Method of Speech Signal Compression in Speaker Identification Systems

Raimy Abdourahmane, Konate Karim, Bykov Nikolaï Maximovitch

Abstract— In this paper we present a technique of efficacy improvement of speech signal compression algorithm without individual features speech production loss. The compression in this case means to delete, from the digital signal, those quantization steps that can be predicted. We propose to decrease the number of those quantization steps using a modified linear predication algorithm with variable order. That allows to decrease compression time and save computer resource.

Index Terms— speech signal compression, quantization steps, linear predication algorithm, computer resource.

1 Introduction

THE task of efficient representation of speech signal is one of the vital tasks in speaker identification problems. For example, an automatic speaker recognition system is installed on a LAN or WAN server, which authorizes a terminal to access the network according to the voice of the subscriber. There are two ways of processing information in this case:

1) get the identity features of the speaker from the speech signal on the subscriber's terminal and transfer them to the server for a decision regarding the possibility of admission;

2) compress the speech signal, without loosing the information about the speaker's identity, in the form of a password wav-file, and transfer it across the network to the server, where the identification procedure is carried out.

One of the advantages of the first approach is the reduction of the transmission time over the network. Its main drawbacks are that it reduces the confidentiality the speaker identification procedures, and there is a need to install on the terminals a system for a primary analysis and description of the speaker signals features. Thus, the second approach is more effective for information processing regarding the number of computations that are required for the compression, and the use of ASP-technologies for the selection of informative features and for decision-making.

Analysis of known works

According to the well known methods of signal compression and given the statistical characteristics of the speech signal, the parameters of the analog-to-digital converters (ADC) are chosen according to the rules presented in [1, 2]: the discretization frequency is determined by the upper limit frequency of the signal, the quantification range – by the dispersion, the quantification step - by the signal to noise ratio and the required precision. Since the speech signal is not stationary, the parameters of the ADC are chosen approximately using the most catastrophic situation, which is rarely encountered. As a result, the inherent redundancy of the speech signal is completed by the redundancy of the discrete transformation. As a result a new problem arises: eliminating the ADC's redundancy. In the numerous variants of pulse modulation and adaptive coding, which are used today to eliminate encoding redundancy, the sample rate remains constant and equals the Nyquist frequency, and redundancy is eliminated by analyzing the values of neighboring signal samples.

The aim of the research

The aim of the research is to increase the efficiency of the algorithm of speech signal compression without loosing the information related to the personnal peculiarities of the speaker, by removing those samples that can be predicted.

2 Theoretical Foundations of the Proposed Method

In this work we propose to reduce the number of signal samples by using the modified method of variable order linear prediction. The peculiarity of the proposed method consists in a two step processing of the speech signal, which allows reducing the time that is necessary for wav-file compression. The process is carried out in two steps:

1. Preliminary compression;
2. Final compression.

At the first stage the wav-file is processed using an original technique, which consists in approximating the speech signal using a polyline, with the possibility to establish the degree of its deviation from the original signal. At the second stage the wav-file areas which were not affected during the initial compression procedure are approximated using a polynomial, whose order is determined according to the accuracy that is required to restore the speech signal from the archive file.

Since the speech signal is a continuous function $S(t)$, whose spectrum is limited by the upper frequency F, it is defined by the succession of his samples, whose time interval is calculated using the following formula:

$$T_k = \Delta t = \frac{1}{2F}.$$

Thus the signal $S(t)$ can be described as follows:

$$S(t) = \sum_{i=-\infty}^{\infty} S(t_i) \cdot \varphi_i(t) = \sum_{i=-\infty}^{\infty} S(i \cdot \Delta t) \cdot \varphi_i(t),$$

where $\varphi_i(t) = \frac{\sin 2\pi F(t - i \cdot \Delta t)}{2\pi F(t - i \cdot \Delta t)}$ is the sample function and i assumes discrete value

$$\varphi_i(t) = \begin{cases} 1; t = i \cdot \Delta t, \\ 0; t = k \cdot \Delta t, k \neq i. \end{cases}$$

For a limited duration τ of the speech signal the number of the signal samples N is defined by the expression:

$$N = \frac{\tau}{\Delta t} = 2\tau F.$$

Taking to account the quazzi stationarity of the signal and also the non critical state of the data collection systems to real time of processing, a method of reduction of the encoding redundancy of the speech signal using the ADC has been developed. Minimization of the error of restored signal consist in the finding those fixed values of the argument $t_0, t_1, t_2, \ldots, t_n$ that ensure convergence of broken plot from the vertices $S_0, S_1, S_2, \ldots, S_n$ towards the function $S(t)$ so that for the entire range of argument changing the absolute error does not exceed permissible values.

The function $S(t)$ in these points can be presented as follows:

$$S(t_1) = S_0 + k_1(t - t_0) \text{ for } t_0 \le t \le t_1,$$

$$S(t_2) = S_0 + k_1(t - t_0) + k_2(t - t_1) \text{ for } t_1 \le t \le t_2,$$

$$S(t_3) = S_0 + k_1(t - t_0) + k_2(t - t_1) + k_3(t - t_2),$$

for $t_2 \le t \le t_3$,

where k_i can be defined as follows :

$$k_1 = tg\gamma_1 = \frac{S_1 - S_0}{t_1 - t_0},$$

$$k_2 = \frac{S_2 - S_1}{t_2 - t_1} \pm \frac{S_1 - S_0}{t_1 - t_0},$$

$$k_3 = \frac{S_3 - S_2}{t_3 - t_2} \pm \frac{S_2 - S_1}{t_2 - t_1} \pm \frac{S_1 - S_0}{t_1 - t_0},$$

In general:

$$S(t) = S_0 + \sum_{j=0}^{n} k_{j+1}(t - t_j) \cdot \psi\left[sign(t - t_j)\right],$$

where $\psi = \begin{cases} 1; (t - t_j) > 0, \\ 0; (t - t_j) < 0. \end{cases}$

Approximation error is determined by the remainder term of interpolation formula. In this case, the segment of line in the within the time interval $[t_j, t_{j+1}]$ is defined by the expression:

$$S(t) = S(t_j) + \frac{(t - t_j) \cdot \left[S(t_{j+1}) - S(t_j)\right]}{t_{j+1} - t_j},$$

and the remaining member of functions expansion at the same interval will be:

$$R(t) = \frac{S''(t)}{2!}(t - t_j)(t - t_{j+1}),$$

where $S''(t)$ - the second derivative of a given function within the interval.

If it is known that $R(t)$ and $S''(t)$ are maximal, then

$$\left|R(t)\right|_{\max} = \frac{\left|S''(t)\right|_{\max}}{2!}\left[\frac{t_{j+1} - t_j}{2}\right]^2.$$

Letting $\Delta S_{\max} = \left|R(t)\right|$, we get the formula for the sampling interval

$$\Delta t^* = t_{j+1} - t_j = \sqrt{\frac{8 \cdot \Delta S_{\max}}{\left|S''(t)\right|_{\max}}}.$$

Asking the upper frequency of signal bandwidth is defined we can determine the deviation of real signal value from predicted. Based on the above, an algorithm to implement the procedure for pre-compression of voice information was created. It includes following steps:

1. Set level of allowable absolute error of the recovery signal Δ;
2. Set the minimum size M of buffer compression;
3. For the current point the coefficient of prediction is determined;
4. If a deviation of the coefficient $\Delta k \le \Delta$, we incorporate current sample in compression buffer, increasing the value m of buffer counter by 1 and go to Item 3, if the inequality is not fulfilled, then check the buffer counter m: if $m < M$ then set $m = 0$ and go to to Item 3; if $m \ge M$ then compression is full field;
5. If end of wav-file not found, then go to Item 3.

Linear prediction used for the realization of the process of the second step of compression [3,4].The signal $S(t)$ is presented in a digital form S_n, $n = 1,2,\ldots,N$, where N is number of signal samples, which is obtained by sampling it at a certain frequency F. This signal S_n, $n = 1,2,\ldots,N$,can be presented as a linear combination of preceding values of the signal and some influence u_n

$$S_n = -\sum_{k=1}^{p} a_k \cdot S_{n-k} + G \cdot u_n$$

ISSN 2229-5518

where G is the amplification coefficient and p is the order of prediction.

Then, knowing the values of signal S_n, the problem reduces to searching the coefficients a_k and G. Concerning the estimate, we will use the least square method assuming the signal S_n as deterministic.

The values of signal S_n will be expressed through his estimating values $\tilde{S}_n$ by the following formula :

$$\tilde{S}_n = -\sum_{k=1}^{p} a_k \cdot S_{n-k} \ .$$

Then the predicion error can be described as follows:

$$e_n = S_n - \tilde{S}_n = S_n - \sum_{k=1}^{p} a_k \cdot S_{n-k}.$$

Using the least square method, the parameters a_k are selected so as to minimize the average or the sum of squares of the prediction error. In order to find the coefficients a_k, let us use the matrix method [5,7] called as Darbin method.

Calculation of the coefficients of linear prediction and the prediction error is performed by the following algorithm: of coefficients of linear prediction and prediction error is:

1. The segmentation of the speech signal at stationary intervals;
2. For separated intervals, a system of linear equations is formed that is solved by matrix method or by Darbin method using the auto-correlation function (method is selected by user);
3. The prediction error is calculated.

3 THE RESULTS OF EXPERIMENTS AND THEIR ANALYSIS

Elaborated algorithms have been realized in MATLAB environment. The software allowed to realize the process of the speech signal compression with the help of proposed method, was elaborated. As an example, the compression of the wav-files (PCM format, F= 8 kHz) which contains the entry word "Priklad" of size 6.54 kbit, was conducted.

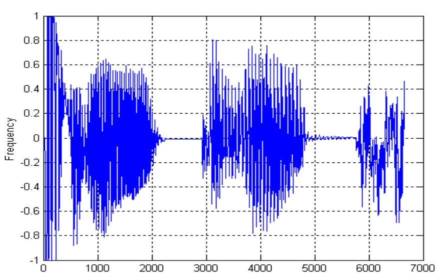

Fig 1 : Wav-file for the Ukrainian word « PRIKLAD »

Finally the working of the program gives us the spectrogram of Fig 2 and the image of intensity of the previous compression in the step (Fig3).

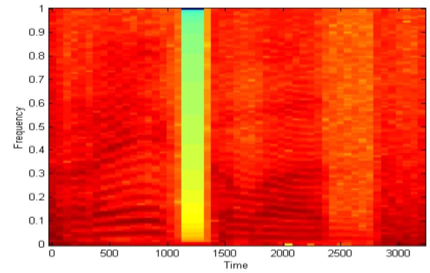

Fig 2 : The spectrogram of wave-file

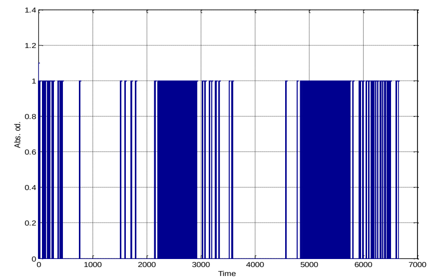

Fig 3: Image of intensity of the previous compression

As shown in fig.3 , at the step 1 of previous compression near 53% of wav-file was treated.

Areas of the original wav-file, which is not subjected to processing at the stage of the previous compression (zero level at the Figure 3) are archived at the final stage of compression, which consist from the splitting into fixed intervals (256 samples duration 20 ms) and calculation for each of them the linear prediction coefficients (using linear prediction order 8).

Size of the obtained archive file in format .mat has dimensions of 3.51 kbit, the compression ratio is 54%.

To illustrate the efficiency of the developed method the study was conducted, in which the comparison of the of archive file sizes from various software products was conducted. Source files of speech signal were stored in the wav-format. Compression of the source file was fulfilled by WinZip 8.0 archiver, by convergence to formats mp3, wma, and by the developed software. The results of the experiments are shown in Table 1.

Listed in Table 1 results shows that the developed software has fulfilled the compression more effectively than WinZip 8.0 archiver and wma codec.

The experiment was implemented with the purpose of displaying the advantages and the drawbacks of using of the previous compression procedure.

As the evaluation criteria were the size of the original mat-file and the time taken to process linguistic material. The experimental results shown in Table 2.

ISSN 2229-5518

Table 1 - Comparison of the efficiency of signal compression methods

Linguistic material	Size of files, compressed by different methods				
	File size in format .wav, Kbit	File size in format .zip, Kbit	File size in format .mp3, Kbit	File size in format .wma, Kbit	File size in format .mat, kb
Numbers 1-10	78,20	34,10	27,40	28,10	30,54
report	5,72	3,78	3,12	3,31	3,19
experiment	12,00	7,42	5,28	5,40	5,52
priklad	6,54	4,77	3,51	3,44	3,51
VNTU	33,90	21,00	17,20	17,30	17,39

Table 2 - Studying of the previous compression procedure efficiency

Linguistic material	Length of wav-file, samples	The size of the .mat file, with pre-compression, Kbit	Size of .mat file, without using pre-compression Kbit	The difference in size,%	Time for which the file was archived, if the use of pre-compression procedures, sec	Time for which the file was archived, without using of pre-compression procedures, sec	Time difference, %
Numbers 1-10	80128	30,54	28,53	7	2,7853	4,2337	34
report	5816	3,19	2,97	7	0,5542	0,8479	35
experiment	12320	5,52	5,14	7	0,8013	1,0737	25
priklad	6656	3,51	3,07	13	0,7109	0,9313	24
VNTU	34752	17,39	17,22	1	1.8130	1,5497	24

The table 2 shows that using of previous compression procedure allows to increase the speed of of speech signal processing by 25-35% with an increase in the output file size by about 10%. Thus, it is advisable to put in software the choice of compression mode.

Tables 1 and 2 show that the efficiency of the compression process depends on the speech material, which allows to define the sensitivity of sounds until the process of compression. The results are illustrated in fig.4

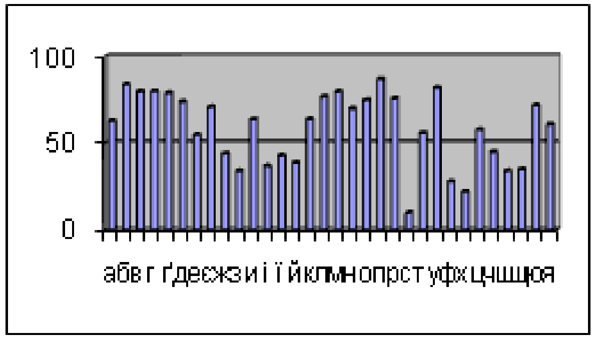

Fig.4: Diagramm of compression of the sounds of Ukrainian alphabet

As we can see on the chart of fig.4 , the most sensitive sounds are observed by means of high louding sounds (the compression coefficient varies from 50 to 80%) whereas the least sensitive are the consonants whose compression coefficient varies from 10 to 40

4. Conclusion

The theoretical and experimental results obtained in the research allowed developing a method of the speech signal compression using of linear prediction. An algorithm and software for this method realization using MATLAB has been elaborated.
In the study we have obtained:

1. Size of compressed by the developed method files corresponds to size of files in formats mp3, wma, obtained by converting the original wav-file, and exceed their quality;
2. Using of previous compression procedure allows to increase the speed of speech signal processing by 25-35% with an increase in the output file size by about 10%.
3. The study of sounds that correspond to the Ukrainian alphabet showed that the most sensitive to compression of the previous procedures were the sounds that correspond to vowels and voiced consonants (50-80% compression ratio), the least sensitive - consonants (10-40% compression ratio).

ISSN 2229-5518

REFERENCES

[1] Rabiner L.R. : Digital Speech Signal Signal Processing, Moscou, 1981.

[2] Makhoil J: Linear prediction //TIIER T.63, 1975

[3] Korotaev G.A.: Analysis and speech synthesis signal using linear prediction method // Zarubejnaya electronika , 1980, n°3, pp:23 – 34.

[4] Barochin S.N. : spectrum analysis and problema of reduction description speech signals //spectrum analysis of sounds. Moscow, 1969 – pp:13 – 30.

[5] Markel J.D. , Gray A.X. Linear predication of speech signal: Moscow , radio and Sviaz , 1980 , 248p.

[6] Kyzmin I.V., Kedrus V.A. Basis of the information theory and coding. Kiev, visha shcola,1986 -238p.

Raimy Abdourahmane, Konate Karim, *Laboratory of Telecommunication and Network, Department of Mathematics and Computer Sciences, Faculty of Sciences and Techniques, University Cheikh Anta Diop of Dakar, Dakar, Senegal,* araimy16@yahoo.fr, kkonate911@yahoo.fr

Bykov Nikolai Maximovitch, PhD, Professor, *Department of Computer Control Systems, Vinnitsa National Technical University, Vinnitsa, Ukraine,* nkbykov@mail.ru