

## SPEAKER IDENTIFICATION METHODS BASED ON PSEUDOSTATIONARY SEGMENTS OF VOICED SOUNDS

Antanas LIPEIKA and Joana LIPEIKIENĖ

Institute of Mathematics and Informatics  
Akademijos 4, 2600 Vilnius, Lithuania  
E-mail: lipeika@ktl.mii.lt

**Abstract.** The problem of speaker identification is investigated. Basic segments - pseudostationary intervals of voiced sounds are used for identification. The identification is carried out, comparing average distances between an investigative and comparatives. Coefficients of the linear prediction model (LPC) of a vocal tract, cepstral coefficients and LPC coefficients of an excitation signal are used for identification as features. Three speaker identification methods are presented. Experimental investigation of their performance is discussed.

**Key words:** speaker identification, likelihood ratio distance, cepstral distance, vocal tract, excitation signal, reliability reserve, vector quantization.

**1. Introduction.** The automatic speaker identification problem (Ramishvili, 1981) is very urgent in the forensic examination. It is more difficult to identify a speaker by his speech phonogram than, for example, by finger-prints (Ramishvili, 1991). The latters are unique and their picture in practice does not change all the life, meanwhile human voice changes in time, it depends on the emotional state and other factors. Besides, a voice phonogram is distorted when recording (influence of an environment noise, imperfect recording equipment, etc.). Therefore this investigation field is being intensively developed. Researchers are seeking for the selection of features, a structure of an identification system and a decision rule that would enable to distinguish speakers with most reliability.

Possibilities of speaker identification by pseudostationary segments of voiced sounds are investigated in this paper. When pronouncing a voiced sound, a vocal tract is fixed for a short period, therefore there occurs a possibility to "measure" parameters of a vocal tract and to identify a speaker using phonograms.

For detection of pseudostationary segments we applied the method (Lowerre, 1980), which is often used in speech recognition. According to this method two neighbouring frames are compared, calculating the likelihood ratio distance between them (Juang *et al.*, 1982). When the distance exceeds the threshold, chosen in advance, it is warned about the end of a pseudostationary segment.

When we want to identify a speaker, we make the assumption that we have phonograms of an unknown speaker (investigative) and  $n$  known suspicious speakers (comparatives). Our purpose is to choose from these comparatives the person that is in some sense closest to the investigative and then to determine, whether this closest person and comparative is the same person. So the problem of identification of the closest suspicious is investigated in this paper and the quality of decision of this problem is evaluated experimentally.

Three methods are used for solution of the identification problem. The first method is based on an average distance between clusters that are formed by vectors of the linear prediction (LPC) coefficients of pseudostationary intervals of phonograms of an investigative and comparatives. The second is called vector quantization method. The point of the method is that clusters of the LPC coefficients (feature vectors) are divided into subclusters and an average distance between centers (centroids) of clusters is calculated. The process when a complicated cluster of features is divided into subclusters is called a clusterization or code-book generation. The third method is based on the use for identification of the LPC coefficients that correspond to a vocal tract of a speaker and the LPC coefficients that correspond to excitation signal of the vocal tract. The main point is that not only parameters of a vocal tract but also parameters of an excitation signal have information about a speaker and the joint use of these parameters should improve an identification quality.

At last the criterium of identification quality is introduced that is called the reliability reserve. It enables to compare identification methods even when the number of identification errors is the same or there are no errors at all.

**2. Detection of pseudostationary segments.** For detection of pseudostationary segments in speaker identification a phonogram is divided into frames (segments) the length of which is  $N$  digital points of speech signal and they are moved with respect to one another by  $M$  points (a step of a frame is  $M$ ). A filtration of a speech signal  $y_t$  for all frames is done according to  $x_t = y_t - 0.94y_{t-1}$  (Tribolet *et al.*, 1979). This filtration enables to suppress

irregular low frequency components. After that a resulted signal is processed using Hamming window (Marple, 1987). The use of Hamming window and low frequency filtration enables to get a stable LPC model of a speech signal. We evaluate parameters of the LPC model by correlation method, using Durbin algorithm (Rabiner *et al.*, 1978; Markel *et al.*, 1976). After that, autocorrelation coefficients of these parameters are calculated, using LPC coefficients. Further, using autocorrelation coefficients of linear prediction parameters of a previous frame and a correlation coefficients of a signal of next frame, divided by square LPC model gain coefficient, we calculate the likelihood ratio distance (Juang *et al.*, 1982) for all neighbouring pairs of frames. If a distance between two neighbouring frames is less than a preassigned threshold (the threshold is chosen experimentally), we draw the conclusion that moving by a frame step does not change a spectral structure of a signal, that means, it is pseudostationary. This condition is checked until the likelihood ratio distance exceeds the threshold. Then we consider that stationary interval terminated. Since we are not interested in very short pseudostationary intervals, we compare them with the threshold of the minimal pseudostationary segment and leave for further investigation only those pseudostationary segments which are longer than this threshold. The likelihood ratio distance has the spectral interpretation (Juang *et al.*, 1982):

$$d_{LR}(\tilde{S}, S) = \int_{-\pi}^{\pi} \frac{\tilde{S}(\theta)/\tilde{b}^2}{S(\theta)/b^2} \frac{d\theta}{2\pi} - 1, \quad (1)$$

where  $S(\theta)$  and  $\tilde{S}(\theta)$  are spectral densities of LPC model of the first and the second frame, respectively,  $\tilde{b}^2$  and  $b^2$  are square gain coefficients of those models.

Due to the great computation amount it is not convenient to calculate the likelihood ratio distance using (1). It is usually calculated in time domain:

$$d_{LR}(\tilde{S}, S) = \left\{ \frac{r_x(0)}{b^2} r_a(0) + 2 \sum_{i=1}^p \frac{r_x(i)}{b^2} r_a(i) \right\} - 1, \quad (2)$$

where  $r_x(i)$  is the autocorrelation function of a signal in the second frame,  $r_a(i)$  are autocorrelations of LPC model parameters for the first frame:

$$r_a(i) = \sum_{k=0}^{p-i} a_{k+i} a_k, \quad i = 0, 1, 2, \dots, p, \quad (3)$$

where  $p$  is the order of LPC model.

**3. Identification method based on a calculation of the average distance between clusters.** Let us have  $N_x$  pseudostationary intervals of investigative speaker ( $X$ ) and  $N_{A_i}$  pseudostationary intervals of comparatives ( $A_i$ ). Let us calculate all possible distances  $d_{ji}(x, A_i)$  between the pseudostationary intervals of investigative  $X$  and comparatives  $A_i$ ,  $i = 1, \dots, n$ . Then the average distance between the cluster, describing investigative  $X$ , and the cluster, describing comparatives  $A_i$ , may be calculated according (Lipeika *et al.*, 1993a; 1993b):

$$D_{X,A_i} = \frac{1}{N_X} \sum_{j \in X} \min_{l \in A_i} d_{jl}(X, A_i) + \frac{1}{N_{A_i}} \min_{l \in X} d_{jl}(X, A_i). \quad (4)$$

Here  $N_X$  and  $N_{A_i}$  are the numbers of frames in phonograms of investigative  $X$  and comparatives  $A_i$ ;  $d_{ji}(X, A_i)$  is the distance between frames.

When detecting the pseudostationary segments we used the likelihood ratio distance. But this measure is not symmetric, i.e.,

$$d_{LR}(\tilde{S}, S) \neq d_{LR}(S, \tilde{S}). \quad (5)$$

It is not shortcoming in the detection of pseudostationary segments because a threshold is not high, meanwhile asymmetry appears when values of distance are large. But when calculating the average distance it is desirable that the distance in the formula (4) would be symmetric. So we make the likelihood ratio distance symmetric:

$$d(\tilde{S}, S) = \frac{d_{LR}(\tilde{S}, S) + d_{LR}(S, \tilde{S})}{2}. \quad (6)$$

After calculating of the average distance between investigative speaker  $X$  and all comparatives  $A_i$  we find "the closest" comparative, comparing the average distances:

$$\hat{I} = \min_{1 \leq i \leq n} D_{X,A_i} \quad (7)$$

The cepstral distance is also often used in speaker identification (Noda, 1989; Xu, 1989; Naik *et al.*, 1989; Bastura *et al.*, 1990). The cepstral coefficients can be calculated from LPC coefficients using formulas (Atal, 1976; 1974):

$$c_0 = \ln b^2,$$

$$c_1 = -a_1,$$

$$c_n = -\sum_{k=1}^{n-1} (1 - k/n) a_k c_{n-k} - a_n, \quad n = 2, \dots, p, \quad (8)$$

$$c_n = -\sum_{k=1}^{n-1} (1 - k/n) a_k c_{n-k}, \quad n = p + 1, \dots, L. \quad (9)$$

The cepstral distance between two frames of a speech signal with corresponding coefficients  $(a_1, \dots, a_p, b)$ ,  $(\tilde{a}_1, \dots, \tilde{p}, \tilde{b})$  may be defined as

$$d_{\text{cep}}(L) = [u(L)]^2 = (c_0 - \tilde{c}_0)^2 + 2 \sum_{k=1}^L (c_k - \tilde{c}_k)^2. \quad (10)$$

It is important to know (Gray *et al.*, 1976) that as  $L$  increases,  $u(L)$  approaches  $d_2$  from below and

$$\lim_{L \rightarrow \infty} u(L) = d_2, \quad (11)$$

where  $d_2^2$  has the following spectral interpretation:

$$d_2^2 = \int_{-\pi}^{\pi} \left| \ln \frac{\tilde{S}(\theta)}{S(\theta)} \right| \frac{d\theta}{2\pi}. \quad (12)$$

It should be mentioned that this distance is symmetric and convenient to use for speaker identification.

Usually we desire that a distance would not depend on gain, so we assume  $b = \tilde{b} = 1$ . Then (12) may be written as

$$d_2^2 = \int_{-\pi}^{\pi} \left| \ln \frac{\tilde{S}(\theta)/\tilde{b}^2}{S(\theta)/b} \right| \frac{d\theta}{2\pi}. \quad (13)$$

After putting the cepstral distance, described by (10), to (4) we carry out speaker identification by cepstral distance.

**4. Speaker identification using vector quantization.** According to a literature (Soong *et al.*, 1985; Buck *et al.*, 1985; Rosenberg *et al.*, 1986; Burton, 1987; Zinke, 1993; Irvine *et al.*, 1993) speaker identification using vector quantization is very popular. Our investigation differs from the known methods in

that generating a code book (carrying out a clusterization) we do not double a number of clusters (centroids) at every step, but increase it by 1. It enables us better to investigate a dependence of identification quality on a codebook length (a number of clusters). In addition we use the average distance (4), described in the previous method, for comparison of centers of an investigative and comparatives.

Further we present the description of the clustering process (Lipeika *et al.*, 1995b).

Let

$$R_j = \{r_j(0), r_j(1), \dots, r_j(p), b_j^2\}, \quad j = 1, \dots, K, \quad (14)$$

be the vector of  $K$  features, got from pseudostationary speech intervals, where  $r_j(0), \dots, r_j(p)$  are values of the autocorrelation function of the  $j$ -th pseudostationary segment.  $b_j^2$  is a square gain of the LPC model.

**Calculation of the zero centroid.** We may calculate a "gravity center" or the so called zero centroid of a cluster that consists of feature vectors  $R_j$ . We update the zero centroid calculating the statistics

$$r^{(0)}(l) = \frac{1}{K} \sum_{j=1}^K r_j(l)/b_j^2, \quad l = 0, 1, \dots, p \quad (15)$$

and estimating the parameters of the linear prediction  $A_0 = (a_1^{(0)}, \dots, a_p^{(0)})$  from it. When estimating LPC parameters according to the Durbin method (Rabiner *et al.*, 1978) we obtain in parallel the reflection coefficients  $k_1^{(0)}, \dots, k_p^{(0)}$  that correspond to the zero centroid.

**Determination of the average distortion while describing features by one reference pattern.** When solving this problem we answer the question what an average error we are making if we describe all features by one reference pattern

$$D(A_0) = \frac{1}{K} \sum_{j=1}^K d(R_j, A_0), \quad (16)$$

where  $d(R_j, A_0)$  is the likelihood ratio distance between the feature vectors  $R_j$  and centroid  $A_0$ . The likelihood ratio distance is calculated according to formula (2).

If the average distortion  $D(A_0)$  exceeds the given threshold  $\delta$ , then we must form two centroids from the zero centroid, which would represent the feature vectors  $R_j$ ,  $j = 1, \dots, k$ , more exactly, to make the average distortion less. Formation of two new centroids. Formation of new centroids is an iterative procedure. The initial point of this process is the reflection coefficients corresponding to the zero centroid. We distort the reflection coefficients, multiplying them by multipliers 0.99 and 1.01, respectively. Thus from the zero centroid we got two new initial centroids, whose coordinates determine two collections of the reflection coefficients  $k_1^{(1)}, \dots, k_p^{(1)}$  and  $k_1^{(2)}, \dots, k_p^{(2)}$ . From the latter, using the recurrent relation (Rabiner *et al.*, 1978), we may calculate LPC model parameters, corresponding to these initial centroids. The LPC model parameters are calculated in such a way:

$$\alpha_i^{(i)}(j) = k_i^{(j)}, \tag{17}$$

$$\alpha_l^{(i)}(j) = \alpha_l^{(i-1)}(j) - k_i^{(j)}\alpha_{i-l}(j), \quad l = 1, \dots, i - 1. \tag{18}$$

When solving (17) and (18) for  $i = 1, \dots, p$ ,  $j = 1, 2$ , we obtain that

$$a_l^{(j)} = \alpha_l^{(p)}(j), \quad l = 1, \dots, p; \quad j = 1, 2.$$

The coordinates of these two centroids expressed by the LPC model coefficients  $(a_1^{(j)}, \dots, a_p^{(j)})$ ,  $j = 1, 2$ , are used to determine the distance of each feature vector  $R_j$ ,  $j = 1, \dots, K$  from these centroids, using formula (8). Further, using the nearest neighbor rule, on the basis of calculated distances we classify the features  $R_j$ ,  $j = 1, \dots, K$ . Every feature is attached to a centroid which is closer to this feature. According to (16), the average distortion is assessed, which caused by the description of  $R_j$ ,  $j = 1, \dots, K$ , by two reference patterns, corresponding to the two initial centroids. For that we rewrite (16) in the following way:

$$D(A^{(1)}, A^{(2)}) = \frac{1}{K} \sum_{j=1}^K d^*(R_j, A^{(l)}), \tag{19}$$

where  $d^*(R_j, A^{(l)}) = \min\{d(R_j^l, A^{(1)}), d(R_j, A^{(2)})\}$ .

As a result of classification by the nearest neighbor rule we obtain that features  $R_j$ ,  $j = 1, \dots, K$  are divided into two initial clusters. As we have already done in the case of the zero centroid, according to (15) we find centers of gravity of these clusters or the so called two improved initial centroids and

their representation by the LPC parameters. Further, according to the same formulas, we again calculate the distances of features  $R_j$ ,  $j = 1, \dots, K$ , from the improved initial centroids and classify the features according to the nearest neighbor rule. On the basis of classification results, the average distortion is calculated according to (19), which is due to the replacement of two reference patterns, describing the features  $R_j$ ,  $j = 1, \dots, K$ , by LPC parameters, corresponding to the two initial centroids. If the average distortion decreases more than the given threshold  $\varepsilon$ , a further specification of centroid position is continued. If it increases less than  $\varepsilon$ , the iterative procedure is terminated. At the same time the procedure of LPC parameter estimation is stopped too. If the average distortion is less than the given threshold  $\delta$ , the cluster, which caused the largest average distortion, is divided into two clusters and the clustering process continues. It terminates only when the average distortion is less than the given quantity  $\delta$  or when the number of centroids coincides with the largest given number of centroids. All calculations are carried out according to the same formulas as in the case of two centroids.

**5. Identification based on the linear prediction parameters of a vocal tract and an excitation signal.** The first two speaker identification methods use parameters of the linear prediction model, which contain information about a vocal tract as features for identification. But they do not have information about an excitation signal of a vocal tract. The excitation signal of a voiced sound has an important parameter that characterizes a speaker. It is a period of pitch of the excitation signal.

We investigated the possibility to use a pitch period as a feature for speaker identification and ascertained that the pitch may be used for speaker identification, though only this one feature can not characterize a speaker entirely. Very often two different persons have the same or very close pitch period. In addition, it is quite difficult to separate a pitch when a speech is disturbed what often happens in forensic examination. That means, a pitch period must be used together with other features, i.e., coefficients of the linear prediction. But there occurs a problem of prescribing of weights for different features. It is not clear what influence a pitch can have for making decisions.

We have solved this problem, describing not only a vocal tract, but also an excitation signal by the linear prediction model (Lipeika *et al.*, 1995a). Then the problem of choosing weights and detection of pitch disappears. The successful



use of vocal tract LPC parameters and cepstral representation of excitation signal in speaker recognition was reported (Thévenaz, 1995; He, 1995). As it was mentioned, we use LPC modeling of the excitation signal instead of cepstral representation.

Let us consider a feature selection when describing an excitation signal of a vocal tract by the linear prediction model. After detection of a pseudostationary segment in a speech signal we estimate the linear prediction model parameters that we use for identification according to the first two methods. When we have the estimated LPC model parameters we can write that the speech signal in the pseudostationary segment is

$$x_t = -a_1x_{t-1} - a_2x_{t-2} - \dots - a_px_{t-p} + \nu_t, \quad (20)$$

where  $a_1, \dots, a_p$  are the LPC model parameters,  $\nu_t$  is the excitation signal of a vocal tract. So we can find the excitation signal of the vocal tract

$$\nu_t = x_t + a_1x_{t-1} + a_2x_{t-2} + \dots + a_px_{t-p}. \quad (21)$$

This signal is low frequency signal if compare it with the speech signal. The excitation signal must be decimated. But before decimation it is necessary to remove high frequency components with the aim not to have frequency aliasing after decimation. As we need to estimate the LPC model parameters of the decimated excitation signal, virtually we estimate not the signal but its correlation function. So we get more exact estimate of the correlation of the excitation signal because when calculating we use more points of the signal. Using the correlation we estimate the LPC model parameters and use them for identification together with the linear prediction model parameters of the vocal tract.

The identification procedure itself is different from the described earlier. First of all we compare the LPC model parameters of the vocal tract. For every collection of the LPC model parameters of an investigative we find a collection of the LPC model parameters of the vocal tract of the comparative speaker that correspond to minimal distance. Then the likelihood ratio distance is calculated between collections of the LPC model parameters of the excitation signal corresponding to these parameters collections.

Similar procedure is carried out replacing the investigative and the comparative. Then all distances are added up and divided by their number. So we get

the average distance, similar to (4), but here are included not only parameters of the vocal tract, but also the parameters of the excitation signal.

Let us try to write this average distance formally. Let we have  $N_x$  pseudostationary segments  $s_j(X)$ ,  $j = 1, \dots, N_x$ , of the investigative and  $N_{A_i}$  pseudostationary segments  $s_l(A_i)$ ,  $l = 1, \dots, N_{A_i}$ , of the  $i$ -th comparative. For every pseudostationary segment of the investigative, comparing the LPC model parameters of the vocal tract corresponding to it with the pseudostationary segments of the comparative speaker according to the likelihood ratio distance we got:

$$d_j^{(1)}(X, A_i) = d_{jl^*}(X, A_i), \quad (22)$$

where

$$l^* = \arg \min_{l=1, \dots, N_{A_i}} d_{jl}(X, A_i), \quad (23)$$

and

$$d_j^{(2)}(X, A_i) = d'_{jl^*}(X, A_i), \quad (24)$$

$$l^* = \arg \min_{l=1, \dots, N_{A_i}} d'_{jl}(X, A_i), \quad (25)$$

where  $d'_{jl^*}(X, A_i)$  is the likelihood ratio distance between the  $j$ -th collection of the LPC model parameters of the investigative speaker's excitation signal and  $l^*$ -th collection of the LPC model parameters of the comparative speaker's excitation signal. Replacing the investigative and the comparative we similarly get the distances  $d_i^{(1)}(A_i, X)$  and  $d_i^{(2)}(A_i, X)$ .

Then the average distance between the investigative and the comparatives is

$$D_{XA_i}^* = \frac{1}{2N_x} \sum_{j=1}^{N_x} ((1 - \alpha)d_j^{(1)}(X, A_i) + \alpha d_j^{(2)}(X, A_i)) + \frac{1}{2N_{A_i}} \sum_{l=1}^{N_{A_i}} ((1 - \alpha)d_l^{(1)}(A_i, X) + \alpha d_l^{(2)}(A_i, X)), \quad (26)$$

where  $\alpha$  is the weight assigned to the influence of the excitation signal ( $0 < \alpha < 1$ ). Then the "closest" comparative we find comparing all average distances according to

$$\hat{I}^* = \min_{1 \leq i \leq n} D_{XA_i}^*. \quad (27)$$

We notice, that if  $\alpha = 0$  we have an identification, based on average distance corresponding to vocal tract (1-st method). If  $\alpha = 1$ , we have an identification,

based on average distance corresponding to pure excitation signal. If  $\alpha = 0.5$ , equal weights are assigned to the average distances corresponding to the vocal tract and excitation signal.

**6. Estimation of identification quality.** For estimation quality of an identification method according to identification errors one needs a great data base which consist of at least one hundred speakers. When a data base is small, using different methods very often the same number of errors is made or there are no errors at all. As we have no possibility to form a great data base of speakers we must search for other ways to estimate identification quality.

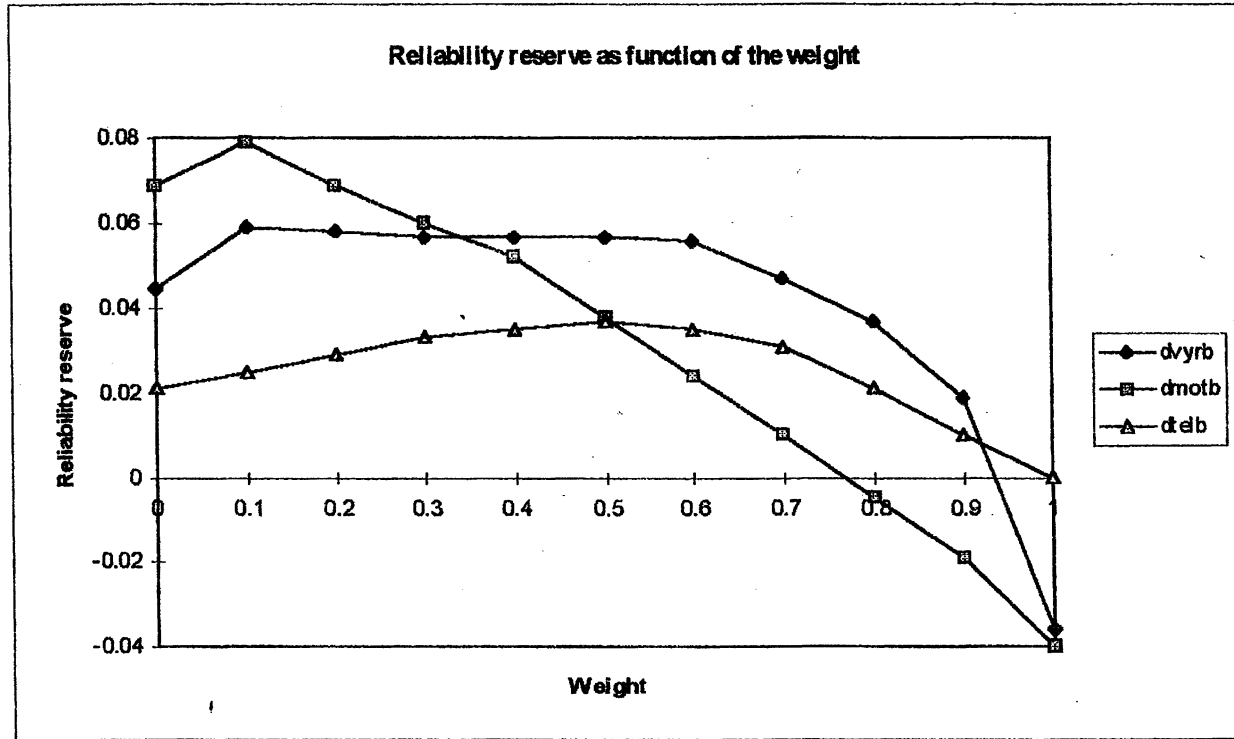
The *reliability reserve* is used as a criterion of identification quality. Let we have in twos phonograms of N speakers. Every of them in turn we regard as investigative and the rest as comparatives and fulfill an identification. Then for every phonogram of an investigative we estimate an average distance to other phonogram of the same person (“to himself”) and an average distance to the rest phonograms. Subtracting the distance “to himself” from the minimal average distance we get the function whose minimum is the reliability reserve. That means the reliability reserve shows how much an average distance to an other person is greater than an average distance to himself in the worst situation of a given data base when other person is “closest” to an investigative. The greater the reliability reserve the better the identification method. When there are identification errors the reliability reserve is negative. A number of negative values of the function is equal to the number of identification errors.

**7. Experiments and conclusions.** For verifying the effectiveness of identification methods three data bases in twos phonograms for each person were formed at the Lithuanian Institute of Forensic Examination:

- ten men;
- five women;
- four men and one woman (records over telephone).

Using these data bases identification experiments were carried out. Results of the third method are presented in Table 1 and Fig. 1 for different  $\alpha$  values.

The results of experiments showed that the using of LPC parameters corresponding to vocal tract and excitation signal may improve identification results. For men and women maximum of reliability reserve was obtained at  $\alpha = 0.1$ , meanwhile for telephone speech – at 0.5. Negative reliability reserve indicates



**Fig. 1.** Reliability reserve as function of the weight.

**Table 1.** Reliability reserve for different  $\alpha$  values

$\alpha$	0.0	0.1	0.2	0.3	0.4	0.5
men	0.045	0.059	0.058	0.057	0.057	0.057
wom.	0.069	0.079	0.069	0.060	0.052	0.038
tel.	0.021	0.025	0.029	0.033	0.035	0.037

$\alpha$	0.6	0.7	0.8	0.9	1.0
men	0.056	0.047	0.037	0.019	-0.036
wom.	0.024	0.010	-0.005	-0.019	-0.040
tel.	0.035	0.031	0.021	0.010	0.001

identification errors.

The second (vector quantization) method gives worse results, evaluating quality according to the reliability reserve:  $RR = -0.187$  (three identification errors) for men;  $RR = 0.041$  for women;  $RR = 0.010$  for telephone speech.

The cepstral distance also was used in the 1-st method.  $RR = 0.054$  for men;  $RR = 0.089$  for women and  $RR = 0.028$  for telephone speech. Unfortunately, we can't compare these RR values with those obtained for other methods, because the reliability reserve is suitable for evaluation the performance only if methods use the same distance for comparing feature vectors.

It should be mentioned that larger data bases should be used for more serious conclusions, but we have no conditions to form them.

**8. Acknowledgment.** The authors wishes to thank Dr. B. Šalna, head of the Department of Phonoscopic Expertise, the Lithuanian Institute of Forensic Expertise for encouragement and valuable discussions and research associate V. Makarevič for support in performing experiments.

#### REFERENCES

- Atal, B.S. (1974). Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification. *J. Acoustic Soc. Amer.*, **55**(6), 1304–1312.
- Atal, B.S. (1976). Automatic recognition of speakers from their voices. *Proc. IEEE*, **64**(4), 460–405.

- Bastura, C., W. Majewski, W. Myslecki and J. Zalewski (1990). *Metody Parametryzacji Sygnału Mowy do Automatycznego Rozpoznawania Głosów*. Wydawnictwo Politechniki Wrocławskiej, Wrocław.
- Buck, T., D. Burton and J. Shore (1985). Text dependent speaker recognition using vector quantization. *Proc. of the ICASSP*, pp. 391–394.
- Burton, D. (1987). Text dependent speaker verification using vector quantization source coding. *IEEE Trans. on Acoustic Speech and Signal Processing*, 391–394.
- Gray, A.H., and J.D. Markel (1976). Distance measures for speech processing. *IEEE Trans. on Acoustic Speech and Signal Processing*, ASSP-24(5), 380–391.
- He, J., L. Liu and G. Palm (1995). On the use of features from prediction residual signals in speaker identification. *Proc. of the 3rd European Conference on Speech Communication and Technology*, Berlin, Germany, 21–23 September. pp. 2275–2278.
- Irvine, D., and F.A. Owens (1993). Comparison of speaker recognition techniques for telephone speech. *Proc. of the 3rd European Conference on Speech Communication and Technology*, Berlin, Germany, 21–23 September. pp. 2275–2278.
- Juang, B.H., D.Y. Wang and A.H. Gray (1982). Distortion performance of vector quantization for LPC voice coding. *IEEE Trans. on Acoustic Speech and Signal Processing*, ASSP-30(2), 294–304.
- Lipeika, A., and J. Lipeikienė (1995a). Recent advances in speaker identification. *Proc. of the Workshop "Integration of Language and Speech: Academia and Industry"*, Moscow.
- Lipeika, A., and J. Lipeikienė (1995b). Speaker identification using vector quantization. *Informatika*, 6(2), 167–180.
- Lipeika, A., and J. Lipeikienė (1993a). Speaker identification. *Informatika*, 4(1), 45–56.
- Lipeika, A., and J. Lipeikienė (1993b). The use of pseudostationary segments for speaker identification. *Proc. of the 3rd European Conference on Speech Communication and Technology*, Berlin, Germany, 21–23 September. pp. 2303–2306.
- Lowerre, B. (1980). The harpy speech understanding system. In Wayne A. Lea (Eds.), *Trends in Speech Recognition*. Prentice-Hall.
- Markel J.D., and A.H. Gray (1976). *Linear Prediction of Speech*. Springer-Verlag.
- Marple, S.L. (1987). *Digital Spectral Analysis with Applications*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey.
- Naik, J.M., L.P. Netsch and G.R. Doddington (1989). Speaker verification over long distance telephone lines. *Proc. of the ICASSP*, pp. 524–527.
- Noda, H. (1989). On the use of the information on individual speaker position in the parameter space for speaker recognition. *Proc. of the ICASSP*, pp. 516–519.
- Rabiner, L.R., and R.W. Schafer (1978). *Digital Processing of Speech Signals*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey.
- Ramishvili, G., and G. Chikoidze (1991). Forensic investigation of speech phonograms and speaker identification. *Mecniereba*, Tbilisi (in Russian).

- Ramishvili, G. (1984). *Automatic Speaker Recognition by Voice*. Radio i sviaz, Moscow (in Russian).
- Rosenberg, A., and F. Soong (1986). Evaluation of the vector quantization talker recognition system in text independent and text dependent models. *Proc. of the ICASSP*. pp. 873–876.
- Soong, F.K., A.E. Rosenberg L.R. Rabiner, and B.H. Juang (1985). A vector quantization approach to speaker recognition. *Proc. of the ICASSP*. pp. 387–390.
- Thévenaz, P., and H. Hügli (1995). Usefulness of the LPC-residue in text-independent speaker verification. *Speech Communication*, **17**, 145–157.
- Tribolet, J.M., L.R. Rabiner and M.M. Sandhi (1979). Statistical properties of LPC distance measure. *IEEE Trans. on Acoustic Speech and Signal Processing*, **ASSP-27(5)**, 550–558.
- Xu, L., J. Oglesley and J.S. Mason (1989). The optimization of perceptually based feature for speaker identification. *Proc. of the ICASSP*. pp. 520–523.
- Zinke, J. (1993). Influence of pattern compression on speaker verification. *Proc. of the 3rd European Conference on Speech Communication and Technology*, Berlin, Germany, 21–23 September. pp. 2267–2270.

**A. Lipeika** is a Doctor of Technical Sciences, a senior researcher of the Recognition Processes Department at the Institute of Mathematics and Informatics. Scientific interests include: processing and recognition of random processes, detection of changes in the properties of random processes, signal processing, speech processing, speaker identification.

**J. Lipeikienė** is a Doctor of Technical Sciences, a senior researcher of the Recognition Processes Department at the Institute of Mathematics and Informatics. Scientific interests include: processing of random signals, robust methods for determination of change-points in the properties of random processes, data compression.

**KALBANČIOJO IDENTIFIKAVIMO METODAI,  
BESIREMIANTYS VOKALIZUOTŲ GARSŲ  
PSEUDOSTACIONARIAIS INTERVALAIS**

Antanas LIPEIKA ir Joana LIPEIKIENĖ

Darbe nagrinėjamas kalbančiojo identifikavimo uždavinys. Baziniai segmentai – vokalizuotų garsų pseudostacionarūs intervalai yra naudojami identifikavimui. Identifikavimas vykdomas, lyginant vidutinį atstumą tarp tiriamojo ir lyginamųjų Balso tiesinės prognozės modelio (LPC) koeficientai, kepstriniai koeficientai ir sužadavimo signalo LPC koeficientai naudojami identifikavimui kaip požymiai. Pateikiami trys kalbančiojo identifikavimo metodai. Aptariamas jų eksperimentiniai tyrimai.