1997 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)

Volume II of V

Speech Processing

April 21-24, 1997
Munich, Germany

Sponsored by
The Institute of Electrical and Electronics Engineers
Signal Processing Society

Co-Sponsored by
DPG, GI, ITG, and TUM

IEEE Computer Society Press
Los Alamitos, California

Washington • Brussels • Tokyo

DTP

IEEE Computer Society Press
10662 Los Vaqueros Circle
P.O.Box 3014
Los Alamitos, CA 90720-1264

IEEE Computer Society Press Order Number PR07919
ISBN 0-8186-7919-0
ISBN 0-8186-7920-4 (case)
ISBN 0-8186-7921-2 (microfiche)
IEEE Order Plan Catalog Number 97CB36052
Library of Congress Number 96-80328

Additional copies may be ordered from:

IEEE Computer Society Press
Customer Service Center
10662 Los Vaqueros Circle
P.O. Box 3014
Los Alamitos, CA 90720-1314
Tel: +1-714-821-8380
Fax: +1-714-821-4641
Email: cs.books@computer.org

IEEE Service Center
445 Hoes Lane
P.O. Box 1331
Piscataway, NJ 08855-1331
Tel: +1-908-981-1393
Fax: +1-908-981-9667
misc.custserv@computer.org

IEEE Computer Society
13, Avenue de l'Aquilon
B-1200 Brussels
BELGIUM
Tel: +32-2-770-2198
Fax: +32-2-770-8505
euro.ofc@computer.org

IEEE Computer Society
Ooshima Building
2-19-1 Minami-Aoyama
Minato-ku, Tokyo 107
JAPAN
Tel: +81-3-3408-3118
Fax: +81-3-3408-3553
tokyo.ofc@computer.org

Editorial production by Bob Werner
Printed in the United States of America by Sony Electronic Publishing Services

The Institute of Electrical and Electronics Engineers, Inc.

Table of Contents

Volume II
Speech Processing

ENHANCED FULL RATE SPEECH CODEC FOR IS-136 DIGITAL CELLULAR SYSTEM

T. Honkanen, J. Vainio, K. Järvinen, P. Haavisto
Nokia Research Center, Tampere, Finland

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

ABSTRACT

In this paper, we describe the enhanced full rate (EFR) speech codec that has recently been standardised for the North American TDMA digital cellular system (IS-136). The EFR codec, specified in the IS-641 standard, has been jointly developed by Nokia and University of Sherbrooke. The codec consists of 7.4 kbit/s speech (source) coding and 5.6 kbit/s channel coding (error protection) resulting in a 13.0 kbit/s gross bit-rate in the channel. Speech coding is based on the ACELP algorithm (Algebraic Code Excited Linear Prediction). The codec offers speech quality close to that of wireline telephony (G.726 32 kbit/s ADPCM used as a wireline reference) and provides a substantial improvement over the quality of the current speech channel. The improved speech quality is not only achieved in error-free conditions, but also in typical cellular operating conditions including transmission errors, environmental noise, and tandeming of speech codecs.

1. INTRODUCTION

During recent years there have been significant advances in speech coding technology to achieve toll quality speech at bit-rates of around 8 kbit/s. One example of this development is the completion of the ITU-T 8 kbit/s speech coding standard G.729.

The current full rate codec of the North American TDMA system employs 7.95 kbit/s for speech coding with additional 5.05 kbit/s used for error protection. The codec is described in the IS-136 standard and also in the earlier IS-54 standard [1]. Due to advances in speech coding technology, TIA (Telecommunications Industry Association) considered the prospects of developing an improved full rate codec for the TDMA system to be promising and launched a standardisation process for an enhanced full rate (EFR) speech codec.

Nokia submitted a proposal for the EFR codec standard in June 1995. The proposed codec was jointly developed by Nokia and University of Sherbrooke (USH). It was based on the same ACELP (Algebraic Code Excited Linear Prediction) technology as the Nokia/USH enhanced codec proposal submitted to PCS 1900 EFR and GSM EFR standardisation [2].

The EFR standardisation activity was finalised in March 1996 when the Nokia/USH EFR codec proposal was approved as the IS-641 standard [3]. Before adopting the codec, an evaluation phase was carried out including verification of implementation complexity and the codec performance for different operating conditions and input signals. In addition to error-free speech quality, improvement was expected in presence of channel errors and background noise.

2. EFR SPEECH CODEC DESCRIPTION

The IS-641 EFR speech codec is based on the ACELP algorithm (Algebraic Code Excited Linear Prediction) [4]. The speech coding (source coding) bit-rate is 7.4 kbit/s and 5.6 kbit/s is used for channel coding. The efficient speech coding enables more bits to be used for error protection still sustaining high speech quality in the error-free channel. The EFR codec therefore provides not only improved basic quality but also improved error robustness compared to the full rate codec. The enhanced full rate codec is optimised for the IS-136 system: the 20 ms frame length fits the IS-136 full rate slot structure, the bit allocation between speech and channel coding is chosen for typical transmission error conditions and the complexity of the codec does not exceed the complexity of the current full rate codec.

The EFR codec operates on speech frames of 20 ms which are divided into four 5 ms subframes. The extracted parameters are typical for the CELP synthesis model: linear prediction filter coefficients, indices for the adaptive and fixed codebooks, and gains for the two codebooks. The codebook gains are jointly vector quantised. The bit allocation among speech coding parameters of the EFR speech codec is shown in Table 1 and a block diagram of the encoder is presented in Figure 1.

Parameter	1st and 3rd subframe	2nd and 4th subframe	Total per 20ms frame
LSP param.			26
ACB delay	8	5	26
Gain VQ	7	7	28
Algebraic code	17	17	68
Total			148

Table 1. Bit allocation of the IS-641 EFR speech codec parameters.

2.1 Linear Prediction Analysis

The input speech signal is first high-pass filtered and a 10th order linear prediction analysis is carried out for each 20 ms frame. A 30 ms asymmetric analysis window with a lookahead of 5 ms (40 samples) is used. The short-term coefficients are calculated with a Levinson-Durbin algorithm using a 60 Hz bandwidth expansion for the autocorrelations. The linear

0-8186-7919-0/97 $10.00

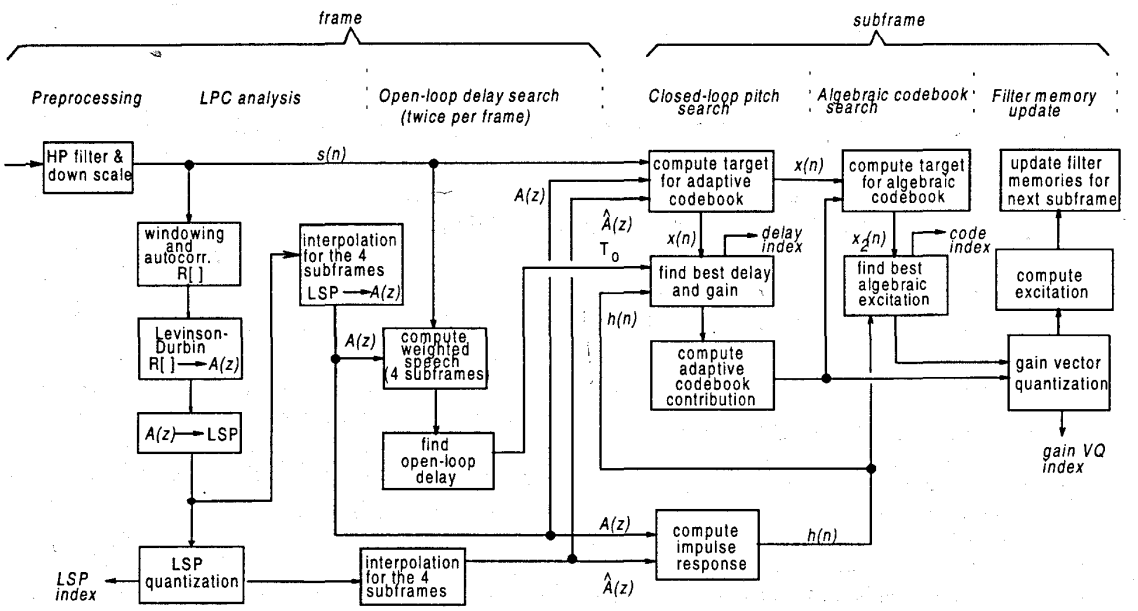

Figure 1. Block diagram of IS-641 EFR speech codec

prediction (LP) coefficients are then converted into line spectrum pairs (LSP).

The LSP parameters are quantised using split vector quantisation (SVQ). First order moving average prediction is applied to the LSP vector. The LSP residual vector is split into three subvectors of dimensions 3, 3, and 4, and the subvectors are quantised with 8, 9, and 9 bits resulting in 26 bits for each 20 ms frame.

The first three subframes use interpolated short term coefficients and the interpolation is carried out for both quantised and unquantised LSP coefficients. The unquantised LP coefficients are utilised in the perceptual weighting filter.

2.2 Pitch Analysis

The adaptive codebook delay (lag) is searched using a combined open-loop/closed-loop search [4]. The optimum delay is searched for the range [19 1/3, 143].

The open-loop search is carried out twice in the 20 ms frame. An optimum integer delay for the weighted input speech signal is searched in three ranges favouring smaller delay values. This is done in order to avoid choosing pitch multiples and to smooth the lag trajectory. This enables efficient differential quantisation.

The closed-loop search is performed on a subframe basis. In the first and third subframes, the search is performed around the found open-loop delays (T_{ol} ±3). Fractional resolution of 1/3 is used for delays smaller than 85 and integer resolution is used in the range [85, 143]. In the second and fourth subframes, a fractional delay is always used. The search is performed around the integer delay selected in the previous subframe [T_{ps} - 5, T_{ps} + 4].

In the first and third subframes the delay is encoded with 8 bits. In the second and fourth subframes the delay is differentially encoded with 5 bits with respect to the integer delay of the previous subframe.

After the closed-loop search, the adaptive codebook gain is calculated and the fixed codebook target is constructed using the unquantised adaptive codebook gain.

2.3 Fixed Codebook Search

An algebraic codebook of 17 bits is used for the fixed codebook excitation and the algebraic codeword is searched for every subframe. The algebraic codebook structure is based on interleaved single-pulse permutation (ISPP) design. Each innovation vector contains 4 non-zero pulses having amplitudes +1 or -1. The allowed positions of each of the four pulses are shown in Table 2. The first three pulse positions are encoded with 3 bits and the fourth pulse is encoded with 4 bits. In addition, one bit is needed to encode the sign of each of the four pulses.

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

Table 2. Allowed pulse positions in algebraic excitation

The optimal pulse positions are determined using a non-exhaustive analysis-by-synthesis search. The pulse positions are organised into five tracks. For each of the tracks, the 4 pulse positions with the largest absolute values of the backward filtered target signal, i.e., the correlation of the target and the impulse response, are searched. Since there are only four pulses, pulse *p3* is set either to track 4 or track 5. The following search procedure is done twice, once for each track of *p3*.

Four iterations are carried out in which the position of pulse *p0* is searched only for the track positions corresponding to the four largest absolute values of the backward filtered target. For each of these four possible positions of *p0*, the pulses *p1*, *p2* and *p3* are searched sequentially for all the 8 positions in their current tracks.

For each iteration, all 4 pulse starting positions are cyclically shifted (*p0=p1, p1=p2, p2=p3, p3=p0*) so that the pulse *p0* is always searched for the four largest values in its current track. The rest of the pulses are searched also for the other positions in the tracks.

The different pulse positions are quantised with 13 bits, so the codeword space consists of 8192 different combinations of pulse positions. The total number of different pulse positions searched are *T*S*n0*(n1 + n2 + n3)*, where *T* is the number of possible tracks for *p3*, *S* is the number of cyclical iterations and *n0*, *n1*, *n2* and *n3* are the numbers of different positions of *p0*, *p1*, *p2* and *p3*. The total number of allowed codewords is 2*4*4*(8+8+8)=768, which means that only about 9 % of the codeword space is searched in the analysis-by-synthesis loop.

For the subframes with delay less than the subframe length, the selected codevector is filtered through an adaptive pitch sharpening filter to emphasise the harmonic spectral components.

After the fixed codebook search, the adaptive codebook gain and the fixed codebook gain are jointly vector quantised using a 7-bit codebook in each subframe. Fourth order moving average prediction with fixed coefficients is applied to the fixed codebook gain before quantisation. The gain quantiser codebook search minimises the weighted error between original and reconstructed speech [4].

3. CHANNEL CODEC

The EFR channel codec uses the same basic structure as in the full rate channel codec. This was an important objective in the codec design. The convolutional code and CRC polynomials used in the full rate channel codec are also used as such in the EFR channel codec. Since the EFR codec allocates more bits to channel coding, fewer bits are transmitted without error protection than in the current full rate codec.

A block diagram of the channel codec is presented in Figure 2. The 148 source coding bits for each 20 ms frame are classified into Class1 (96 bits) and Class2 (52 bits). A 7-bit CRC checksum is calculated over the 48 most important bits from Class1 to check the correctness of the received frame in the decoder. The Class1 bits and the CRC bits are then protected by rate 1/2 convolutional code with a constraint length of 6. The less important Class2 bits are transmitted without protection.

Slight puncturing (8 out of 216 convolutionally coded bits) makes it possible to protect all the bits except the bits defining the fixed codebook pulse positions. The division of bits into the two classes is based on their relative importance to speech quality, which was determined using both objective bit-error sensitivity analysis and expert listening. The most sensitive bits include the bits defining the LSP, delay, and gain parameters. The least sensitive bits are those defining the pulse positions. These bits are Gray-coded and left unprotected.

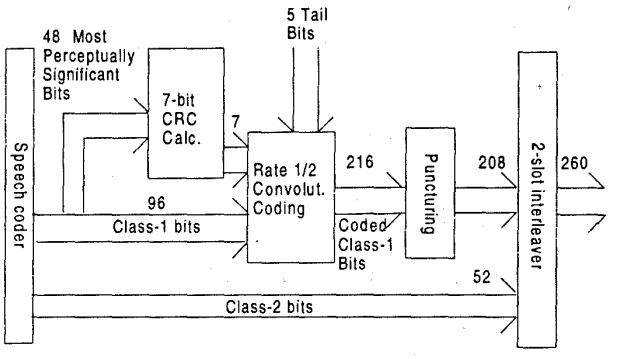

Figure 2. Block diagram of IS-641 EFR channel codec

4. COMPLEXITY AND DELAY

The IS-641 EFR speech codec is defined in fixed point arithmetic using a set of basic operations defined by ETSI [5]. This allows calculation of the complexity of the codec as a number of WMOPS (Weighted Million Operations per Second) . The worst case complexity figure for the speech EFR codec is estimated to be 13.54 WMOPS [6]. The peak memory consumption figures for scratch and static RAM are 2366 and 1129 16-bit words, respectively. In addition, 5493 words of data ROM is needed. The complexity of the EFR codec does not exceed the complexity of the existing full rate speech codec.

The complexity of the IS-641 EFR channel codec is approximately the same as that of the current full rate channel codec. The complexity and the memory consumption of the channel codec depend on the implementation of the channel decoder which is not defined to be bit-exact.

The buffering delay of the speech codec is 25 ms including a 5 ms delay due to the LPC lookahead. In addition, the interleaving of two adjacent frames increases the delay by 20 ms.

5. SUBJECTIVE TEST RESULTS

During the evaluation phase, subjective tests of the EFR codec performance were carried out in COMSAT laboratories. The tests included the performance in typical channel impairment conditions (transmission errors), performance in typical background (environmental) noise, tandeming of codecs, and talker dependency. In the tests, the performance of the EFR codec was compared to the current full rate codec based on VSELP (Vector Sum Excited Linear Prediction) speech coding [7]. The G.726 (32 kbit/s ADPCM) codec was included in the tests as a reference.

Figure 3 shows the results from the channel impairment test. Transmission errors were inserted using error patterns generated by the channel simulator software provided by TIA. The tested channels covered typical cellular conditions including C/I ratios 11dB, 15dB, 19dB, as well as the error-free case. The source speech was weighted by the ITU modified IRS filtering. The G.726 codec was included as an error-free reference only.

The MOS results show that the EFR codec performs better than the full rate codec in each channel impairment condition. The performance in low error-rate condition (C/I=19dB) is very close to the error-free quality. Also, the performance in medium

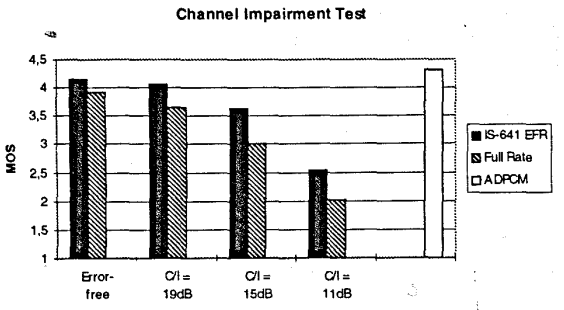

Figure 3. Results from channel error test

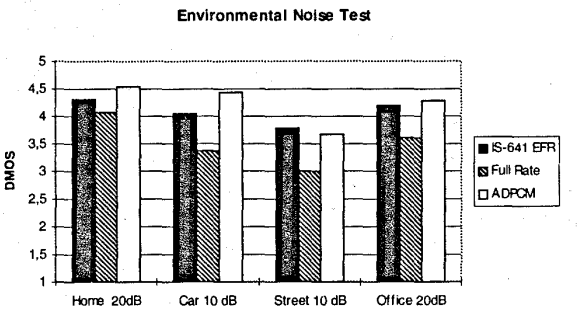

Figure 4. Results from background noise test

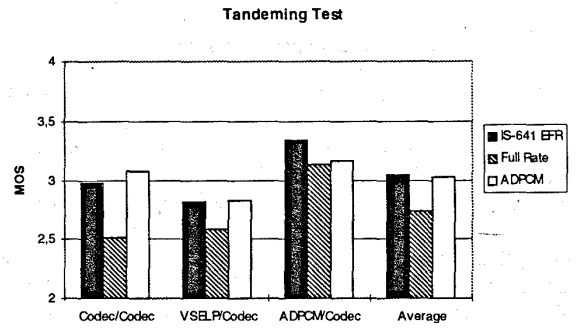

Figure 5. Results from tandeming test

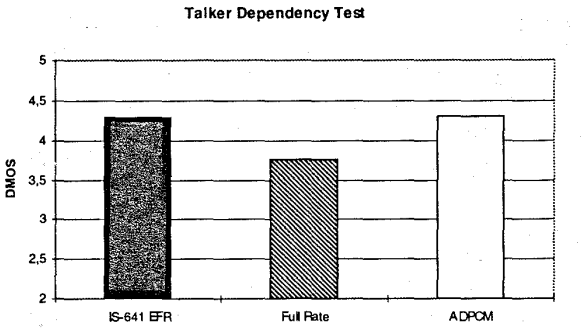

Figure 6. Results from talker dependency test

error-rate channel (C/I=15dB) is equivalent to the performance of the full rate codec in low error-rate channel.

Figure 4 shows results from the background noise test with four types of noise. The test was carried out as a DMOS test with flat input characteristics. The EFR codec performs significantly better than the full rate codec in each background noise condition.

The results of the tandeming test are presented in Figure 5. Again, the IS-641 EFR codec performs statistically better than the full rate codec.

Figure 6 shows the results from the talker dependency test. The IS-641 EFR codec performed statistically equivalently to the ADPCM reference codec and significantly better than the full rate codec.

6. CONCLUSIONS

The IS-641 EFR codec offers toll quality, or near toll-quality, basic performance and also very good performance in typical mobile environments with channel errors and background noise. The 20 ms frame length fits the IS-136 full rate slot structure, the speech and channel codec bit rates have been optimised for the total channel capacity, and the complexity of the codec does not exceed the complexity of the full rate codec. The new codec brings a significantly improved speech service to the North American TDMA system.

ACKNOWLEDGEMENTS

The authors acknowledge the efforts of M. Fatini and T. Rautava in IS-641 EFR codec standardisation, as well as the algorithmic contribution of B. Bessette.

REFERENCES

[1] IS-136.1-A TDMA Cellular/PCS - Radio Interface - Mobile Station - Base Station Compatibility - Digital Control Channel, Revision A, Aug 1996.
IS-136.2-A TDMA Cellular/PCS - Radio Interface - MobileStation - Base Station Compatibility - Traffic Channels and FSK Control Channel, Revision A, Aug 1996.

[2] K. Järvinen *et al.* , "GSM Enhanced Full Rate Codec," Proc. ICASSP'97, Munich, Germany, 20-24 April, 1997.

[3] TIA/EIA/IS-641, Interim Standard, TDMA Cellular/PCS radio interface - Enhanced Full-Rate Speech Codec, May 1996.

[4] 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.

[5] Complexity Evaluation Subgroup of ETSI TCH-HS (Alcatel Radiotelephone, Analog Devices, Italtel, 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.

[6] "Complexity calculation of the Nokia/USH EFR speech codec, version 1.2," TIA TR45.3.5/95.12.11.08.

[7] I.A. Gerson and M.A. Jasiuk, "Vector Sum Excited Linear Prediction (VSELP) Speech Coding at 8 kb/s," Proc. ICASSP'90, Albuquerque, April 1990, pp. 461-464