БЪЛГАРСКА АКАДЕМИЯ НА НАУКИТЕ . BULGARIAN ACADEMY OF SCIENCES

ПРОБЛЕМИ НА ТЕХНИЧЕСКАТА КИБЕРНЕТИКА И РОБОТИКАТА, 51
PROBLEMS OF ENGINEERING CYBERNETICS AND ROBOTICS, 51

София . 2001 . Sofia

A Wavelet-Packet Based Speech Coding Algorithm[1]

Grigor Marchokov, Atanas Gotchev[2], Zdravko Nikolov

Institute of Information Technologies, 1113 Sofia

1. Introduction

The trend toward real-time, low-bit-rate speech coders dictates current research efforts in speech compression. Such coders are desirable for a number of applications including transmission of digital speech signals and multimedia applications. Multimedia and video conferencing, dynamic web-site access with voice and video introduces the idea of using voice over the Internet. This idea also opens up new commercial opportunities in the area of self-service and other service applicability [1]. Another area of interest, where speech compression has gained widespread attention is the area of cellular and mobile stations. Both Internet and wireless communications channels are characterized by limited bandwidth conditions, which determine the research efforts of designing low bit-rate speech coders with admissible quality. There have been a number of speech coders developed for lossy coding of speech signals, and recently linear predictive coders (LPC) have been widely used for achieving low bit-rate speech [2]. They are based on the speech production mechanism and widely use the periodicity and the autoregressive structure of the speech signal. They differ mainly in the methods, aimed at coding the LPC residual carrying information about the exciting source. The RLP based GSM standard [3] for mobile communications and the G.728 CELP based standard for computer network communications are most commonly applied [4]. They have fixed bit-rate based on the linear predictive models used.

An alternative, offering variable bit-rate speech coders is based on certain decorrelating linear transform and successive transform coefficients quantization. The wavelet transform (WT) has been most widely used because of its nice properties. It is

[1] This work is supported by the Ministry of Education and Science, National Science Council, under the Grand No: MU-I-003/96.

[2] Atanas Gotchev is currently with the Tampere International Center for Signal Processing, Tampere University of Technology, P.O.Box 553, 72101 Tampere, Finland, Email: agotchev@cs.tut.fi

closest to the optimum Karhunen-Loeve transform and almost decorrelates a number of signal classes [5]. The basic functions (the wavelets) are well localized in time and frequency (scale) that gives a natural possibility to deal with the transform coefficients in an effective application-oriented manner.

There are several crucial parts when applying WT for compression. First step is an appropriate choice of the basis. The famous Daubechies family of compactly supported and orthogonal wavelets has been extremely used for compression [6]. While applied on digital images it gives excellent results based on their compaction properties and maximal number of vanishing moments, but for audio signals its application us limited because of the lack of linear phase. It is well known that the orthogonality, the compact support and the linear phase can not coexist [6]. Fortunately nice bi-orthogonal and compactly supported wavelets, most of them based on B-splines has been designed [5, 6] and corresponding speech coding methods have been introduced.

The second crucial step in designing a wavelet-based coder is the appropriate transform coefficients quantization. It has been proved that the zero-tree coders (ZTC) perform better than other quantizators, taking advantage of the hierarchical tree structure of the wavelet coefficients [7]. And the third step is an appropriate loss-less coder, most often entropy based.

In this paper, good answers for the three steps mentioned above are given. We argue the usage of wavelets based on B-spline basic functions. In order to obtain the best time-frequency tilling possible we apply Wavelet Packet Transform and determine the best basis by using a perceptual entropy measure. Then we introduce a modified Zero Tree Coder (ZTC) related with the speech signal's properties.

The paper is organized as follow: Section 2 gives a brief overview of the concept of wavelets, wavelet packets and best basis selection. Section 3 argues the usage of B-splines as generating functions for wavelet bases and describes some of their nice properties. Section 4 details the modification of the ZTC, aimed at effective usage of the preliminary knowledge of speech signal nature. Section 5 introduces the loss-less entropy coder for a higher degree of compression. Section 6 describes various issues involved in a real-time implementation. Section 7 compares the performance of this coder with respect to G.728 standard and others unified lossy coders. We summarize in Section 8 with conclusions and directions for future work.

2. Wavelet packet transform

The main advantage of WT over the other linear transforms (e.g. Fourier or DCT) is its ability to represent the signal in both time and frequency within the Heisenberg's uncertainty principle limits [6]. It has been shown that the wavelets can approximate time-varying non-stationary signals in a better way that the Fourier T. [6]. More recently, a number of decomposition, leading to an optimized time-frequency tilling has been proposed, e. g. wavelet packets (WP); frames, local overlapped transforms (LOT), etc. [6]

2.1. Theory in brief

We consider the Hilbert space L_2 of finite energy functions. The wavelet packet for such space is well localized in time and frequency. It is parameterized by three parameters, describing its scale, position and frequency. Fast wavelet packets can be established by a pair of quadrature-mirror filters. Let $h=\{h_j\}$ is a low-pass filter possessing the following properties:

$$(a)\ \text{For}\ \varepsilon > 0,\ \sum_j |h_j||j|^{\varepsilon} < \infty;$$

(1) $$(b)\ \sum h_{2j+1} = \frac{1}{\sqrt{2}}\ \text{for}\ i = 0,1;$$

$$(c)\ \sum h_{2j+1} = \delta_k,\ \text{where}\ \delta\ \text{is the Kroneker symbol.}$$

The property (*a*) represents the filter decay, and (*b*) and (*c*) – its orthogonality.

Let $g=\{g_j\}$ depends on h as follows:

(2) $$g_j = (-1)^{1-j} h_{1-j}\ .$$

The two discrete sequences form a quadrature-mirror pair. Two filtering-decimation operator then can be defined:

(3) $$\mathbf{H}x(t) = \sum_j h_j x(2t - j)\ \text{and}\ \mathbf{G}x(t) = \sum_j g_j x(2t - j)$$

together with their conjugates as well

(4) $$\mathbf{H}^*x(t) = \frac{1}{2}\sum_j h_j x\left(\frac{t}{2} + \frac{j}{2}\right)\ \text{and}\ \mathbf{G}^*x(t) = \frac{1}{2}\sum_j g_j x\left(\frac{t}{2} + \frac{j}{2}\right).$$

Assuming h и g are with finite length, we define: $\phi = \lim_{n\to\infty} \mathbf{H}^n\aleph$, where $\aleph$ is an indicator function for the interval $[-1/2, 1/2]$ and it is the unique fixed point in the equation $\phi = \mathbf{H}\phi$.

The wavelet packet is a projection of ϕ with successive applications of **H** and **G** and with some possible translations and dilatations. The wavelet packets as obtained, are orthogonal with respect to their translated and dilated versions. We can arrange the three parameters and index the wavelet packets $w_{f,s,p}$ as follows:

$$w_{0,0,0}(t)=f(t);\ w_{2f,0,0}(t)=\mathbf{H}w_{f,0,0}(t);\ w_{2f+1,0,0}(t)=\mathbf{G}w_{f,0,0}(t),\ \text{etc.}$$

By wavelet packets we can approximate a continuous-time function $x \in L^2(\mathbf{R})$ with accuracy $O(2^{-L})$ by the l^2 sequence of inner products $x_i = \langle x, w_{0,-L,i}\rangle$, for i integer. The following equation allow recursive computation of the wavelet packets:

(5) $$\langle x, w_{2f,s+1,p}\rangle = \sum_j h_j \langle x, w_{f,s,2p+j}\rangle,$$
$$\langle x, w_{2f+1,s+1,p}\rangle = \sum_j g_j \langle x, w_{f,s,2p+j}\rangle.$$

The **H** and **G** operators are applicable on discrete sequences (signals):

(6) $$\mathbf{H}: l^2 \to l^2,\ \mathbf{H}x_n = \sum_j h_j x_{2n+j}\ ,$$
$$\mathbf{G}: l^2 \to l^2,\ \mathbf{G}x_n = \sum_j g_j x_{2n+j}\ .$$

The wavelet packets form the so-called dictionary of bases. Considering vectors in R^N, there are $N\log N$ bases and more than 2^N orthonormal bases exist in R^N, then. The basic vector and the corresponding coefficients are located in nodes of a binary tree. Nodes from one level correspond to a particular scale. They differ in frequency positions. The coefficients in each vector differ in their time position (Fig. 1). Each node is a cartesian sum of its descendants. Starting from the root we can divide the nodes forming in such way certain basis from the bases dictionary.

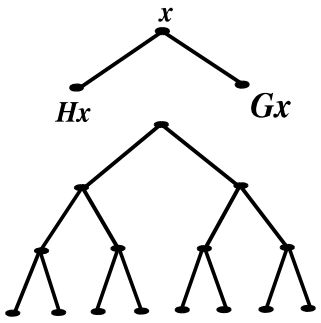

Fig. 1. Tree structured WPT

As we argued, there is a redundancy of bases.

A question arises: how we can find the best basis. We need an appropriate information measure. Once established, it can be searched for a minimum through all possible bases. This best basis search can be done by fast algorithms of "divide and conquer" type [8, 9]. Obviously in order to be applicable on tree structured bases, this measure has to be additive. It can be expressed as a functional $\boldsymbol{M}$ mapping the sequence $\{x_i\}$ into R^N, where $\mathbf{M}(\{x_i\}) = \sum_j M(x_i)$. Since the vector space R^N can be factorized on a product of one-dimensional spaces, the finding of a minimum of $\boldsymbol{M}$ requires $O(N)$ operations.

In our investigations we use the perceptual entropy measure, as defined in [10].

2.2. The perceptual entropy as a best basis information measure

Our best-basis information measure is based on the psycho-acoustic model. It is well known that the human hearing performs space selection of different frequency bands called critical bands [11]. Hence two sounds in the same frequency band are undistinguishable. This property is known as frequency masking effect. Using this effect one can determine how many bits are sufficient to code the subband signal. The number of bits (quantization levels) influences on the quantization noise. The perceptual entropy (PE) determines the threshold representing the maximum level of the injected quantization noise, being inaudible while added to the input signal. We calculate the PE using Model II of the ISO-MPEG [12]. Hence the determined by the PE number of bits per subband can serve as an information measure $\boldsymbol{M}$. Instead of building the complete tree structure of the WPT, we start from the level 0 and at every stage, a decision is made whether to decompose the subband further, based on the $\boldsymbol{M}$. If the decomposition results in a smaller $\boldsymbol{M}$, it is carried out. Otherwise, it stops. In this manner we adapt the WPT tree to approach the critical bands, determined by the PE, as close as possible.

3. Wavelet bases, derived from B-splines

Here we briefly argue our choice to use wavelets, generated by B-spline basic functions. More theoretical details can be found in [13, 14].
For the classical B-spline case $\varphi(x) = \beta''(x)$ is a central B-spline of degree n:

$$\beta''(x) = \beta^0(x)\beta^{n-1}(x), \tag{7}$$

$$\beta^0(x) = \begin{cases} 1, \text{ if } x \in [-1/2, 1/2], \\ 0 \text{ otherwise.} \end{cases} \tag{8}$$

They are in the function class C^{n-1} i. e. they are the most regular functions of degree n with a support of $n+1$ and an approximation order $L=n+1$ [13]. When sampled at integers, the B-splines are symmetrical finite-length sequences, whose values are binomial coefficients.

The human hearing system can be modeled as successive convolutions with Gaussian kernels with different scales [15]. B-splines are good approximations of the Gaussian kernel. By the numerical computations, it was shown that the cubic B-spline is already near optimal in terms of time-frequency localization in the sense that its variance product is within 2% of the limit specified by the uncertainty principle [16]. Another significant property of the B-spline of a given order n is that it is the unique compactly supported refinable spline function of order n which can provide a stable hierarchical representation of a signal at different scales [17]. Hence, a compactly

supported spline is m-refinable and stable *if and only if* it is a shifted B-spline:

$$S_h^n = \{\sum_{k\to-\infty}^{\infty} c_m(k)\beta_h^n(x - hk) : c_m \in l^2(Z)\}. \quad (9)$$

Then

$$S_{lm}^n \subset S_m^n \ , \ \forall\ i \in Z_+ \text{ and } \bigcup_{h>0} S_h^n = L^2(R). \quad (10)$$

Since B-splines provide a stable multiresiolution representation of a signal at multiple scales, it is preferable to select B-splines as smoothing kernels to extract multiscale information inherent in a signal. If we choice a B-spline of certain order as a smoothing (scaling) functions, the corresponding wavelet is easy to construct by an approximation with a linear combination of B-splines:

$$\varphi(x) = \sum g(k)\beta^n(x - k). \quad (11)$$

The weighting coefficients can be chosen with different assumptions, e.g. taking first or second derivatives of B-splines. By using such types of wavelets, we can represent a signal by its multiscale maxima or zero-crossings [18]. Higher order of derivatives of B-splines can represent signal transients with higher singularity and the coefficients g are the binomial-Hermite sequences. This leads to construction of fast algorithms [5]. For most compact representation, needed in compression tasks, the m-scale relation is simplified to two-scale relation, which leads exactly to critically sampled wavelet (wavelet packet) scheme (), where the QMF are with binomial coefficients.

4. Modified Zero-tree coder

The original version of the zero-tree coder was found useful in wavelet coding of still images [7]. The basic assumption is that most of the signal's energy is concentrated in the lover frequency bands. Under the above assumption there is a high probability that if the energy of some frequency band is lover than a certain threshold, the energies of the higher bands will remain below the threshold as well.

We can adapt this assumption for speech signals as well. The speech signal's energy is also concentrated in the relatively lover frequencies. In the terms of the WPT those are the frequency bands with higher scale parameters. They contain the pitch frequency and the first two high-energy formants. Another reason is that we, adapting the decomposition by means of the PE, had found the appropriate frequencies bands, whish the hearing system is much sensitive at. Hence, we modify the thresholding operation by inserting two different thresholds: one (lower) for the four most significant low-frequnecy bands, and second, two times higher, for the rest bands.

5. Huffman entropy coder

After the ZTC we apply a loss-less coning based on an adaptive zero-order Huffman algorithm. The first data pass includes checking the counts for each symbol in the alphabet. The Huffman table is then built using a simple yet elegant procedure in which the individual symbols are laid out as a string of weighted leaf nodes to be joined ad a binary tree. The weight of each node is set by the frequency count of the symbol it represents. The binary tree structure is built as follows:

- The two nodes with the lowest weight are allocated.
- A parent node for these two nodes is created. It is assigned a weight equal to the sum of the two child nodes.
- The parent node replaces the two child nodes in the list of free nodes.

- One of the child nodes is designed as the path taken from the parent node when decoding a 0 bit, while the other is set as the path when decoding a 1 bit.

- The previous steps are repeated until only one free node remains in the list: this last node is therefore the root of the tree.

Longer length words are allocated to symbols with lower counts and shorter length codes are given to symbols with higher counts.

6. Real-time implementation

6.1. Implementation of the WT

The wavelet transform module was originally implemented in C. The program's assembly code was then optimized to eliminate unnecessary address load and branch instructions. Block repeat and single repeat instructions were also used wherever possible in this version of the program to reduce the number of cycles required for each program run. The block size employed by the program can be anywhere between 512 samples and 4096 samples. Large block sizes typically result in superior compression and larger run-time memory and time-delay requirements. We chose a block-length of 1024 samples, which is acceptable form the time-consuming point of view and also is well-related with the second-order stationary statistics of the speech signals.

We exploit the main differences between wavelet-based methods and linear-prediction based methods. In most window-based vocoders, such as those based on linear-predictive methods, some samples from the previous block are attached to the boundaries of the current block to avoid edge discontinuities. In this wavelet based implementation, boundary artifacts are reduced by symmetric extensions the block boundaries. This requires symmetrical wavelets to be applied.

The program was originally written to test diverse wavelet families, which were specified by the length and impulse response coefficients of the low-pass and high-pass filters required. Also some possibilities to test different decomposition techniques (e.g. DWT, frames, WPT) were included. As it turned out, this approach requires too processor overhead from a speed perspective so the program was redesigned to handle only a particular set of wavelet filters. In our final version we deal with the bi-orthogonal set of B-spline based wavelets. A final optimization technique involved replacing branched loop structures with repeated instructions wherever possible. This avoided several types of instructions and their associated overhead at the expense of program memory.

6.2. Implementation of the modified ZTC

In the zero-three algorithm the ordering of the transmission of coefficients values in not done by sending the indices, but by sending coefficient significance information. A coefficient c_n is called significant with respect to a given l, if it satisfies $|c_n| \geq 2^l$; otherwise it is called an insignificant coefficient. A subset S_m is defined to be a significant subset with respect to a given k, if it contains at least one significant coefficient with respect to a given k; otherwise it is defined as an insignificant subset. For the ordering process we have split the set of WP coefficients into subsets and check for the significance of each subset. If a subset is found to be insignificant then all of its members are insignificant; if a subset is significant the decoder needs more information about the significant members. This is done by an appropriate subset splitting. The process is repeated until a magnitude check is applied to all significant subsets that include only one member. Since the splitting rules are common for the coder and the decoder, there is no need to send the indices of the transmitted coefficients.

The algorithm can be represented in a pseudo-code as follows:

```
Treshold := max(abs(Ck)) / 2, k = 0..N-1
While(Treshold > LimitTreshold)
Begin
      DominantPass
      SubordinatePass
      Treshold := Treshold / 2
End
```

The initial threshold is taken to be two times lower the maximal element in the subset.

For efficient compression the significant-subset splitting rules should satisfy the following: a subset that is expected to be insignificant should have as many coefficients, as possible, while a subset that is expected to be significant should have the lowest possible number of coefficients (preferable only one – the significant coefficient). That is the reason we use the WPT decomposition since the frequency bands are expected to have the best possible energy compaction, related with the psycho-acoustic reception. We take advantage also of the features of the speech signal and use two different thresholds for low-frequency and high-frequency bands.

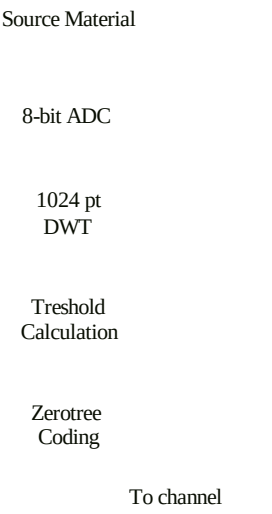

Fig. 2. Flow-diagram of the proposed algorithm

6.3. Implementation of the entropy coder

As was described in the previous section, an adaptive order-0 Huffman coder was implemented in assembly language. This coder function employs a two-pass algorithm: The first pass counts each symbol while building the Huffman table, and the second pass generates the coded bitstream from the input symbol data. The coder was written in assembly. Each node of the Huffman code is stored using a data structure with symbol count, parent, child-0 and child-1 as its members. Existing symbols are counted and included in the header of each block of data sent to the decoder, which then uses these counts to build the required Huffman table for block processing. This table is employed in the second pass of data processing to decode the bitstream.

7. Experimental results

The performance of the proposed techniques was compared with the existing speech coding technique LD-CELP G.728 [4]. The compression results were compared in terms of peak-signal-to-noise ratio (PSNR) and the compression ratio. The output bit-rate of the G.728 standard-based coder is 16 kbits/s. This implies that an 8 bit/sample μ-law signal would achieve a compression ratio 4:1, and an 8 bit/sample speech signal would achieve a compression ratio of only 2.5:1.

Four different test sets separated at segments of 1024 samples each were tested:

'First sentence' – male voice English

'Second sentence' –female voice English

'Third sentence' – male voice Bulgarian

'Fourth sentence' – female voice Bulgarian.

Table 1 summarizes the results got. A segment of the 'First sentence' signal is presented in Fig. 3. Four segments decoded with different quality are given on Figs. 4–7. It can be seen that the energy behavior of the speech segment has been preserved

Table 1. Experimental results of the proposed algorithm and the LD-CELP (G.728) algorithm over four sentences with different compression rates

	Compression rate bits/sample	PSNR, [dB]	PSNR for LD-CELP, fixed compression rate, [dB]
Bulgarian Male Speaker	1.336	27.06	
Bulgarian Male Speaker	2.399	30.24	
Bulgarian Male Speaker	3.648	33.80	
Bulgarian Male Speaker	4.948	34.35	31.03
English Male Speaker	1.164	29.50	
English Male Speaker	2.129	31.04	
English Male Speaker	3.251	32.67	
English Male Speaker	4.515	40.14	37.00
Bulgarian Female Speaker	1.898	36.45	
Bulgarian Female Speaker	3.314	42.19	
Bulgarian Female Speaker	4.794	46.25	
Bulgarian Female Speaker	6.290	48.47	36.00
English Female Speaker	1.467	33.19	
English Female Speaker	2.551	37.36	
English Female Speaker	3.766	39.93	
English Female Speaker	5.143	41.51	31.70

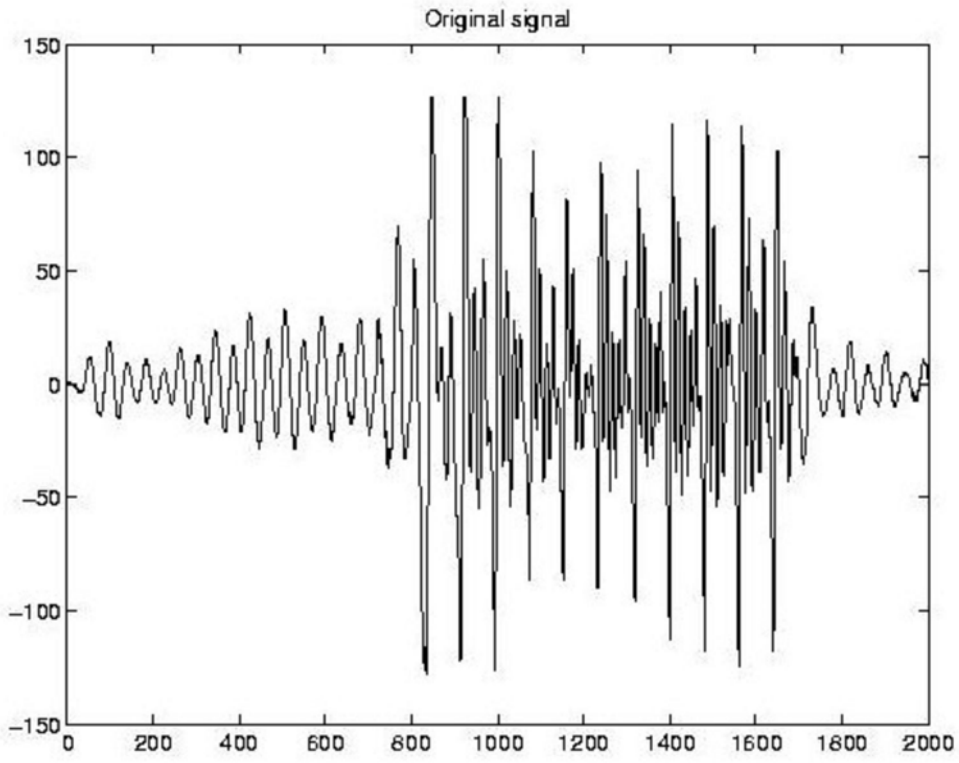

Fig. 3. A segment from the 'First sentence'

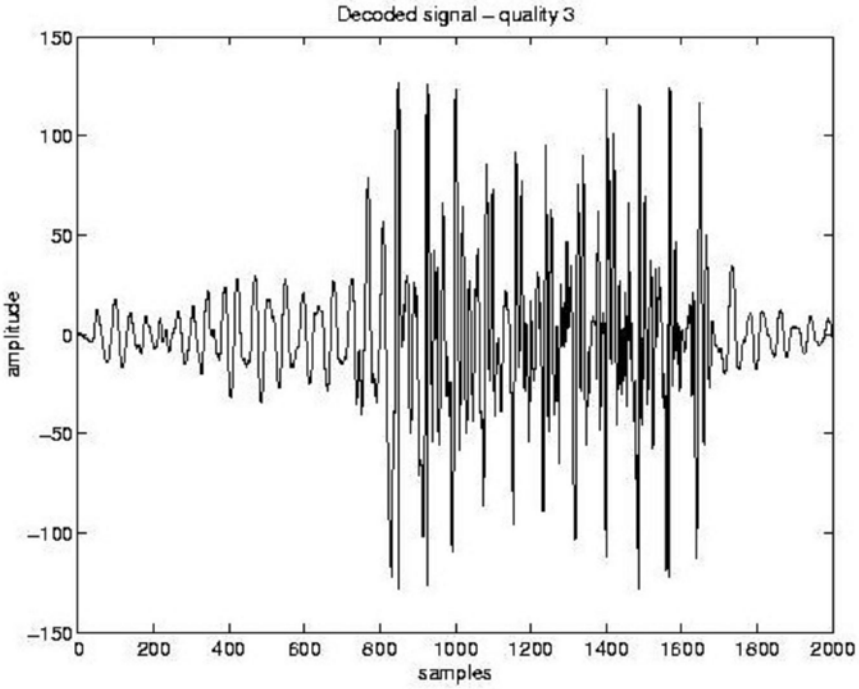

Fig. 4. The same segment as in Fig. 3 compressed with a compression ratio 1.336 bits/sample and then decompressed

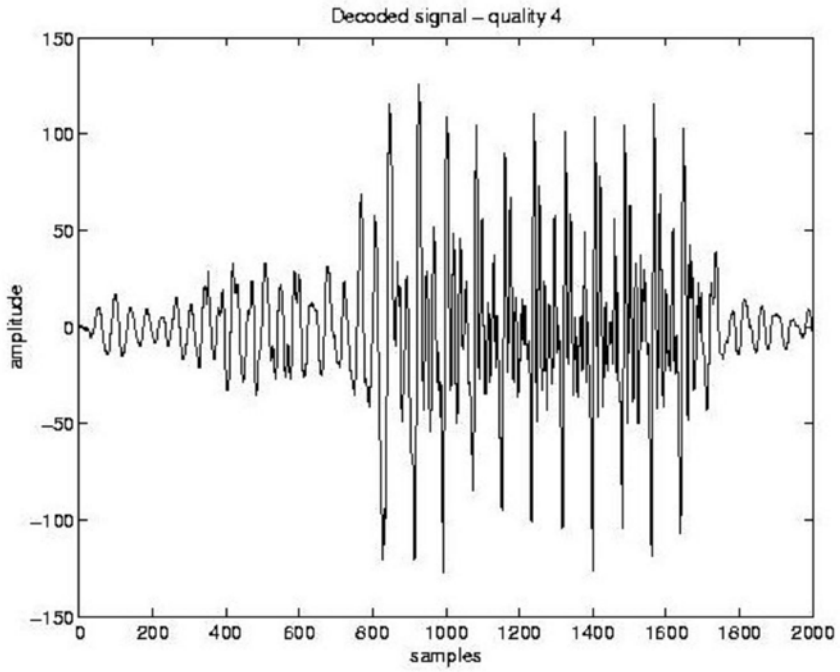

Fig. 5. The same segment as in Fig. 3 compressed with a compression ratio 2.399 bits/sample and then decompressed

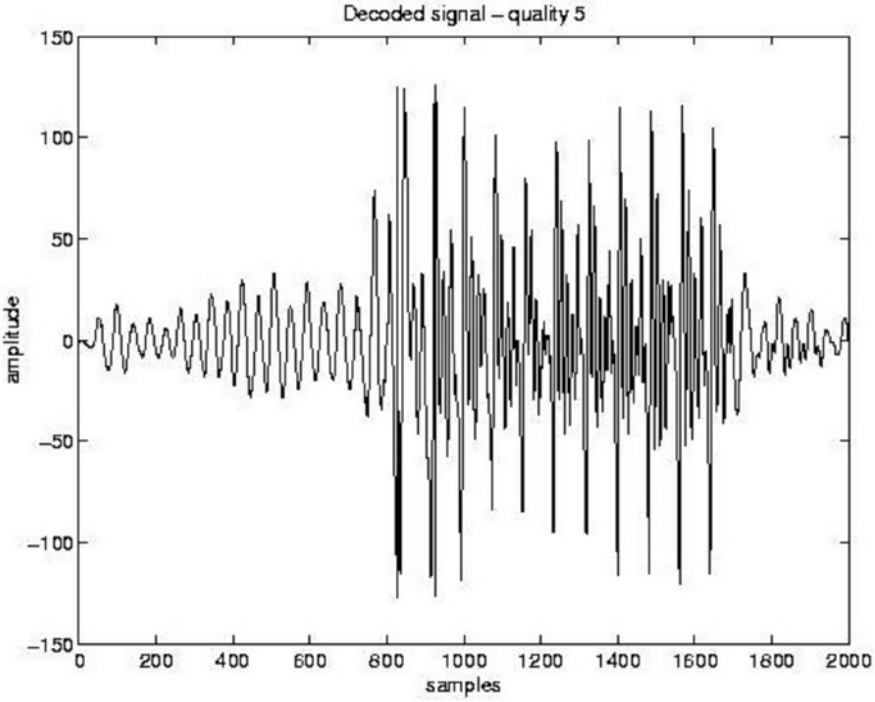

Fig. 6. The same segment as in Fig. 3 compressed with a compression ratio 3.648 bits/sample and then decompressed

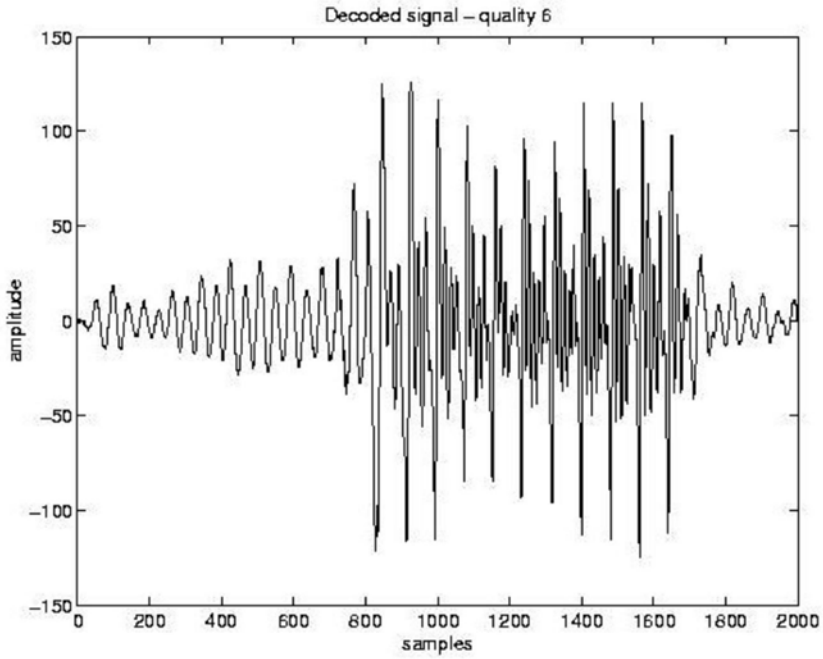

Fig. 7. The same segment as in Fig. 3 compressed with a compression ratio 4.948 bits/sample and then decompressed

for all four compression rates.

Subjective tests were also carried out on the decompressed samples. It was observed that for the LD-CELP coder, the decompressed sound samples had a background noise, which was objectionable to the listener, though a post-filter would have improved its subjective quality by removing these artifacts. For the rates up to 1bit per samples our wavelet-based coder has shows good subjective quality.

8. Conclusions

A wavelet-based speech coder has been proposed, and a program has been realized and tested. The speech signals were sampled at 8 KHz at 8 bits/sample. The proposed vocoder was compared with the G.728 standard LD-CELP vocoder. Results indicate that the proposed vocoder performed best in terms of SNR, PSNR and in terms of subjective tests.

Adaptive model to choose the number of quantization levels for each of the subbands would be warranted.

Using wavelets, the compression ratio can be easy varied while other compression schemes have fixed compression rate. This fact makes the algorithm highly useful for Internet and other bandwidth-dependent applications.

The algorithm can be further improved by incorporating the modified ZTC or other advanced quantification techniques [19] into the best-basis selection step.

References

1. Minoli, D., E. Minolu. Delivering Voice over IP Networks. John Wiley and Sons, 1998.
2. Chui, S., C. Fan. Real-time implementation of an 8kbps low-delay CELP coder. - In: Proc. of Int. Conf. Signal Processing, Applications, Technology, Dalas, October, 1994, 1588-1593.
3. Degener, J. Digital Speech Compression – Putting the GSM 06.10 RPL-LTP algorithm to work. - Dr. Dobb's Journal, December, 1994.
4. Chen, J., N. Jayant, R. Cox. Improving the performance of the 16 kbps LD-CELP speech coder. - In: Proc. IEEE Int. Conf. ASSP, San Francisco, March, 1992, 69-72.
5. Veterli, M., J. Kovacevic. Wavelets and Subband Coding. Prentice-Hall, 1995.
6. Daubechies, I. Ten lectures on wavelets. Philadelphia, SIAM, 1992.
7. Shapiro, J. Embedded image coding using zerotrees of wavelet coefficients. – In: IEEE Trans. SP, **41**, Dec. 1993.
8. Coifman, R., Y. Meyer, S. Quake, M. Wickerhauser. Signal processing and compression with packets. Preprint, Yale University.
9. Coifman, R. R., M. V. Wickerhauser. Entropy-based algorithms for best-basis selection. – In: IEEE Trans. Info. Theory Vol. **38**, 1992, 713-718.
10. Srinivasan, P., L. Jamieson. High-quality audio compression using an adaptive wavelet packet decomposition and psychoacoustic modeling. – In: IEEE Trans. SP, **46**, 1998, 1085-1093.
11. Johnston, J. Transform coding of audio signals using perceptual noise criteria. – In: IEEE J. Sel. Areas in Communications, **6**, 1988, 314-323.
12. ISO/IEC IS11172-3, Coding of Moving Pictures and Associated Audio for Digital Storage Media at u to about 1.5 Mbit/s – Part 3: Audio, 1993.
13. Aldroubi, A., M. Unser, M. Eden. B-spline signal processing". - In: IEEE Trans. SP, **41**, 1993, 821-849.
14. Gotchev, A., J. Vesma, T. Saramaki, K. Egiazarian. Modified B-spline interpolators: theory and efficient realizations. –TICSP series, **4**, to appear.
15. Fastl, H. Temporal masking effects: 2 Critical noise maskers. – Acoustica, **36**, 1976.
16. Unser, M., A. Aldroubi, M. Eden. On the asymptotic convergence of B-spline wavelets to Gabor functions. – In: IEEE Trans. IT, 1992, 38:864-872.

17. Lawton, W., S. Lee, Z. Shen. Characterization of complactly supported refinable splines. – Advances in Comp. Math., **3**, 1995, 137-145.
18. Mallat, S. Singularity detection and processing with wavelets. – In: IEEE Trans. IT, **32**, 1992, 617-643.
19. Said, A., W. Pearlman. A new fast and efficient image codec based on set partitioning in hierarchical trees. – In: IEEE Trans. Circuits, Systems, Video Tech. **6**, 1996.

Алгоритм кодирования речевого сигнала при помощи пакетных волн

Г. Марчоков, А. Гочев, З. Николов

Институт информационных технологий, 1113 София

(Резюме)

Предлагается новый алгоритм для эффективного кодирования речи. В нем реализуется пакетно-волновая декомпозиция речевого сигнала при помощи пакетных волн и на второй степени– энтропийное кодирование Huffman. Применяется модифицированное волновое кодирование с нулевым деревом, которое использует факт, что большая часть энергии речевого сигнала концентрирована в первых 4-ех нискочастотных обхватах. Так устанавливаются два порога в кодере нулевого дерева, которые разные для ниских и высокочастотных обхватах. Существенные коэффициенты кодированы при помощи ентропийного кодера нулевого порядка.

Обсуждены некоторые аспекты практического применения алгоритма. Показанные эксперименты показывают лучшее объективное и субъективное качество по сравнении с LD-CELP кодером.