GSM ENHANCED FULL RATE SPEECH CODEC

K. Järvinen, J. Vainio, P. Kapanen, T. Honkanen, P. Haavisto
Nokia Research Center, Tampere, Finland
kari.jarvinen@research.nokia.com

R. Salami, C. Laflamme, J-P. Adoul
University of Sherbrooke, Sherbrooke, Canada

ABSTRACT

This paper describes the GSM enhanced full rate (EFR) speech codec that has been standardised for the GSM mobile communication system. The GSM EFR codec has been jointly developed by Nokia and University of Sherbrooke. It provides speech quality at least equivalent to that of a wireline telephony reference (32 kbit/s ADPCM). The EFR codec uses 12.2 kbit/s for speech coding and 10.6 kbit/s for error protection. Speech coding is based on the ACELP algorithm (Algebraic Code Excited Linear Prediction). The codec provides substantial quality improvement compared to the existing GSM full rate and half rate codecs. The old GSM codecs lack behind wireline quality even in error-free channel conditions, while the EFR codec provides wireline quality not only for error-free conditions but also for the most typical error conditions. With the EFR codec, wireline quality is also sustained in the presence of background noise and in tandem connections (mobile to mobile calls).

1. INTRODUCTION

The background for introducing wireline speech quality to GSM is the increasing use of the GSM system in communications environments where it competes with fixed or cordless systems. To be competitive also with respect to speech quality, GSM must provide wireline speech quality which is robust to typical usage conditions such as background noise and transmission errors.

The standardisation of an enhanced full rate (EFR) codec for GSM started in European Telecommunications Standards Institute (ETSI) in 1994 with a pre-study phase. The pre-study phase was undertaken to set essential requirements for the EFR codec and to assess the technical feasibility of meeting them. During the pre-study phase, wireline speech quality was set as a development target for the EFR codec [1]. Wireline quality (with ITU-T G.728 16 kbit/s LD-CELP as a reference codec) was required not only for error-free transmission, but also in low error-rate conditions (C/I=13 dB) as well as in background noise (error-free conditions). Wireline performance was required also for speaker independence and for speaker recognisability. For more severe error conditions (C/I=10 dB and C/I=7 dB) significant improvement to the existing GSM full rate (FR) codec was required. In extreme error conditions (C/I<7 dB) the requirement was to provide graceful degradation without annoying effects. In error-free self-tandem (mobile to mobile calls) the EFR codec should perform equal to G.728 in tandem. In erroneous tandem at C/I=10 dB, the EFR codec was required to perform significantly better than the FR codec.

In addition, essential requirements were set for bit-rate, complexity, and delay. The same channel bit-rate of 22.8 kbit/s was required as used in the existing FR codec. The complexity was not to exceed the complexity of the GSM half rate codec. The requirement for one way end-to-end delay was to be no more than in the FR channel.

A competitive EFR codec selection process was launched in ETSI in 1995. Altogether six EFR candidate codecs were submitted into the first phase of testing (pre-selection tests) which started in August 1995. Based on the pre-selection test results and also results from complementing verification tests the EFR candidate codec jointly developed by Nokia and University of Sherbrooke (USH) was selected for GSM in October 1995. A few months earlier the same codec had been chosen as the EFR codec for the US PCS 1900 system which is based on the GSM technology. The advantage of using the same codec in PCS 1900 and in GSM was one more factor in favour of this particular solution. During 1996, verification tests for the EFR codec have been completed and the GSM specifications have been finalised.

2. SPEECH CODEC

The GSM EFR speech codec is based on the ACELP algorithm (Algebraic Code Excited Linear Prediction) [2]. The speech coding (source coding) bit-rate is 12.2 kbit/s. For channel coding (error protection) 10.6 kbit/s is used resulting in 22.8 kbit/s channel bit-rate. The EFR codec employs 0.8 kbit/s more error protection than the FR codec where speech coding bit-rate is 13.0 kbit/s. The bit allocation of the GSM EFR codec is shown in Table I. A block diagram of the encoder is shown in Figure 1.

Parameter	1st and 3rd subframe	2nd and 4th subframe	Total per frame
2 LSP sets			38
ACB index (lag)	9	6	30
ACB gain	4	4	16
FCB pulses	35	35	140
FCB gain	5	5	20
Total			244

Table I: The parameter bit-allocation of the GSM EFR codec.

The EFR codec operates on 20 ms speech frames which are divided into four 5 ms subframes. In the encoder, the speech signal is analysed and the parameters of the CELP speech synthesis model are extracted. Two sets of linear prediction filter coefficients are calculated for each frame. The indices for the adaptive (ACB) and fixed codebooks (FCB) as well as their gains are searched for each subframe. In the decoder, a spectral post-filter is used for quality enhancement.

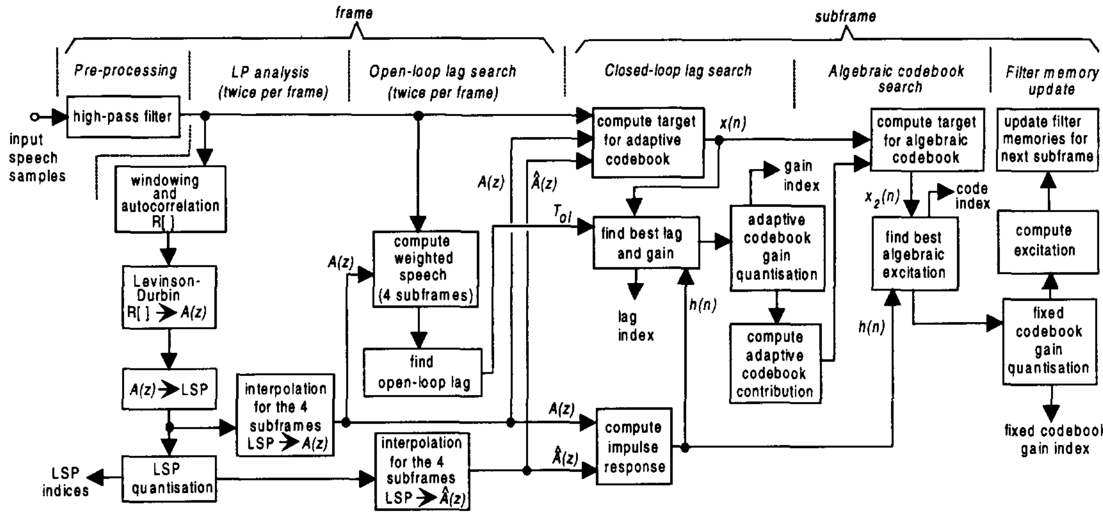

Figure 1: Block diagram of the GSM EFR encoder.

2.1 Linear Prediction

A 10th order linear prediction (LP) analysis is carried out twice for each 20 ms frame using two different asymmetric windows of length 30 ms. Both LP analyses are performed for the same set of speech samples without using any samples from future frames (no lookahead). The two sets of LP parameters are converted into line spectrum pairs (LSP). First order moving average prediction is used for both LSP sets. The LSP residual vectors are jointly quantised using split matrix quantisation (SMQ) with 5 submatrices of dimension 2x2 (two elements from both sets). The submatrices are quantised with 7, 8, 9, 8, and 6 bits, respectively. A total of 38 bits are used for LSP quantisation. For the 1st and 3rd subframes, LP parameters interpolated from the adjacent subframes are used in the codec.

2.2 Pitch Analysis

The adaptive codebook is searched for the lag range [17 3/6, 143] with a combined open-loop/closed-loop search [2]. A fractional lag with 1/6th resolution is used for lag values below 95 in the 1st and 3rd subframes and for all lag values in the 2nd and 4th subframes. The codebook search consists of the following steps:

- An open-loop search for integer lag values is carried out once every 10 ms from the weighted original speech. Small lag values are preferred to avoid pitch multiples.
- A closed-loop search for integer lag values is performed on subframe basis. For the 1st and 3rd subframe the search is carried out in the vicinity of the found open-loop lag [T_{ol} - 3, T_{ol} + 3] and for the 2nd and 4th subframe in the vicinity of the lag found for the previous subframe [T_{ps} - 5, T_{ps} + 4].
- Fractions are searched around the closed-loop lag if it is less than 95 (and always in the 2nd and 4th subframes).

The lag is quantised with 9 bits for the 1st and 3rd subframes and with 6 bits for the other two subframes where the lags are coded differentially. The codebook gain is quantised to 4 bits.

2.3 Fixed Codebook

An algebraic codebook with 35 bits is used as the fixed codebook. Each excitation vector contains 10 non-zero pulses, with amplitudes +1 or -1. The 40 positions in each subframe are divided into 5 tracks where each track contains two pulses. In the design, the two pulses for each track may overlap resulting in a single pulse with amplitude +2 or -2. The allowed positions for pulses are shown in Table II.

Track	Pulses	Positions
1	*p0, p1*	0, 5, 10, 15, 20, 25, 30, 35
2	*p2, p3*	1, 6, 11, 16, 21, 26, 31, 36
3	*p4, p5*	2, 7, 12, 17, 22, 27, 32, 37
4	*p6, p7*	3, 8, 13, 18, 23, 28, 33, 38
5	*p8, p9*	4, 9, 14, 19, 24, 29, 34, 39

Table II: Allowed pulse positions for each track.

The optimal pulse positions are determined using a non-exhaustive analysis-by-synthesis search:

- For each of the five tracks, the pulse positions with maximum absolute values of the sum of normalised backward filtered target and normalised long-term prediction residual are searched. From these the global maximum value for all the pulse positions is selected. The first pulse *p0* is always set in the position corresponding to the global maximum value.
- Five iterations are carried out in which the position of pulse *p1* is set to one of the five track maxima. The rest of the pulses are searched in pairs by sequentially searching each of the pulse pairs {*p2, p3*}, {*p4, p5*}, {*p6, p7*}, {*p8, p9*} in nested loops. For each iteration, all 9 initial pulse positions are cyclically shifted so that the pulse pairs are changed and the pulse *p1* is placed at a local maximum of a different track. The rest of the pulses are searched also for the other positions in the tracks. In the search, at least one pulse position is located corresponding to the global maximum and one pulse is located in a position corresponding to one of the 5 local maxima.

For subframes with a lag less than the subframe length, adaptive pitch sharpening filter is used. The two pulse positions in each track are encoded with 6 bits and the sign of the first pulse in each track is encoded with one bit. The fixed codebook gain is coded using moving average prediction and quantised to 5 bits.

2.4 Error Concealment

Error concealment in the EFR codec is based on partially replacing the parameters of the received bad frame with values extrapolated from the previous good frames:

- The LSPs of the previous good frame are used but shifted towards their mean values.
- The codebook gains are replaced by attenuated values derived from the previous subframes using median filtering. The amount of attenuation is different for the two codebooks and depends on which state the error concealment is in. The lag values are replaced by the lag value from the 4th subframe of the previous good frame.
- The received excitation pulses of the fixed codebook are used as such.

In case a good frame is preceded by a bad frame, the codebook gains for the good frame are limited below values used for the last good subframe.

3. CHANNEL CODEC

The EFR channel codec is almost the same as the FR channel codec because the design aim was to keep it as unchanged as possible. During the GSM EFR codec standardisation, the use of the existing FR channel codec (or any existing GSM generator polynomials) was encouraged since this minimises hardware changes in the GSM base stations and speeds up the introduction of the EFR codec. In the PCS 1900 EFR codec standardisation, the use of the existing FR channel codec was an essential requirement. Therefore, the FR channel codec was included in the EFR channel codec as a module together with additional error protection. The additional error protection consists of an 8-bit CRC parity check and a repetition code. The FR channel codec module protects the 182 most important bits with 1/2-rate convolution code and it uses a 3-bit CRC that covers the 50 most important bits. The bits in the EFR codec are divided into protected and unprotected bits according to their subjective importance to speech quality. Only 66 bits are left unprotected. These consist of the least significant bits of pulse positions in the algebraic code. The 8-bit CRC covers the 65 most important bits. It was included in the EFR codec to achieve reliable bad frame detection and, consequently, to reduce the number of undetected bad frames. These are a major source of audible degradations in current digital cellular systems.

4. VAD/DTX

The EFR codec contains also the functions of discontinuous transmission (DTX) and voice activity detection (VAD). DTX allows the radio transmitter to be switched off during speech pauses in order to save power in the mobile station and also to reduce the overall interference level over the air interface. VAD is used on the transmit side to detect speech pauses, during which characteristic parameters of the background acoustic noise are transmitted to the receive side, where similar noise, referred to as comfort noise, is then generated. The comfort noise parameters in the EFR codec consist of averaged LSP parameters and an averaged fixed codebook gain. Locally generated random numbers are used on the receive side as excitation pulses. During comfort noise generation, the adaptive codebook is switched off. The estimated average speech channel activity for the EFR codec is 64% [3].

5. COMPLEXITY AND DELAY

The complexity of the EFR codec has been estimated from a C-code implemented with a fixed point function library in which each operation has been assigned a weight representative for performing the operation on a typical DSP [3]. The theoretical worst case complexity of the EFR codec has been estimated during ETSI EFR verification phase to be 18.1 WMOPS (weighted million operations per second) [3]. This is below that of the GSM half rate codec (21.2 WMOPS) [3], [4]. Memory consumption estimated for data RAM (4.7k 16-bit words), data ROM (5.9k 16-bit words) and program ROM are each below those of the GSM half rate codec.

The delay of the EFR codec is approximately the same as that of the FR codec. Both codecs have a buffering delay of 20 ms without any lookahead. The round-trip delay (uplink delay + downlink delay) for EFR taking into account all system and processing delays of the GSM network is 191.0 ms while for the FR codec it is estimated to be 188.5 ms [3]. The difference is unnoticeable.

6. SUBJECTIVE TEST RESULTS

The most extensive subjective tests of the performance of the EFR codec are from the PCS 1900 EFR codec validation tests [5]. These were carried out to characterise the performance of the EFR codec after selection for PCS 1900. The standardisation process in ETSI also included pre-selection tests which were carried out in six laboratories, but no common analysis averaging the scores over all laboratories exists [3]. The results from both tests are well in line with each other. Both show that the EFR codec has basic speech quality at least equal to that of a wireline reference (G.721 32 kbit/s ADPCM in PCS 1900 tests and G.728 16 kbit/s LD-CELP in ETSI tests).

The PCS 1900 EFR codec validation test results are discussed first. The EFR codec was tested in three channel error conditions with C/I-ratios 13, 10 and 7 dB. The channel bit error-rates for these are approximately 2%, 5% and 8%, respectively. The two lowest error-rate conditions correspond to operating well inside a cell while the 7 dB C/I condition corresponds to operating at a cell boundary. Figure 2 shows the results from channel error test. These show that the EFR codec performs much better than the FR codec in the error-free case and in all the tested error conditions. In error-free and low error-rate conditions (C/I=13 dB), the performance of the EFR codec is statistically better (based upon 95% confidence intervals) than the performance of the error-free reference ADPCM codec. At medium error-rate conditions (C/I=10 dB) the EFR codec performs equally well to (error-free) ADPCM. Only at medium to high error-rates (C/I<10 dB) the EFR performance falls below wireline quality. Figure 3 shows the results from background noise test (home noise at 20 dB, car noise at 15 dB and 25 dB, street noise at 10 dB, and office noise at 20 dB). For all of these, the EFR codec performs clearly better than the FR codec and equal to or better than ADPCM. For scores averaged over all background noise conditions, the EFR codec performs statistically better than ADPCM. Figure 4 shows test results from tandem test for self-tandem and for tandeming with either FR or ADPCM codec. The EFR codec performs statistically equivalent to ADPCM in tandems with FR and ADPCM and statistically better than ADPCM for self-tandem. Figure 5 shows the results from talker dependency test (with 12 talkers). The

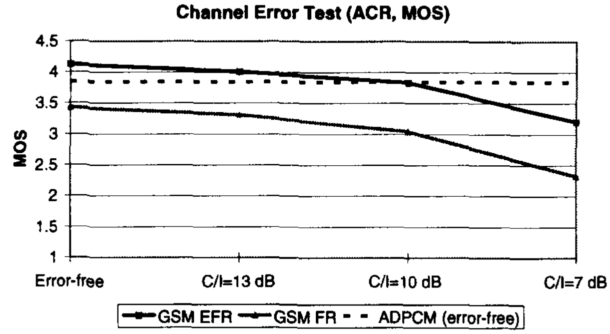

Figure 2: Results from channel error test.

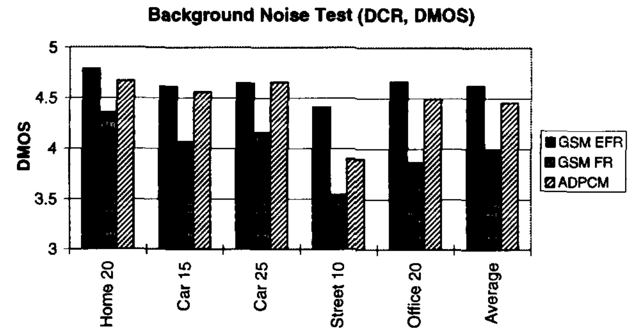

Figure 3: Results from background noise test.

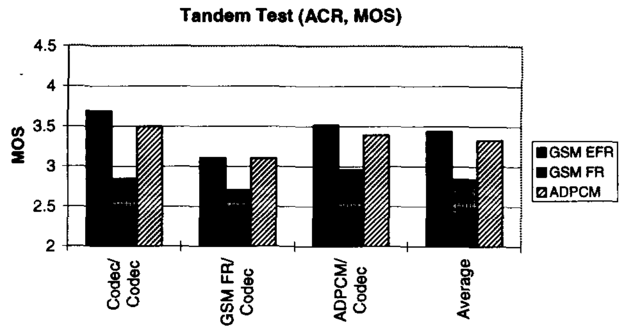

Figure 4: Results from tandem test.

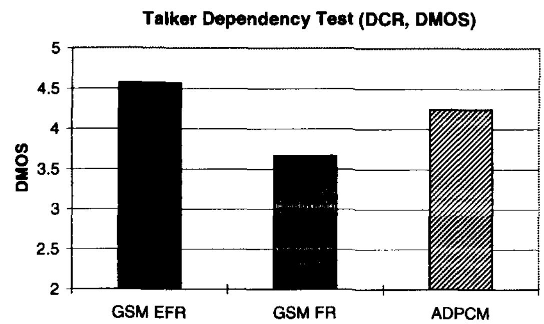

Figure 5: Results from talker dependency test.

EFR codec was found statistically better than ADPCM.

The ETSI pre-selection tests consisted of five experiments: transmission errors (C/I=10 and 7 dB), tandeming (C/I=10 dB), background noise (music 20 dB and vehicle 10 dB), talker dependency and high error conditions (C/I=4 dB). The performance of the EFR codec was found equal to G.728 for error-free transmission, speech in background noise (for both noise types) and talker dependency. No testing was carried out for the low error-rate condition C/I=13 dB. In erroneous transmission at C/I=10 dB and 7 dB, the EFR codec was found clearly better than the FR codec. At C/I=10 dB, the EFR codec performed equal to or better than MNRU 24 dB in half of the tests. For the outside-a-cell error condition of C/I=4 dB, the results show that the EFR codec has approximately the same performance as the FR codec. The EFR codec was tested in error-free self-tandem in one laboratory and was found equivalent to G.728. In self-tandem at C/I=10 dB, the EFR codec performed clearly better than the FR codec and equal performance to the FR codec in single encoding at C/I=10 dB was demonstrated in all laboratories except one.

Verification tests complementing the pre-selection tests were carried out in ETSI for such items as DTMF and network information tones, frequency response, complexity, delay, different languages, music and special input signals (e.g., different input levels, sine waves, noise signals etc.). In all the verification tests, the EFR codec performed well [3]. For DTMF-tones, the EFR codec was found 100% transparent.

7. CONCLUSIONS

The GSM EFR codec has met and even exceeded the essential requirements set for the development of the EFR codec. It provides substantial improvement in speech quality compared to the existing GSM full rate codec and brings high speech quality associated previously only with fixed networks to the end users of mobile communication systems. The GSM EFR specifications have been completed in 1996 and the codec is expected to be deployed in GSM, DCS 1800 and PCS 1900 networks in 1997-1998.

ACKNOWLEDGEMENTS

The authors acknowledge the efforts of M. Fatini and J. Hagqvist in PCS 1900 EFR codec standardisation, the algorithmic contribution of B. Bessette and the efforts of members of ETSI SMG2 SEG, especially those who have carried out pre-selection and verification tests for the EFR codec.

REFERENCES

[1] "Enhanced Full-Rate Speech Codec - Study Phase Report," Report of the ETSI SMG1 sub-group on the EFR codec, Version 0.5.3, December 19, 1994.

[2] R. Salami, C. Laflamme, J-P. Adoul, and D. Massaloux, "A toll quality 8 kb/s speech codec for the personal communications system (PCS)," IEEE Trans. Veh. Technol., vol 43, no. 3, pp. 808-816, Aug. 1994.

[3] "Digital cellular telecommunications system; Performance characterization of the GSM enhanced full rate speech codec (GSM 06.55)," ETSI Technical Report (in preparation), Draft ETR 305, version 1.0.0.

[4] Complexity Evaluation Subgroup of ETSI TCH-HS (Alcatel Radiotelephone, Analog Devices, Italtel-SIT, Nokia), "Complexity Evaluation of the Full-Rate, ANT and Motorola Codecs for the Selection of the GSM Half-Rate Codec," Proceedings of the 1994 DSPx Exposition & Symposium, June 13-15, San Fransisco (USA), 1994.

[5] "COMSAT Test Report on the PCS1900 EFR Codec," Source: Nokia, Temporary Document 47/95 of ETSI SMG2-SEG, June 1995.