SLOVENSKI STANDARD
SIST EN 301 245 V4.1.1:2003

01-december-2003

8][ ]HʹbŮ]ʹW̌Ý] b]ʹhʹYʹY_ca i b]_UÝ]ʹg_]ʹg]ghÝa ʹfÉ£UnUʹ&Ł Ë=nVc`ʹýUbcʹdfY_cX]fUb·Y [cjcfUʹgʹdcʹbcʹ\ʹ]fcgh·cʹfÐ:FŁ£fƈ GA ʹ$*ʹ'*ʹ$žfUbʈ]W̌Uʹ("%%Ł

Digital cellular telecommunications system (Phase 2) (GSM); Enhanced Full Rate (EFR) speech transcoding (GSM 06.60 version 4.1.1)

iTeh STANDARD PREVIEW
(standards.iteh.ai)

Ta slovenski standard je istoveten z: **EN 301 245 Version 4.1.1**

ICS:

33.070.50	Globalni sistem za mobilno telekomunikacijo (GSM)	Global System for Mobile Communication (GSM)

SIST EN 301 245 V4.1.1:2003 en

iTeh STANDARD PREVIEW
(standards.iteh.ai)

ETSI EN 301 245 V4.1.1 (2000-08)

European Standard (Telecommunications series)

Digital cellular telecommunications system (Phase 2); Enhanced Full Rate (EFR) speech transcoding (GSM 06.60 version 4.1.1)

iTeh STANDARD PREVIEW

(standards.iteh.ai)

GLOBAL SYSTEM FOR MOBILE COMMUNICATIONS

ETSI

Reference
REN/SMG-110660PR1

Keywords
EFR, digital cellular telecommunications system,
Global System for Mobile communications
(GSM), speech

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00   Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

iTeh STANDARD PREVIEW
(standards.iteh.ai)

Important notice

Individual copies of the present document can be downloaded from:
http://www.etsi.org

The present document may be made available in more than one electronic version or in print. In any case of existing or
perceived difference in contents between such versions, the reference version is the Portable Document Format (PDF).
In case of dispute, the reference shall be the printing on ETSI printers of the PDF version kept on a specific network drive
within ETSI Secretariat.

Users of the present document should be aware that the document may be subject to revision or change of status.
Information on the current status of this and other ETSI documents is available at http://www.etsi.org/tb/status/

If you find errors in the present document, send your comment to:
editor@etsi.fr

ETSI

Contents

Intellectual Property Rights

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (http://www.etsi.org/ipr).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Foreword

This European Standard (Telecommunications series) has been produced by ETSI Technical Committee Special Mobile Group (SMG).

The present document describes the detailed mapping between input blocks of 160 speech samples in 13-bit uniform PCM format to encoded blocks of 244 bits and from encoded blocks of 244 bits to output blocks of 160 reconstructed speech samples within the digital cellular telecommunications system.

National transposition dates	
Date of adoption of this EN:	30 October 1998
Date of latest announcement of this EN (doa):	31 January 1999
Date of latest publication of new National Standard or endorsement of this EN (dop/e):	31 July 1999
Date of withdrawal of any conflicting National Standard (dow):	31 July 1999

1   Scope

The present document describes the detailed mapping between input blocks of 160 speech samples in 13-bit uniform PCM format to encoded blocks of 244 bits and from encoded blocks of 244 bits to output blocks of 160 reconstructed speech samples. The sampling rate is 8 000 sample/s leading to a bit rate for the encoded bit stream of 12,2 kbit/s. The coding scheme is the so-called Algebraic Code Excited Linear Prediction Coder, hereafter referred to as ACELP.

This EN also specifies the conversion between A-law PCM and 13-bit uniform PCM. Performance requirements for the audio input and output parts are included only to the extent that they affect the transcoder performance. This part also describes the codec down to the bit level, thus enabling the verification of compliance to the part to a high degree of confidence by use of a set of digital test sequences. These test sequences are described in GSM 06.54 [7] and are available on disks.

In case of discrepancy between the requirements described in this EN and the fixed point computational description (ANSI-C code) of these requirements contained in GSM 06.53 [6], the description in GSM 06.53 [6] will prevail.

The transcoding procedure specified in this EN is applicable for the enhanced full rate speech traffic channel (TCH) in the GSM system.

In GSM 06.51 [5], a reference configuration for the speech transmission chain of the GSM enhanced full rate (EFR) system is shown. According to this reference configuration, the speech encoder takes its input as a 13-bit uniform PCM signal either from the audio part of the Mobile Station or on the network side, from the PSTN via an 8-bit/A-law to 13-bit uniform PCM conversion. The encoded speech at the output of the speech encoder is delivered to a channel encoder unit which is specified in GSM 05.03 [3]. In the receive direction, the inverse operations take place.

2   References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non-specific.

- For a specific reference, subsequent revisions do not apply.

- For a non-specific reference, the latest version applies.

- A non-specific reference to an ETS shall also be taken to refer to later versions published as an EN with the same number.

[1]   GSM 01.04 (ETR 100): "Digital cellular telecommunications system (Phase 2); Abbreviations and acronyms".

[2]   GSM 03.50 (ETS 300 540): "Digital cellular telecommunications system (Phase 2); Transmission planning aspects of the speech service in the GSM Public Land Mobile Network (PLMN) system".

[3]   GSM 05.03 (ETS 300 575): "Digital cellular telecommunications system (Phase 2); Channel coding".

[4]   GSM 06.32 (ETS 300 580-6): "Digital cellular telecommunications system (Phase 2); Voice Activity Detection (VAD)".

[5]   GSM 06.51 (EN 301 243): "Digital cellular telecommunications system (Phase 2); Enhanced Full Rate (EFR) speech processing functions General description".

[6]   GSM 06.53 (EN 301 244): "Digital cellular telecommunications system (Phase 2); ANSI-C code for the GSM Enhanced Full Rate (EFR) speech codec".

[7]   GSM 06.54 (EN 301 250): "Digital cellular telecommunications system (Phase 2); Test vectors for the GSM Enhanced Full Rate (EFR) speech codec".

[8]　　　ITU-T Recommendation G.711 (1988): "Coding of analogue signals by pulse code modulation Pulse code modulation (PCM) of voice frequencies".

[9]　　　ITU-T Recommendation G.726: "40, 32, 24, 16 kbit/s adaptive differential pulse code modulation (ADPCM)".

3　　Definitions, symbols and abbreviations

3.1　　Definitions

For the purposes of the present document, the following terms and definitions apply:

adaptive codebook: The adaptive codebook contains excitation vectors that are adapted for every subframe. The adaptive codebook is derived from the long term filter state. The lag value can be viewed as an index into the adaptive codebook.

adaptive postfilter: This filter is applied to the output of the short term synthesis filter to enhance the perceptual quality of the reconstructed speech. In the GSM enhanced full rate codec, the adaptive postfilter is a cascade of two filters: a formant postfilter and a tilt compensation filter.

algebraic codebook: A fixed codebook where algebraic code is used to populate the excitation vectors (innovation vectors).The excitation contains a small number of nonzero pulses with predefined interlaced sets of positions.

closed-loop pitch analysis: This is the adaptive codebook search, i.e., a process of estimating the pitch (lag) value from the weighted input speech and the long term filter state. In the closed-loop search, the lag is searched using error minimization loop (analysis-by-synthesis). In the GSM enhanced full rate codec, closed-loop pitch search is performed for every subframe.

direct form coefficients: One of the formats for storing the short term filter parameters. In the GSM enhanced full rate codec, all filters which are used to modify speech samples use direct form coefficients.

fixed codebook: The fixed codebook contains excitation vectors for speech synthesis filters. The contents of the codebook are non-adaptive (i.e., fixed). In the GSM enhanced full rate codec, the fixed codebook is implemented using an algebraic codebook.

fractional lags: A set of lag values having sub-sample resolution. In the GSM enhanced full rate codec a sub-sample resolution of 1/6th of a sample is used.

frame: A time interval equal to 20 ms (160 samples at an 8 kHz sampling rate).

integer lags: A set of lag values having whole sample resolution.

interpolating filter: An FIR filter used to produce an estimate of sub-sample resolution samples, given an input sampled with integer sample resolution.

inverse filter: This filter removes the short term correlation from the speech signal. The filter models an inverse frequency response of the vocal tract.

lag: The long term filter delay. This is typically the true pitch period, or a multiple or sub-multiple of it.

Line Spectral Frequencies: (see Line Spectral Pair).

Line Spectral Pair: Transformation of LPC parameters. Line Spectral Pairs are obtained by decomposing the inverse filter transfer function A(z) to a set of two transfer functions, one having even symmetry and the other having odd symmetry. The Line Spectral Pairs (also called as Line Spectral Frequencies) are the roots of these polynomials on the z-unit circle).

LP analysis window: For each frame, the short term filter coefficients are computed using the high pass filtered speech samples within the analysis window. In the GSM enhanced full rate codec, the length of the analysis window is 240 samples. For each frame, two asymmetric windows are used to generate two sets of LP coefficients. No samples of the future frames are used (no lookahead).

LP coefficients: Linear Prediction (LP) coefficients (also referred as Linear Predictive Coding (LPC) coefficients) is a generic descriptive term for describing the short term filter coefficients.

open-loop pitch search: A process of estimating the near optimal lag directly from the weighted speech input. This is done to simplify the pitch analysis and confine the closed-loop pitch search to a small number of lags around the open-loop estimated lags. In the GSM enhanced full rate codec, open-loop pitch search is performed every 10 ms.

residual: The output signal resulting from an inverse filtering operation.

short term synthesis filter: This filter introduces, into the excitation signal, short term correlation which models the impulse response of the vocal tract.

perceptual weighting filter: This filter is employed in the analysis-by-synthesis search of the codebooks. The filter exploits the noise masking properties of the formants (vocal tract resonances) by weighting the error less in regions near the formant frequencies and more in regions away from them.

subframe: A time interval equal to 5 ms (40 samples at an 8 kHz sampling rate).

vector quantization: A method of grouping several parameters into a vector and quantizing them simultaneously.

zero input response: The output of a filter due to past inputs, i.e. due to the present state of the filter, given that an input of zeros is applied.

zero state response: The output of a filter due to the present input, given that no past inputs have been applied, i.e., given the state information in the filter is all zeroes.

3.2 Symbols

For the purposes of the present document, the following symbols apply:

$A(z)$ The inverse filter with unquantized coefficients

$\hat{A}(z)$ The inverse filter with quantified coefficients

$H(z) = \dfrac{1}{\hat{A}(z)}$ The speech synthesis filter with quantified coefficients

a_i The unquantized linear prediction parameters (direct form coefficients)

$\hat{a}_i$ The quantified linear prediction parameters

m The order of the LP model

$\dfrac{1}{B(z)}$ The long-term synthesis filter

$W(z)$ The perceptual weighting filter (unquantized coefficients)

γ_1, γ_2 The perceptual weighting factors

$F_E(z)$ Adaptive pre-filter

T The nearest integer pitch lag to the closed-loop fractional pitch lag of the subframe

β The adaptive pre-filter coefficient (the quantified pitch gain)

$H_f(z) = \dfrac{\hat{A}(z/\gamma_n)}{\hat{A}(z/\gamma_d)}$ The formant postfilter

γ_n	Control coefficient for the amount of the formant post-filtering
γ_d	Control coefficient for the amount of the formant post-filtering
$H_t(z)$	Tilt compensation filter
γ_t	Control coefficient for the amount of the tilt compensation filtering
$\mu = \gamma_t k_1'$	A tilt factor, with k_1' being the first reflection coefficient
$h_f(n)$	The truncated impulse response of the formant postfilter
L_h	The length of $h_f(n)$
$r_h(i)$	The auto-correlations of $h_f(n)$
$\hat{A}(z/\gamma_n)$	The inverse filter (numerator) part of the formant postfilter
$1/\hat{A}(z/\gamma_d)$	The synthesis filter (denominator) part of the formant postfilter
$\hat{r}(n)$	The residual signal of the inverse filter $\hat{A}(z/\gamma_n)$
$h_t(z)$	Impulse response of the tilt compensation filter
$\beta_{sc}(n)$	The AGC-controlled gain scaling factor of the adaptive postfilter
α	The AGC factor of the adaptive postfilter
$H_{h1}(z)$	Pre-processing high-pass filter
$w_I(n)$, $w_{II}(n)$	LP analysis windows
$L_1^{(I)}$	Length of the first part of the LP analysis window $w_I(n)$
$L_2^{(I)}$	Length of the second part of the LP analysis window $w_I(n)$
$L_1^{(II)}$	Length of the first part of the LP analysis window $w_{II}(n)$
$L_2^{(II)}$	Length of the second part of the LP analysis window $w_{II}(n)$
$r_{ac}(k)$	The auto-correlations of the windowed speech $s'(n)$
$w_{lag}(i)$	Lag window for the auto-correlations (60 Hz bandwidth expansion)
f_0	The bandwidth expansion in Hz
f_s	The sampling frequency in Hz
$r'_{ac}(k)$	The modified (bandwidth expanded) auto-correlations
$E_{LD}(i)$	The prediction error in the ith iteration of the Levinson algorithm
k_i	The ith reflection coefficient

$a_j^{(i)}$ The jth direct form coefficient in the ith iteration of the Levinson algorithm

$F_1^{'}(z)$ Symmetric LSF polynomial

$F_2^{'}(z)$ Antisymmetric LSF polynomial

$F_1(z)$ Polynomial $F_1^{'}(z)$ with root $z = -1$ eliminated

$F_2(z)$ Polynomial $F_2^{'}(z)$ with root $z = 1$ eliminated

q_i The line spectral pairs (LSPs) in the cosine domain

$\mathbf{q}$ An LSP vector in the cosine domain

$\hat{\mathbf{q}}_i^{(n)}$ The quantified LSP vector at the ith subframe of the frame n

ω_i The line spectral frequencies (LSFs)

$T_m(x)$ A mth order Chebyshev polynomial

$f_1(i), f_2(i)$ The coefficients of the polynomials $F_1(z)$ and $F_2(z)$

$f_1^{'}(i), f_2^{'}(i)$ The coefficients of the polynomials $F_1^{'}(z)$ and $F_2^{'}(z)$

$f(i)$ The coefficients of either $F_1(z)$ or $F_2(z)$

$C(x)$ Sum polynomial of the Chebyshev polynomials

x Cosine of angular frequency ω

λ_k Recursion coefficients for the Chebyshev polynomial evaluation

f_i The line spectral frequencies (LSFs) in Hz

$\mathbf{f}^t = [f_1\, f_2 \ldots f_{10}]$ The vector representation of the LSFs in Hz

$\mathbf{z}^{(1)}(n), \mathbf{z}^{(2)}(n)$ The mean-removed LSF vectors at frame n

$\mathbf{r}^{(1)}(n), \mathbf{r}^{(2)}(n)$ The LSF prediction residual vectors at frame n

$\mathbf{p}(n)$ The predicted LSF vector at frame n

$\hat{\mathbf{r}}^{(2)}(n-1)$ The quantified second residual vector at the past frame

$\hat{\mathbf{f}}^k$ The quantified LSF vector at quantization index k

E_{LSP} The LSP quantization error

$w_i, i = 1, \ldots, 10,$ LSP-quantization weighting factors

d_i The distance between the line spectral frequencies f_{i+1} and f_{i-1}

$h(n)$	The impulse response of the weighted synthesis filter
O_k	The correlation maximum of open-loop pitch analysis at delay k
$O_{t_i}, i=1,\ldots,3$	The correlation maxima at delays $t_i, i=1,\ldots,3$
$(M_i, t_i), i=1,\ldots,3$	The normalized correlation maxima M_i and the corresponding delays $t_i, i=1,\ldots,3$

$$H(z)W(z) = \frac{A(z/\gamma_1)}{\hat{A}(z)A(z/\gamma_2)}$$ The weighted synthesis filter

$A(z/\gamma_1)$	The numerator of the perceptual weighting filter
$1/A(z/\gamma_2)$	The denominator of the perceptual weighting filter
T_1	The nearest integer to the fractional pitch lag of the previous (1st or 3rd) subframe
$s'(n)$	The windowed speech signal
$s_w(n)$	The weighted speech signal
$\hat{s}(n)$	Reconstructed speech signal
$\hat{s}'(n)$	The gain-scaled post-filtered signal
$\hat{s}_f(n)$	Post-filtered speech signal (before scaling)
$x(n)$	The target signal for adaptive codebook search
$x_2(n)$, $\mathbf{x}_2^t$	The target signal for algebraic codebook search
$res_{LP}(n)$	The LP residual signal
$c(n)$	The fixed codebook vector
$v(n)$	The adaptive codebook vector
$y(n) = v(n) * h(n)$	The filtered adaptive codebook vector
$y_k(n)$	The past filtered excitation
$u(n)$	The excitation signal
$\hat{u}(n)$	The emphasized adaptive codebook vector
$\hat{u}'(n)$	The gain-scaled emphasized excitation signal
T_{op}	The best open-loop lag
t_{min}	Minimum lag search value
t_{max}	Maximum lag search value
$R(k)$	Correlation term to be maximized in the adaptive codebook search

b_{24} The FIR filter for interpolating the normalized correlation term $R(k)$

$R(k)_t$ The interpolated value of $R(k)$ for the integer delay k and fraction t

b_{60} The FIR filter for interpolating the past excitation signal $u(n)$ to yield the adaptive codebook vector $v(n)$

A_k Correlation term to be maximized in the algebraic codebook search at index k

C_k The correlation in the numerator of A_k at index k

E_{Dk} The energy in the denominator of A_k at index k

$\mathbf{d} = \mathbf{H}^t \mathbf{x}_2$ The correlation between the target signal $x_2(n)$ and the impulse response $h(n)$, i.e., backward filtered target

$\mathbf{H}$ The lower triangular Toepliz convolution matrix with diagonal $h(0)$ and lower diagonals $h(1), \ldots, h(39)$

$\Phi = \mathbf{H}^t \mathbf{H}$ The matrix of correlations of $h(n)$

$d(n)$ The elements of the vector $\mathbf{d}$

$\phi(i,j)$ The elements of the symmetric matrix Φ

$\mathbf{c}_k$ The innovation vector

C The correlation in the numerator of A_k

m_i The position of the i th pulse

ϑ_i The amplitude of the i th pulse

N_p The number of pulses in the fixed codebook excitation

E_D The energy in the denominator of A_k

$res_{LTP}(n)$ The normalized long-term prediction residual

$b(n)$ The sum of the normalized $d(n)$ vector and normalized long-term prediction residual $res_{LTP}(n)$

$s_b(n)$ The sign signal for the algebraic codebook search

$d'(n)$ Sign extended backward filtered target

$\phi'(i,j)$ The modified elements of the matrix Φ, including sign information

$\mathbf{z}^t$, $z(n)$ The fixed codebook vector convolved with $h(n)$

$E(n)$ The mean-removed innovation energy (in dB)