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

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, comparable to the one presented in [12] 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 56 kbit/s.

1. INTRODUCTION

Wideband speech coding techniques, i.e. coding of wideband 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. During the last two years several papers were published representing wideband speech codecs operating at bitrates between 16 kbit/s and 32 kbit/s [3, 4, 5, 6, 7]. Summarizing the published results, the quality of the CCITT standard G.722 can be achieved with an analysis-by-synthesis coding structure operating at 32 kbit/s. Compared to G.722 at 48 kbit/s currently 16 kbit/s coders do not provide the same performance.

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. 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 best speech quality at 56 kbit/s .

In the following we will present a description of the main parts of the codec. The paper conludes with simulation results and a discussion of complexity.

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.

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 as proposed by Barnwell [2]. During the analysis the input speech is weighted with the reversed impulse response of a 2nd order IIR filter having the z-transform

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

The window is positioned such that a delay of only 10ms (160 samples, f_s=16kHz) is introduced. In comparison to conventional block-oriented LPC analysis techniques this approach does not only reduce delay but it also provides 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 in combination with subjective speech quality. The order of the LPC filter is 20. The quantization of the prediction coefficients is based on the arcsin of the reflection coefficients, according to [8]. This decision was taken in contrast to some papers, which propose to quantize the line spectral pairs (LSPs). Informal listening tests indicated, no significant advantage of LSPs, which also correspond to the results given in [9] for the narrowband case. In addition, the LSP method is less attractive due to its high computational effort.

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 sampling periods down to half of the subframelength, i.e. 20 samples. For lags lower than the full framelength the recursive search procedure [11] 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 [12]. Each of 9 basic excitation vectors consists of two non zero samples with unity amplitude and different pulse spacing. Different classes are formed by simply shifting both pulses one position at a time through the vector. By this procedure a codebook with 128 different excitation vectors is build 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 build 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

The contribution of the different blocks of the codec to the final bitrate of 24kbit/s is given in Table 1.

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 1: Bit allocation of the proposed wideband-CELP-codec

Table 2 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 only. The codec was implemented in floating point using the program language C on a SUN Sparc station ELC. In spite of the computational reduction introduced in the adaptive codebook search, it still consumes most of the simulation time. Further improvements are necessary to reduce the computational complexity of the LTP.

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

Table 2: Complexity distribution of the main codec routines in terms of CPU-time (SUN Sparc station ELC) and bitrate

6. SIMULATION RESULTS

The presented codec was tested with speech material taken from the European Broadcasting Union database [10]. The material consists of nearly 100 sec 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.

The extension of the pitch lag of the adaptive codebook down to half of the subframelength resulted in a better quality of the higher frequency components. This is especially audible with female speakers. As a result of the simulations 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. 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] 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, Bruessels, Belgium*, 1992, pp. 483–485.

[4] 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.

[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] 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.

[7] 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.

[8] 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.

[9] 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.

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

[11] 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.

[12] 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.