Speech Codec for the European Mobile Radio System

S6.1

P. Vary, K. Hellwig, R. Hofmann
Philips
Kommunikations Industrie AG
Nürnberg, Fed. Rep. of Germany

R.J. Sluyter
Philips
Research Laboratories
Eindhoven, The Netherlands

C. Galand, M. Rosso
IBM
Research Laboratories
La Gaude, France

ABSTRACT

In 1991 a digital mobile radio system will be introduced in Europe. The speech codec to be used as the standard will be presented. The coding scheme which has been selected by the CEPT Groupe-Speciale-Mobile (GSM) as a result of formal subjective listening tests, is based on the Regular-Pulse Excitation LPC technique (RPE-LPC) combined with Long-Term Prediction (LTP). The so-called RPE-LTP codec has a net bit rate of 13 kbit/s. The algorithm and the experimental implementations based on different VLSI signal processors will be decribed and demonstrated by tape recordings.

1. INTRODUCTION

In the context of the standardisation of the future Pan-European digital mobile radio system, the CEPT Groupe Speciale Mobile (GSM) has carried out recently subjective codec tests using different proposals /1/. As a result of these tests, a Regular-Pulse Excitation LPC scheme was preselected /2/ on the basis that it produced the highest average speech quality. A conclusion from these tests was that the speech quality during transmission with errors could be even further improved. This led to the 'Franco-German Compromise Codec' which combines the features of the preselected German proposal /2/ with features of the French proposal /3/. The resulting RPE-LTP codec /4/ will be recommended as the final CEPT standard.

The basic RPE-scheme is related firstly to the well known baseband RELP-codec (RELP = Residual Excited Linear Prediction, /5/) and secondly to the Multi-Pulse-Excitation-LPC technique (MPE-LPC, /6/). The advantage of the baseband RELP-codec is its relative low complexity whereas the speech quality is limited due to the tonal noise which is introduced by the process of high frequency regeneration. In contrast to this, the MPE-LPC technique produces excellent speech quality but the complexity is rather high. A compromise between both techniques is the Regular-Pulse Excitation LPC (RPE-LPC, /7/), especially in its simplified version (/7/, /2/).

2. GSM CODEC TEST /1/

Initially, more than 20 different codec proposals from 9 European countries were under consideration. For the international formal listening tests this number was reduced by national tests to 6 codecs from 6 countries. The gross bit rate including error protection was 16 kbit/s. After an initial evaluation of the test results, two sub-band codecs were withdrawn. Thus, in the final test evaluation two sub-band codecs and two residual excited LPC codecs were included:

RPE-LPC: Regular-Pulse Excitation
Linear Predictive Coding
Fed. Rep. of Germany / Philips

MPE-LTP: Multi-Pulse Excitation
Long-Term Prediction
France / IBM

SBC-APCM: Sub-Band Coding
block adaptive PCM in 14 sub-bands
Sweden / ELLEMTEL

SBC-ADPCM Sub-Band Coding
adapt. different. PCM in 6 sub-bands
England / British Telecom Research

The codecs were tested in 7 languages under various transmission conditions:

* 3 input levels: 12, 22, 32 dB below overload
* 3 bit error rates: 0, 1%, 0.1% (random)
* 1 and 2 transcodings
* 2 different forms of environmental noise

Several reference conditions were included, such as a companded FM with carrier to noise ratios of 18 and 26 dB combined with simulated fading at a vehicle speed of 10 m/s (36 km/h).

The average speech quality taken over all test conditions in terms of the five-point mean opinion score (MOS), the net bit rates, and the computational complexities are given in table 1.

Codec	Speech Quality	Net Bit Rate kbit/s	Complexity MOPs/s
RPE-LPC	3.54	14.77	1.5
MPE-LTP	3.27	13.20	4.9
SBC-APCM	3.14	13.00	1.5
SBC-ADPCM	2.92	15.00	1.9
FM	1.95		

Table 1: Result of the codec test
Quality: 1 = bad 5 = excellent

It is noted that each of the codecs exceeded the speech quality of the analogue reference system (FM). The RPE-LPC codec obtained the highest average quality score. The analysis of the individual test conditions revealed that the quality advantage of the RPE codec decreased with an increasing bit error rate while the quality of the MPE-LTP codec was not very much affected. This motivated the 'Franco-German Codec Team' to investigate possibilities of combining the features of the RPE-LPC codec /2/ with features of the MPE-LTP codec /3/. The most important modification was the addition of a long-term prediction loop (LTP). The result was that the net bit rate could be reduced from 14.77 kbit/s to 13.0 kbit/s while maintaining the same level of quality. This increased the bit-rate capacity to improve error protection. The modified scheme is called RPE-LTP codec /4/.

3. BLOCK DIAGRAM OF THE ENCODER

The block diagram of the encoder is subdivided in 5 sections as shown in figure 1 /9/.

Preprocessing: Offset compensation is applied, to prevent a DC component being translated into an annoying side tone by the process of high frequency regeneration in the decoder. A first order FIR preemphasis filter is used for numerical reasons.

LPC Analysis: In the segmentation buffer, the speech signal is divided into non-overlapping segments having a length of T_0 = 20 ms (160 samples). A new LPC analysis is performed for each segment by calculating 8 reflection coefficients r(i) using the Schur recursion algorithm /8/. Due to the favourable quantisation characteristics, the reflection coefficients are converted into Log.Area Ratios (LARs). A piecewise linear approximation is utilised. Due to their different dynamic ranges and different asymmetric amplitude distributions, the transformed coefficients LAR(i) have different limits and are quantised uniformly as shown in table 2.

LAR No. i	1&2	3&4	5&6	7&8
bits/LAR	6	5	4	3

Table 2: Bit assignment of the LAR coefficients

Short-Term Analysis Filtering: The 8 coefficients of the short-term analysis filter are preprocessed as follows: First, the quantised and coded Log.-Area-Ratios are decoded. Then the most recent and the previous set of LAR coefficients are interpolated linearly within a transition period of 5 ms to avoid spurious transients. Finally, the interpolated Log.-Area Ratios are reconverted into the coefficients r'(i) of the FIR lattice filter.

The computation cycle outlined so far is repeated every 20 ms and produces 160 samples of the prediction error signal d.

Long-Term Predictor Loop: The LTP loop is used to compute the estimate d" of the residual signal d from the reconstructed excitation signal e'. The LTP filter is characterised by the gain b and the delay N according to

$$d''(k) = b' \cdot d'(k-N) \qquad (1)$$

where b' denotes the quantised versions of b. The parameters b and N are calculated every 5 ms (40 samples). Each segment of the 160 samples of the residual d, beginning with $d(k_0)$, is subdivided into four sub-segments $d(k_0+j\cdot 40+i)$ (j=0,1,2,3; i=0,..39). Then the cross-correlation functions $R(\lambda)$ is calculated according to

$$R(\lambda) = \sum_{i=0}^{39} d(k_j+i)\cdot d'(k_j+i-\lambda); \quad k_j=k_0+j\cdot 40 \quad j=0,1,2,3 \quad \lambda=40..120 \qquad (2)$$

The optimum delay value N then is searched, for which this function has its maximum

$$R(N) = \max \{ R(\lambda);\ \lambda = 40...120 \}. \qquad (3)$$

Due to the lower limit of λ = 40, N does not necessarily correspond to one pitch period of the speech signal, but at least to a multiple of this period. The long-term predictor gain b for the j-th sub-segment is calculated as

$$b = R(N) / \sum_{i=0}^{39} d'^2(k_j+i-N). \qquad (4)$$

The LTP parameters b and N are encoded with 2 and 7 bits, respectively.

RPE Encoding: A FIR 'block filter' algorithm is applied to each sub-segment of 40 samples of the second residual signal e. Only the 40 centrical samples of the convolution of the 40 input samples with the 11-tap impulse response are calculated. For notational convenience the block filtered version of each sub-segment is denoted by x(k), k=0...39. For the next step the filtered signal x is down-sampled by a ratio of 3 resulting in three interleaved sequences of lengths 14, 13, and 13, which are divided again into 4 sequences xm of length 13

$$x_m(i) = x(k_j+m+3\cdot i); \qquad m = 0,1,2,3 \quad i = 0,1..12 \quad k_j = k_0 + j\cdot 39 \qquad (5)$$

with k_j defining the beginning of the j-th sub-segment and with m denoting the phase of the decimation grid. The optimum candidate sequence $x_M(i)$ is selected being the one with the maximum energy /7,2/

$$E = \max_m \sum_{i=0}^{39} x_m^2(i); \qquad m=0,1,2,3. \qquad (6)$$

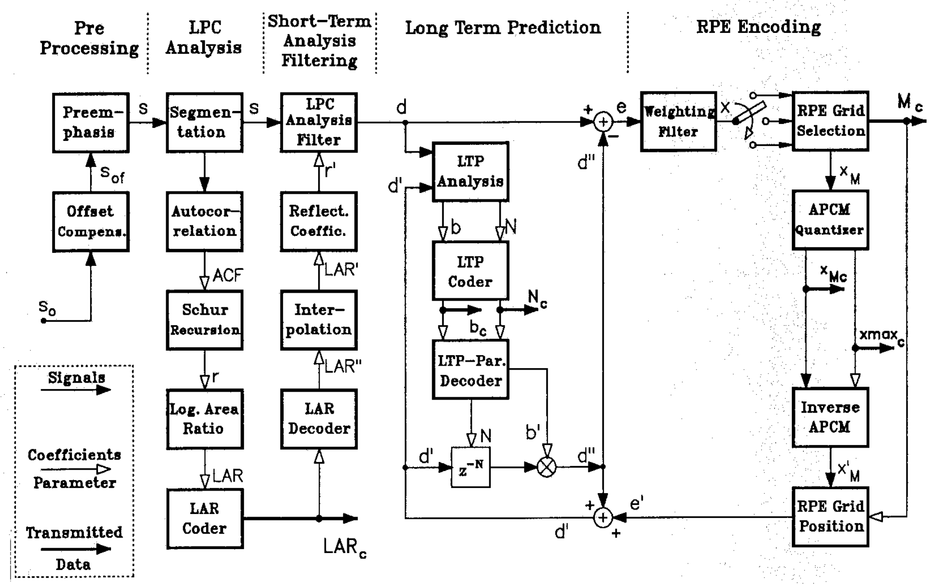

Figure 1: Block diagram of the RPE-LTP encoder (b_c, b' = coded and quantised version of b, respectively)

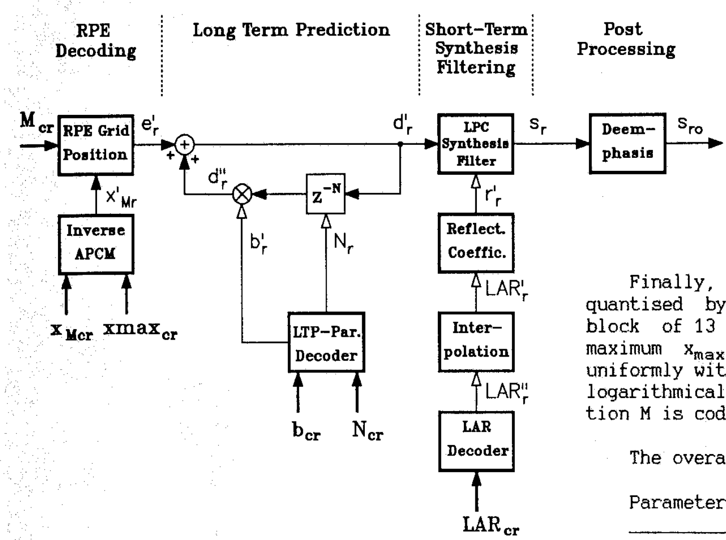

Figure 2: Block diagram of the RPE-LTP decoder (br' = received/quantised version of b)

Thus the explicit solution of the RPE-approximation of the prediction error signal e requires only energy calculations and can be interpreted as a generalisation of the sample rate decimation process in a baseband RELP-coder.

Finally, the selected RPE-sequence $x_M(i)$ is quantised by block adaptive PCM (APCM). Each block of 13 samples is normalised by its block maximum x_{max}. The samples are then quantised uniformly with 3 bits, the block maximum is coded logarithmically with 6 bits, and the grid position M is coded with 2 bits.

The overall bit allocation is given in tab. 3

Parameter		Number of bits
8 LPC coefficients	LAR(i)	36
4 LTP coefficients	b	8
4 LTP delays	N	28
4 RPE grids	M	8
4 block maxima	x_{max}	24
52 RPE samples	x_M	156
bits per frame (20 ms)		260

Table 3: Bit allocation (bit rate = 13.0 kbit/s)

4. BLOCK DIAGRAM OF THE DECODER

The decoder is shown in fig. 2. The RPE parameters M_{cr}, x_{Mcr}, and x_{maxcr} are decoded and used to reconstruct the excitation e_r' of the long-term synthesis filter which produces the excitation signal d'_r for the short-term synthesis filter. The sample rate of the denormalised RPE samples x_{Mr}' is increased by a factor of 3 by inserting zero samples and by placing the non-zero samples in the correct temporal grid position M.

5. ERROR PROTECTION

Due to the radio transmission scheme utilising time division multiplex, GMSK and (optionally) frequency hopping, the gross bit rate including error protection is 22.8 kbit/s. At the time when this paper was written, different versions of the error protection scheme were in the discussion. Convolutional codes and Viterbi decoding will be used. The frame of 260 bits produced every 20 ms by the speech encoder will probably be sub-divided into 3 classes according to the different bit error sensitivities. These three classes will be protected differently. From preliminary listening tests based on simulations of the error protection schemes, it can be concluded that the RPE-LTP codec is fairly tolerant to errors. At a bit error rate on the radio channel of 5%, no severe degradation of the speech quality was recognised.

6. IMPLEMENTATIONS

The codec algorithm will be specified on bit level as an obligatory recommendation. Meanwhile, several implementations have been completed, using different 16 bit VLSI signal processors as shown in table 4.

processor: implementation by:	PCB5011 Philips	DSP-16 IBM	TMS320C25 CNET /10/
instruction cycle	125 ns	100 ns	100 ns
computational load	60 %	40 %	45 %
program memory	2 K	2 K	3 K
external data RAM	1 K	1 K	1 K

Table 4: Comparison of the codec implementation

For the implementation using the Philips signal processor PCB5011, a signal delay of only 28 ms has been measured.

The DSP-16 is an IBM proprietary processor which has been described in /11/.

7. CONCLUSIONS

The specification of the speech codec to be used as the standard in the Pan-European digital mobile radio system is almost complete. The net and the gross bit rates are 13 and 22.8 kbit/s, respectively. The speech quality is far superior than that obtained from present day analogue mobile radio systems. The speech coding algorithm can be implemented using just a single 16 bit VLSI signal processor with external data memory of about 1Kx16 bits.

Acknowledgements: Many colleagues have contributed to this work. The authors would like to mention especially H. Bauer, W. Koch, D. Lorenz, P. Patrick (Philips, Nuernberg), D. Aubert, Ph. Elie, J. Paturet (IBM, La Gaude), H.J. Braun (FTZ, Darmstadt), P. Combescure, J.P. Petit, and Mdm. D. Massaloux (CNET, Lannion).

References

/1/ J. E. Natvig: Speech Coding in the Pan-European Digital Mobile Radio System. Special Issue 'SPEECH COMMUNICATION' No. 1,1988

/2/ K.Hellwig, R.Hofmann R.J.Sluyter, P.Vary: MATS-D Speech Codec: Regular-Pulse Excitation LPC. Special Issue 'SPEECH COMMUNICATION' on Medium Bit Rate Coding, No. 1, 1988

/3/ C. Galand, E. Lancon, M. Rosso, Ph. Elie: MPE/LTP Algorithm for Mobile Radio Applications. 'SPEECH COMMUNICATION', No. 1, 1988

/4/ P. Vary, R.J. Sluyter, C. Galand, M. Rosso: RPE-LPC Codec - The Candidate for the GSM Radio Communication System. International Conf. on Digital Land Mobile Radio Communic. 30.06.-03.07.1987, Venice, Italy, pp 507-516

/5/ C.K. Un, D.T. Magill: The Residual-Excited Linear Prediction Vocoder with Transmission Rate Below 9.6 kBit/s. IEEE Trans. COM-23, Dec. 1975, pp. 1466-1474

/6/ B.S. Atal, J.R. Remde: A New Model of LPC Excitation for Reproducing Natural-Sounding Speech at Low Bit Rates. ICASSP-82, 614-617

/7/ P. Kroon, E.F. Deprettere, R.J. Sluyter: Regular-Pulse Excitation: a Novel Approach to Effective and Efficient Multi-Pulse Coding of Speech. IEEE Transact. ASSP, October 1986, pp. 1054-1063

/8/ J. Schur: Über Potenzreihen, die im Innern des Einheitskreises beschränkt sind. Journal für die reine und angewandte Mathematik, 1917, vol. 147, pp.205-232

/9/ GSM Recommendation 06.10: Full rate speech encoding and decoding.

/10/ P. Combescure, J.P. Petit: personal communication . December 1987

/11/ C. Galand, C. Couturier, G. Platel, R. Vermot-Gauchy: Voice excited predictive coder (VEPC) implementation on a 10 MIPS signal processor. IBM Journal of Research and Development, March 1985