3rd European Conference on
Speech Communication and Technology
EUROSPEECH'93
Berlin, Germany, September 19-23, 1993

HIGH QUALITY CODING OF WIDEBAND SPEECH AT 24 KBIT/S

Jürgen Paulus, Christiane Antweiler and Christian Gerlach

Institute for Communication Systems and Data Processing (IND)
RWTH Aachen, University of Technology
52056 Aachen, Germany
e-mail: paulus@ggind.ind.rwth-aachen.de

ABSTRACT

This paper proposes a Wideband-CELP-Coding scheme (bandwidth 7kHz) at 24 kbit/s. The codec introduces a delay of just 10 ms. This fulfills the requirements of a possible codec candidate for wideband speech coding within DECT or video applications [1].

The analysis-by-synthesis structure of the proposed Wideband-CELP-Codec includes an alternative LPC analysis concept, where the autocorrelation function is calculated recursively [2]. *This special LPC scheme provides an improved speech quality and a reduction of computational complexity in comparison to conventional algorithms for the LPC analysis. In addition a stochastic sparse codebook with extremely low computational effort is presented resulting in a neglectable amount of storage.*

The CCITT G.722 standard was applied as reference codec, in order to compare the new coding scheme in terms of subjective quality. With the proposed Wideband-CELP a speech quality is achieved, which is equivalent to the reference codec operating at 56kbit/s.

1. INTRODUCTION

Wideband speech coding techniques, i.e. coding of speech within the bandwidth of 0.05–7 kHz, are currently being studied for application to videophone, multimedia and digital mobile telephone with high speech quality.

In 1986 the CCITT recommended a standard for wideband speech coding, the CCITT G.722, providing high speech quality at 64 kbit/s. This codec uses a quadratur mirror filter (QMF) to split up the signal into two subbands. In each subband an ADPCM coding structure is used working with 6 bits for the lower and 2 bits for the upper band. The gross bitrate of 64 kbit/s can be split up according to Table 1, to allow the combined transmission of speech and data. The different transmission modes are realized by decreasing the accuracy in the lower band from 6 bits to 5 or 4 bits.

mode	speech [kbit/s]	data [kbit/s]
1	64	0
2	56	8
3	48	16

Table 1: The wideband transmission modes of G.722

During the last two years several papers were published representing wideband speech codecs operating at 32 kbit/s [3–5], 24 kbit/s [6] and 16 kbit/s [7–10]. Summarizing the published results, the quality of G.722 at 64 kbit/s can be achieved with an analysis-by-synthesis coding structure operating at 32 kbit/s. 16 kbit/s coders currently provide a slight worse performance compared to G.722 at 48 kbit/s [9].

The development of the wideband coding scheme was preceded by various investigations concerning the basic codec structure. These investigations included the option of splitting the original speech signal into sub-bands and various forms of analysis-by-synthesis schemes according to Figure 1.

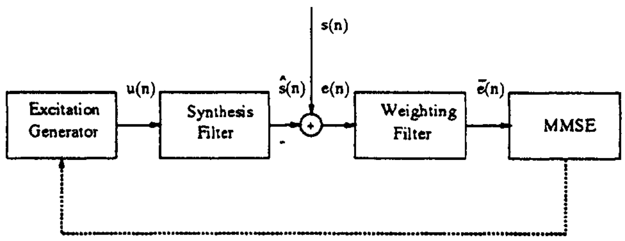

Figure 1: General analysis-by-synthesis coding structure
$s(n)$: speech signal, $\hat{s}(n)$: synthesized speech signal
$u(n)$: excitation, $e(n)$: error signal
MMSE: minimum mean square error

Especially different forms of excitation signals, regular pulse (RPE), multi pulse (MPE) or code excitation (CELP) were considered. As a result of our examinations a Wideband-CELP-Coding scheme, as full-band analysis-by-synthesis structure, provided the required high speech quality [11–13].

In the following we will present a description of the main parts of the codec as there are LPC and LTP analysis and the stochastic codebook structure. The paper concludes with simulation results and a discussion of complexity.

2. LPC ANALYSIS

The short-term linear prediction analysis is performed using the autocorrelation method in combination with the Levinson-Durbin recursion. The autocorrelation function is calculated recursively according to Barnwell [2].

During the analysis the input speech is weighted with the impulse response of a 2nd order IIR filter

$$w(z) = \frac{1}{(z-\alpha)^2} \tag{1}$$

The window is positioned such that a delay of only 10 ms (160 samples) is introduced as depicted in Figure 2. The use of conventional LPC analysis structures such as autocorrelation with windowing or stabilized covariance method requires overlapping frames to achieve high speech quality. However, these techniques also result in an increased delay. In addition, our choice of the LPC analysis method was confirmed by an improvement of the subjective quality. The applied IIR filter has a double-pole at α=0.9795, which was optimized for a given framelength according to SNR and subjective speech quality.

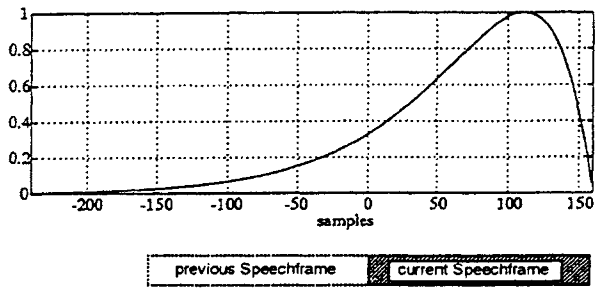

Figure 2: Frame position of 2nd order IIR filter with double-pole at α=0.9795 (normalized to its maximum)

The order of the LPC filter is 20. The quantization of the prediction coefficients is based on the arcsine of the reflection coefficients, according to [14]. This decision was taken in contrast to some papers, which propose to quantize the line spectral pairs (LSPs), for example [15–17]. Informal listening tests indicated, that for this application there is no significant advantage of LSPs, which also corresponds to the results given in [18]. In addition, the LSP method is less attractive due to its high computational effort.

Coefficient	number of Bits
1	7
2	6
3 - 5	5
6 - 14	4
15 - 18	3
19 - 20	2
	Σ = 80 (8kbit/s)

Table 2: Nonlinear quantization table of ARCSIN coefficients

Table 2 lists the bit allocation for nonlinear quantization of the LPC coefficients. The quantization levels are obtained using 1-dimensional LBG codebooks [19] for each ARCSIN filter coefficient. The speech material for training and testing the codebooks is taken out of the European Broadcasting Union (EBU) database [20] consisting of nearly 100 s of three different languages (English, German and French) including male and female speakers.

3. LTP ANALYSIS

The long-term-prediction (LTP) is carried out in a closed loop pitch search i.e. using an adaptive codebook filled with previous computed excitation signals.

The pitch lag is searched from 275 down to half of the subframelength, i.e. 20 samples. For lags lower than the full framelength the recursive search procedure [21] is utilized. The search is performed once per subframe, i.e. every 40 samples.

The fact that the LPC parameters are updated only every 4th subframe can be exploited for the adaptive codebook search in order to reduce the computational complexity. Since the adaptive codebook changes only partly from subframe to subframe and the synthesis filter remains constant for 4 subframes, intermediate results can be stored and reused for the search in the next subframe.

This measure decreases the complexity of the adaptive codebook search by about 30% and the overall complexity by about 18%.

4. CODEBOOK

A 9–bit sparse codebook with an extremely low computational effort is developed for the search of the stochastic excitation vector, comparable to the one proposed in [22], as shown in Table 3. Each vector consists of two non zero samples with unity amplitude. The vectors have different pulse spacing, and different classes are formed by simply shifting both pulses one position at a time through the vector.

basic excitation vector	spacing	number of possible vectors
1100 0000 0000 0000 0000	1	19
1010 0000 0000 0000 0000	2	18
1001 0000 0000 0000 0000	3	17
1000 1000 0000 0000 0000	4	16
1000 0100 0000 0000 0000	5	15
1000 0010 0000 0000 0000	6	14
1000 0001 0000 0000 0000	7	13
1000 0000 0100 0000 0000	9	11
1000 0000 0000 0001 0000	15	5
		Σ = 128

Table 3: Stochastic codebook structure showing basic excitation vectors and the number of possible vectors obtainable by simply shifting the excitation vectors

By this procedure a codebook with 128 different excitation vectors is built up. The codebook is then doubled in size by changing the sign of the first pulse. Using one further bit the codebook is doubled again by changing the sign of the whole codeword resulting in a codebook with in total 512 entries. Thus starting from 9 keywords, a 9–bit sparse codebook is built up. This kind of stochastic codebook ensures a fast search, because the pulse positions are known, and the different excitations can be calculated recursively. Furthermore, the codebook requires a very small storage.

5. BIT ALLOCATION AND COMPLEXITY

In the previous chapters the main components of the wideband-CELP codec were presented. They consist of an recursive LPC analysis method, LTP and a 9-bit sparse codebook with extremely low computational complexity.

These components build up a codec resulting in a final bitrate of 24 kbit/s, according to Table 4.

Parameter		Bits	Update time [ms]	Bitrate [kbit/s]
LPC	coefficients	80	10	8.0
LTP	delay	8	2.5	3.2
	gain	4	2.5	1.6
CB	index	9	1.25	7.2
	gain	5	1.25	4.0
				$\Sigma = 24.0$

Table 4: Bit allocation for the proposed wideband-CELP-codec

The LPC coefficients are updated every 10 ms. Consequently fast changes in the spectral behaviour of the speech signal can be tracked.

The LTP analysis, i.e the calculation of the adaptive codeword, is performed four times during the LPC intervall.

The stochastic excitation is updated every 1.25 ms, resulting in short excitation vectors of only 20 samples. Figure 3 provides an overview of the different update rates.

Table 5 summarizes the complexity and bitrate distribution in percent. As a result of the high update rate for the stochastic excitation the codebook search requires nearly half of the bitrate, while it takes 12% of CPU-time. The codec is implemented in floating point using the program language C on a SUN Sparc station ELC.

Routine	Complexity [%] (CPU-time)	Bitrate [%]
LPC analysis	11	33
LTP analysis	77	20
Codebook search	12	47

Table 5: Complexity distribution of the main codec routines in terms of CPU-time (SUN Sparc station ELC) and bitrate

In spite of the computational reduction introduced in the adaptive codebook search, it still consumes most of the simulation time to code 20 % of the main bitrate. Further improvements are necessary to reduce the computational complexity of the LTP.

6. SIMULATION RESULTS

The presented codec was tested with speech material taken out of the European Broadcasting Union database [20]. This material consists of various languages (English, German and French), each with male and female speakers.

With the low delay (recursive) LPC analysis, an improved speech quality compared to conventional LPC analysis algorithms was obtained. In addition this concept provides a reduction of computational complexity [2].

The extension of the pitch lag of the adaptive codebook down to half of the subframe-length resulted in a better quality of the higher frequency components. This is especially audible with female speakers.

As a result of the simulation it can be stated that speech material of male speakers could be better reconstructed than female speakers. The obtained speech quality for these sentences was judged in an informal subjective listening test to be equivalent to the CCITT G.722 wideband codec operating at 56 kbit/s.

7. CONCLUSION

In this paper a wideband-CELP-coding scheme working at 24 kbit/s has been presented. The achieved speech quality was judged to be equivalent to the CCITT G.722 wideband codec operating at 56 kbit/s. With this bitrate the codec fulfills the requirements of a possible codec candidate for wideband speech coding within DECT or video applications [1].

ACKNOWLEDGEMENT

This work has been supported by Philips Kommunikations Industrie AG Nürnberg, Germany. The authors would like to thank Karl Hellwig for stimulating and fruitful discussions.

REFERENCES

[1] ETSI/STC-TM5-(90)04, "Coding, Speech Processing and associated Network Issues," tech. rep., ETSI Sub Technical Commitee TM5, September 1990.

[2] T. Barnwell, "Recursive Windowing for Generating Autocorrelation Coefficients for LPC Analysis," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 29, No. 5, pp. 1062–1066, October 1981.

[3] S. Quackenbush, "A 7 kHz Bandwidth, 32kbps Speech Coder for ISDN," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, Toronto, Canada, 1991, pp.1–4, S1.1.

[4] E. Ordentlich and Y. Shoham, "Low-Delay Code-Excited Linear-Predictive Coding of Wideband Speech at 32 kbps," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, Toronto, Canada, 1991, pp.9–12, S1.3.

[5] V. Sanchez-Calle, C. Laflamme, R. Salami, and J. Adoul, "Low-Delay Algebraic CELP Coding of Wideband Speech," in *Signal Processing VI, Proc. of EUSIPCO*, Brussels, Belgium, 1992, pp. 495–498.

[6] A. Le_Guyader, B. Lozach, and N. Moreau, "Embedded Algebraic CELP Coders for Wideband Speech Coding," in *Signal Processing VI, Proc. of EUSIPCO*, Brussels, Belgium, 1992, pp. 527–530.

[7] C. Laflamme, J. Adoul, R. Salami, S. Morisette, and P. Mabilleau, "16 kbps Wideband Speech Coding Technique Based on

Algebraic CELP," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, Toronto, Canada, 1991, pp. 13–16, S1.4.

[8] G. Roy and P. Kabal, "Wideband CELP Speech Coding at 16 kbit/sec," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, Toronto, Canada, 1991, pp. 17–20, S1.5.

[9] A. Fuldseth, E. Harborg, F. Johansen, and J. Knudsen, "A Real-Time Implementable 7kHz Speech Coder at 16kbit/s," in *IEEE Workshop on Speech Coding*, Whistler, British Columbia, Canada, 1991.

[10] R. De_Iacovo, R. Montagna, and D. Sereno, "A Low-Delay Wideband Speech Codec at 16 kbit/s," in *Signal Processing VI, Proc. of EUSIPCO*, Brussels, Belgium, 1992, pp. 483–485.

[11] F. Rodewald, "Entwicklung eines breitbandigen Sprachcodecs mit gehörbezogener Fehlergewichtung," *diploma thesis*, Institut für Nachrichtengeräte und Datenverarbeitung IND, RWTH Aachen, May 1992.

[12] J. Schröer, "Untersuchungen zur Breitbandcodierung mittels CELP-Verfahren," *diploma thesis*, Institut für Nachrichtengeräte und Datenverarbeitung IND, RWTH Aachen, July 1992.

[13] J. Paulus, "WISP-Wideband Speech Coding at 16 to 32 kBit/s," *Internal Report*, Institut für Nachrichtengeräte und Datenverarbeitung IND, RWTH Aachen, December 1992.

[14] A. Gray and J. Markel, "Quantization and Bit Allocation in Speech Processing," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 24, No. 6, pp. 459–473, December 1976.

[15] F. Soong and B. Juang, "Line Spectrum Pair (LSP) and Speech Data Compression," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, San Diego, CA, 1984, pp. 1.10.1–1.10.4.

[16] G. Kang and L. Fransen, "Low-Bit Rate Speech Encoders Based on Line-Spectrum Frequencies (LSFs)," tech. rep., Naval Research Laboratory, January 24, 1985.

[17] S. Saoudi, J. Boucher, and A. L. Guyader, "Optimal Scalar Quantization of the LSP and the LAR for Speech Coding," in *Proc. Int. Conf. Spoken Language Processing*, Kobe, Japan, 1990, pp. 101–104.

[18] B. Atal, R. Cox, and P. Kroon, "Spectral Quantization and Interpolation for CELP Coders," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, Glasgow, Scotland, 1989, pp. 69–72.

[19] Y. Linde, A. Buzo, and R. Gray, "An Algorithm for Vector Quantizer Design," *IEEE Trans. Communications*, vol. 28, No. 1, pp. 84–95, January 1980.

[20] European Broadcasting Union (EBU), "*Sound Quality Assesment Material (Recordings for Subjective Test)*, no. 422 204–2 ed.

[21] W. Kleijn, D. Krasinski, and R. Ketchum, "Analysis and Improvement of the Vector Quantization in SELP," in *Signal Processing IV, Proc. of EUSIPCO*, Grenoble, France, 1988, pp. 1043–1046.

[22] L. Cellario, G. Ferraris, and D. Sereno, "A 2 ms Delay CELP Coder," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, Glasgow, Scotland, 1989, pp. 73–76.

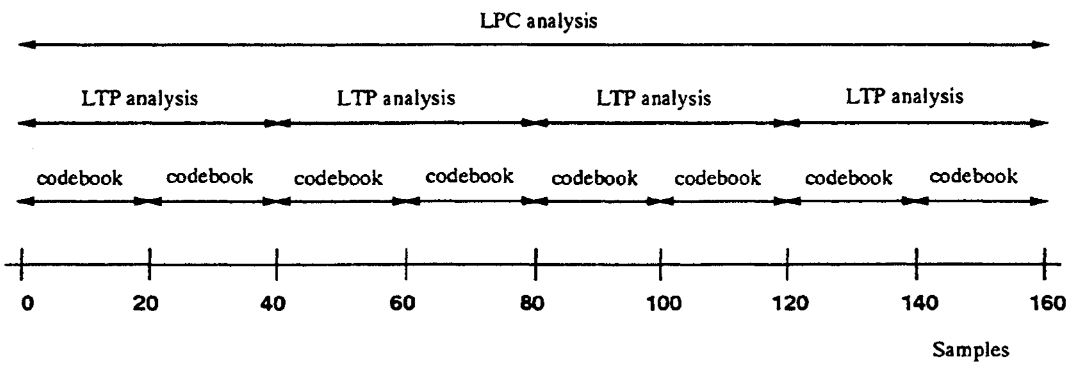

Figure 3: Update of the codec parameters