INFORMATICA, 2010, Vol. 21, No. 1, 1–12

The Use of Group Delay Features of Linear Prediction Model for Speaker Recognition

Algirdas BASTYS[1], Andrej KISEL[1], Bernardas ŠALNA[2]
[1] *Mathematics and Informatics Faculty, Vilnius University*
Naugarduko 24, LT-03225 Vilnius, Lithuania
[2] *Phonoscope Expertise Department, Lithuanian Forensic Expertise Institute*
Lvovo 19a, LT-09313 Vilnius, Lithuania
e-mail: algirdas.bastys@mif.vu.lt

Received: May 2009; accepted: September 2009

Abstract. New text independent speaker identification method is presented. Phase spectrum of all-pole linear prediction (LP) model is used to derive the speech features. The features are represented by pairs of numbers that are calculated from group delay extremums of LP model spectrum. The first component of the pair is an argument of maximum of group delay of all pole LP model spectrum and the second is an estimation of spectrum bandwidth at the point of spectrum extremum. A similarity metric that uses group delay features is introduced. The metric is adapted for text independent speaker identification with general assumption that test speech channel may contain multiple speakers. It is demonstrated that automatic speaker recognition system with proposed features and similarity metric outperforms systems based on Gaussian mixture model with Mel frequency cepstral coefficients, formants, antiformants and pitch features.

Keywords: linear prediction model, group delay, features, information theory, similarity metric, speaker recognition.

1. Introduction

Automatic speaker recognition quality still remains pretty low in comparison with other biometric identification methods based on fingerprints (Kisel *et al.*, 2008), irises and even faces analysis (Struc and Pavesic, 2009; Ribaric *et al.*, 2008). Conventionally, the front-end of the recognition system uses features such as cepstral, Bark or Mel frequency cepstral coefficients (Alsteris and Paliwal, 2007). The features are based on spectrum amplitude of the speech frames or their residual parts of linear prediction model (Lipeika and Lipeikiene, 1999). In our opinion mainstream of speaker recognition algorithms underestimate information contained in phase spectrum. The idea that spectrum phase can contain valuable information fo speaker recognition is not very surprising because it is known that traditional power spectrum resonant characteristics can be derived exclusively from the phase spectrum (Jinhai *et al.*, 1993; Murthy and Yeganarayana, 1991). To resolve stability problem, the phase spectrum of traditional Linear Prediction (LP) model is used. In Thiruvaran *et al.* (2007) the group delay features for speaker recognition were

derived directly from the Fourier spectrum of the speech frames. Such approach requires special techniques dealing with instabilities of unwarped Fourier spectrum. We combine Jinhai *et al.* (1993) and Yu and Wang (2003) techniques to extract group delay features of LP model. In Jinhai *et al.* (1993) third order derivatives of the LPC phase spectrum were used to extract speech formants. We explore only first and second derivatives of LPC phase spectrum. The zero-crossings of the second derivative provides information about formants positions. LPC phase first derivative at formants frequencies gives simple approximations of the formants bandwidth. In Yu and Wang (2003) a connection between LPC phase and Line Spectrum Frequencies (LSF) is described. That inspired us to construct a symmetrized form of LPC phase representation which saves features computation cost and gives simple formulas for approximation of LPC spectrum poles. Gaussian Mixture Model (GMM; see, Reynold and Rose, 1995; Kamarauskas, 2008) becomes a standard technique for modelling of distributions of speakers features and their comparison. Since our features are restricted to the rectangle $(0, \pi) \times (0, 1)$ we estimated features distribution using histogram technique and constructed an information theory based similarity measure for comparison of speech utterances.

2. Group Delay Features of All-Pole LP Model

2.1. *Linear Prediction*

In Linear prediction (LP) model (Itakura and Saito, 1968) samples of a speech frame are represented in the form

$$x_n = \sum_{i=1}^{P} a_i x_{n-i} + G e_n, \tag{1}$$

where $a_1, a_2, \ldots, a_P$ are the Linear Prediction Coefficients (LPCs), P is the model order, G and e_n are the excitation gain and source, respectively. The LPCs are derived adaptively for each 20–30 ms speech frame by minimisation of excitation mean square energy. For simplicity, we will assume that the order of LP model is uneven, i.e., $P = 2M - 1$. The *LPC spectrum* or the *transfer function* of the LP filtering is defined by

$$H(z) = \frac{G}{A(z)}, \tag{2}$$

where

$$A(z) = 1 - \sum_{i=1}^{2M-1} a_i z^{-i} \tag{3}$$

is the inverse filter. The LPC spectrum represents an envelope of the speech spectrum.

2.2. *Phase of Spectrum of LP Model*

Let us define symmetrical polynomial $p(z)$ and antisymmetrical polynomial $q(z)$ by the formulas

$$p(z) = \frac{z^M A(z) + z^{-M} A(z^{-1})}{2}, \tag{4}$$

$$q(z) = \frac{z^M A(z) - z^{-M} A(z^{-1})}{2i}, \quad i = \sqrt{-1}. \tag{5}$$

The $p(z)$ and $q(z)$ polynomials are related to the symmetrical polynomial $P(z)$ and antisymmetrical polynomial $Q(z)$ of Line Spectrum Frequencies (LSF) analysis (Yu and Wang, 2003) by the following formulas

$$P(z) = A(z) + z^{-2M} A(z^{-1}) = 2z^{-M} p(z), \tag{6}$$

$$Q(z) = A(z) - z^{-2M} A(z^{-1}) = 2iz^{-M} q(z). \tag{7}$$

On the unit circle $p(z)$ and $q(z)$ are real-valued,

$$|A(z)|^2 = p(z)^2 + q(z)^2, \tag{8}$$

and

$$p(z) + q(z)i = z^M A(z). \tag{9}$$

Equations (8) and (9) show that the frequency response and the phase of the transfer function of the LP model satisfy the equations

$$|H(z)| = \frac{G}{\sqrt{p(z)^2 + q(z)^2}}, \tag{10}$$

and

$$(\arg H)\big(e^{i\omega}\big) = \Phi(\omega) = M\omega - \arctan\left(\frac{q(e^{i\omega})}{p(e^{i\omega})}\right), \quad \omega \in [0, 2\pi). \tag{11}$$

2.3. *LPC Phase Spectrum Features*

The LPC spectrum in the all-pole representation has the following form:

$$H(z) = \frac{G}{\prod_{m=1}^{P}(1 - r_m e^{i\alpha_m} z^{-1})}, \tag{12}$$

where $r_m e^{i\alpha_m}$ is location of the mth pole of the LPC spectrum, and $\alpha_m \in [0, 2\pi)$ is the angular frequency of the pole. From Eq. (12) follows that the mth pole contributes to the

LPC phase spectrum with the additive term

$$\arctan\left(\frac{r_m \sin(\omega-\alpha_m)}{1-r_m\cos(\omega-\alpha_m)}\right).$$

Therefore for the first and second phase spectrum derivative we have

$$\frac{d\Phi(\omega)}{d\omega}=\sum_m \frac{r_m(\cos(\omega-\alpha_m)-r_m)}{1-2r_m\cos(\omega-\alpha_m)+r_m^2} \tag{13}$$

and

$$\frac{d^2\Phi(\omega)}{d\omega^2}=-\sum_m \frac{r_m(1-r_m^2)\sin(\omega-\alpha_m)}{(1-2r_m\cos(\omega-\alpha_m)+r_m^2)^2}. \tag{14}$$

The negative derivative of phase of the LP spectrum is called group delay of LP model. The poles locations are not estimated and the phase spectrum derivatives are calculated by numerical differentiation of Eq. (11) identity to reduce calculation time.

Equation (13) gives that for the strong pole with r_m close to 1 one can expect local maximum of the group delay at a point ω_m close to the angular frequency α_m. The local maximum ω_m can be found as second derivative zero-crossing point which is closest to the α_m. Equation (13) gives

$$\Phi'(\omega_m)\approx\frac{r_m}{1-r_m} \tag{15}$$

and

$$r_m\approx\frac{\Phi'(\omega_m)}{1+\Phi'(\omega_m)}. \tag{16}$$

Considering the provided observations, we define the group delay features of a speech frame as a set of pairs

$$\left(\omega_m,\frac{1}{1+\Phi'(\omega_m)}\right)=(\omega_m,\delta_m), \tag{17}$$

where $\{\omega_m\}_m$ is the set of all zero crossings of the phase spectrum second derivative that belong to the radian frequency interval $(0,\pi)$ and

$$\delta_m=1-\frac{\Phi'(\omega_m)}{1+\Phi'(\omega_m)}=\frac{1}{1+\Phi'(\omega_m)} \tag{18}$$

defines a bandwidth of a formant of the speech frame.

3. Speech Utterance Similarity Measure for Speaker Identification

Suppose there are two sampled speech utterances $\{x_n\}$ and $\{y_n\}$ and similarity between them must be measured. Let's assume that $\{x_n\}$ samples belong to a speaker X of the training set and $\{y_n\} = Y$ samples belong to a speech utterance of one, two or even more test speakers. The similarity measure should estimate the probability that speaker X of the training set speaks in Y speech utterances. Such speaker recognition scenario occurs in forensic evaluation of the evidence using automatic speaker recognition systems. In forensic evaluation speech utterance of a training speaker can be recorded in a separate channel or manually segmented from multi speakers speech utterances, and test speech utterances may consists of natural records of persons under investigation.

3.1. *Features Statistics*

In previous section we introduced LPC phase spectrum variation features which for the kth speech frame consist of (f_m^k, δ_m^k) pairs where f_m^k is the frequency position of the mth local maximum of the group delay and δ_m^k a bandwidth of the extremum point. The speech utterances are divided into short time intervals of 1 s. Duration and distribution of the group delay features of their frames is estimated. Since distance between two neighbour frames is 0.01 s, we have about $100(M-1)$ pairs (f_m^k, δ_m^k) of features in 1 s. duration utterance. Distribution of $(f_m^k, \delta_m^k) \in (0, \frac{FS}{2}) \times (0,1)$ is estimated by division of $(0, \frac{FS}{2}) \times (0,1)$ into $N \times L$ rectangular boxes and calculating number of pairs (f_m^k, δ_m^k) that belong to the boxes. Warping parameter $\lambda = \lambda(FS)$ is adapted to sampling frequency FS so that division of frequency range $(0, \frac{FS}{2})$ in equal width intervals corresponds roughly to the Bark frequency scale. Possible bandwidth interval (0,1) is divided into increasing width intervals of total number 10.

3.2. *On Mutual Information Based Similarity Measure of Two Short Speech Utterances*

Similarity measure between two speech utterances is defined as a mutual information of the two group delay feature distributions. Let $I = N \times L$ is total number of all possible rectangular boxes $\{B_i\}_{i=1}^I$ and $C_X^x = \{c_i^x\}_{i=1}^I$ and $C_Y^y = \{c_i^y\}_{i=1}^I$ are feature vectors which components are numbers of group delay features belonging to boxes B_i. By definition, all the c_i^x and c_i^y correspond to $[x, x+1)$ and $[y, y+1)$ seconds time intervals of X and Y speech utterances respectively. Let H_X^x and H_Y^y are Shannon's entropies of the C_X^x and C_Y^y counts, i.e.,

$$H_X^x = -\sum_{i=1}^{I} c_i^x / |C_X^x| \log_2 \left(c_i^x / \left| C_X^x \right| \right), \tag{19}$$

$$H_Y^y = -\sum_{i=1}^{I} c_i^y / |C_Y^y| \log_2 \left(c_i^y / \left| C_Y^y \right| \right), \tag{20}$$

$$|C_X^x| = \sum_{i=1}^{I} c_i^x, \qquad |C_Y^y| = \sum_{i=1}^{I} c_i^y. \tag{21}$$

Let $C^{x,y}_{X,Y} = \{c^x_i + c^y_i\}^I_{i=1}$ denotes conjoint counts of C^x_X and C^y_Y and

$$H^{x,y}_{X,Y} = -\sum_{i=1}^{I} c^{x,y}_i / \left|C^{x,y}_{X,Y}\right| \log_2 \left(c^{x,y}_i / \left|C^{x,y}_{X,Y}\right|\right) \tag{22}$$

is the Shannon's entropy of the $C^{x,y}_{X,Y}$. It is easy to prove the following statement about a relation between the three entropies.

Theorem 1. *For any counts C^x_X and C^y_Y and their conjoint count $C^{x,y}_{X,Y}$ the following inequalities hold true*:

$$pH^x_X + qH^y_Y \leqslant H^{x,y}_{X,Y} \leqslant pH^x_X + qH^y_Y + H_{p,q}, \tag{23}$$

where

$$p = \frac{|C^x_X|}{|C^{x,y}_{X,Y}|}, \qquad q = \frac{|C^y_Y|}{|C^{x,y}_{X,Y}|} = 1 - p, \tag{24}$$

and

$$H_{p,q} = -p \log_2 p - q \log_2 q. \tag{25}$$

Proof. The Gibbs' inequality (MackKay, 2003) for any two distributions p_i and q_i gives

$$-\sum_{i=1}^{I} p_i \log_2 p_i \leqslant -\sum_{i=1}^{I} p_i \log_2 q_i.$$

Applying this inequality for $p_i = c^x_i/|C^x_X|$ or $p_i = c^y_i/|C^y_Y|$ and $q_i = c^{x,y}_i/|C^{x,y}_{X,Y}|$ we have

$$\begin{aligned} pH^x_X + qH^y_Y &= -\sum_{i=1}^{I} c^x_i/|C^{x,y}_{X,Y}| \log_2(c^x_i/|C^x_X|) - \sum_{i=1}^{I} c^y_i/|C^{x,y}_{X,Y}| \log_2(c^y_i/|C^y_Y|) \\ &\leqslant -\sum_{i=1}^{I} c^{x,y}_i/|C^{x,y}_{X,Y}| \log_2(c^{x,y}_i/|C^{x,y}_{X,Y}|) = H^{x,y}_{X,Y} \end{aligned}$$

that proves the left-hand side inequality of Eq. (23).

The right-hand side inequality of Eq. (23) can be justified by information theory reasoning. $H^{x,y}_{X,Y}$ is the average Shannon's information for appearance of a text letter of the text with $C^{x,y}_{X,Y}$ letters counts. The information about a letter of the text with conjoint $C^{x,y}_{X,Y}$ counts can be obtained using the following procedure. At first the question is

asked *"is this letter from the text with C_X^x or C_Y^y counts"*? Then, depending on the answer to the first question, the second question is asked *"which letter is from the text with C_X^x counts?"* or *"which letter is from the text with C_Y^y counts?"* with probability p and $q = 1 - p$ respectively. Answer to the first question contains $H_{p,q} = -p \log_2 p - q \log_2 q$ bits of information and the second one contains H_X^x or H_Y^y bits of information with probability p and q respectively. Since the strategy of provided two questions is not optimal in general, we have the right-hand side inequality of Eq. (23).

To provide a formal proof of he right-hand side inequality of Eq. (23) let us consider continuous function

$$f(x) = -x \log_2(x), \quad x \geqslant 0.$$

It is easy to check that this function is subadditive, that is

$$f(x + y) \leqslant f(x) + f(y) \quad \forall x, y \geqslant 0.$$

Really, if $y \geqslant 0$ is fixed than

$$\frac{d(f(x + y) - f(x) - f(y))}{dx} = \log_2 \left(\frac{x}{x + y} \right) \leqslant 0,$$

and

$$f(x + y) = f(x) + f(y) \quad \text{at } x = 0.$$

Therefore $f(x + y) \leqslant f(x) + f(y) \ \forall x, y \geqslant 0$. Applying this inequality we have

$$\begin{aligned} H_{X,Y}^{x,y} &= -\sum_{i=1}^{I} \left(pc_i^x / \left|C_X^x\right| + qc_i^y / \left|C_Y^y\right| \right) \log_2 \left(pc_i^x / \left|C_X^x\right| + qc_i^y / \left|C_Y^y\right| \right) \\ &\leqslant -\sum_{i=1}^{I} pc_i^x / \left|C_X^x\right| \log_2 \left(pc_i^x / \left|C_X^x\right| \right) - \sum_{i=1}^{I} qc_i^y / \left|C_Y^y\right| \log_2 \left(qc_i^y / \left|C_Y^y\right| \right) \\ &= pH_X^x + qH_Y^y + H_{p,q}. \end{aligned}$$

DEFINITION 1. Similarity of ρ of $[x, x+1)$ time interval (in seconds) of speech utterance of the X speaker to the $[y, y + 1)$ time interval of the Y speaker(s) is the number

$$\rho(X_{[x,x+1)}, Y_{[y,y+1)}) = 1 + \frac{pH_X^x + qH_Y^y - H_{X,Y}^{x,y}}{H_{p,q}}. \tag{26}$$

Proposition 1 gives that the similarity of any two speech utterances $X_{[x,x+1)}$ and $Y_{[y,y+1)}$ is always non-negative and not greater than 1. The next definition gives similarity of $Y_{[y,y+1)}$ short speech utterance to all the X utterances.

DEFINITION 2. Similarity of $Y_{[y,y+1)}$ short speech utterance to the X utterances is the number

$$\rho(X, Y_{[y,y+1)}) = \frac{\sum_{x=0}^{T_X-1} \rho(X_{[x,x+1)}, Y_{[y,y+1)})}{T_X}, \tag{27}$$

where T_X is the amount of seconds in X speech utterance.

In other words, similarity $\rho(X, Y_{[y,y+1)})$ is the average similarity of $Y_{[y,y+1)}$ utterance to the set of all of one second duration utterances $X_{[x,x+1)}$.

The last definition combines short segments similarities to an integrated similarity of X and Y utterances.

DEFINITION 3. Similarity of X speech utterances to the Y utterances is the number

$$\rho(X, Y) = \text{“average value of half biggest” } \rho(X, Y_{[y,y+1)}), \quad y = 0, 1, \ldots, T_Y - 1. \tag{28}$$

The provided similarity measure $\rho(X, Y)$ is asymmetrical (in general $\rho(X, Y) \neq \rho(Y, X)$). This is explained by the asymmetry in X and Y data: X consists of utterances of one speaker and Y may contain utterances of two or even more speakers. If a priori Y contains speech utterances of only one speaker too, the $\rho(X, Y)$ can be modified to symmetrical similarity by skipping *“half biggest”* words in Definition 3. All provided speech similarity measures are based on mutual information, are non-negative, and do not exceed 1. If X and Y are totally different, i.e., the X and Y group delay features points belong to non-intersecting sets of boxes B_i, then, with all x and y, $H_{X,Y}^{x,y} = pH_X^x + qH_Y^y + H_{p,q}$ and $\rho(X, Y) = 0$. In opposite case, when the all counts are proportional ($\forall x, y, i$: $c_i^x = \text{const } c_i^y$), $H_{X,Y}^{x,y} = pH_X^x + qH_Y^y = H_X^x$ and $\rho(X, Y) = 1$. Consequently, the similarity measure $\rho(X, Y)$ has a probabilistic interpretation: $\rho(X, Y)$ is a probability that X speaker participates in Y dialogue.

4. Experimental Results

4.1. *Preprocessing of Initial Data*

The following standard steps of initial data preprocessing were used in all our experimentations:

- Silent or low energy speech intervals were detected and removed from the further analysis.
- Sound data was pre-emphasized with first order filter of the form $1 - 0.95z$.
- Speech utterance was segmented into 30 ms frames with 20 ms overlapping.
- Frame samples were windowed with Hanning window.
- First order all-pass filter with warping parameter $\lambda \approx 0.5$ (Strube, 1980) was applied to the windowed speech data.

4.2. *A Graphical Illustration of Group Delay Features*

A speech frame with LPC log power spectrum represented in Fig. 1 illustrates ideas about group delay features. The first derivative of LPC phase spectrum of the same speech frame is presented in Fig. 2. Comparing LPC log power spectrum and LPC phase spectrum variation, one can notice that the last has two additional formants (maximums of the spectrum). The rest five formants of both spectrums have similar positions at frequency axis, however, peaks of the first derivative of LPC phase spectrum are more prominent than that of LPC log power spectrum.

Equation (16) approximation gives a "pole distance" of a chosen formant $f = f_m$ to the unit circle. Fig. 3 presents $-\log$ of the distances with marked points that correspond

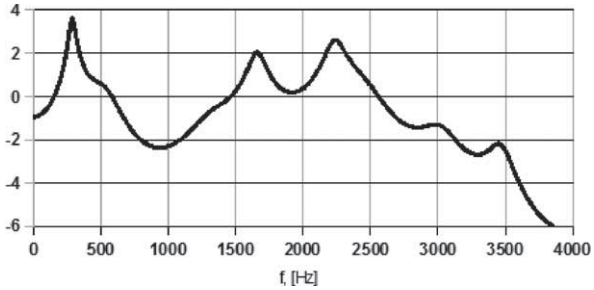

Fig. 1. LPC log power spectrum of a speech frame.

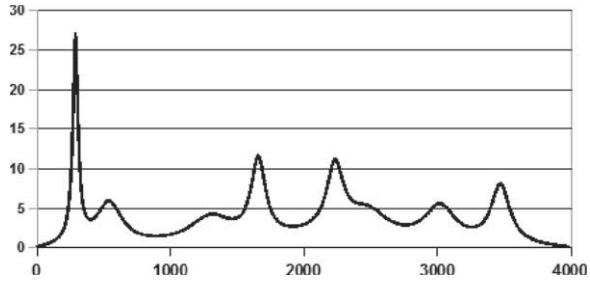

Fig. 2. First derivative of LPC phase spectrum of the same speech frame.

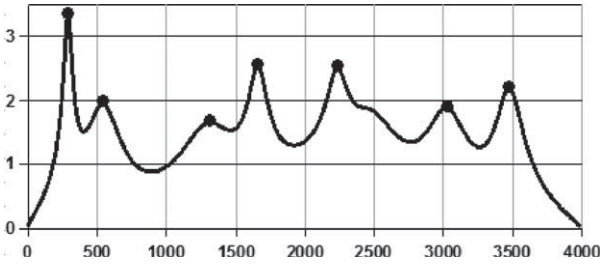

Fig. 3. $-\log$ band width with marked features points $(f_m, -\log(\delta_m))$ of the speech frame.

to formats. The coordinates of the marked points define pairs $(f_m, -\log(\delta_m))$ that form features vector of the speech frame.

4.3. *Experimentation Data Sets and Results*

Different speaker recognition techniques were compared using Russian Speech Data voice (RUSBASE) database which is distributed by ELRA (European Language Resources Association; ELRA-S0050, 1998) and data from the Netherlands Forensic Institute Speaker Recognition Evaluation (NFISRE). The NFISRE was conducted in 2004–2005 in order to compare the methods used by different forensic institutes belonging to the European network of forensic science institutes. NFISRE has two reference recordings containing speech utterances of a known and suspected speaker. Other test recordings contain from 20 s to 10 min speech utterances of two speakers. The NFISRE task was to determine if suspected speaker participates in provided test utterances. Correct training set was constructed by manual segmentation of 2 training recordings leaving only utterances of suspected speaker and recordings that were fully automatically checked. Ideal recognition was obtained by comparing with ground truth released by N FI (Gambier-Langeveld, 2005), that is – all impostor and genuine speakers were correctly classified.

RUSBASE is divided into 5 cases with approximately 15 sessions for each case. It contains 44 men and 35 women voices with total size of speech recordings about 500 Mb. First three sessions were used as a training set. Remaining sessions were used for testing.

RUSBASE recognition based group delay features were compared with Gaussian Mixture Model (GMM) that uses Mel Frequencies Cepstral Coefficients (MFCC), Formants and Antiformants (F&A), pitch value F0. Table 1 (Šalna and Kamarauskas, 2008) gives results of Equal Error Rates (EER) for speaker recognition on RUSBASE, case 1, men voices, using MFCC, F&A, and F0 features and Vector Quantisation (VQ; see, Lipeika and Lipeikiene, 1995) and GMM recognition methods. The EER ranges from 2.32 to 8.8% (see Table 1). On the group delay and mutual information based speaker recognition algorithm for the same data gives EER = 0,042%. Table 2 provides full results of speaker recognition of RUSBASE. Here FAR0 and FRR0 are Zero False Acceptance and Zero False Rejection rates, respectively.

Table 1

Recognition of RUSBASE speaker, case 1, voice man, using different methods and features

Method	Features	EER [%]
VQ	MFCC	8.8
GMM	MFCC	5.8
GMM	F&A	5.1
Phonemic	F&A	2.32

Table 2

Speaker recognition using phase spectrum features and of mutual information type similarity. RUSBASE data set, cases 1–5

Case	Voice	FAR0 [%]	EER0 [%]	FRR0
1	Man	1.8	0.042	0.12
1	Woman	1.96	0.042	0.07
2	Man	0.8	0.084	0.12
2	Woman	2.17	0.2	1.37
3	Man	3.19	0.058	0.09
3	Woman	1.96	0.033	0.06
4	Man	0.6	0.01	0.02
4	Woman	4.6	0.112	0.15
5	Man	2.79	0.199	0.59
5	Woman	0.44	0.007	0.01

5. Conclusions

It is shown that phase of transfer function defined by linear prediction model can be used for derivation of features of utterances. The features represent extremes of the group delay of the LP model. Similarity measure between two speech utterances was defined as a mutual information of the two group delay feature distributions. The performance of group delay features and their similarity metric was tested on two speakers datasets that contain text-dependent and text-independent utterances. The new speaker recognition technique showed up a reduction of equal error rate up to twenty times in comparison to traditional methods that use features derived exclusively from the amplitude of the power spectrum.

References

Alsteris, L.D., Paliwal, K.K. (2007). Short-time phase spectrum in speech processing: A review and some experimental results. *Digital Signal Processing*: *A Review Journal*, 17, 578–616.

ELRA-S0050 Russian speech database (STC). `http://www.linguistlist.org/issues/9/9-891.html`.

Gambier-Langeveld, T. (2005). NFD, speaker recognition fake case evaluation. In: *8th Meeting of ENFSI Expert Working Group for Forensic Speech and Audio Analysis*. Netherlands Forensic Institute.

Itakura, F., Saito, S. (1968). Analysis synthesis telephony based upon the maximum likelihoood method. In: Kohasi, Y. (Ed.), *Reports on 6th Int. Cong. Acoust.*. Tokyo. C-5-5, C17-20.

Jinhai, C., Gangji, J., Lihe, Z. (1993). New method for extracting speech formants using LPC phase spectrum. *Electronic letters*, 29(24), 2081–2082.

Kamarauskas, J. (2008). Speaker recognition using aussian mizture model. *J. Electronics and Electrical Enginiering*, 5(85), 29–32.

Kisel, A., Kochetkov, A., Kranauskas, J. (2008). Fingerprint minutiae matching without global alignment using local structures. *Informatica*, 19(1), 31–44.

Lipeika, A., Lipeikienė, J. (1995). Speaker identification using vector quantization. *Informatica*, 6(2), 167–180.

Lipeika, A., Lipeikienė, J. (1999). Speaker recognition based on the use of vocal tract and residue signal LPC parameters. *Informatica*, 10(4), 377–388.

MackKay, D.J.C. (2003). *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press.

Murthy, H.A., Yeganarayana, B. (1991). Speech processing using group delay functions. *Signal Processing*, 22, 259–267.

Reynold, D.A., Rose, R.C. (1995). Robust text-independent speaker identification using Gaussian mixture speakers models. *IEEE Transactions on Speech and Audio Processing*, 3(1), 72–83.

Ribaric, S., Fratric, I., Kis, K. (2008). A novel biometric personal verification system based on the combination of palmprints and faces. *Informatica*, 19(1), 81–100.

Šalna, B., Kamarauskas, J. (2008). Voice biometrics – Evaluation of effectiveness of different methods in speaker recognition. In: *Proc. of IEEE Workshop on Bio-inspired Signal and Image Processing*. Warshaw, pp. 1–6.

Strube, H.W. (1980). Linear prediction on a warped frequency scale. *J. Acoust. Soc. Am.*, 68(4), 1071–1076.

Struc, V., Pavesic, N. (2009). Gabor-based kernel partial-least-squares discrimination features for face recognition. *Informatica*, 20(1), 115–138.

Thiruvaran, T., Ambikairajah, E., Epps, J. (2007). Group delay features for speaker recognition. In: *Proc. Information, Communications & Signal Processing*. Singapore, pp. 1546–1550.

Yu, A.-T., Wang, H.-Ch. (2003). Channel effect compensation in LSF domain. *EURASIP Journal on Applied Signal Processing*, 9, 922–929.

A. Bastys finished Lomonosov State University in 1980, PhD doctor of mathematics in 1983. He is assitent professor of Faculty of Mathematics and Informatics of Vilnius University. His research interests include wavelets, harmonic analysis, biometrics.

A. Kisel received his BS and MS degree in computer science from Vilnius University, Lithuania in 2003 and 2005 respectively. He is currently a PhD student in Vilnius University. His research interests include biometric algorithms, image processing and synthetic fingerprint generation. From 2002 he is a research scientist at "Neurotechnology".

B. Šalna in 1977 graduated from Radio-Technical Faculty of Kaunas Polytechnic University and got the qualification of a radio engineer. In 1977 he started working at Vilnius University. In 1990 he defended a thesis for a degree of doctor in engineering science "The creation and research of language signal processing methods based on linear prognosis". Since 1991 he has been the head of Phonoscope Expertise Department of the Lithuanian Forensic Expertise Institute as well as the lecturer of the Criminalities Department of Mykolas Romeris University. His main scientific interests: criminology research of speech, voice and acoustical signals and their record means, speech signals analysis, the methods of the identification of a person according to his voice, biometrics.

Tiesinės prognozės modelio gupinės delsos požymių panaudojimas atpažinti asmens balsą

Algirdas BASTYS, Andrej KISEL, Bernardas ŠALNA

Darbe pasiūlyta panaudoti tiesinės prognozės modelio perdavimo funkcijos spektro fazę kalbos požymiams apibrėžti. Požymiai išreiškiami skaičių poromis, kurios aprašo tiesinės prognozės modelio grupinės delsos ekstremumo taškus. Pirmoji poros komponentė yra grupinės delsos ekstremumo taško abscisė, o antroji yra spektro ekstremumo taške juostinio pločio įvertis. Pasiūlyta balsų panašumo įvertinimo metrika, kuri apibrėžiama panaudojant įvestus grupinės delsos požymius. Metrika adaptuota nepriklausomam nuo teksto balso atpažinimui laikantis nuostatos, kad pateikiamą kalbos signalą gali sudaryti kelių asmenų balsai. Atlikti tyrimai parodė, kad automatinis asmens balso atpažinimas besiremiantis pasiūlytais požymiais ir jų panašumo metrika kokybės prasme lenkia atpažinimą naudojantį Gauso mišinių modelius ir Mel kepstro, formančių, antiformančių ar pagrindinio tono požymius.