6th European Conference on
Speech Communication and Technology
(EUROSPEECH'99)
Budapest, Hungary, September 5-9, 1999

A MULTI-RATE CODEC FAMILY BASED ON GSM EFR AND ITU-T G.729

Balázs Kövesi[‡], *Claude Lamblin*[‡], *Catherine Quinquis*[‡], *Philippe Thiérion*[†], *William Navarro*[†]
[‡]France Telecom CNET, DIH/DIPS, F-22307 Lannion Cedex, France
balazs.kovesi/claude.lamblin/catherine.quinquis@cnet.francetelecom.fr
[†]Nortel Matra Cellular, F-78042 Guyancourt Cedex, France
pthierio/navarro@nortel.com

10.21437/Eurospeech.1999-392

ABSTRACT

This paper gives the description of the multi-rate speech codec family which was the basis of the candidate developed by France Telecom and Nortel Networks for the ETSI GSM AMR normalization. This codec family is built on the ACELP technology. For each type of parameters derived from CELP analysis like LP filter coefficients, adaptive and fixed codebooks and gains, a set of vector quantizers has been designed. This allows to select the bit-rate for any parameter and consequently to vary the global speech coder bit-rate gradually in the range of 4.7 – 12.2 kbit/s. Formal subjective listening test results of the GSM AMR candidate codec performed by two world wide recognized listening laboratories in two languages (English and French) are also given.

Keywords: GSM AMR, ACELP

1. INTRODUCTION

In the present standards of the European Telecommunications Standard Institute (ETSI) for Global System for Mobile Communication (GSM) the coding rates of the speech and the channel codecs are fixed:

Standard	Gross bit-rate	Speech codec bit-rate	Channel codec bit-rate
GSM FR (RPE-LTP)	22.8 kbit/s	13.0 kbit/s	9.8 kbit/s
GSM HR (EVSELP)	11.4 kbit/s	5.6 kbit/s	5.8 kbit/s
GSM EFR (ACELP)	22.8 kbit/s	12.2 kbit/s	10.6 kbit/s

The bit-rate trade-off between source coding and channel coding was chosen as a compromise in a way to reach good performance both in error-free and in severe transmission conditions (C/I = 4 dB). Obviously, this choice is not optimal in most channel conditions. In case of bad channel conditions in the full rate (FR, 22.8 kbit/s) speech traffic channel it would be better to choose a speech codec with slightly worst intrinsic quality but needing less bit-rate to be able to reinforce the error protection. On the other hand, in an error-free half rate (HR, 11.4 kbit/s) speech traffic channel an intrinsically better speech codec could be used.

According to this idea, ETSI started a standardization program to develop an Adaptive Multi Rate (AMR) codec for GSM. Both in case of full rate or half rate channel modes this codec should adapt the source coding and the error protection bit-rates according to the quality of the radio channel to approach the optimal performance.

Two main goals were aimed: to improve error robustness of the full rate channel and to improve speech quality of the half rate channel especially in good channel conditions.

This paper is organized as follows: first a brief description of the proposed codec family is given, then the resulted GSM AMR candidate is presented. The last section gives several formal subjective listening test results.

2. MULTI RATE SPEECH CODEC

The codec family proposed by France Telecom and Nortel Networks both derives from the GSM Enhanced Full Rate (EFR) codec [1] and the ITU-T G.729 codec [2]. This codec family is built on the well-known ACELP technology (Algebraic Code-Excited Linear Predictive Coding), and operates on speech frames of 20 ms. The bit-rate can vary from 12.2 kbit/s to 4.7 kbit/s gradually. The highest bit-rate mode is identical to the EFR speech codec while the other codecs of the family can be seen as lower bit-rate extensions of this codec scheme. Using the same basic algorithm favors the fact that switching between coding modes does not cause annoying audible effects. Furthermore, it reduces the amount of ROM program needed since many modules are shared by the various modes.

In fact, for each type of parameters derived from CELP analysis like linear prediction (LP) filter coefficients, adaptive and fixed codebooks indices and gains, a set of vector quantizers (VQ) has been designed. This allows selecting the bit-rate for any parameter and consequently to vary the global speech coders bit-rate gradually.

2.1 Linear Prediction

2.1.1 Analysis

The 20 ms frames are divided into four 5 ms subframes. Always two 10^{th} order LP filters are computed for each frame centered on the second and the fourth subframes. The LP coefficients are transformed to the LSF (Line Spectral Frequencies) domain before quantization. In FR mode, the LP analysis is made as in the EFR codec while in HR mode a look-ahead of 5 ms has been introduced.

2.1.2 Quantization of LP parameters

Two vector quantizers are available:

- The first one, which will not be described here, is identical to that of the EFR codec i.e. the two LP filters are jointly quantized on 38 bits with split matrix quantization of the first order MA (Moving Average) prediction LSF residuals.

- The second vector quantizer needs 22 bits to encode the two LSP sets. The LSP sets corresponding to the second and fourth subframes will be denoted by $f^{(1)}$ and $f^{(2)}$ respectively. The LSF coefficients are quantized either by using a first order MA prediction, as in the 38-bit vector quantizer, or without any prediction. This second mode is generally chosen when the spectrum changes radically.

Both in predictive and non predictive modes, each second LSF set $f^{(2)}$ is quantized using split vector quantization. The 10 dimension vectors are split into 3 subvectors of dimension 3,3 and 4. These subvectors are respectively vector quantized using 6,7 and 7 bits. The chosen codeword for the second LSF set $f^{(2)}$ depends on the corresponding first LSF set $f^{(1)}$ as well. Let us note $\hat{f}^{(1)}$ and $\hat{f}^{(2)}$ the quantized LSF sets. The codebook indices are chosen to minimize the following weighted squared distance:

$$E_1 = \sum_{i=1}^{10} w_i^{(2)}\left(f_i^{(1)} - \hat{f}_i^{(2)}\right)^2 + 2\sum_{i=1}^{10} w_i^{(2)}\left(f_i^{(2)} - \hat{f}_i^{(2)}\right)^2$$

where $w_i^{(2)}$ denotes the IHM (Inverse Harmonic Mean [4]) weighting factor of the i^{th} coefficient of the second LSF set.

To choose between the two quantization modes, the one that gives the smallest distortion is taken. This information is then transmitted with one bit/frame.

The quantized first LSF set $\hat{f}^{(1)}$ is obtained by linear interpolation of the previous and the following quantized second LSF sets. Two interpolated vectors are computed, one with the weights (0.5, 0.5), the other with the weights (0.125, 0.875). The one that minimizes the distance

$$E_2 = \sum_{i=1}^{10} w_i^{(1)}\left(f_i^{(1)} - \hat{f}_i^{(1)}\right)^2$$

is selected and this information is quantized on 1 bit.

Finally, the quantized LSF parameters of the first and third subframes are obtained by linear interpolation of the quantized LSF vectors of the neighboring subframes, using 0.5-0.5 as weighting factors.

The indices of all codebooks of the 38-bit and 22-bit quantizers are reordered with a simulating annealing type algorithm in a way to achieve better intrinsic robustness to channel errors.

Informal listening test performed by trained listeners has shown that replacing the 38-bit vector quantizer of the EFR codec by this 22-bit vector quantizer does not involve any audible degradation.

2.2 Adaptive Codebook

A combined open-loop/closed-loop search is employed to find the fractional pitch lag as in the EFR codec. The adaptive codebook lag is absolute coded for the 1^{st} and 3^{rd} subframes and differentially coded for the remaining subframes. Three fractional resolutions were implemented from the $1/6^{th}$ sample (coded on 9+6 bits) through the $1/3^{rd}$ sample (coded on 8+5 bits, like in ITU-T G.729 codec) to the $1/2^{nd}$ sample (coded on 7+3 bits) for the lowest bit-rates.

The adaptive codebook gain is coded using a 4, 3 or 2 bits non-uniform scalar quantization.

2.3 Fixed Codebook

In the EFR codec, the ACELP codebook represents the largest part of the bit-stream, 140 bits from 244 bits. To achieve bit-rate reduction both the number of pulses in the codevector and the number of the their possible positions have to be reduced. Various ACELP codebooks have been designed for this codec family, the number of pulses in each subframe of 40 samples varies from 10 to 1 and the bit-rate varies from 140 bit/frame to 28 bit/frame.

To encode the algebraic codebook gain, a 4^{th} order MA predictive quantization is employed. Besides the 5-bit/subframe quantizer of the EFR codec a 4-bit/subframe quantizer was designed as well.

2.4 Voiced – unvoiced decision

In lower bit-rates, only 1 or 2 pulses can be coded for each 40 samples, which yields quite audible

artifacts in the reconstructed signal, especially for unvoiced segments [3]. By introducing a voicing decision, for unvoiced segments the LTP (Long Term Prediction) parameters do not need to be coded. The saved bits are used to reinforce the fixed excitation. Algebraic codebooks, more suitable to model noise-like excitation, have also been designed for this purpose. These codebooks are also ternary but not pulse-like.
The voicing decision, made twice per frame, is based on the prediction gain obtained in the open-loop pitch analysis, on the energy and on the correlation function of the input signal.

3 GSM AMR CANDIDATE

3.1 Source Codec

The source codec of the GSM AMR candidate of France Telecom and Nortel Networks is based on the presented codec family with the following modifications in half rate channel mode [5]:

- The 20 ms frame is divided into 3 subframes of 53,53 and 54 samples.
- A third 19-bit LSF quantizer has been designed. Note that this quantizer is also used for the FR lowest rate.
- The time warping technique [6] has been used for LTP coding.

The employed source coding bit-rates are 12.2 (A), 9.2 (B), 7.8 (C) and 6.05 (D) kbit/s in the GSM full rate channel and 7.7 (A), 6.35 (B), 5.45 (C) and 4.55 (D) kbit/s in the half rate channel.

3.2 Complexity

The complexity requirement for the GSM AMR speech coder/decoder was to be less complex than 8 times the GSM FR speech codec. The complexity evaluation methodology was based on ETSI fixed point basic operators. The presented AMR candidate fulfills the complexity requirement.
For each parameter analysis and quantization module, the complexity is equal or less than the corresponding EFR or G.729 ones. For instance, the algebraic codebooks searches are similar to the EFR fixed codebook search [1]. However, to further speed up the search procedure while drastically reducing memory storage, the procedure used in G.729 to compute and store the elements of the autocorrelation matrix has been adapted to the new codebooks and the EFR-like search [7].

3.3 Performance Evaluation

In the qualification phase of testing, the subjective listening experiments of the candidate proposal were performed by Nortel and France Telecom/CNET respectively in their Ottawa and Lannion listening labs using two languages (English and French).
The performance evaluation has been carried out following ETSI's documentation [8].
Four experiments were realized both for full rate and half rate codecs. For clean speech, the codecs performances were evaluated in case of transmission errors in static or dynamic C/I conditions as well as the effect of switching and tandeming. The MOS (Mean Opinion Score) rating scale was used in these experiments.
Figure 1 and *Figure 2* illustrate the performance of the four codec modes as function of the static C/I level of the transmission channel respectively for full rate and half rate channel conditions as well as the quality required by the specifications.

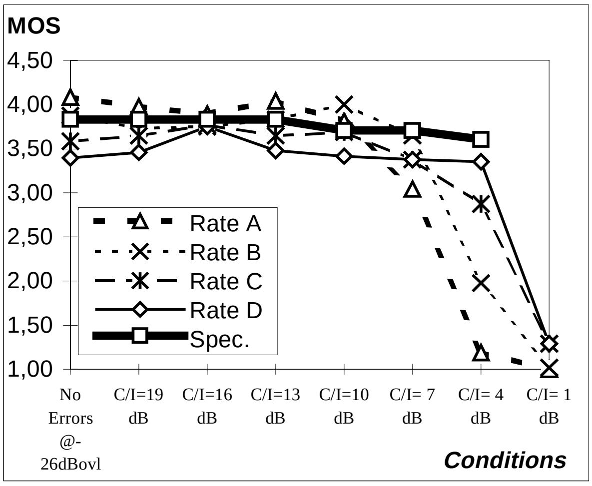

Figure 1. Effect of errors in clean speech conditions for Full Rate channel

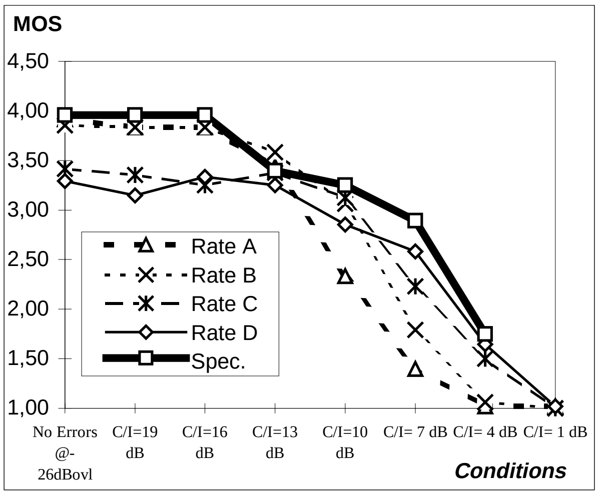

Figure 2. Effect of errors in clean speech conditions for Half Rate channel

Figure 3 and *Figure 4* summarize the results obtained in five different dynamically varying channel conditions (DEC1-DEC5) that represent typical real scenarios. The adaptive AMR approach outperforms the EFR codec and any given fixed rate codec as well.

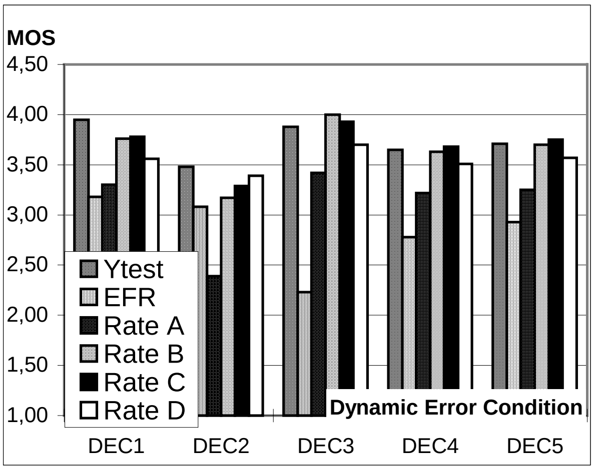

Figure 3. Performances in dynamic C/I conditions for Full Rate channel

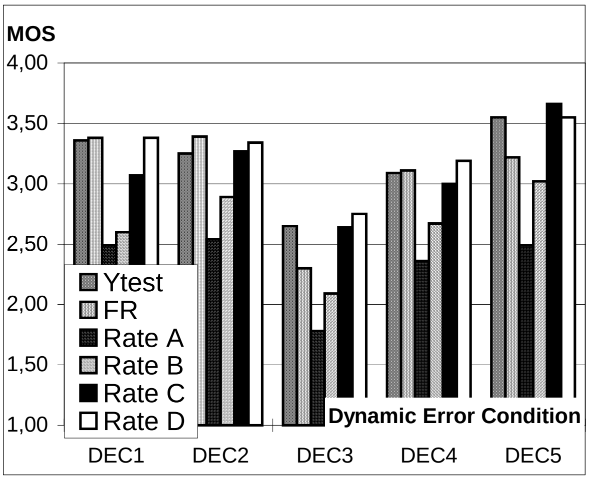

Figure 4. Performances in dynamic C/I conditions for Half Rate channel

In all clean speech experiments the AMR candidate passed the requirements.

The background noise conditions were tested with street noise at 15 S/N = dB and car noise at S/N = 10 and 15 dB in constant C/I conditions. In this experience, the DMOS (Degradation MOS) rating scale was applied. The AMR candidate showed similar performance improvement compared to non-adaptive coders as in clean speech experiments, but the required specifications were not reached neither in full rate channel mode in poor channel conditions (C/I =4 dB) nor in half rate channel modes. Note that these requirements were very ambitious and no AMR candidate fulfilled all the background noise requirements.

4. CONCLUSION

This paper presented the GSM AMR candidate of France Telecom and Nortel Networks. Formal listening tests have shown the superiority of this multi rate codec to the non-adaptive reference codecs with its improved error robustness in full rate channel conditions and its nearly wireline quality in good half rate channel conditions.

Acknowledgment

The authors would like to thank Dominique Pascal and Leigh Thorpe for the subjective test results.

5. REFERENCES

[1] K. Järvinen et al. “GSM Enhanced Full Rate Speech Codec”, *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Munich, Germany, 20-24 April, 1997.

[2] R. Salami, C. Laflamme, J.-P. Adoul, A. Kataoka, S. Hayashi, T. Moriya, C. Lamblin, D. Massaloux, S. Proust, P. Kroon, and Y. Shoham, “Design and Description of CS-ACELP: A Toll Quality 8 kb/s Speech Coder”, *Trans. on Speech and Audio Processing*, Vol.6, Num.2, pp: 116-130, March, 1998.

[3] R. Hagen, E. Ekudden, B. Johansson, and W.B. Kleijn, “Removal of Sparse-Excitation Artifact in CELP”, *Proc. of IEEE International conference on Acoustics, Speech, and Signal Processing*, Seattle, Washington, USA, 12-15 May, 1998.

[4] J. Pan and T.R. Ficher, “Vector quantization – lattice quantization of speech LPC coefficients” *Proc. of IEEE International conference on Acoustics, Speech, and Signal Processing*, Adelaide, Australia, April, 1994.

[5] M.Foodeei, H. Zarrinkoub, R. Matmti, R. Rabipour, F. Gabin and S. Gosné, “A low bit rate codec for AMR standard”, *IEEE Speech Coding Workshop*, 1999.

[6] W.B. Kleijn, P. Kroon and D. Nahumi, “The RCELP speech coding algorithm”, *European transactions in Telecommunications*, vol 5, Number 5 Sep/Oct 1994, pp 573-582.

[7] IUT-T, “Recommendation G.729 - Coding of speech at 8 kbit/s using conjugate structure algebraic code excited linear prediction (CS-ACELP)”, 1996.

[8] ETSI AMR Qualification Phase Documentation, 1998.