10% for Mahalonabis'es distance (Table 4).

3) The normalization of sets of parameters $ZP2$ and $ZP3$ resulted in an increase of voice identification correctness by several percent on the average for all algorithms except for Camberr's distance function (see point 1) which had exactly the same effectiveness as for $ZP1$.

4) Greatest differentiation of effectiveness occured for individual similarity functions in case of normalized parameters. Distance functions such as d^{EUK}, d^{MIN3} and neamess function b^{TAN} were distinguished, and the NN algorithm was distinguished as for $ZP1$.

5) Positively worst results (for the NM algorithm especially) were achieved with the square of Mahalanobis'es distance (enc. 10). This conclusion confirms results and

conclusions presented in TADEUSIEWICZ'es paper [5]; namely, that in certain cases better results can be reached with less complex similarity functions.

To recapitulate we can accept a general conclusion that it is advisable to use Camberr's distance function for natural parameters (directly from measurements). While it is sufficient to use Euklides'es distance function or Tanimoto's nearness function when parameters are normalized. The application of the square of Mahalanobis'es distance is not recommended, for short learning series especially. As to the evaluation of algorithms, the nearest mean *NM* algorithm achieves the positively lowest general rating. It is understandable that presented results can not (this concerns exact numerical values) be transferred directly for experiments with sets of parameters with different dimensions and structure. This finds confirmation in the differentiations achieved for *ZP*2 and *ZP*3. In order to achieve exact numerical values some experiments out of these presented above should be repeated at random at least.

References

[1] Cz. BASZTURA, *Sources, signals and acoustic images* (in Polish), WKiŁ, Warszawa 1988.

[2] Cz. BASZTURA, J. JURKIEWICZ, *Analysis of zero-crossings of a speech signal in a short-term model of automatic speaker identification* (in Polish) Arch. Akustyki **13**, 3, 203–214 (1978).

[3] Cz. BASZTURA, *Similarity functions of acoustic images as indicators of objective evaluation of speech quality transmission* (in Polish) Arch. Akustyki **22**, 3, 217–233 (1987).

[4] A. J. GRAY, J. D. MARKEL, *Distance measures for speech processing*, IEEE ASSP–24, **5**, 380–391 (1976).

[5] R. TADEUSIEWICZ, *Usability evaluation of chosen metric in minimal distance methods of speach recognition* (in Polish) Arch. Akustyki, **18**, 3, 275–284 (1983).

[6] J. ZALEWSKI, *A comparison of the effectiveness of some distances measures in speaker recongnition experiments*. Paper on the Speaker Recognition Working Group on the X-th International Congress of Phonetic Sciences, Utrecht 1983.

Received June 6, 1989