TELKOMNIKA Indonesian Journal of Electrical Engineering
Vol.12, No.4, April 2014, pp. 3214 ~ 3223
DOI: http://dx.doi.org/10.11591/telkomnika.v12i4.4934

Perceptual Speech Hashing Authentication Algorithm Based on Linear Prediction Analysis

Yi-bo Huang*[1], Qiu-yu Zhang[2], Zhan-ting Yuan[1]
[1]College of Electrical and Information Engineering, Lanzhou University of Technology, Lanzhou, 730050, PR China;
[2]School of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, China
*Corresponding author, e-mail: huang_yibo@foxmail.com

Abstract

According to the situation that traditional speech authentication algorithms are not appropriated for the present speech communication, we proposed a speech authentication algorithm of perceptual hashing which is based on linear forecast analysis (LPC), it can satisfy the requirement of the efficiency and the robustness for the speech authentication. Firstly, the LPC coefficients are optimized, based on the principle of LPC for speech signal, a new prediction coefficients is constituted through combing the energy and the LPC coefficients of a frame; then dividing the prediction coefficient matrix into smaller blocks, and do singular value decomposition (SVD) to the coefficients of each block. Finally, quantifying the formed singular value and getting perceptual hashing sequences. Experiments show that the proposed algorithm has good robustness for content preserving operations, and it doesn't reduce the efficiency while meeting robustness, it can satisfy the real-time requirement of speech communication.

Keywords: *speech authentication, perceptual hashing, robustness, LPC, energy*

1. Introduction

Speech messages as human communication means is the most natural, effective and convenient way of communication, With the development of mobile communications and personal communications, people around the world can be anywhere at any time on the phone, and people connect with each other more and more closely. On the other hand, existing forms of speech are also extended the acoustic wave to analog and digital signals, it can be unlimited dissemination and preservation [1].

At the same time in the convenience of human life, what followed was vast amounts of speech information processing, information security and social security issues [2]. speech content authentication technology is one of the most effective technical means what which can protect speech data integrity and speech authenticity, It can detect the received speech data in the transition without a third party in the course of malicious editing and tampered. Due to the particularity of speech, traditional speech signature verification algorithm can not meet the certification requirements, first, because of the requirements of robustness, speech in the transmission process is often subjected to various disturbances, and the speech message will not affect the overall operation of the content holding auditory understanding, Therefore, the certification process should be contented to remain operating induction certified range, the speech authentication algorithm robustness made high demands. On the other hand, real-time speech transmission and voice mobile terminal resources make speech authentication algorithm for computing not only efficiency but also has a very high demand. Summary of the traditional authentication algorithm is all the data as a bit stream, so that the original data, minor changes will produce a summary of the changes, poor robustness, not for speech authentication, and of high demand, large amount of calculation, and not the mobile terminal for speech.

In 2001, Ton Kalker was first proposed "perception hash" concept [3-4]. Hash function[5] is based on the perception of human psychology multimedia information processing theory, multimedia data is unidirectional mapped to a data set of multimedia perception summary sets, will have the same perception content of multimedia data sequence is mapped to a unique digital abstract, resulting digital Abstract is only, and it is unidirectional. So far, the

Received September 22, 2013; Revised November 21, 2013; Accepted December 10, 2013

hash function algorithm design can be divided into three major categories: standard, based on packet encryption type and based on modular arithmetic type [4]. Only the perception of simple and fast hash functions to meet multimedia massive data analysis applications. Random hash function Distinguish and compressibility is better than security, so it can rely on the key settings to protect security, a hash function for these properties is very suitable for the information security field, especially in the field of speech authentication. Currently, there are many methods for speech perception hash feature extraction and processing, mostly are around the human auditory system optimization [6], including the front-end of human auditory model to simulate nonlinear filtering sequence extraction method [7]. Including the use of Temporal modulation normalization and Gammatone filter noise reduction modulation of speech information using fractional Fourier transform characteristic value to improve the accuracy of extracting characteristic values, etc[8]. Burges CJ [11] uses MCLT to calculat the logarithmic spectrum, and using the PCA methods analysis the spectral parameters. Chen N [12] first obtain the speech Mel cepstral coefficients, and then use the nonnegative matrix decomposition to analyze spectrum coefficient. M. Pavithra [13] using sparse kernel principal component analysis to maximize the reduction of the model data, the original data can reduce the amount of data, thereby improving operation efficiency. Ekin Olcan Sahin [14] is applied hidden Markov model to robust hashing algorithms. The literature [15] Hilbert transform spectrum estimation method is used to implement robust speech feature extraction, construct hash function perception.

As a basic speech parameters in estimation, and transmit or store with low rate speech aspects of core technology, linear prediction parameters are available, effective and correct performance properties of speech waveform and its frequency spectrum, and high calculation efficiency, flexible and convenient in application [16].

Based on voice in the certification process robustness and real-time problem, put forward a kind of based on linear prediction analysis of speech perception hash authentication algorithm, the algorithm for content maintain operation of robustness, and sensitivity to malicious attacks. Firstly, for each frame by computing the energy and LPC parameter to extracted perception characteristic parameters, and then block the characteristic parameters, singular value decomposition processing, the final processing of the obtained quantization parameters, to obtain a perception hash value sequence.

2 Speech Authentication Algorithm

2.1. Linear Prediction Analysis

Linear prediction thoughts by US scientists N.Wiener was put forward in 1947, which initially applied to the automatic control of artillery By 1967, Japanese scientists Itakura and others for the first application of linear prediction technique in speech, in the various speech processing technology, linear prediction can be applied to a variety of basic speech parameter estimates, such as pitch period, spectral characteristics function, etc.

Linear prediction analysis is one of the most effective methods of speech signal analysis. Linear predictive analysis in speech application of the basic idea is a speech piece (frame) value, use the past several (order of the linear prediction) weighted linear combination of speech segments. In composition with the process of linear prediction, the weighted coefficients can be referred to as the predictor coefficients, based on the linear weighted voice segment and the actual speech fragments consisting of difference approximation to the minimum, to identify a set of weighted coefficient values [17, 18].

Set s(n),n=1,2…, n is the speech signal sampling sequence, s (n) is the speech signal sampling values of n time, that is we want to predict the current sampling points. p is the order of the linear predictor, is based on past weighted sums sample values of p to predict the current sample value and s (n), in which the predictor is called p-th order predictor.

$$\hat{s}(n) = \sum_{i=1}^{p} a_i s(n-i) \quad (1)$$

Where the s (n) of the predicted value, αi , i = 1,2, ..., p is called the linear prediction coefficients, formula (1) is called the p–th order linear predictor.

Speech signal sampling values s (n) in time n and the linear predictive value of n time difference is called linear prediction error, represented by e (n).

$$e(n) = s(n) - \hat{s}(n) = s(n) - \sum_{i=1}^{p} a_i s(n-i) \tag{2}$$

Let:

$$A_p = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}, \quad R_p = \begin{bmatrix} R(0) & R(1) & \cdots & R(p-1) \\ R(2) & R(0) & \cdots & R(p-2) \\ \vdots & \vdots & & \vdots \\ R(p-1) & R(p-2) & \cdots & R(0) \end{bmatrix}, \quad R_p^a = \begin{bmatrix} R(1) \\ R(2) \\ \vdots \\ R(p) \end{bmatrix} \tag{3}$$

Then,

$$R_p^a - R_p A_p = 0 \tag{4}$$

According to the principle of the linear predictive analysis, in order to get the minimum prediction error, it is necessary to make the minimum mean square error, according to equation (2) can get the error of e(n) equation, therefore, it can be obtained:

$$\begin{aligned} E_p &= E[e^2(n)] = E\{e(n)[s(n) - \sum_{i=1}^{p} a_i s(n-i)]\} \\ &= E[e(n)s(n)] - \sum_{i=1}^{p} a_i E[s(n)s(n-i)] \end{aligned} \tag{5}$$

The orthogonal equation is equal to 0, so formula (4) available:

$$E_p = R(0) - \sum_{i=1}^{p} a_i R(i) \tag{6}$$

By the formula (6) and the formula (3) can be obtained:

$$\begin{bmatrix} R(0) & R(1) & \cdots & R(p) \\ R(1) & R(0) & \cdots & R(p-1) \\ R(2) & R(1) & \cdots & R(p-2) \\ \vdots & \vdots & & \vdots \\ R(p) & R(p-1) & \cdots & R(0) \end{bmatrix} \begin{bmatrix} 1 \\ -a_1 \\ -a_2 \\ \vdots \\ -a_p \end{bmatrix} = \begin{bmatrix} E_p \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{7}$$

In the linear prediction coding, in order to improve the robustness of LPC coefficients, we raise many of LPC coefficients equivalent representation method, such as linear prediction cepstral coefficients (LPCC), LSP characteristics (LSP), and ISP partial correlation coefficient, etc.

2.2. Short-time Average Energy

Speech signal energy along with the time change relatively obvious, while general clean part of the energy is much smaller than the voiced sound energy. Short-time average energy of speech signal analysis gives a suitable description method to reflect these changes. Definition of short-time average energy:

$$E_n = \sum_{m=-\infty}^{+\infty} \left[x(m)w(n-m)\right]^2 = \sum_{m=n-(N-1)}^{n} \left[x(m)w(n-m)\right]^2 \tag{8}$$

Speech short energy mainly has several aspects of the application:

(1) It can be used to distinguish between short-time average energy of the voiceless and voiced, the voiced is much more energetic than the energy of the voiceless

(2) We can use the short-time energy of sound and the silence are determined.

(3) In speech recognition, authentication system can be used as a one dimensional characteristic value to describe the characteristics of speech signals.

In practice, short time can not only average energy, it can also be buck energy [19], and energy ratio of wavelet coefficient [20], spectrum energy coefficient [21], as speech perception characteristics of hash values.

2.3. Singular Value Decomposition

Singular value decomposition [22, 23] is the diagonalization of matrix to numerical algorithm. For the singular value decomposition of an arbitrary matrix A, i.e.

$$A = USV^T \tag{9}$$

Where, S is a diagonal matrix, the diagonal element is the singular values of matrix A, it can be called the characteristic values, U and V are M*M and N*N matrix, can be called the base matrix. U and V each column are called left singular vectors and right singular vectors.

$$S = \begin{bmatrix} s_r & 0 \\ 0 & 0 \end{bmatrix}_{m*n} \tag{10}$$

Image data matrix can be used to indicate, for binary image, can be directly to singular value decomposition, the image matrix singular value decomposition is applied in image has three salient features [22]:

(1) Singular value can exhibit the inherent characteristics of the image.

(2) Singular values of robustness is better, after the image is imposing disturbance, singular value will be no big change, to better respond to the original image.

(3) When the image decomposition, the singular value sequence of the first value in the form of much more than other values.

These properties by using the singular value, we try to use the SVD is applied to speech, carried on the forecast to the speech signal, generates a linear prediction matrix, in this way, we have to produce the matrix SVD, and can also get speech singular value.

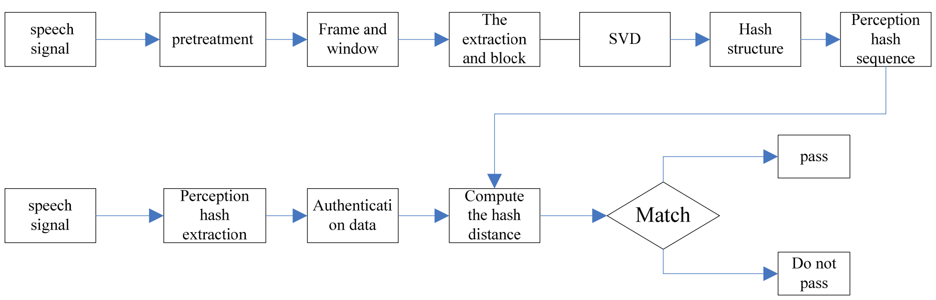

Figure 1. Algorithm Flow Chart

3. The Optimized Perceptual Hashing Algorithm

It can be seen from Equation (19), get the first coefficient of predicted is 1, the coefficient of prediction:

$$A_p = \begin{bmatrix} 1 & 1 & \cdots\cdots & 1 \\ -a_{11} & -a_{12} & \cdots\cdots & -a_{1n} \\ -a_{21} & -a_{22} & \ddots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{p1} & -a_{p2} & \cdots\cdots & -a_{pn} \end{bmatrix}^T \tag{11}$$

Where, n is the total number of frames of speech signal, p is the order of the linear predictor.

The first 1 in the subsequent SVD will have an impact on the matrix decomposition, so that when the decomposition characteristic value of distinction, we can remove the first line, then produce new prediction coefficient matrix.

$$A_p = \begin{bmatrix} -a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & -a_{22} & \cdots & -a_{2n} \\ \vdots & \vdots & & \vdots \\ -a_{p1} & -a_{p2} & \cdots & -a_{pn} \end{bmatrix}^T \tag{12}$$

Just now, we have to voice energy features are described, so before the decomposition coefficient of linear prediction, we can know that each frame energy E (n) is added to the front of linear prediction coefficient of each frame, namely replace 1 position. In this way, using the short-time average energy of each frame and linear prediction coefficient are used to describe the characteristics of each speech frame. Constituted new linear prediction matrix for:

$$A_p = \begin{bmatrix} E(1) & E(2)\cdots & & E(n) \\ -a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & -a_{22} & \cdots & -a_{2n} \\ \vdots & \vdots & & \vdots \\ -a_{p1} & -a_{p2} & \cdots & -a_{pn} \end{bmatrix}^T \tag{13}$$

After receiving a new matrix, we won't directly on SVD of matrix, the first will be partitioned matrix, here, we will small matrix, we will be divided into matrix m p*p small matrices.

$$m = \lceil n / p \rceil \tag{14}$$

$$A_P = \left[A_1, A_{2,} \cdots A_m \right] \tag{15}$$

Thus, the speech signal can be affected by the local interference or against part of the scope, the control in one or several small matrix, not affecting the whole speech segment, improve the robustness of voice authentication.

In the experiments, we use the U, V; constitute a new base matrix, a new matrix w as follows:

$$w = \begin{bmatrix} U & V \\ 0 & 0 \end{bmatrix}_{\max(M,N)*(M+N)} \tag{16}$$

Calculate the new matrix w the sum of elements of each column:

$$s(i) = \sum_{j=1}^{r} H_{ij} \quad 1 \le i \le k \tag{17}$$

To quantify the formation coefficient and the rows of the matrix, and form a speech segment hash value,

$$h(i) = \begin{cases} 1, & s(i) > \hat{s} \quad 1 \le i \le k \\ 0, & 其他 \end{cases} \tag{18}$$

Where, $\hat{s}$ is the median.

4. Experimental Analysis

4.1. The Experimental Environment

In the experiment, we used the TIMIT speech database and studio recorded speech, the speech of the length 4seconds 1189 segment, in which different content of voice contains Chinese and English and the same content of different people read speech. Speech parameters used for the sampling rate 16000Hz, the bit rate is 256kbps, the number of channels is mono, sampling precision is 16bit, format is wav, frame length is 20ms, frame shift is 10ms.

The perception of the binary sequence hash value has much outstanding which are simple structure, the small volume has advantages of easy calculation and analysis. At present, we more often adopt the method of compute hash value perception of distance such as bit error rate, two vector norms, hamming distance method. This experiment adopts the hash range evaluation parameters for Bit error rate (BER), is pointed out that error bits percentage in the total number of bits, the normalized hamming distance, the calculation method is as follows:

$$BER = \frac{\sum_{i=1}^{N} \left(hash_{new} \oplus hash_{origin} \right)}{N} \tag{19}$$

4.2. Distinguish Ability

To perception of 1189 a speech segment hash value of the two, obtained 706266 bit error rate of data, the bit error rate of the normal distribution as seen:

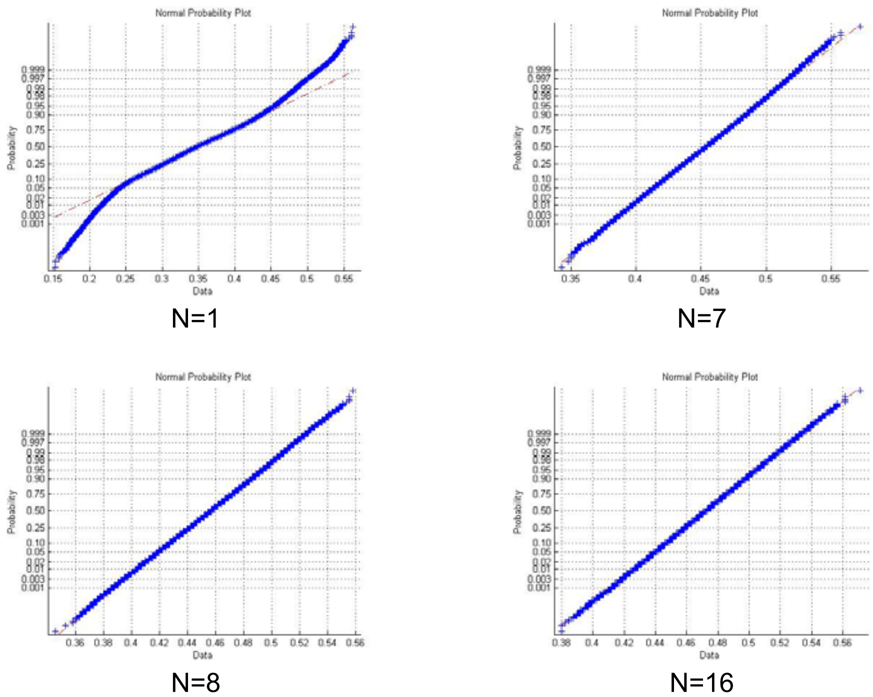

Figure 2. The Optimized Algorithm Normal Distribution Diagram of Different Orders

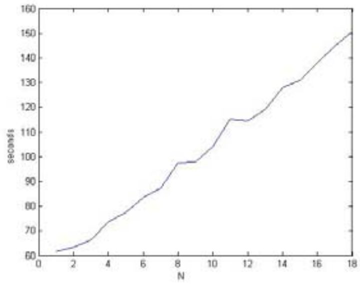

Figure 3. The Optimized Algorithm Order and Computation Time Diagram

Table 1. The Optimization Algorithm Parameter Estimation

N	μ	σ	N	μ	σ
1	0.3483	0.0645	9	0.4632	0.0224
2	0.3636	0.0480	10	0.4616	0.0313
3	0.4002	0.0346	11	0.4633	0.0209
4	0.4204	0.0307	12	0.4702	0.0210
5	0.4360	0.0268	13	0.4698	0.0205
6	0.4471	0.0252	14	0.4723	0.0202
7	0.4533	0.0240	15	0.4764	0.0204
8	0.4564	0.0226	16	0.4721	0.0197

Here N is the order of the LPC predictor, it can be seen from Figure 2, LPC predictor order of the N=8, the distinguish curves have been able to overlap the normal lines, the normal distribution curve in accordance with the order number size along the X-axis translation, can be drawn from Figure 3, the order number N and the operational time are linear relationship, the greater the predictor order, the longer the calculation time. Can be drawn from Table 2, the predictor order N is greater, the anti-collision algorithm is stronger, according to the actual needs of the application, select the predictor order, require authentication speed, accuracy is poor, we can choose N ratio is small value; if it required high precision, authentication speed is less demanding, we can choose the value of N is relatively large. Here we choose N=8 as a follow-up test LPC predictor order.

In the experiments compared with LPC algorithm of optimization is not optimized LPC, non-optimized 8-order normal distribution does not completely coincide with the expected slash. Therefore, the optimized discriminative is superior to LPC discriminative.

When N=8, from Figure 4, the speech content of different perception Hash's bit error rate obeys the normal distribution, it is probability distribution parameters mean μ = 0.4564, standard deviation σ = 0.0226, according to the formula (27) to get the algorithm error rate.

$$FAR(\tau) = \int_{-\infty}^{\tau} f(\alpha \,|\, \mu, \sigma) d\alpha = \int_{-\infty}^{\tau} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(\alpha-\mu)^2}{2\sigma^2}} d\alpha \tag{20}$$

Table 2. The Optimized Algorithm Error Rate

τ	FAR
0.25	2.7786e-020
0.30	2.0186e-012
0.35	1.1843e-006

Compared with the optimized LPC algorithm, the algorithm is not optimal probability distribution parameters as mean u = 0.4921, the standard deviation σ=0.0249.

Table 3. LPC Algorithm Error Rate

τ	FAR
0.25	1.4181e-022
0.30	6.6981e-015
0.35	6.0701e-009

4.3. Robustness

The speech in the library for the following operation:

(1) Decrease volume: the original speech volume decreased by 50%

(2) Increases volume: the original speech volume increases by 50%;

(3) Resampling: speech signal sampling frequency reduced to 8kHz, and up to 16kHz.

(4) Echo: stack attenuation was 60%, the time delay for 300ms, the echo of the initial intensity of 20% and 10% respectively.

(5) Narrowband noise: the speech signal with the center frequency distribution in 0~4KHz narrowband Gaussian noise.

(6) Cut: Cut random speech piece, and then by inserting silence frame approach enables speech piece back to the original length of the shear rate of 10%.

According to the attack Obtained BER, draw the FRR and FAR curve, as shown in Figure 5, extracted from the content of the same speech perception of the hash value and BER are below the threshold value of 0.35, the experimental results show that this algorithm has higher robustness. And FRR-FAR curve in the diagram is not crossed, this algorithm also has good distinguishability and robustness, can accurately identify their content and content of malicious operation. According to Table 3 shows that when the threshold τ=0.35, FAR=1.1843e-006.

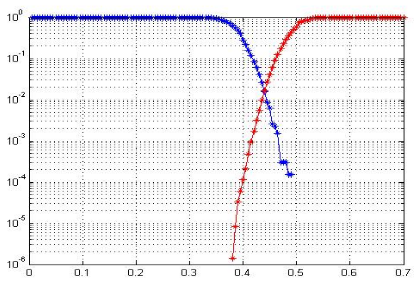

Figure 4. LPC Algorithm FRR-FAR Curve

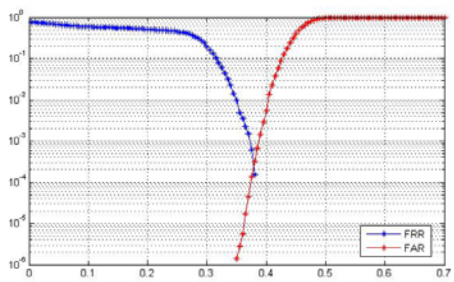

Figure 5. The Optimized Algorithm FRR-FAR Curve

Table 4. The Average Bit Error Rate of the Optimized Algorithm

Operation method	The average bit error rate
Decrease volume	0.1267
Increases volume	0.2294
Echo	0.3138
Resampling	0.0995
Narrowband noise	0.3185
Cut	0.0698

As can be seen from Table 5, the above several attacks average bit error rate are below decision threshold 0.35, increasing and decrease the volume does not change the channel model. Therefore, the optimized LPC coefficients greater change will not happen, thus adjust the volume does not produce the error rate. During the cut, the block operate for speech frame. Therefore, the impact will be limited to the shear in the local scope, cut the error rate is still low. The algorithm for echo, low pass filtering, adding noise robustness is relatively poor, but the average error rate is still controlled by the decision threshold range, showing the contents of this algorithm is to keep operating with better robustness.

4.4. The Efficiency Analysis

This algorithm's features for authentication data volume is small, but high efficiency. Features and characteristics are calculated from the algorithm efficiency considerations. Randomly selected 50 speech from the speech database, statistical algorithm running time. The algorithm of pre-processing and certification time compared with the LPC algorithm, the computation efficiency in improving the robustness of the cases, the speed is not a great loss, authentication efficiency is very high, can satisfy the real-time application requirements.

Table 5. Operation Time

	The optimized algorithm	LPC algorithm
Feature extraction	0.048955s	0.044249s
Hash structure	0.014295s	0.008149s
total	0.06325s	0.05238s
Certification time	0.000038s	0.000029s

5. Conclusion

Speech energy combined with linear prediction coding speech perception hash algorithm is proposed in this paper, as energy can be relatively good performance characteristics of the speech, the using of speech energy replace the first prediction coefficient of linear predictive coding, and then the coefficients constituting the block, so that the scope of the attack can be controlled within a few frames, and then most of the speech frame is not affected, and the block after the matrix SVD, the speech feature is finally quantified for characteristics of speech and the speech perception obtained hash value. Experimental results show that the proposed algorithm can get better distinguish, robustness of the compromise, and the algorithm is the simple, efficient operation. The hash value data rate can be a good and completed voice content integrity certification.

Acknowledgements

Project supported by the National Nature Science Foundation of China (No. 61363078), the Natural Science Foundation of Gansu Province of China (No. 1212RJZA006).

References

[1] Jing Wu. Research Of Audio Retrieval Scheme Based On Perceptual. Harbin: Harbin Institute of Technology. 2008.
[2] Jin Gu. Research on Key Technologies of Speech Perceptual. Hefei: University Of Science and Technology of China. 2009.
[3] Yuhua Jiao. Research on Perceptual Audio Hashing. Harbin：Harbin Institute of Technology. 2010.
[4] NIU Xia-mu, JIAO Yu-hua. An Overview of Perceptual Hashing. *Acta Electronica Sinica*. 2008; 36(7): 1405-1411.
[5] Li Jianwei, Chen Huijie. A Dynamic Hashing Algorithm Suitable for Embedded System. *TELKOMNIKA Indonesia Journal of Electrical Engineering*. 2013; 11(6): 3220-3227.
[6] Gu Jin, Guo Li, Liang Hui, Cheng Long. Effective Robust Speech Authentication Algorithm Based On Perceptual Characteristics. *Journal of Chinese Computer Systems.* 2010; 4(7): 1461-1466.
[7] Ben Salem A, Selouani SA, Hamam H, Caelen J. *A highly robust audio hashing system using auditory-based front-end processing*. Acoustics, Speech and Signal Processing. ICASSP 2009. IEEE International Conference. 2009: 1413-1316.
[8] Xugang Lu, Shigeki Matsuda, Masashi Unoki. Temporal modulation normalization for robust speech feature extraction and recognition. *Multimedia Tools and Applications.* 2011; 1(52): 187-199.

[9] Hui Yin, Volker Hofmann, Climent Nedeu. Acoustic Features for Speech Recognition Based on Gammatione Filterbank and Instantaneous Frequency. *Speech Communication.* 2011; 5(53): 707-715.
[10] Xi Ji, Zhao Li, Liang Rui-yu. A Scheme of Transparent Speech Information Hiding Based on Fractional Fourier Transform. *Internationnal Journal of Advancements in Compution Technology.* 2011; 3(11): 284-290.
[11] Burges CJ, Patt JC, Jana S. Distortion discriminant analysis for audio fingerprinting. *IEEE Trans. Speech Audio Processing*. 2003; 11(3): 165-174.
[12] Chen N, Xiao HD, Wan WG. Audio hash function based on non-negative matrix factorizations of Mel-frequency cepstral coefficients. *Information Security, IET*. 2011; 5(1): 19-25.
[13] M Pavithra, G Chinnasamy, Azha Periasamy, S Muruganand. Feature Matching by Skpca with Unsupervised Algorithm and Maximum Probability in Speech Recognition. *Journal of Management and Science*. 2011; 1(1): 11-15.
[14] Ekin Olcan Sahin. Robust Speech Hashing. Bogazici University. 2009.
[15] Huan Zhao, He Liu, Kai Zhao, Yong Yang. Robust Speech Feature Extraction Using the Hilbert Transform Spectrum Estimation Method. *International Journal of Digital Content Technology and its Applications.* 2011; 5(12): 85-95.
[16] Xiangyang Xue, Wei Li, Yue Yin. *Towards Content-based Audio Fragment Authentication.* MM '11 Proceedings of the 19th ACM international conference on Multimedia. 2011; 28: 1249-1252.
[17] Xiao Qiang, Chen Liang, Zhu Tao, Huang Jin-Jun. Efficient Compressed Sensing Quantization Of LSP Parameters Based On The Approximate KLT Domain. *Journal of Electronics &Information Technology*. 2011; 33(9): 2062-1466.
[18] Gao Yue, Chen Yan-Pu, Min Gang, Du Jia. Compressed Sensing of Speech Signals Based On Linear Prediction Coefficients And Difference Transformation. *Journal of Electronics &Information Technology*. 2012; 34(6): 1408-1413.
[19] Mathieu Ramona, Geoffroy Peeters. *Audio identification based on spectral modeling of bark-bands energy and synchronization through onset detection*. Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference on. 2011: 477-480.
[20] Wang, Jingzhong, Xu, Tianling, Yuan, Kaiguo. Audio Digital Watermarking Algorithm Based on Energy Ratio of Wavelet Coefficient. *Computer Applications and Software*. 2011; 28(5): 288-291.
[21] Jijun Deng, Wanggen Wan, Swaminathan, Ram, Xiaoqing Yu, Xueqian Pan. *An audio fingerprinting system based on spectral energy structure*. Smart and Sustainable City (ICSSC 2011), IET International Conference. 2011: 1-4.
[22] ZHOU Li, LIU Chunhua, JIANG Tianfa. A algorithm of image watermarking based on DWT & SVD. *Engineering Journal of Wuhan University*. 2011; 44(1): 120-123.
[23] Wang Guang Yan, Geng Yan Xiang, Zhao Xiao Qun. A Signal Subspace Speech Enhancement Method for Various Noises. *TELKOMNIKA Indonesia Journal of Electrical Engineering*. 2013; 11(2): 726-735.