O. M. Tkachenko, Cand. Sc. (Eng.), Assist. Prof.

METHOD OF VECTOR DELTA-QUANTIZATION OF SPEECH SIGNAL PARAMETERS

The paper suggests the method of vector quantization of LSF-parameters of speech signal with the prediction of the next value. Main idea of the method is that instead of quantization of LSF real vector the difference between real and predicted values is coded. This approach enables to reduce dynamic range of input values and, correspondingly, quantization error. Iterative procedure of codebooks design for the realization of the suggested method is elaborated. Experimental results of method testing for various rates of speech signal coding are suggested.

Key words: *speech signal, parametrization, code books, separate and multistage vector quantization, vector of LSF-parameters, clustering error, correlation coefficient, spectral distortion.*

Introducton

Moor's low of 2 fold increase of processors' performance every two years perfectly corresponds to the practice during previous 40 years [1]. Volumes of multimedia information, transferred and stored in computer systems increased approximately at the same rate. However, availability of fundamental restrictions, resulted from monatomic nature of the substance and velocity of light-propagation, made H. Moor in 2007 admit that the law soon would not act. At the same time similar restrictions concerning the increase of information volume are not revealed nowadays.

Thus, there exists the problem of underproductivity of computer systems, designed for processing of multimedia, particularly, speech information. Decrease of volumes of data, needed for correct reproduction of digital speech signal, allows to increase considerably the performance of such systems .

Nowadays the highest degree of compression is provided by parametric methods with further quantization of the obtained parameters. At the stage of parametrization the signal is split into frames of the same duration, as a rule 10 – 30 ms, and a certain vector of parameters is calculated for each of these frames. In modern systems of speech signal compression computation of parameters, based on the model of linear prediction of the tenth order became standard *de-factor*. At the stage of quantization the obtained vector of parameters is replaced by so-called quantized vector – the nearest neighbour to the input vector from the codebook. In [2] it was shown:

1. It is expedient to use as quantization parameters, line spectral frequences (LSF) computed by the values of linear prediction coefficients (LPC).
2. Optimal quantization of complete LSF-vector is computationally complex that is why it is split into 2 or 3 subvectors, each of which is quantized separately. The alternative variant is multistage quantization of the complete vector by means of codebooks (CB) of smaller volume, at each next stage, the error, left after quantization at the previous stage is being quantized.
3. Quantization error for i^{th} frame is evaluated by spectral distortion SD_i :

$$SD_i^2 = \frac{1}{F_s}\int_0^{F_s}\left[10\log_{10}(P_i(f)) - 10\log_{10}\left(\hat{P}_i(f)\right)\right]^2 df,$$

where $P_i(f) = 1/\left|A_i(\exp(j2\pi f/F_s))\right|^2$, $\hat{P}_i(f) = 1/\left|\hat{A}_i(\exp(j2\pi f/F_s))\right|^2$ –are power spectral densities of original and quantized vectors for i -th frame, $A_i(z), \hat{A}_i(z)$ – are original and quantized LPC-polynomials, that correspond to i -th frame, F_s – sampling frequency.

4. To meet the conditions of transparence average value of quantization error SD_{av} for all the frames must be approximately 1 dB; the number of frames, for which $SD > 2dB$, must not exceed 2%, there must not be frames, for which $SD > 4dB$.

5. The given conditions can be achieved while splitting of LSF-parameter vector, into two subvectors and using for quantization the input vector of 24 bits per frame.

A number of publications was devoted to the problem of further reduction of data volume for the description of spectral information [3, 4]. However, the problem of bit rate reduction, necessary for speech information transfer, on condition of preserving quantization transparency, as well as realization of existing limitations by the volumes of memory, complexity of computations and coding delay, remains actual.

The given paper suggests the method of speech information compression, enabling to reduce data volumes for description of spectral information up to 20 bits per frame, maintaining suitable quality of speech signal, being reproduced.

Method of vector delta-quantization with the prediction

Any compression method is based on usage of redundancy, inherent to speech signal. This redundancy remains after the transition to parametric representation of the signal. Quantitative measure of redundancy is correlation coefficient. Table 1 contains values of correlation coefficient between separate LSF parameters within the frame, obtained for training sequence from 90000 vectors.

Table 1

LSF correlation coefficients within the frame

№	1	2	3	4	5	6	7	8	9	10
1	1.00	0.45	–0.05	–0.07	–0.17	–0.27	–0.28	–0.28	–0.24	–0.10
2	0.45	1.00	0.57	0.23	0.11	–0.02	0.03	0.06	–0.04	–0.01
3	-0.05	0.57	1.00	0.56	0.32	0.39	0.38	0.38	0.19	0.06
4	-0.07	0.23	0.56	1.00	0.58	0.46	0.44	0.26	0.19	0.03
5	-0.17	0.11	0.32	0.58	1.00	0.62	0.42	0.24	0.01	0.01
6	0.27	–0.02	0.39	0.46	0.62	1.00	0.62	0.45	0.20	0.01
7	-0.28	0.03	0.38	0.44	0.42	0.62	1.00	0.63	0.25	0.22
8	-0.28	0.06	0.38	0.26	0.24	0.45	0.63	1.00	0.58	0.30
9	0.24	0.04	0.19	0.19	0.01	0.20	0.25	0.58	1.00	0.56
10	-0.15	–0.01	0.06	0.03	0.01	0.01	0.22	0.30	0.56	1.00

The values of correlation coefficient between corresponding LSF-parameters for serial frames are given in Table 2.

Table 2

LSF correlation coefficients for neighbouring frames

№	1	2	3	4	5	6	7	8	9	10
	0.68	0.73	0.76	0.81	0.86	0.84	0.82	0.81	0.76	0.75

As it is seen from the Table, rather high correlation is observed both between separate parameters within the frame, and between components of LSF vector for neighbouring frames. The following variants of usage of these two types of correlation are possible:

Transition from scalar quantization (SQ) of parameters to vector quantization (VQ);

Application of prediction methods with subsequent elimination of predicted values in order to reduce dynamic range of input values.

There exists fundamental possibility of the given approaches application for usage of both interframe correlation and correlation within the frame. However, scalar quantization (SQ) of LSF-parameters with prediction within the frame requires minimum computational efforts, it is far less efficient than vector quantization (VQ), that is proved by the increase of spectral distortion [3].

At the same time quantization of LSF-parameters for serial frames is connected with the introduction of additional delay that is inadmissible for certain applications. Thus, the expedient is the variant of VQ of LSF parameters for certain frames with further prediction of values for further frames. This approach, called by the author vector delta-quantization with prediction (VDQP) is suggested in the given paper.

The basis of VQ with the help of CB is simple idea: instead of transfer of real value of input speech signal parameters index of nearest CB vector to the input signal is computed and transferred. The most representative values of parameters, obtained as a result of clustering of training sequence of LSF input vectors are stored in CB.

However, according to VDQP method, values of parameters, obtained as a result of clustering of the difference (delta) $\mathbf{e}_i$ of real $\mathbf{l}_i$ and predicted value $\tilde{\mathbf{l}}_i$ of LSF are stored in CB. On condition of successful prediction the dynamic range of input values, used for CB design and, correspondingly, clustering error must decrease, that, in its turn, will lead to the reduction of spectral distortion in the process of quantization.

Thus, in the process of CB design the input sequence will consist of vectors $\mathbf{e}_i = \mathbf{l}_i - \tilde{\mathbf{l}}_i$. Using autoregressive model of linear prediction of the first order the predicted value $\tilde{l}_i(m)$ of m -th component of LSF for the i-th frame is calculated by the formula:

$$\tilde{l}_i(m) = \alpha(m)\hat{l}_{i-1}(m) + \beta(m), \tag{1}$$

where parameters $\alpha(m)$ and $\beta(m)$ are computed, proceeding form the condition of minimization of mean square error of prediction:

$$\alpha(m) = \frac{COV(l_i(m), \hat{l}_{i-1}(m))}{VAR(l_i(m))}, \tag{2}$$

$$\beta(m) = E(l_i(m)) - \alpha(m)E(\hat{l}_{i-1}(m)), \tag{3}$$

where E, COV, VAR denote, correspondingly, expectation, covariance, variance, $\hat{l}_{i-1}(m)$ – is quantized value of m -th component of LSF vector at $i-1$ -th frame; $\hat{l}_{i-1}(m) = l_{i-1}(m)$.

The procedure of LSF-parameters quantization, applying VDQP method is shown in Fig. 1.

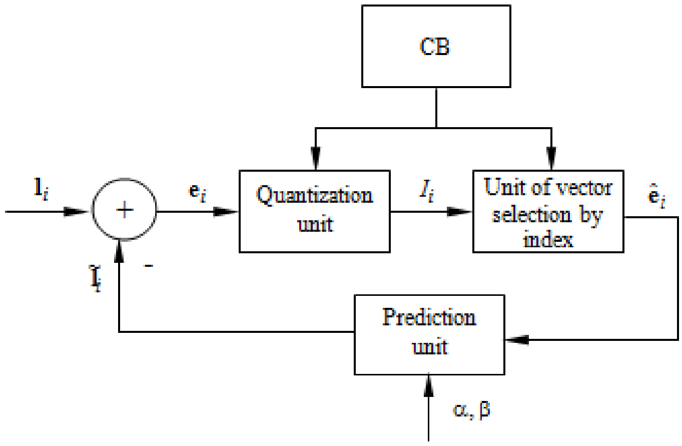

Fig. 1. Quantization block diagram, applying VDQP method

Codebooks design applying VDQP method

Parameters α and β are computed for training sequence of LSF vectors. It assumes the impossibility of direct application of formulas (2) and (3) in the process of CB design, since the values $\hat{l}_{i-1}(m)$ are computed only in the process of quantization, which, in its turn, can be performed only if CB is available.

Thus, practical application of VDQP requires the realization of iterative procedure of CB design.

Step 1. For $i = 1, 2, \ldots, N,\quad m = 1, 2, \ldots, M$ assign $\hat{l}_{i-1}(m) = l_{i-1}(m)$,

Step 2. By the formulas (2) and (3) calculate parameters α, β.

Step 3. By the formula (1) calculate predicted values of LSF parameters $\tilde{\mathbf{l}}_i$.

Step 4. Calculate values of error vectors $\mathbf{e}_i = \mathbf{l}_i - \tilde{\mathbf{l}}_i$.

Step 5. For obtained error vectors $\mathbf{e}_i$ according to the method of K-means create CB $\mathbf{Y}$.

Step 6. Using parameters α, β and created CB $\mathbf{Y}$, perform quantization of training sequence vectors $\mathbf{l}_i$ and obtain quantized values of vectors $\hat{\mathbf{l}}_i$.

Step 7. Calculate quantization error. If the change is minor as compared with the previous iteration, then complete the procedure. If not, return to step 2.

Thus, the procedure of CB design comprises three stages: computation of prediction parameters, quantization and clustering. It is expedient to perform the design of initial CB, applying the improved methods of k-means [5]. At the next iterations classical algorithm of k-means can be applied, using the available vectors $\mathbf{Y}$ for initialization of centroides. Schematically the procedure of CB design is shown in Fig. 2:

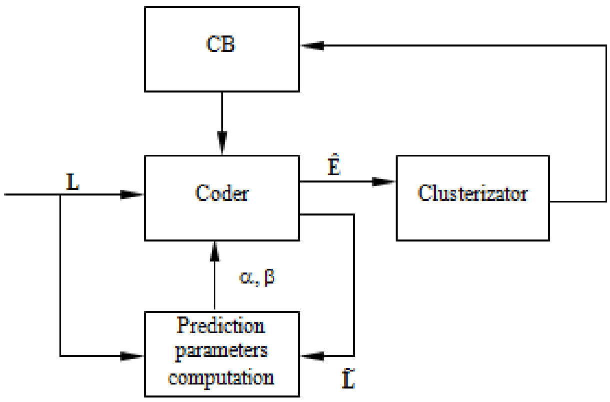

Fig. 2. Block diagram of CB design for VDQP method

The procedure of CB design, considered above, can be used both for split and multistage VQ. In the latter case the procedure of CB design may require additional iterations, stipulated by the necessity of matching of separate parts of CB [6]. However, performance increase, expressed in the reduction of spectral distortion in the process of quantization, is sufficient substantiation of additional time expenditures at the stage of CB design, which is performed once at preparation stage.

Experimental results

For experimental research of the developed method accessible part of English speaking acoustic corpus TIMIT [7] was used.

Training sequence comprised 90000 LSF vectors, obtained on the basis of linear prediction model of the tenth order. Test sequence comprised 15000 LSF vectors, differing from training sequence vectors. The duration of the frame was 20 msec. Distance was measured by weighted Eucledian metric, using digit weights, computed by spectral sensitivity [8].

The scheme of split vector quantization (SVQ) with the split of 10-D LSF vector into two subvectors of 5-D was chosen as the basis for comparison. The same splitting was used while split vector quantization with prediction (SVQP). In the process of multistage quantization the coding took place in two stages, the dimensions of both CB were equal. Two variants of search were modelled: serial, where, for the search at the second stage one nearest vector, obtained at the first stage was used, and tree search, when N_1 vectors, obtained at the first stage, were used at the second stage. Corresponding schemes are designated as MSVDQ with SS and MSVDQ with TS.

Tables 3 – 5 contain the results, obtained while using input vector of 24, 22 and 20 bits per one frame for quantization, correspondingly.

Performance of quantization was evaluated by spectral distortion.

Table 3

Spectral information quantization performance by24 bits per frame

Quantization scheme	SD, dB	% of exceedings	
		SD dB > 2	SD > 4 dB
SVQ	1.18	1.96	0.00
SVDQP	0.93	1.45	0.00
MSVDQP with SP	0.94	1.45	0.00
MSVDQP with TP, $N_1 = 10$	0.92	1.27	0.00
MSVDQP with TP, $N_1 = 30$	0.90	1.12	0.00

Table 4

Spectral information quantization performance by22 bits per frame

Quantization scheme	SD, dB	% exceedings	
		SD > 2 dB	SD > 4 dB
SVQ	1.34	6.28	0.06
SVQP	1.06	3.56	0.03
MSVDQP with SP	1.06	3.57	0.03
MSVDQP with TP, $N_1 = 10$	1.01	2.23	0.01
MSVDQP with TP, $N_1 = 30$	0.98	2.04	0.00

Table 5

Spectral information quantization performance by20 bits per frame

Quantization scheme	SD, dB	% exceedings	
		SD > 2 dB	SD > 4 dB
SVQ	1.51	15.23	0.16
SVQP	1.22	7.11	0.10
MSVDQP with SP	1.23	7.13	0.11
MSVDQP with TP, $N_1 = 10$	1.10	3.63	0.05
MSVDQP with TP, $N_1 = 30$	1.08	3.42	0.03

As it is seen form tables, the conditions of transparency (see p.4, Introduction) are met when 20 bits per one frame of the signal are used for the description of spectral information.

It should be noted, that computational complexity of VDQP method practically coincides with SVQ. Only when tree search is used, the amount of operations increased N_1 times, that is stipulated by the peculiarities of search scheme and not VDQP method. The possibility of reduction of computational complexity of the nearest vector search is considered in [9].

Fig. 3 shows quantization error as the function from the number of iteration in the process of CB design. The error was calculated as the value of distance between input vector and the nearest vector of CB, and was averaged by all the frames.

As it can be seen, already at the fifth iteration quantization error changes only 0.7%.

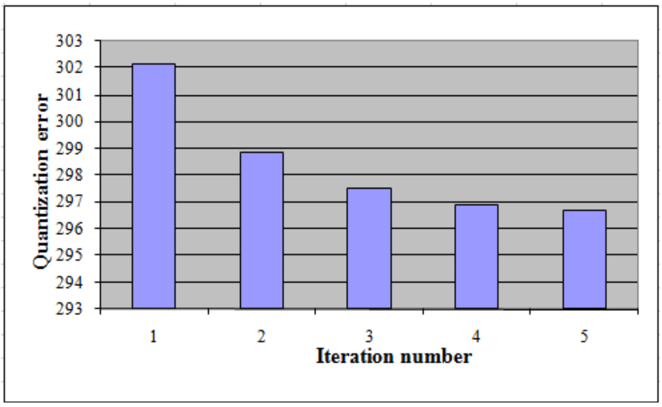

Fig. 3. Dependence of quantization error on the number of iteration

Conclusions

The developed VDQP method allows to increase the performance of speech signal parameters quantization due to the decrease of dynamic range of input vectors, used for CB design and, correspondingly, for decrease of clustering error. This is achieved at the expense of complication of preparatory stage, that requires iterative procedure for CB design. Experimental verification of the developed method showed that the transparency conditions are met for two-stage VDQP with tree search while usage for quantization 20 bits per one frame of speech signal.

REFERENCES

1. Родоначальник [Електронний ресурс] / Ю. Ревич // Компьютерра-Онлайн. – 2012. – Режим доступу до журн.: http://www.computerra.ru/xterra/homo/21659/.

2. Paliwal K. K. Efficient vector quantization of LPC parameters at 24 bits/frame / K. K. Paliwal, B. S. Atal. // IEEE Transaction on Speech and Audio Processing. – 1993. – No. 2, vol. 1. – P. 3 – 14.

3. Lahouti F. Quantization of LSF parameters using a trellis modeling / F. Lahouti, A. K. Khandani // IEEE Transaction Speech and Audio Processing. – Sep. 2003. – No. 5, vol. 11. – P. 400 – 412.

4. Hai Le Vu. Efficient Distance Measure for Quantization of LSF and Its Karhunen–Loeve Transformed

Parameters / Hai Le Vu and Laszlo Lois // IEEE Transactions on speech and audio processing. – Nov. 2000. – No. 6, vol. 8. – P. 744 – 746.

5. Ткаченко О. М. Метод кластеризації на основі послідовного запуску k-середніх з обчисленням відстаней до активних центроїдів / О.М. Ткаченко, Н. О. Біліченко, О. Ф. Грійо Тукало, О. В. Дзісь // Реєстрація, зберігання і обробка даних. – 2012. – № 1, Т. 14. – С. 25 – 34.

6. Chu W. C. Speech Coding Algorithms: Foundation and Evolution of Standardized Coders / Wai C. Chu // NY.: John Wiley & Sons, Inc. – 2003 – 558 p.

7. Fisher William M. The DARPA Speech Recognition Research Database: Specifications and Status / William M. Fisher, George R. Doddington, Kathleen M. Goudie-Marshall // Proceedings of DARPA Workshop on Speech Recognition. – Feb. 1986. – P. 93 – 99.

8. Gardner W. R. Theoretical analysis of the high-rate vector quantization of LPC parameters / W. R. Gardner, B. D. Rao // IEEE Transaction Speech and Audio Processing. – Sep. 1995. – Vol. 3. – P. 367 – 381.

9. Ткаченко О. М. Двоетапна стратегія пошуку у векторних кодових книгах для ущільнення мовлення / О. М. Ткаченко, О. Ф. Грійо Тукало // Вісник Вінницького політехнічного інституту. – 2011. – № 6. – С. 194 – 201.

Tkachenko Olexandre – Cand. Sc. (Eng.), Assistant Professor, Department of Computing Engineering, tel. 59-84-13, e-mail: ant@vstu.vinnica.ua.

Vinnytsia National Technical University.