Journal of Fundamental and Applied Sciences

ISSN 1112-9867

Available online at http://www.jfas.info

Research Article

Special Issue

ABOUT SPEECH DATA COMPRESSION

Evgeniy G. Zhilyakov, Sergey P. Belov*, Aleksandr S. Belov, Aleksandra A. Medvedeva

Belgorod State University Russia, 308015, Belgorod, Pobedy st., 85

Published online: 15 February 2017

ABSTRACT

The continuous increase of voice message flow in the information and telecommunications systems (ITS), as the most natural form of information exchange for a person, led to the almost full occupation of the time-frequency resources of modern ITS. In this regard, the methods were created to reducing the amount of speech data bit representations (compression), both by removal of the pauses between individual words, and by the compression of speech signals generated by the actual speech sounds, which allowed to reduce the cost of ITS time-frequency resources significantly during the transfer of this type of traffic. However, existing approaches of speech sounds compression, based on the use of a psychoacoustic model with the use of coarse quantization in terms of so-called subband transformation levels concerning speech signal sample (vector) transformations are not optimal.

The paper proposes the method of speech data compression without pauses, created on the basis of the mathematical apparatus application concerning the eigenvectors of subband matrices, which allows to formulate variational conditions and to solve the optimization problems of speech data processing adequately.

Key words: speech signal segment, speech data, energy distribution, subband matrix, eigenvectors of a subband matrix, information frequency intervals, compression ratio, codebooks of quasi-optimal quantizers.

Author Correspondence, e-mail: belov@bsu.edu.ru

doi: http://dx.doi.org/10.4314/jfas.v9i1s.783

INTRODUCTION

The problem of bit representation volume reduction concerning speech data during their storage and transmission is considered in the works of many authors, especially by the experts in the field of telecommunications, which is confirmed by the results of scientific and technical literature analysis [1-10].

At that two main aspects are noted: the need to detect pauses with their subsequent coding [4,7-9]. Such pauses arise between separate words and in a dialogue mode, occupying up to 60% of the original sound record duration, and the reduction of the bit representations concerning the actual sounds of speech without pauses [4,7-9] [1-2, 10-11].

The existing methods of speech sound compression without pauses using a coarse quantization by level are based on the psychoacoustic model, which leads to the necessity of so-called subband transformation application concerning the segments of speech signal samples (vectors) that allow to obtain other vectors whose subvectors reflect the frequency properties of a source vector in the selected ranges of frequency axis [12-14]. The components of these sub-vectors are quantized in terms of the level with different steps, and the frequency-selective properties of human hearing are considered by this. Currently, it is customary to use the decimation procedure for output sequences of FIR filters (filters with finite impulse response) tuned to the corresponding sections of a frequency axis for subband conversion [13-14]. This subband conversion procedure is not an optimal one in the sense of a minimum of spectrum approximation errors of the original vectors in the selected frequency ranges, which leads to the increase of data recovery errors by quantized values and, consequently, to the deterioration of reproduced speech quality [14, 15].

In this regard, the authors propose the method of speech data compression without pauses, created on the basis of the mathematical apparatus application with the eigenvectors of subband matrices, which allows to formulate variational conditions and to solve the optimization problems of speech data processing adequately.

Main part

Mathematical bases of the method

The processing of a speech signal separate segments (indication vectors) is carried out:

$$\vec{x} = (x_1,...,x_N)^T \quad (1)$$

In accordance with the chosen uniform division of normalized frequency band $0 \le \omega \le \pi$ into R of intervals V_r with the following form:

$$V_{r2} - V_{r1} = \pi / R;\ V_{r1+1} = V_{r2};\ r = 1;2 \ldots\ R-1 \tag{2}$$

of the same width. The data compression is based on the property of speech signal energy concentration in a small fraction of a frequency band, which allows us to use the approximation of the following form

$$\hat{\vec{x}} = a \sum_{r \in R_m} \vec{x}_r\ , \tag{3}$$

where

$$\vec{x}_r = A_r \cdot \vec{x}, \tag{4}$$

and A_r is the subband matrix, defined by the following elements:

$$A_r = \left\{ a_{i,k}^r \right\},\ i,k = 1, \ldots, N\ ;\ a_{ik}^r = \frac{\sin\left(V_{2,r}(i-k)\right) - \sin\left(V_{1,r}(i-k)\right)}{\pi(i-k)};\ i \neq k\ .$$

The coefficient a of the sum (3) is chosen from the condition $\| \hat{\vec{x}}_r \|^2 = m\|\vec{x}\|^2$, which results in:

$$a = \frac{\sqrt{m}\|\vec{x}\|}{\left\| \sum_{r \in R_m} \vec{x}_r \right\|}, \tag{5}$$

where R_m is the set of frequency intervals of the minimum total width, for which the following conditions are implemented

$$\sum_{r \in R_m} P_r(\vec{x}) \cong m \| \vec{x} \|^2; \tag{6}$$

$$0{,}85 \leq m \leq 0{,}98; \tag{7}$$

where m is the set of information frequency intervals.

The components of the form (4) corresponding to certain frequency intervals have the optimality property in the following sense:

$$\int_{\omega \in V_r} |X(\omega) - X_r(\omega)|^2\ d\omega + \int_{\omega \notin V_r} |X_r(\omega)|^2\ d\omega \Big/ 2\pi = \min$$

i.e. the best approximation of the Fourier transformant segments $X_r(\omega)$ of the original vector in the corresponding frequency intervals, and admit the representation of the following form

$$\vec{x}_r = \sum_{i=1}^{J_r} \lambda_i^r \alpha_{ir} \vec{q}_i^{\,r}\ ;\ \alpha_{ir} = (\vec{x}, \vec{q}_i^{\,r})\ , \tag{8}$$

where λ_i^r are the eigenvalues of the eigenvectors $\vec{q}_i^{\,r}$ of the subband matrix, taking the values $0 < \lambda_i^r \leq 1$. The substitution of (8) into (3) gives the expansion by the set of eigenvectors:

$$\hat{\hat{x}} = \sum_{r \in R_m} \sum_{i=1}^{J} \beta_{ir} \vec{q}_i^{\,r} \,, \tag{9}$$

where

$$\beta_{ir} = a \lambda_i^r \alpha_{ir} \,, \mathrm{i} = 1,..,J_r \,. \tag{10}$$

Since the sets of eigenvectors $\vec{q}_i^{\,r}$ are assumed to be known, it is sufficient to store information on the corresponding coefficients of expansion to restore an initial segment. The conducted studies showed that the power of frequency interval set $(\mathrm{int}\, R_m)$ satisfies the following relation for almost all sounds of Russian speech

$$\mathrm{int}\, R_m \approx 0{,}3R. \tag{11}$$

Therefore, taking into account the equality $J \cong N / R$ we obtain the compression ratio due to the use of approximation (3) (by the number of stored numbers):

$$CH = N /(\mathrm{int}\, R_m \bullet J) \approx 3. \tag{12}$$

The next step is the application of quantization coefficient decomposition by level with their small number. In general, the quantization procedure is described as follows if the following condition is satisfied:

$$\beta_{ir} \in \hat{\mathrm{O}}_m = [\varphi_{m-1}, \varphi_m), \tag{13}$$

then let

$$\beta_{ir}^* = d_m, m = 1,..., K, \tag{14}$$

where K is the number of used quantization levels.

The problem lies in the optimal choice of segment boundaries in (13) and the values d_m in (14) within the sense of error reduction concerning the approximation of the original data by quantized values:

$$\varepsilon_i^2 = \sum_{m=1}^{K} \sum_{\beta_{ir} = S_m} (\beta_{ir} - d_m)^2, (i = 1,..., J), r \in R_m, \tag{15}$$

where S_m is the set of values β_{ir}, satisfying the condition (13). The performed studies showed that at given intervals $\hat{\mathrm{O}}_m$ in (13) the minimum of the right part of (15) is achieved at the set of quantization levels equal to the corresponding mean values:

$$d_m = \sum_{\beta_{ir} = S_m} (\beta_{ir} / \mathrm{int}\, S_m), m = 1,..., K, \tag{16}$$

where $\mathrm{int}\, S_m$ is the power of the set S_m (the number of corresponding β_{ir} values).

Let's introduce a positive non-decreasing sequence:

$$0 \le z_k < z_{k+1}, k = 1; 2 \ldots NK - 1, \tag{17}$$

$$NK = J \bullet \operatorname{int} R_m, \tag{18}$$

at that

$$z_k \in \{|\beta_{ir}|\} / \gamma, i = 1,\ldots, J; \;\; r \in R_m;$$

$$\gamma = \max|\beta_{ir}|, \forall i \text{ и } \; r \in R_m. \tag{19}$$

It was shown that the performance of the following conditions

$$\sum_{m=1}^{K} I_m \vec{z}_m^2 = \max, \tag{20}$$

$$\sum_{m=1}^{K} I_m = NK, \tag{21}$$

where

$$\bar{z}_m = \sum_{i=1}^{I_m} z_{L_{m-1}+i} / I_m \;\; , \tag{22}$$

$$L_m = \sum_{i=1}^{m-1} I_i, L_0 = 0,$$

as well as the choice of quantization levels in the form

$$\hat{d}_m = \bar{z}_m, m = 1,\ldots, K, \tag{23}$$

Gives the minimum of approximation error z_k by quantized values

$$z_k^* = \hat{d}_m, \tag{24}$$

When the following condition is performed:

$$z_{L_{m-1}+1} \le z_k \le z_{L_{m-1}+I_m}. \tag{25}$$

Actually, instead of operation (24), one should use the following encoding

$$codz_k = \log_2 m, \tag{26}$$

bearing in mind that the numbers of the quantization levels can be denoted appropriately by the binary digit numbers p, so that

$$K = 2^p. \tag{27}$$

Thus, the number of quantization levels should be chosen from the set (2; 4; 8 ...). In accordance with this, during the process of the study carrying out, the algorithm for the problem (20), (21) solution was developed with a successive division of subsequences into two parts, each of which satisfies these conditions with its parameters I_m and d_m (since the division into two sequences of any length is easily realized by successive search). The use of

standardized sequences of the form (17) - (19) allows not to store the values of the levels (23), and to use the levels from a pre-formed codebook (to restore the data) that satisfies the following condition

$$\sum_{m=1}^{K} I_m \bullet (\bar{z}_m - d_m^l)^2 = \min \forall D_l, \tag{28}$$

where

$$D_l = \{d_1^l, ..., d_K^l\}, d_1^l < d_2^l < ... < d_K^l. \tag{29}$$

Such codebooks are formed at = 2; 4; 8 taking into account all the sounds of Russian speech with the averaging over the set of announcers. Tables 1-3 show some of the resulting codebooks, as well as the average value of the quantization levels for all sounds and the mean square deviation (MSD) calculated by the following formula

$$\sigma = \sqrt{\frac{1}{L}\sum_{i=1}^{L}\left(Nk_i - \overline{Nk}\right)^2}, \tag{30}$$

Where L is the number of Russian speech sounds; Nk is the level of quantization; $\overline{Nk}$ is the average value of the quantization level.

Table 1. Code books at quasioptimal quantization for two levels

	1-st level	**2-nd level**
А	0.2875	0.7065
Б	0.2138	0.7845
В	0.2262	0.7623
Г	0.2437	0.7497
Д	0.2277	0.7749
Е	0.2425	0.7465
Ё	0.2507	0.7676
Ж	0.2852	0.7143
З	0.2347	0.7726
И	0.2099	0.7838
Й	0.2212	0.7451
К	0.3247	0.6721
Л	0.224	0.7656
М	0.2104	0.7801

Н	0.1955	0.794
О	0.2801	0.7313
П	0.2904	0.7223
Р	0.2839	0.7168
С	0.3282	0.6911
Т	0.3357	0.6718
У	0.2313	0.7641
Ф	0.3149	0.712
Х	0.2916	0.6782
Ц	0.3155	0.6648
Ч	0.3197	0.68
Ш	0.336	0.6802
Щ	0.3209	0.6771
Ы	0.2335	0.7397
Э	0.2672	0.7201
Ю	0.2354	0.7534

Table 1 (continued)

Я	0.2681	0.7262
Mean value	0,2661	0,7306
MSD	0,0434	0,0396

Table 2. Code books at quasi-optimal quantization for 4 levels

	1-st level	**2-nd level**	**3-rd level**	**4-th level**
А	0.0911	0.3554	0.9133	0.9133
Б	0.0444	0.2847	0.9597	0.9597
В	0.0486	0.3058	0.952	0.952
Г	0.0558	0.3213	0.9344	0.9344
Д	0.0533	0.3003	0.9528	0.9528
Е	0.0558	0.3256	0.9481	0.9481
Ё	0.0654	0.342	0.9453	0.9453
Ж	0.09	0.355	0.9097	0.9097

З	0.0544	0.306	0.95	0.95
И	0.0356	0.2847	0.965	0.965
Й	0.0522	0.2983	0.946	0.946
К	0.1266	0.3917	0.8778	0.8778
Л	0.0431	0.2972	0.9543	0.9543
М	0.0381	0.2779	0.9619	0.9619
Н	0.0303	0.2602	0.9641	0.9641
О	0.0788	0.3603	0.929	0.929
П	0.1028	0.3744	0.8969	0.8969
Р	0.0832	0.3495	0.9148	0.9148
С	0.1239	0.3907	0.8799	0.8799
Т	0.1471	0.3991	0.8557	0.8557
У	0.0563	0.3112	0.9523	0.9523
Ф	0.1191	0.3758	0.8925	0.8925
Х	0.1012	0.3631	0.8983	0.8983
Ц	0.1228	0.3841	0.8587	0.8587
Ч	0.1321	0.3921	0.8705	0.8705
Ш	0.1327	0.3991	0.8708	0.8708
Щ	0.1235	0.3857	0.8641	0.8641
Ы	0.0495	0.3093	0.9488	0.9488
Э	0.0666	0.342	0.9272	0.9272
Ю	0.0599	0.3399	0.9399	0.9399
Я	0.075	0.3469	0.9116	0.9116
Mean value	0,0793	0,3396	0,9208	0,9208
MSD	0,0346	0,04047	0,0353	0,0353

Table 3. Code books at quasi-optimal quantization for 8 levels

	1-st level	**2-nd level**	**3-rd level**	**4-th level**	**5-th level**	**6-th level**	**7-th level**	**8-th level**
А	0.0134	0.1334	0.3973	0.3973	0.6254	0.6697	0.8711	0.9878
Б	0.005	0.0722	0.3207	0.3207	0.6998	0.7408	0.9327	0.996

								2
В	0.008	0.0788	0.3452	0.3452	0.6747	0.7172	0.9195	0.991 5
Г	0.0061	0.0881	0.3621	0.3621	0.6543	0.6962	0.902	0.993 2
Д	0.0079	0.083	0.3401	0.3401	0.6926	0.7324	0.9256	0.994 3
Е	0.0056	0.0887	0.3671	0.3671	0.6419	0.6849	0.9148	0.994
Ё	0.0043	0.103	0.3877	0.3877	0.6657	0.6993	0.9139	0.996 2
Ж	0.0147	0.1304	0.396	0.396	0.6291	0.6681	0.8705	0.987
З	0.0091	0.0832	0.3462	0.3462	0.6823	0.719	0.924	0.993 7
И	0.0032	0.0599	0.3204	0.3204	0.6954	0.7397	0.9406	0.997 4
Й	0.0065	0.082	0.3371	0.3371	0.6579	0.7112	0.9096	0.993 7
К	0.0249	0.1745	0.4354	0.4354	0.5901	0.6226	0.8304	0.978 5
Л	0.0044	0.0712	0.3362	0.3362	0.6793	0.7236	0.9238	0.995 4
М	0.0049	0.0642	0.3125	0.3125	0.6967	0.742	0.9371	0.996 9
Н	0.0019	0.0525	0.2931	0.2931	0.7156	0.7623	0.9411	0.997 9
О	0.0111	0.1217	0.404	0.404	0.6395	0.6772	0.8875	0.991 2
П	0.0173	0.1494	0.4202	0.4202	0.6111	0.6457	0.8485	0.982 6
Р	0.0123	0.12	0.3906	0.3906	0.6333	0.6735	0.878	0.987 4
С	0.0279	0.1672	0.4297	0.4297	0.6157	0.6443	0.8376	0.973
Т	0.0367	0.1953	0.4419	0.4419	0.5881	0.6148	0.8059	0.963

								2
У	0.0054	0.0907	0.3538	0.3538	0.6679	0.7113	0.9209	0.995 7
Ф	0.0238	0.166	0.4194	0.4194	0.6291	0.6589	0.8497	0.98
Х	0.0189	0.146	0.4019	0.4019	0.5885	0.628	0.8526	0.980 9
Ц	0.0278	0.1644	0.4264	0.4264	0.5786	0.6111	0.8101	0.966 6
Ч	0.0298	0.1846	0.4346	0.4346	0.5867	0.6206	0.82	0.970 2
Ш	0.0304	0.1859	0.4425	0.4425	0.5984	0.6287	0.8213	0.973 1
Щ	0.0241	0.1695	0.4315	0.4315	0.5926	0.627	0.8154	0.972 4
Ы	0.0049	0.0816	0.3489	0.3489	0.6466	0.6964	0.9171	0.996 4
Э	0.009	0.1052	0.3895	0.3895	0.6251	0.6709	0.8852	0.990 8
Ю	0.0065	0.0971	0.3805	0.3805	0.6402	0.6866	0.8939	0.992 5
Я	0.0112	0.111	0.394	0.394	0.6288	0.6717	0.8666	0.989 7
Mean value	0,0134	0,1168	0,3808	0,3808	0,641	0,6805	0,8828	0,987 1
MSD	0,0097	0,0427	0,0429	0,0429	0,0389	0,0431	0,0434	0,010 2

In order to illustrate the efficiency and the effectiveness of the developed method and algorithm, experimental studies were conducted that showed that speech intelligibility is preserved even at K = 2. Thus, taking into account the need to preserve the sign bit and the value γ the maximum achieved compression ratio can be equal to

$$CH_{\max} = 12N/(N+12), \tag{31}$$

(Assuming 8-bit initial samples). That is, if N is sufficiently large, then

$$CH_{\max} \cong 12 . \quad (32)$$

SUMMARY

The conducted studies established that the proposed method of speech data compression without pauses based on the optimal quantization by the level of decomposition ratios in respect of speech signals by eigenvectors of subband matrices from m-information frequency intervals using the codebooks of quasi-optimal quantizers allows to provide the compression ratio up to 12 times, depending on the bit depth of these signals initial samples.

CONCLUSION

The use of the developed method will make it possible to achieve an overall compression ratio of the speech data both by pause detection and coding, which can make more than 60% of a dialogue duration, and by quantizing the coefficients of speech segment decomposition by the eigenvectors of subband matrices from m-information value frequency intervals in 20-25 times.

REFERENCES

1. Sergienko V.S., Barinov V.V. Data, speech, sound and image compression in telecommunication systems. Moscow: Radio Soft, 2009. - 360 p.
2. The compression of data in the systems of information collection and transmission. Ed. by V.A. Sviridenko. M.: Radio and Communication, 1985. - 184 p.
3. Salomon D. Compression of data, images and sound. M.: TECHNOSPHERE, 2004. - 368 p.
4. Digital processing and voice transmission. Ed. by O.I. Shelukhin. M.: Radio and communication, 2000. - 456 p.
5. Ashwin R., Kumaresan R. On decomposing speech into modulated components. IEEE Transactions on Speech and Audio Processing, vol. 8, № 3, May 2000. pp. 240-254.
6. Seki H. A new method of speech transmission by frequency division and multiplication. The Journal of the Acoustical Society of Japan, vol.14, 1958. pp. 138- 142
7. Kalintsev Yu.K. Intelligibility of speech in digital vocoders [Text] / Yu.K. Kalintsev. - Moscow: Radio and Communication, 1991. - 220 p. Ill.
8. Zhilyakov E.G. On the effectiveness of various approaches to the segmentation of speech signals based on the detection of pauses [Text] / E.G. Zhilyakov, S.P. Belov, A.S. Belov, A.A. Firsova // Scientific statements of BelSU Ser.: Computer science. - Belgorod:

Publishing house of BelSU, 2010. - № 7 (78), Issue. 14/1. - pp. 187-1193.

9. Bykov S.F. Digital telephony [Text] / S.F. Bykov, V.I. Zhuravlev, I.A. Shalimov. M.: Radio and Communication, 2003. 144 p.: ill.
10. Aldoshina I. Fundamentals of psychoacoustics. Hearing and speech. Part 1 [Text] / I. Aldoshina // Information and Technical Journal "Sound producer". 2002. №1. - pp. 38-44.
11. Kovalgin Yu.A. Digital coding of audio signals [Text] / Yu.A. Kovalgin, E.I. Vologdin. - St. Petersburg: Crown print, 2004. - 240 p.: ill.
12. Gusinskaya, E.I. Optimization of filter bank in subband coding problems: a thematic review [Text] / E.I. Gusinskaya, A.A. Zaitsev // Scientific and technical journal "Digital signal processing". - 2004. - No. 3 (12). - pp. 18-29.
13. Sergienko A.B. Digital signal processing [Text] / St. Petersburg: Peter, 2003. - 604 p.: ill.
14. Sinilnikov A.M. Band coding of audio signals with orthogonal transformation [Text] // Telecommunications. - 1988. - №9. - pp. 34-36.
15. Belov S.P. New method of optimal linear filtering for processing of speech data processing on a computer [Text] / S.P. Belov, E.G. Zhilyakov, E.I. Prokhorenko // The issues of radio electronics. Ser. «Electronic computing equipment (ECE)». - M., 2008. - Issue. 1. - pp. 132-143.

How to cite this article:
Evgeniy G. Zhilyakov, Sergey P. Belov, Aleksandr S. Belov, Aleksandra A. Medvedeva. About speech data compression. J. Fundam. Appl. Sci., 2017, *9(1S), 1301-1312.*