ISSN: 0974-6471, Vol. 10, No. (4) 2017, Pg. 736-744

Oriental Journal of Computer Science and Technology

Journal Website: www.computerscijournal.org

Methods and Algorithms of Speech Signals Processing and Compression and Their Implementation in Computer Systems

FADI ALKALANI[1]* and RAED SAHAWNEH[2]

[1]Shaqra University, Saudi Arabia.
[2]Irbed National University, Jordan.

Abstracts

The review and comparative analysis of the methods of compression and recognition of speech signals is carried out. The result of the carried out analysis of the existing recognition methods indicates, that all of them are based on the use of "inflexible" algorithms, which are badly adapted to the characteristic features of speech signals, thus degrading the efficiency of the operation of the whole recognition system. The necessity of the use of algorithms for determination of recognition features along with the use of the wavelet packet analysis as one of the advanced directions of the creation of the effective methods and principles of the development of the speech signals recognition systems is substantiated. Analysis of the compression methods with the use of the orthogonal transformations at the complete exception of minimal decomposition factors is conducted; a maximal possible compression degree is defined. In this compression method the orthogonal transformation of the signal segment with the subsequent exception of the set of the smallest modulo decomposition factors, irrespective of the order of their distribution, is conducted. Therefore the additional transfer of the information on the factors distribution is required. As a result, two information streams appear, the first one corresponds to the information stream on the decomposition factors, and the second stream transfers information on the distribution of these factors. Method of the determination of the speech signals recognition features and the algorithm for nonlinear time normalization is proposed and proved. Wavelet-packet transformation is adaptive, i.e. it allows adapting to the signal features more accurately by means of the choice of the proper tree of the optimal decomposition form, which provides the minimal number of wavelet factors at the prescribed accuracy of signal reconstruction, thus eliminating the information-surplus and unnecessary details of the signals. Estimation of the informativeness of the set of wavelet factors is accomplished by the entropy. In order to obtain the recognition factors, the spectral analysis operation is used. In order to carry out the temporary normalization, the deforming function is found, the use of which minimizes the discrepancy between the standard and new words realization. Dedicated to the determination of admissible compression factors on the basis of the orthogonal transformations use at the incomplete elimination of the set of minimal decomposition factors, to the creation of the block diagram of the method of the recognition features formation, to the practical testing of the software- methods. In order to elevate the compression factor, the adaptive uniform quantization is used, where the adaptation is conducted for all the decomposition factors. The program testing of the recognition methods is carried out by means of determination of the classification error probability using Mahalanobis (Gonzales) distance.

Article History

Received: 04 December 2017
Accepted:11 December 2017

Keywords

Recognition of speech signals,
Compression,
Recognition features.

CONTACT Fadi Alkalani falkilani@su.edu.sa Shaqra University, Saudi Arabia.

To link to this article: http://dx.doi.org/10.13005/ojcst/10.04.06

Introduction

The increased interest to theprogress of the computer engineering, computer and telecommunication networks can be observed lately. This caused the creation of large information resources, which are to be stored, processed and transferred. Therefore the tasks of protecting computer-based information become of great importance. These tasks are generally solved by means of deriving information by performing a cryptographic transformation. Recently the protection of information in computer systems is supported on the basis of knowledge about the biometrical features of human voice, signature, eye's iris and other biometrical parameters.

A lot of methods and systems based on biometrical principles for reliable protection of fast changing information streams are developed lately. They include development of the effective methods of compression and recognition of speech signals.

The most of the known algorithms of determination of biometrical parameters features are based on the methods which are not enough accurate to describe the specific changes of signal (pulses, peaks etc). Therefore the received data contain only overall information about the signal. There is as a rule a great shortcoming of information to conduct the effective recognition. Through their variety the abovementioned algorithms adapt themselves badly to the analysis of speech signals.

There is a great necessity of designing new methods for reliable recognition on the basis of wavelet-packet analysis and speech signals compression using Karhunen-Loev transformation for development of the biometrical systems for real-time protection of information streams[1].

Result and Analysis

The specific compression methods deserve the attention. In them the orthogonal conversion of a signal segment is conducted with the next exception of a collection of the least by absolute value factors of expansion. The additional transfer of information about arrangement of factors is necessary. In the elementary case the transmission of this information is conducted with a bit sequence and for each factors of expansion just one bit needs to be allocated. Therefore the total number of bits of the additional information is equal to the size of the transformation window and factor of compression κ is calculated with equation[2]:

$$K = BN / (BU + N)$$

Where B is a quantity of bits allocated for each factor; N is a quantity of selections in a segment of processing; U is a quantity of the transferred factors of expansion.

As a result two information flows are formed: the first one corresponds to the information stream about factors of expansion and the second one transfers the information about arrangement of these factors.

The results of the mathematical simulation of the speech signal compression and decompressionsystem with the help of the application in the MATLAB environment (figure 1) have permitted to present graphics of dependence of signal/distortion (s/dis) ratio from factor of compression for segments of conversion in 64, 128, 256 samples.

Comparing the presented results it is possible to make a conclusion that at double increase of a size of conversion segment, for example from 64 up to 128 samples, it is possible to achieve only very small increase of quality of compression (in the best case 4 Db), and in some cases even the decrease of the quality.

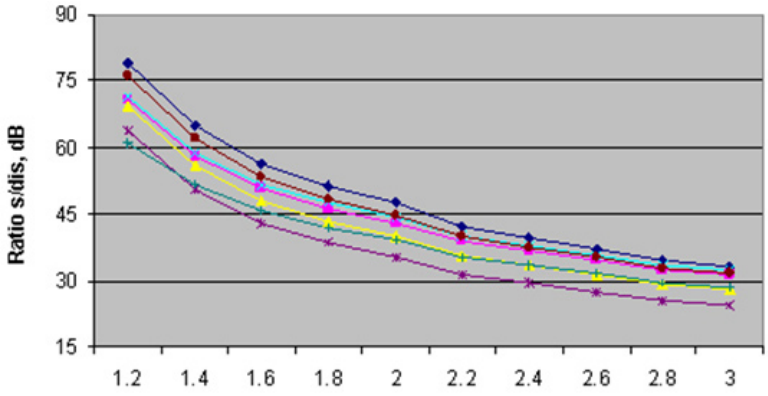
The form of the graphs for all explored conversions is similar. The very positive property is decrease of a droop steepness of s/dis when increasing the compression factor value. Therefore even the small improvement of orthogonal functions base will lead to increase of all dependence and as a result - to appreciable increase of possible compression factor[3].

Let's compare the valid compression factor for orthogonal conversions having fixed the ratio of s/dis at the level of 30 dB[4]. It is often impossible to expose precisely 30 Db. Then with the help of linear interpolation between two adjacent values we conduct the elimination of conversions factors

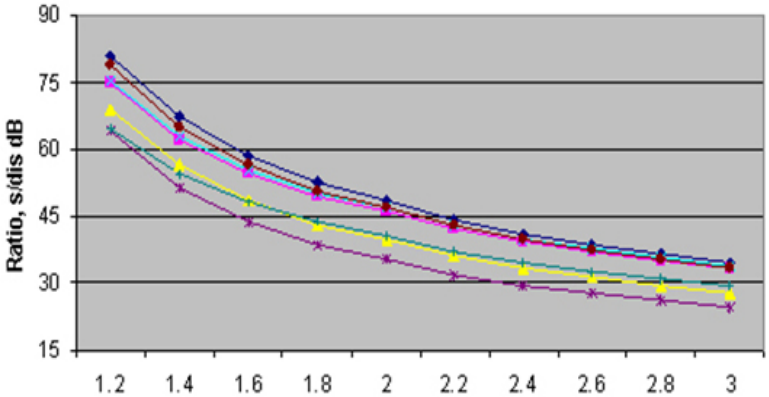
of compression factor determination. The results of calculations for different selections are presented in the table.

At subjective comparison of qualities of speech signal transmission (by ear), the distortions under the usage of inclined and Adamar transformations look like broadband noise. That can be explained by their specific structure (the Adamar transformation-two-level digital functions; the inclined transformation - triangular-like form functions). When using more harmonic transformation – the sine, cosine and Karhunen-Love ones, the signal becomes unnaturally "metal" and under considerable increase of compression factor the unnaturalness increases and reaches the complete loss of clearness of perception of some sound fragments of a phrase. At an increase of the size of the transformation window distortions become more extended and capture a few sounds, but their level does not decrease and the clearness of the speech does not improve.

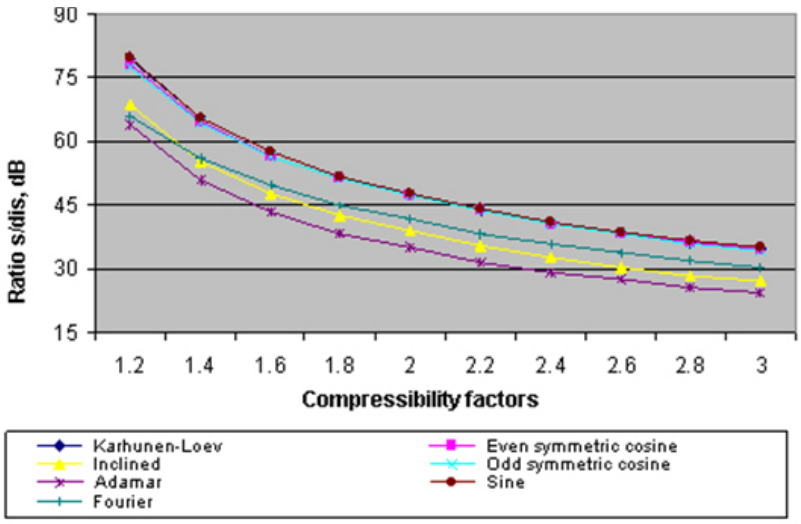
It is studied that the human perception of these distortions depends on the last segment of the system, namely from the signal-sound converter, because the distortions under the high-quality converter along with the good sound the distortions are clearly heard. Simple membrane converters are used mainly in telephony. Therefore there is a poor quality of speech playback and the subjective



a



b



c

Fig. 1: Dependence of the s/dis relation by the compression factor for the transformation segments: a) 64 samples; b) 128 samples; c) 256 samples

Table 1: Possible factors of compression for the explored orthogonal transformations at the window of transformation (64, 128, 256) of sample

	Type of Transformation									
Samples	**Kar.**	**Kar. -Loev g=0.1**	**Kar. -Loev g=0.4**	**Fourier -Loev g=0.9**	**Sine**	**Adamar**	**Odd symmetric cosine**	**Inclined**	**Even symmetric cosine**	**Kar.-Loev for linguistic signal**
64	3,227	3,232	3,252	2,788	3,439	2,353	3,469	2,752	3,217	3,332
128	3,43	3,445	3,432	2,937	3,439	2,377	3,469	2,773	3,429	3,501
256	3,523	3,492	3,485	3,057	3,612	3,106	3,52	2,671	3,571	3,553

level of distortions essentially diminishes on this background.

In the research paper the method of characteristics determination of the speech signals recognition and the algorithm of nonlinear time normalization are offered and substantiated. Basing on the high performance of the wavelet-packet analysis, a conclusion is made about possibility of its effective application to speech signals recognition and compression.

In the wavelet-packet algorithm the rapid wavelet-transformation (RWT) in the form of frequency sequential splitting operation is applied both for low frequency and high-frequency (detailed) factors. In this case the wavelets of each next level are emerged from the wavelet of the previous level by splitting into two new wavelets:

$$\Psi_1 = \sum_n h_n \Psi(t-n)\,; \quad \Psi_2 = \sum_n g_n \Psi(t-n)$$

where h_n, g_n are the appropriate weight factors; t is the time.

New wavelets are also localized but on two-times broader interval. Accordingly the complete set of

wavelet expansion functions is called the wavelet -packet.

Wavelet-packet transformation is adaptive[5,6]. This fact allows fitting more precisely to the features of signals by the choice of an appropriate tree of the optimal expansion form which provides a minimum quantity of wavelet-factors at the given accuracy of the signal reconstruction. Thus the informational redundancy and unnecessary details of signals are switching-off[7].

The estimation of information level of a collection of wavelet-factors is carried out by the entropy under which understands the next value:

$$E = \exp\left(-\sum_n p_n \cdot \log(p_n)\right), \qquad p_n = |x_n|^2 \| x \|^{-2}$$

where x, x_n are signal expansion in the n-node and previous one.

Any effort of factors averaging multiplies the entropy. During the analysis of tree the entropy of knots and its spited parts are calculated. If during splitting the node the entropy does not decrease there is no necessity of the subsequent expansion of this node (such nodes are called the last or terminal ones).

The searching of optimal terminal nodes for creation the recognition class can be divided into the following steps. The first step consists in definition of a list of all terminal nodes that are contained in the optimized wavelet-trees of each image of the properly recognized class. The second step is the comparison of terminal nodes of identical numbers for different images of the given class. Speed of change of amount of transitions through zero is used in a role of comparison criterion. It is necessary to mark a case when numbers of terminal nodes for different images of one class do not fit. In such case a comparison of terminal node with nodes that are not a terminal one and are stored in other images of the same class but the numbers of them are equal to the numbers of terminal nodes.

The third step is definition of node number from the given list (can be terminal one) that presents a given image in the class. The maximum fitting of value of the function of transitions through zero speed change with the same functions obtained for other images is the criterion of choice.

On the next step over the obtained function the packing of a volume range is carried out. In the role of operation the nonlinear one of taking the logarithm can be used.

To receive the recognition factors the operation of spectral analysis is used.

$$C_s(m) = \frac{1}{N}\sum_{n=0}^{N-1} In|S(n)|^2 e^{i\frac{2\pi}{N}nm}, \quad m = 0,1,...,2N-1$$

Where m is the amount of expansion factors, S(n) is the amplitude spectrum of the signal, N is a selection.

To conduct the temporal normalization a strain function is searched. The implementation of this function minimizes disagreement between the standard and new realizations of words. The two functions are searched more precisely:

$\omega_x: \{1,...,l\} \rightarrow \{1,...,m\}$, $\omega_y: 1,\ldots,l\} \rightarrow \{1,\ldots,n\}$
$(\max\{m,n\} \leq l < m + n)$

such that

$\omega_x(1) = 1$, $\omega_y(1) = 1$, $\omega_x(l) = m$, $\omega_y(l) = n$,

$\omega_x(i+1) = \omega_x(i)$, i = 1,..., m-1,,

$\omega_y(j+1) = \omega_y(j)$, j = 1,..., n-1,,

and, in addition $\sum_{k=1}^{l} \rho_{\omega x(k)\omega x(k)}$ is minimal. Here $\rho_{i,j} = (S_x(i) - S_y(j))^2$, where $S_x(i) - S_y(j)$ - value of segmenting function from the proper contours.

Segmenting function should characterize an aggregate modification of parameters of a speech signal that are used by it and it depends on two frames: current one and previous one. The energy distribution of a signal on frequency groups is used in a role of parameters of a speech signal.

The procedure of determination of distortion functions ω_x and ω_y is implemented by the method of

dynamic programming and enables carrying out the interior nonlinear time smoothing of implementations of words.

Knowing the distortion functions ω_x and ω_y we can for any area of standard realization of word to find the proper area of a new realization. We will apply it for separation of the new implementation of a word on sound dyads. Sound dyad is a transient from phoneme to phoneme, which maps reorganization of a means of an articulation. Unlike realization of phoneme, realization of a sound dyad is much less inclined to influence of a context and map correlation of adjacent phonemes of a speech stream. The centers of pseudo-steady areas of phonemes are boundaries of dyad. Thus, dyad consists of the second half of the first phoneme and the first half of the second phoneme.

The standard realization of the word is divided by sound dyads manually: numbers of $a_0,...,a_L$ frames are stressed. These frames are the centers of pseudo-steady areas of phonemes. Then the points n_l, $l = 0,...,L$ such, that $\omega_x(n_l) = a_l$ are selected. Now, with the help of the function it is possible to spot numbers of $b_0,...,b_L$ frames, that are the centers of pseudo-steady areas of phonemes in the new realization of word: $b_l = \omega_y(n_l)$, $l = 0,...,L$.

The resulted procedure allows passing from the comparison of realization of words to comparison of realization of sound dyads.

For providing of identical quality of transmission, both loud and weak signals, the adaptive quantization of decomposed factors are used, which takes place due to the change of an amplification factor of link that precedes to the compressor, therefore together with decomposed factors there is the necessity to pass an amplification factor k(n) . The amount of bits on the decomposed factor is defined under the equation:

$$q(n) = ceil\left(\log_2\left(2^{q_{MAX}} \cdot W(n)\right)\right)+1, \qquad n = 1,...,N,$$

where W(n) is average level of decomposed factors,
q_{max} isthe maximal quantity of bits on a factor,
ceil()the operation of rounding to greater whole (at the negative value gives a zero),
+1 - consideration of sign bit.

The greatest possible value of the unit of decomposed factor is defined under the formula:

$$K_{\max}(n) = 2^{q(n)-1} - 1, \qquad n = 1,...,N.$$

The amount of bits necessary for the transmission of complete aggregate of decomposed factors cannot decrease more than twice. In order to promote the factor of compression is used adaptive uniform quantization, in which the adapting will be carried out for all decomposed factors (segment of factors). We can really obtain the good results of compression using less than eight bits on factor.

The factor of compression at usage of this method is defined under the formula:

$$K = \frac{B \cdot N}{\sum_{n=1}^{N} q(n) + K_B - K_{nul}}$$

where K_B - quantity of bits selected for the transmission of amplification factor,
B - quantity of bits on a factor on the entrance of compressor,
K_{nul} - quantity of zero factors for which sign bits are not needed.

For comparison of orthogonal conversions it is possible to use the complex quality index, but evidently it does not give the picture of such parameters as a factor of compression or correlation of s/dis, which is more better for comparison of compression with the usage of one of the orthogonal conversion.

On the figure 2 and figure 3 the value of factors of compression is compared for different transformations for the fixed relation of s/dis at the level of 30 dB.
The program testing methods of recognition is carried out by definition of probability of an classification error with the help of Mahalanobisdistance (Gonzales). The connection between Mahalanobis distance and identification error is presented with the next equation (Gonzales)[8]:

$$p(e)=\frac{1}{2}\Phi\left(-\frac{1}{2}\sqrt{r_{ij}}\right)+\frac{1}{2}\left[1-\Phi\left(\frac{1}{2}\sqrt{r_{ij}}\right)\right]=\frac{1}{2}\int_{\frac{1}{2}\sqrt{r_{ij}}}^{\infty}\frac{1}{\sqrt{2\pi}}\exp(-\frac{y^2}{2})dy+$$

$$+\frac{1}{2}\left(1-\int_{-\infty}^{\frac{1}{2}\sqrt{r_{ij}}}\frac{1}{\sqrt{2\pi}}\exp(-\frac{y^2}{2})dy\right)=\int_{\frac{1}{2}\sqrt{r_{ij}}}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{y^2}{2}\right)dy,$$

where p(e) is the probability of error r_{ij} is the Mahalanobis distance.

In the process of testing the 6 classes of images were used. One of them was the correct on. So, Mahalanobisdistance was calculated between the correct class and the five wrong classes. Estimation of error for the worst class makes 30 %, and for the best one approximately 5 %. Estimation of the average error of detection for all classes makes 20 %.

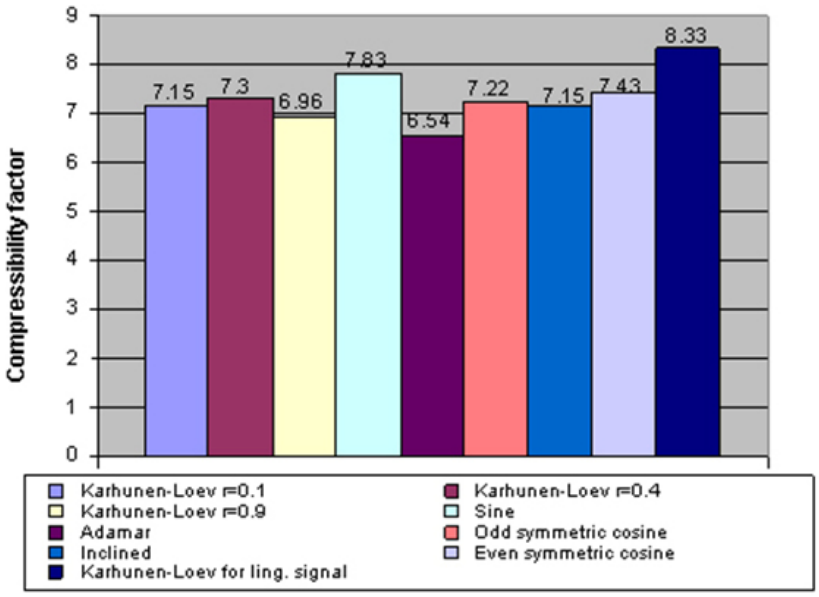


Fig. 2: Possible factors of compression for explored orthogonal transformations at the window of transformation (64 samples)

The selection in our case is smaller in the statistical value. In the general case the range of small selection as a result of many researches of number sequences makes from 10-15 to 200. There is no sense to hope for a regular of statistical performances of average value and variance of probability of a correct (wrong) recognition at such circumstances[9]. If to use deciding rule, which is founded on a principle of the nearest neighbor, in this case reliability of a correct (wrong) recognition is equal 1 (0). It is precisely known that the probability of a correct recognition is value less 1 for selections of arbitrary size. Nevertheless classical statistical approaches have no sufficient sensitivity that in condition of small selections reliably to avoid a single (zero) event in relation to probability of correct (wrong) recognition. The arbitrary classical statistical estimation in conditions of small selections is characterized by the large values of dispersions, which extremely strongly aggravate reliability of the most statistical estimation.

The methods of the differential estimation of reliability of correct (wrong) recognition are applied to avoid the problems of estimation with the help of the statistical approaches. Also it is proved that the differential reliability averaged behind the database of correct (wrong) recognition equals probabilities of correct (wrong) recognition of algorithm as a whole. Thus dispersion of the constructed estimation of reliability of a correct (wrong) recognition in some times can be smaller than in case of the classical statistical estimations. On completion it is necessary to mark that the estimation of the average reliability of wrong recognition at the level of 20 % responds to such class of biometric systems.

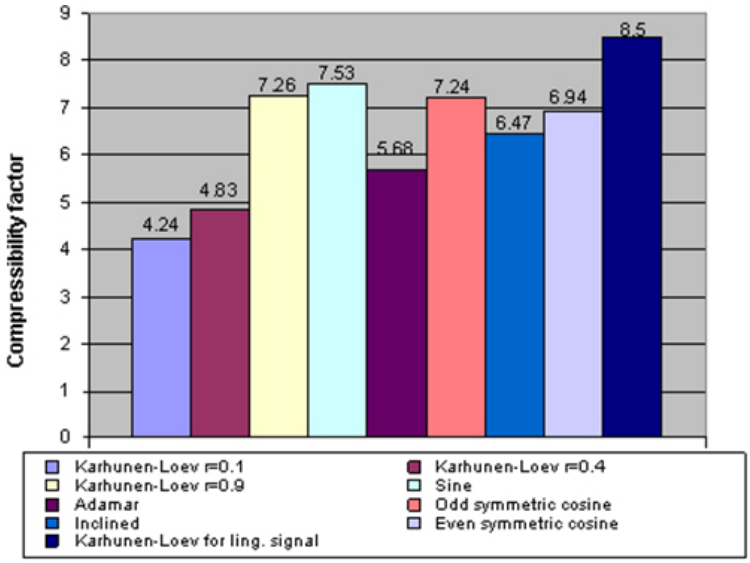


Fig. 3: Possible factors of compression for explored orthogonal transformations at the window of transformation (128 samples)

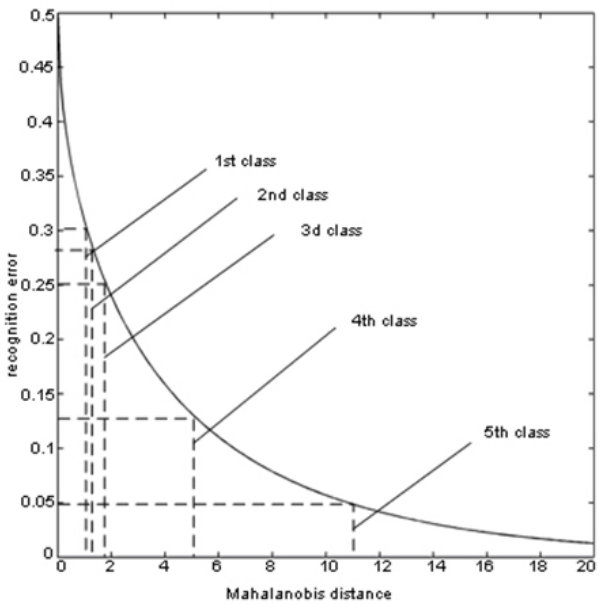


Fig. 4: Dependence of error probability on Mahalanobis distance

Conclusion

The new method of signals recognition features determination is developed with the use of wavelet-packet analysis is designed and the analysis of efficiency of methods of speech signals compression is conducted. The scientific and practical results are as following:

1. The comparative analysis and classification of the known algorithms of speech signals recognition and compression is conducted. It is established that the existent methods of recognition signs determination are based on the use of "rigid" algorithms which badly adapt to the characteristic features of speech signals. This feature badly decreases the

efficiency of all performance of the recognition system.

2. The algorithm of nonlinear time normalization that in contrast to the known algorithms takes into account the phenomenon of co-articulation is designed. The method of ranking is offered.
3. The new and structurally adaptive algorithm for determination of signs of speech signals recognition based on the wavelet-packet analysis is proposed. The main advantage of the proposed algorithm in comparison to the known methods is the possibility of adaptation to the characteristic features of speech signals. This leads to the improvement of information level of the received signs and increase of recognition efficiency of the whole computer system.
4. Recommendations for practical application of the developed systems of forming of information features are created. The method of the algorithms comparison provides the possibility to build schemata of the speech signals recognition and compression systems and improve efficiency of their performance.
5. Dependence of recognition error is established on the base of the Mahalanobis distance. That indicates the indexes of probability of correct recognition of the offered biometrical system are much better than existing ones and is equal R=0.95.
6. The algorithms of speech signals compression are programmed with the use of wavelet-packet algorithms. Descriptions of signals are defined and dependence of compression factor is studied on q_{max} for different types of orthogonal transformations. Necessity of the use of Karhunen-Loev transformation is shown.

References

1 Kinney, A., and J. Stevens. "Wavelet packet cepstral analysis for speaker recognition." Conference Record of the Thirty-Sixth Asilomar Conference on Signals, Systems and Computers, 2002.

2 Klein, ShmuelTomi. "Data Compression in Information Retrieval Systems." *Database and Data Communication Network Systems*, 2002, pp. 573–633.

3 Fadi Alkalani,Compression of speech signal based on orthogonal transformations, *selection and processing of information,* **20**(96),P. 137-142, 2004.

4 Car, J. "Improving quality and safety of telephone based delivery of care: teaching telephone consultation skills." *Quality and Safety in Health Care*, vol. **13**, no. 1, Jan. 2004, pp. 2–3.

5 Burrus, C.S., Gopinath, R.A. and Guo, H. (1998) Introduction to Wavelet and Wavelet Transforms. Prentice Hall, New Jersey. - References - Scientific Research Publishing.

6 Rao, Raghuveer M.; BorosTibor. "Wavelet Transforms: Introduction to Theory and Applications, J*ournal of Electronic Imaging." DeepDyve, SPIE*, 1 Oct. 1999.

7 https://www.krsu.edu.kg/vestnik/index.html

8 Sydow, A. "Tou, J. T./Gonzalez, R. C., Pattern Recognition Principles, Publishing Company. 1974. *Journal of Applied Mathematics and Mechanics*, 22 Nov. 200.

9 Kapustii, B. E., et al. "Features in the design of optimal recognition systems." *Automatic Control and Computer Sciences*, vol. **42**, no. 2, 2008, pp. 64–70.