3GPP TS 26.445 V16.2.0 (2021-12)

Technical Specification

3rd Generation Partnership Project;

Technical Specification Group Services and System Aspects;

Codec for Enhanced Voice Services (EVS);

Detailed Algorithmic Description

(Release 16)

 

The present document has been developed within the 3rd Generation Partnership Project (3GPP TM) and may be further elaborated for the purposes of 3GPP.  
The present document has not been subject to any approval process by the 3GPPOrganizational Partners and shall not be implemented.  
This Specification is provided for future development work within 3GPPonly. The Organizational Partners accept no liability for any use of this Specification.  
Specifications and Reports for implementation of the 3GPP TM system should be obtained via the 3GPP Organizational Partners' Publications Offices.

Keywords

UMTS, LTE, EVS, telephony

***3GPP***

Postal address

3GPP support office address

650 Route des Lucioles - Sophia Antipolis

Valbonne - FRANCE

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

Internet

http://www.3gpp.org

***Copyright Notification***

No part may be reproduced except as authorized by written permission.  
The copyright and the foregoing restriction extend to reproduction in all media.

© 2021, 3GPP Organizational Partners (ARIB, ATIS, CCSA, ETSI, TSDSI, TTA, TTC).

All rights reserved.

UMTS™ is a Trade Mark of ETSI registered for the benefit of its members

3GPP™ is a Trade Mark of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners  
LTE™ is a Trade Mark of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners

GSM® and the GSM logo are registered and owned by the GSM Association

Contents

Foreword [14](#__RefHeading___Toc90383920)

1 Scope [14](#__RefHeading___Toc90383921)

2 References [14](#__RefHeading___Toc90383922)

3 Definitions, abbreviations and mathematical expressions [16](#__RefHeading___Toc90383923)

3.1 Definitions [16](#__RefHeading___Toc90383924)

3.2 Abbreviations [17](#__RefHeading___Toc90383925)

3.3 Mathematical Expressions [18](#__RefHeading___Toc90383926)

4 General description of the coder [19](#__RefHeading___Toc90383927)

4.1 Introduction [19](#__RefHeading___Toc90383928)

4.2 Input/output sampling rate [19](#__RefHeading___Toc90383929)

4.3 Codec delay [19](#__RefHeading___Toc90383930)

4.4 Coder overview [20](#__RefHeading___Toc90383931)

4.4.1 Encoder overview [20](#__RefHeading___Toc90383932)

4.4.1.1 Linear Prediction Based Operation [21](#__RefHeading___Toc90383933)

4.4.1.2 Frequency Domain Operation [21](#__RefHeading___Toc90383934)

4.4.1.3 Inactive Signal coding [22](#__RefHeading___Toc90383935)

4.4.1.4 Source Controlled VBR Coding [22](#__RefHeading___Toc90383936)

4.4.2 Decoder overview [22](#__RefHeading___Toc90383937)

4.4.2.1 Parametric Signal Representation Decoding (Bandwidth Extension) [22](#__RefHeading___Toc90383938)

4.4.2.2 Frame loss concealment [23](#__RefHeading___Toc90383939)

4.4.3 DTX/CNG operation [23](#__RefHeading___Toc90383940)

4.4.3.1 Inactive Signal coding [23](#__RefHeading___Toc90383941)

4.4.4 AMR-WB-interoperable option [23](#__RefHeading___Toc90383942)

4.4.5 Channel-Aware Mode [23](#__RefHeading___Toc90383943)

4.5 Organization of the rest of the Technical Standard [24](#__RefHeading___Toc90383944)

5 Functional description of the encoder 25

5.1 Common processing 25

5.1.1 High-pass Filtering 25

5.1.2 Complex low-delay filter bank analysis 25

5.1.2.1 Sub-band analysis 25

5.1.2.2 Sub-band energy estimation 26

5.1.3 Sample rate conversion to 12.8 kHz 27

5.1.3.1 Conversion of 16, 32 and 48 kHz signals to 12.8 kHz 27

5.1.3.2 Conversion of 8 kHz signals to 12.8 kHz 27

5.1.3.3 Conversion of input signals to 16, 25.6 and 32 kHz 29

5.1.4 Pre-emphasis 29

5.1.5 Spectral analysis 30

5.1.5.1 Windowing and DFT 30

5.1.5.2 Energy calculations 31

5.1.6 Bandwidth detection 32

5.1.6.1 Mean and maximum energy values per band 32

5.1.7 Bandwidth decision 34

5.1.8 Time-domain transient detection 37

5.1.9 Linear prediction analysis 38

5.1.9.1 LP analysis window 38

5.1.9.2 Autocorrelation computation 38

5.1.9.3 Adaptive lag windowing 39

5.1.9.4 Levinson-Durbin algorithm 39

5.1.9.5 Conversion of LP coefficients to LSP parameters 40

5.1.9.6 LSP interpolation 41

5.1.9.7 Conversion of LSP parameters to LP coefficients 41

5.1.9.8 LP analysis at 16kHz 42

5.1.10 Open-loop pitch analysis 43

5.1.10.1 Perceptual weighting 43

5.1.10.2 Correlation function computation 44

5.1.10.3 Correlation reinforcement with past pitch values 45

5.1.10.4 Normalized correlation computation 46

5.1.10.5 Correlation reinforcement with pitch lag multiples 46

5.1.10.6 Initial pitch lag determination and reinforcement based on pitch coherence with other half-frames 47

5.1.10.7 Pitch lag determination and parameter update 48

5.1.10.8 Correction of very short and stable open-loop pitch estimates 49

5.1.10.9 Fractional open-loop pitch estimate for each subframe 51

5.1.11 Background noise energy estimation 52

5.1.11.1 First stage of noise energy update 52

5.1.11.2 Second stage of noise energy update 54

5.1.11.2.1 Basic parameters for noise energy update 54

5.1.11.2.2 Spectral diversity 55

5.1.11.2.3 Complementary non-stationarity 55

5.1.11.2.4 HF energy content 56

5.1.11.2.5 Tonal stability 56

5.1.11.2.6 High frequency dynamic range 60

5.1.11.2.7 Combined decision for background noise energy update 60

5.1.11.3 Energy-based parameters for noise energy update 62

5.1.11.3.1 Closeness to current background estimate 62

5.1.11.3.2 Features related to last correlation or harmonic event 62

5.1.11.3.3 Energy-based pause detection 63

5.1.11.3.4 Long-term linear prediction efficiency 63

5.1.11.3.5 Additional long-term parameters used for noise estimation 64

5.1.11.4 Decision logic for noise energy update 65

5.1.12 Signal activity detection 68

5.1.12.1 SAD1 module 69

5.1.12.1.1 SNR outlier filtering 71

5.1.12.2 SAD2 module 72

5.1.12.3 Combined decision of SAD1 and SAD2 modules for WB and SWB signals 75

5.1.12.4 Final decision of the SAD1 module for NB signals 75

5.1.12.5 Post-decision parameter update 76

5.1.12.6 SAD3 module 77

5.1.12.6.1 Sub-band FFT 77

5.1.12.6.2 Computation of signal features 78

5.1.12.6.3 Computation of SNR parameters 81

5.1.12.6.4 Decision of background music 83

5.1.12.6.5 Decision of background update flag 83

5.1.12.6.6 SAD3 Pre-decision 84

5.1.12.6.7 SAD3 Hangover 86

5.1.12.7 Final SAD decision 86

5.1.12.8 DTX hangover addition 88

5.1.13 Coding mode determination 90

5.1.13.1 Unvoiced signal classification 91

5.1.13.1.1 Voicing measure 92

5.1.13.1.2 Spectral tilt 92

5.1.13.1.3 Sudden energy increase from a low energy level 93

5.1.13.1.4 Total frame energy difference 94

5.1.13.1.5 Energy decrease after spike 94

5.1.13.1.6 Decision about UC mode 95

5.1.13.2 Stable voiced signal classification 96

5.1.13.3 Signal classification for FEC 96

5.1.13.3.1 Signal classes for FEC 97

5.1.13.3.2 Signal classification parameters 97

5.1.13.3.3 Classification procedure 98

5.1.13.4 Transient signal classification 99

5.1.13.5 Modification of coding mode in special cases 100

5.1.13.6 Speech/music classification 101

5.1.13.6.1 First stage of the speech/music classifier 101

5.1.13.6.2 Scaling of features in the first stage of the speech/music classifier 103

5.1.13.6.3 Log-probability and decision smoothing 104

5.1.13.6.4 State machine and final speech/music decision 105

5.1.13.6.5 Improvement of the classification for mixed and music content 108

5.1.13.6.6 Second stage of the speech/music classifier 112

5.1.13.6.7 Context-based improvement of the classification for stable tonal signals 114

5.1.13.6.8 Detection of sparse spectral content 118

5.1.13.6.9 Decision about AC mode 120

5.1.13.6.10 Decision about IC mode 120

5.1.14 Coder technology selection 120

5.1.14.1 ACELP/MDCT-based technology selection at 9.6kbps, 16.4 and 24.4 kbps 121

5.1.14.1.1 Segmental SNR estimation of the MDCT-based technology 121

5.1.14.1.2 Segmental SNR estimation of the ACELP technology 127

5.1.14.1.3 Hysteresis and final decision 128

5.1.14.2 TCX/HQ MDCT technology selection at 13.2 and 16.4 kbps 129

5.1.14.3 TCX/HQ MDCT technology selection at 24.4 and 32 kbps 131

5.1.14.4 TD/Multi-mode FD BWE technology selection at 13.2 kbps and 32 kbps 134

5.2 LP-based Coding 135

5.2.1 Perceptual weighting 135

5.2.2 LP filter coding and interpolation 136

5.2.2.1 LSF quantization 136

5.2.2.1.1 LSF weighting function 136

5.2.2.1.2 Bit allocation 139

5.2.2.1.3 Predictor allocation 140

5.2.2.1.4 LSF quantizer structure 140

5.2.2.1.5 LSFQ for voiced coding mode at 16 kHz internal sampling frequency : BC-TCVQ 145

5.2.2.1.6 Mid-frame LSF quantizer 152

5.2.3 Excitation coding 153

5.2.3.1 Excitation coding in the GC, VC and high rate IC/UC modes 153

5.2.3.1.1 Computation of the LP residual signal 154

5.2.3.1.2 Target signal computation 155

5.2.3.1.3 Impulse response computation 155

5.2.3.1.4 Adaptive codebook 155

5.2.3.1.5 Algebraic codebook 158

5.2.3.1.6 Combined algebraic codebook 168

5.2.3.1.7 Gain quantization 182

5.2.3.2 Excitation coding in TC mode 187

5.2.3.2.1 Glottal pulse codebook search 187

5.2.3.2.2 TC frame configurations 191

5.2.3.3 Excitation coding in UC mode at low rates 196

5.2.3.3.1 Structure of the Gaussian codebook 196

5.2.3.3.2 Correction of the Gaussian codebook spectral tilt 197

5.2.3.3.3 Search of the Gaussian codebook 198

5.2.3.3.4 Quantization of the Gaussian codevector gain 199

5.2.3.3.5 Other parameters in UC mode 200

5.2.3.3.6 Update of filter memories 200

5.2.3.4 Excitation coding in IC and UC modes at 9.6 kbps 200

5.2.3.4.1 Algebraic codebook 201

5.2.3.4.2 Gaussian noise generation 202

5.2.3.4.3 Gain coding 202

5.2.3.4.4 Memory update 204

5.2.3.5 Excitation coding in GSC mode 204

5.2.3.5.1 Determining the subframe length 205

5.2.3.5.2 Computing time-domain excitation contribution 205

5.2.3.5.3 Frequency transform of residual and time-domain excitation contribution 206

5.2.3.5.4 Computing energy dynamics of transformed residual and quantization of noise level 207

5.2.3.5.6 Find and encode the cut-off frequency 207

5.2.3.5.7 Band energy computation and quantization 209

5.2.3.5.8 PVQ Bit allocation 209

5.2.3.5.9 Quantization of difference signal 210

5.2.3.5.10 Spectral dynamic and noise filling 210

5.2.3.5.11 Quantized gain addition, temporal and frequency contributions combination 210

5.2.3.5.12 Specifics for wideband 8kbps 211

5.2.3.5.13 Inverse DCT 212

5.2.3.5.14 Remove pre-echo in case of onset detection 212

5.2.4 Bass post-filter gain quantization 213

5.2.5 Source Controlled VBR Coding 213

5.2.5.1 Principles of VBR Coding 213

5.2.5.2 EVS VBR Encoder Coding Modes and Bit-Rates 214

5.2.5.3 Prototype-Pitch-Period (PPP) Encoding 214

5.2.5.3.1 PPP Algorithm 214

5.2.5.3.2 Amplitude Quantization 216

5.2.5.3.3 Phase Quantization 216

5.2.5.4 Noise-Excited-Linear-Prediction (NELP) Encoding 216

5.2.5.5 Average Data Rate (ADR) Control for the EVS VBR mode 216

5.2.6 Coding of upper band for LP-based Coding Modes 220

5.2.6.1 Bandwidth extension in time domain 220

5.2.6.1.1 High band target signal generation 221

5.2.6.1.2 TBE LP analysis 222

5.2.6.1.3 Quantization of linear prediction parameters 224

5.2.6.1.4 Interpolation of LSF coefficients 227

5.2.6.1.5 Target and residual energy calculation and quantization 229

5.2.6.1.6 Generation of the upsampled version of the lowband excitation 229

5.2.6.1.7 Non-Linear Excitation Generation 230

5.2.6.1.8 Spectral flip of non-linear excitation in time domain 231

5.2.6.1.9 Down-sample using all-pass filters 231

5.2.6.1.10 Adaptive spectral whitening 232

5.2.6.1.11 Envelope modulated noise mixing 232

5.2.6.1.12 Spectral shaping of the noise added excitation 234

5.2.6.1.13 Post processing of the shaped excitation 235

5.2.6.1.14 Estimation of temporal gain shape parameters 236

5.2.6.1.15 Estimation of frame gain parameters 239

5.2.6.1.16 Estimation of TEC/TFA envelope parameters 241

5.2.6.1.17 Estimation of full-band frame energy parameters 244

5.2.6.2 Multi-mode FD Bandwidth Extension Coding 246

5.2.6.2.1 SWB/FB Multi-mode FD Bandwidth Extension 246

5.2.6.2.2 WB Multi-mode FD Bandwidth Extension 257

5.2.6.3 Coding of upper band at 64 kb/s 261

5.2.6.3.1 Coding in normal mode 261

5.2.6.3.2 Coding in transient mode 265

5.3 MDCT Coding Mode 268

5.3.1 General description 268

5.3.2 Time-to-frequency transformations 268

5.3.2.1 Transform sizes and MDCT configurations 268

5.3.2.2 Long block transformation (ALDO window) 268

5.3.2.2.1 Folding and on-the-fly window decimation 270

5.3.2.2.2 eDCT 272

5.3.2.3 Transient location dependent overlap and transform length 274

5.3.2.4 Short block transformation 275

5.3.2.4.2 Short window transform for MDCT based TCX 279

5.3.2.5 Special window transitions 280

5.3.2.5.1 ALDO to short transition 280

5.3.2.5.2 Short to ALDO transition 280

5.3.2.6 Modified Discrete Sine Transform 280

5.3.3 MDCT based TCX 281

5.3.3.1 General description 281

5.3.3.1.1 High level overview 281

5.3.3.1.2 Rate dependent configuration 281

5.3.3.2 General encoding procedure 282

5.3.3.2.1 LPC parameter calculation 282

5.3.3.2.2 Temporal Noise Shaping 286

5.3.3.2.3 LPC shaping in MDCT domain 288

5.3.3.2.4 Adaptive low frequency emphasis 291

5.3.3.2.5 Spectrum noise measure in power spectrum 292

5.3.3.2.6 Low pass factor detector 293

5.3.3.2.7 Uniform quantizer with adaptive dead-zone 293

5.3.3.2.8 Arithmetic coder 293

5.3.3.2.9 Global gain coding 305

5.3.3.2.10 Noise Filling 306

5.3.3.2.11 Intelligent Gap Filling 309

5.3.3.2.12 Memory updates 324

5.3.3.2.13 Global Gain Adjuster 324

5.3.4 High Quality MDCT coder (HQ) 325

5.3.4.1 Low-rate HQ coder 325

5.3.4.1.1 Tonality Estimation 326

5.3.4.1.2 Grouping of spectral coefficients 326

5.3.4.1.3 Energy Envelope coding 331

5.3.4.1.4 MDCT coefficients quantization 338

5.3.4.2 High-rate HQ coder 374

5.3.4.2.1 Normal Mode 376

5.3.4.2.2 Transient Mode 386

5.3.4.2.3 Generic, Harmonic and HVQ mode detector 387

5.3.4.2.4 Harmonic Mode 388

5.3.4.2.5 HVQ 389

5.3.4.2.6 Generic Mode 392

5.3.4.2.7 Pyramid Vector Quantization (PVQ) and indexing 400

5.4 Switching of Coding Modes 410

5.4.1 General description 410

5.4.2 MDCT coding mode to CELP coding mode 410

5.4.2.1 MDCT to CELP transition 1 (MC1) 411

5.4.2.2 MDCT to CELP transition 2 (MC2) 411

5.4.2.3 MDCT to CELP transition 3 (MC3) 411

5.4.3 CELP coding mode to MDCT coding mode 411

5.4.3.1 CELP coding mode to MDCT based TCX coding mode 411

5.4.3.2 CELP coding mode to HQ MDCT coding mode 412

5.4.3.2.1 Constrained CELP coding and simplified BWE coding 412

5.4.3.2.2 HQ MDCT coding with a modified analysis window 413

5.4.4 Internal sampling rate switching 414

5.4.4.1 Reset of LPC memory 414

5.4.4.2 Conversion of LP filter between 12.8 and 16 kHz internal sampling rates 414

5.5.4.1.1 Modification of the Power Spectrum 414

5.5.4.1.2 Computation of the Power Spectrum 415

5.5.4.1.3 Computation of the Autocorrelation 416

5.4.4.3 Extrapolation of LP filter 417

5.4.4.4 Buffer resampling with linear interpolation 417

5.4.4.5 Update of CELP input signal memories 417

5.4.4.6 Update of MDCT-based TCX input signal memories 418

5.4.4.7 Update of CELP synthesis memories 418

5.4.5 EVS primary and AMR-WB IO 418

5.4.5.1 Switching from primary modes to AMR-WB IO 418

5.4.5.2 Switching from AMR-WB IO mode to primary modes 419

5.4.6 Rate switching 419

5.4.6.1 Rate switching along with internal sampling rate switching 419

5.4.6.2 Rate switching along with coding mode switching 419

5.5 Frame erasure concealment side information 420

5.5.1 Signal classification parameter 420

5.5.2 Energy information 420

5.5.3 Phase control information 421

5.5.4 Pitch lag information 422

5.5.5 Spectral envelope diffuser 422

5.5.6 Tonality flag information 423

5.6 DTX/CNG operation 424

5.6.1 Overview 424

5.6.1.1 SID update 425

5.6.1.2 Spectral tilt based SID transmission 426

5.6.1.3 CNG selector 427

5.6.2 Encoding for LP-CNG 427

5.6.2.1 LP-CNG CN parameters estimation 428

5.6.2.1.1 LP-CNG Hangover analysis period determination 428

5.6.2.1.2 LP-CNG filter parameters evaluation for low-band signal 428

5.6.2.1.3 LP-CNG CNG-LSF quantization for low-band signal 429

5.6.2.1.4 LP-CNG synthesis filter computation for local CNG synthesis 430

5.6.2.1.5 LP-CNG energy calculation and quantization 431

5.6.2.1.6 LP-CNG energy smoothing for local CNG synthesis 432

5.6.2.1.7 LP-CNG LF-BOOST determination and quantization 432

5.6.2.1.8 LP-CNG high band analysis and quantization 433

5.6.2.2 LP-CNG local CNG synthesis 435

5.6.2.3 LP-CNG CNG Memory update 435

5.6.3 Encoding for FD-CNG 435

5.6.3.1 Spectral partition energies 436

5.6.3.1.1 Computation of the FFT partition energies 436

5.6.3.1.2 Computation of the CLDFB partition energies 436

5.6.3.1.3 FD-CNG configurations 436

5.6.3.2 FD-CNG noise estimation 437

5.6.3.2.1 Dynamic range compression for the input energies 437

5.6.3.2.2 Noise tracking 438

5.6.3.2.3 Dynamic range expansion for the estimated noise energies 442

5.6.3.3 Adjusting the first SID frame in FD-CNG 442

5.6.3.4 FD-CNG resetting mechanism 442

5.6.3.5 Encoding SID frames in FD-CNG 442

5.6.3.6 FD-CNG local CNG synthesis 444

5.6.3.6.1 SID parameters interpolation 445

5.6.3.6.2 LPC estimation from the interpolated SID parameters 445

5.6.3.6.3 FD-CNG encoder comfort noise generation 445

5.6.3.6.4 FD-CNG encoder memory update 445

5.7 AMR-WB-interoperable modes 446

5.7.1 Pre-processing 446

5.7.2 Linear prediction analysis and quantization 446

5.7.2.1 Windowing and auto**‑**correlation computation 446

5.7.2.2 Levinson**‑**Durbin algorithm 446

5.7.2.3 LP to ISP conversion 446

5.7.2.4 ISP to LP conversion 447

5.7.2.5 Quantization of the ISP coefficients 447

5.7.2.6 Interpolation of the ISPs 447

5.7.3 Perceptual weighting 447

5.7.4 Open-loop pitch analysis 447

5.7.5 Impulse response computation 447

5.7.6 Target signal computation 447

5.7.7 Adaptive codebook search 448

5.7.8 Algebraic codebook search 448

5.7.9 Quantization of the adaptive and fixed codebook gains 448

5.7.10 Memory update 448

5.7.11 High-band gain generation 448

5.7.12 CNG coding 448

5.8 Channel Aware Coding 448

5.8.1 Introduction 448

5.8.2 Principles of Channel Aware Coding 449

5.8.3 Bit-Rate Allocation for Primary and Partial Redundant Frame Coding 450

5.8.3.1 Primary frame bit-rate reduction 450

5.8.3.2 Partial Redundant Frame Coding 450

5.8.3.2.1 Construction of partial redundant frame for Generic and Voiced Coding modes 450

5.8.3.2.2 Construction of partial redundant frame for Unvoiced Coding mode 450

5.8.3.2.3 Construction of partial redundant frame for TCX frame 451

5.8.3.2.4 RF\_NO\_DATA partial redundant frame type 451

5.8.3.3 Decoding 451

5.8.4 Channel aware mode encoder configurable parameters 451

6 Functional description of the Decoder 453

6.1 LP-based Decoding 453

6.1.1 General LP-based decoding 453

6.1.1.1 LSF decoding 453

6.1.1.1.1 General LSF decoding 453

6.1.1.1.2 LSF decoding for voiced coding mode at 16 kHz internal sampling frequency 456

6.1.1.2 Reconstruction of the excitation 458

6.1.1.2.1 Reconstruction of the excitation in GC and VC modes and high rate IC/UC modes 458

6.1.1.2.2 Reconstruction of the excitation in TC mode 465

6.1.1.2.3 Reconstruction of the excitation in UC mode at low rates 465

6.1.1.2.4 Reconstruction of the excitation in IC/UC mode at 9.6 kbps 467

6.1.1.2.5 Reconstruction of the excitation in GSC 468

6.1.1.3 Excitation post-processing 469

6.1.1.3.1 Anti-sparseness processing 469

6.1.1.3.2 Gain smoothing for noise enhancement 469

6.1.1.3.3 Pitch enhancer 470

6.1.1.3.4 Music post processing 471

6.1.2 Source Controlled VBR decoding 478

6.1.3 Synthesis 478

6.1.4 Post-processing 479

6.1.4.1.1 Long-term post-filter 479

6.1.4.1.2 Short-term post-filter 480

6.1.4.1.3 Post-filter NB parameters 480

6.1.4.1.4 Post-filter WB and SWB parameters 481

6.1.4.1.5 Tilt compensation 481

6.1.4.1.6 Adaptive gain control 482

6.1.4.2 Bass post-filter 482

6.1.5 Decoding of upper band for LP-based Coding Modes 485

6.1.5.1 Decoding Time-domain Bandwidth Extension 485

6.1.5.1.1 Generation of the upsampled version of the lowband excitation 485

6.1.5.1.2 Non-Linear Excitation Generation 486

6.1.5.1.3 De-quantization of high band parameters 487

6.1.5.1.4 LSP interpolation 487

6.1.5.1.5 Spectral flip in time domain 487

6.1.5.1.6 Down-sample using all-pass filters 487

6.1.5.1.7 Adaptive spectral whitening 488

6.1.5.1.8 Envelope modulated noise mixing 488

6.1.5.1.9 Spectral shaping of the noise added excitation 490

6.1.5.1.10 Post processing of the shaped excitation 490

6.1.5.1.11 Gain shape update 490

6.1.5.1.12 SHB synthesis 491

6.1.5.1.13 Core-switching and high-band memory updates 492

6.1.5.1.14 TEC/TFA post processing 493

6.1.5.1.15 Full-band synthesis 495

6.1.5.2 Multi-mode FD Bandwidth Extension decoding 495

6.1.5.2.1 SWB multi-mode FD BWE decoding 495

6.1.5.2.2 WB multi-mode FD BWE decoding 503

6.1.5.3 Decoding of upper band at 64 kb/s 511

6.1.5.3.1 Decoding in normal mode 511

6.1.5.3.2 Decoding in transient mode 516

6.1.5.3.3 Windowing and frequency-to-time transformation 518

6.1.5.3.4 Post-processing in temporal domain 518

6.2 MDCT Coding mode decoding 521

6.2.1 General MDCT decoding 521

6.2.2 MDCT based TCX 521

6.2.2.1 Rate dependent configurations 521

6.2.2.2 Init module parameters 521

6.2.2.2.1 TCX block configuration 521

6.2.2.2.2 LPC parameter 521

6.2.2.2.3 PLC Wavefrom adjustment 522

6.2.2.2.4 Global Gain 522

6.2.2.2.5 Noise fill parameter 523

6.2.2.2.6 LTP 523

6.2.2.2.7 TNS parameter 523

6.2.2.2.8 Harmonic model 523

6.2.2.2.9 IGF bit stream reader 523

6.2.2.2.10 Spectral data 526

6.2.2.2.11 Residual bits 526

6.2.2.3 Decoding process 527

6.2.2.3.1 Arithmetic decoder 527

6.2.2.3.2 Adaptive low frequency de-emphasis 528

6.2.2.3.3 Global gain decoding 529

6.2.2.3.4 Residual bits decoding 529

6.2.2.3.5 TCX formant enhancement 530

6.2.2.3.6 Noise Filling 531

6.2.2.3.7 Apply global gain and LPC shaping in MDCT domain 532

6.2.2.3.8 IGF apply 533

6.2.2.3.9 Inverse window grouping (TCX5 separation) 538

6.2.2.3.10 Temporal Noise Shaping 539

6.2.2.3.11 IGF temporal flattening 539

6.2.3 High Quality MDCT decoder (HQ) 540

6.2.3.1 Low-rate HQ decoder 540

6.2.3.1.1 Mode decoding 540

6.2.3.1.2 Energy Envelope decoding 540

6.2.3.1.3 Spectral coefficients decoding 543

6.2.3.2 High-rate HQ decoder 553

6.2.3.2.1 Normal Mode 553

6.2.3.2.2 Transient Mode 561

6.2.3.2.3 Harmonic Mode 562

6.2.3.2.4 HVQ 564

6.2.3.2.5 Generic Mode 565

6.2.3.2.6 PVQ decoding and de-indexing 570

6.2.4 Frequency-to-time transformation 572

6.2.4.1 Long block transformation (ALDO window) 572

6.2.4.1.1 eDCT 572

6.2.4.1.2 Unfolding and windowing 572

6.2.4.1.3 Overlap-add 573

6.2.4.1.4 Pre-echo attenuation 574

6.2.4.2 Transient location dependent overlap and transform length 580

6.2.4.3 Short block transformation 580

6.2.4.3.1 Short window transform in TDA domain 580

6.2.4.3.2 Short window transform for MDCT based TCX 581

6.2.4.4 Special window transitions 581

6.2.4.4.1 ALDO to short transition 581

6.2.4.4.2 Short to ALDO transition 582

6.2.4.5 Low Rate MDCT Synthesis 582

6.3 Switching coding modes in decoding 582

6.3.1 General description 582

6.3.2 MDCT coding mode to CELP coding mode 583

6.3.2.1 MDCT to CELP transition 1 (MC1) 583

6.3.2.2 MDCT to CELP transition 2 (MC2) 584

6.3.2.3 MDCT to CELP transition 3 (MC3) 584

6.3.3 CELP coding mode to MDCT coding mode 584

6.3.3.1 CELP coding mode to MDCT based TCX coding mode 585

6.3.3.2 CELP coding mode to HQ MDCT coding mode 586

6.3.3.2.1 Constrained CELP decoding and simplified BWE decoding 587

6.3.3.2.1.1 Optimized cubic interpolation 587

6.3.3.2.2 HQ MDCT decoding with a modified synthesis window 588

6.3.3.2.3 Cross-fading 588

6.3.4 Internal sampling rate switching 588

6.3.4.1 Reset of LPC memory 588

6.3.4.2 Conversion of LP filter between 12.8 and 16 kHz internal sampling rates 589

6.3.4.3 Extrapolation of LP filter 589

6.3.4.4 Update of CELP synthesis memories 589

6.3.4.5 Update of CELP decoded past signal 589

6.3.4.6 Post-processing 589

6.3.4.6.1 Adaptive post-filtering 589

6.3.4.6.2 Bass post filter 589

6.3.4.7 CLDFB 589

6.3.5 EVS primary modes and AMR-WB IO 590

6.3.5.1 Switching from primary modes to AMR-WB IO 590

6.3.5.2 Switching from AMR-WB IO mode to primary modes 590

6.3.6 Rate switching 590

6.3.6.1 Rate switching along with internal sampling rate switching 590

6.3.6.2 Rate switching along with coding mode switching 590

6.3.6.3 Adaptive post-filter reset and smoothing 590

6.3.7 Bandwidth switching 590

6.3.7.1 Bandwidth switching detector 590

6.3.7.2 Super wideband switching to wideband 591

6.3.7.2.1 TBE mode 591

6.3.7.2.2 Multi-mode FD BWE mode 593

6.3.7.2.3 MDCT core 594

6.3.7.3 Wideband switching to super wideband 594

6.4 De-emphasis 594

6.5 Resampling to the output sampling frequency 594

6.6 Decoding of frame erasure concealment side information 594

6.7 Decoding in DTX/CNG operation 595

6.7.1 Overview 595

6.7.2 Decoding for LP-CNG 595

6.7.2.1 LP-CNG decoding Overview 595

6.7.2.1.1 CNG parameter updates in active periods 595

6.7.2.1.2 DTX-hangover based parameter analysis in LP-CNG mode 596

6.7.2.1.3 LP-CNG low-band energy decoding 598

6.7.2.1.4 LP-CNG low-band filter parameters decoding 598

6.7.2.1.5 LP-CNG low-band excitation generation 598

6.7.2.1.6 LP-CNG low-band synthesis 600

6.7.2.1.7 LP-CNG high-band decoding and synthesis 600

6.7.2.2 Memory update 602

6.7.3 Decoding for FD-CNG 602

6.7.3.1 Decoding SID frames in FD-CNG 602

6.7.3.1.1 SID parameters decoding 603

6.7.3.1.2 SID parameters interpolation 603

6.7.3.1.3 LPC estimation from the interpolated SID parameters 604

6.7.3.2 Noise tracking during active frames in FD-CNG 604

6.7.3.2.1 Spectral partition energies 604

6.7.3.2.2 FD-CNG noise estimation 605

6.7.3.2.3 Noise shaping in FD-CNG 605

6.7.3.3 Noise generation for SID or zero frames in FD-CNG 607

6.7.3.3.1 Update of the noise levels for FD-CNG 607

6.7.3.3.2 Comfort noise generation in the frequency domain 607

6.7.3.3.3 Comfort noise generation in the time domain 607

6.7.3.3.4 FD-CNG decoder memory update 609

6.8 AMR-WB-interoperable modes 610

6.8.1 Decoding and speech synthesis 610

6.8.1.1 Excitation decoding 610

6.8.1.2 Excitation post-processing 611

6.8.1.2.1 Anti-sparseness processing 611

6.8.1.2.2 Gain smoothing for noise enhancement 611

6.8.1.2.3 Pitch enhancer 611

6.8.1.3 Synthesis filtering 611

6.8.1.4 Music and Unvoiced/inactive Post-processing 611

6.8.1.4.1 Music post processing 611

6.8.1.4.2 Unvoiced and inactive post processing 612

6.8.1.5 Synthesis filtering and overwriting the current CELP synthesis 615

6.8.1.6 Formant post-filter 615

6.8.1.7 Comfort noise addition 615

6.8.1.8 Bass post-filter 615

6.8.2 Resampling 616

6.8.3 High frequency band 616

6.8.3.1 Preliminary estimation steps 616

6.8.3.1.1 Estimation of tilt, figure of merit and voice factors 617

6.8.3.1.2 Estimation of sub-frame gains based on LP spectral envelopes 618

6.8.3.2 Generation of high-band excitation 621

6.8.3.2.1 DCT 621

6.8.3.2.2 High band generation 621

6.8.3.2.3 Extraction of tonal and ambiance components 623

6.8.3.2.4 Recombination 624

6.8.3.2.5 Filtering in DCT domain 625

6.8.3.2.6 Inverse DCT 626

6.8.3.2.7 Gain computation and scaling of excitation 626

6.8.3.3 LP filter for the high frequency band 628

6.8.3.4 High band synthesis 628

6.8.4 CNG decoding 629

6.9 Common post-processing 629

6.9.1 Comfort noise addition 629

6.9.1.1 Noisy speech detection 629

6.9.1.2 Noise estimation for CNA 630

6.9.1.2.1 CNA noise estimation in DTX-on mode when FD-CNG is triggered 630

6.9.1.2.2 CNA noise estimation in DTX-on mode when LP-CNG is triggered 630

6.9.1.2.3 CNA noise estimation in DTX-off mode 630

6.9.1.3 Noise generation in the FFT domain and addition in the time domain 630

6.9.1.4 Noise generation and addition in the MDCT domain 631

6.9.2 Long term prediction processing 631

6.9.2.1 Decoding LTP parameters 631

6.9.2.2 LTP post filtering 632

6.9.3 Complex low delay filter bank synthesis 634

6.9.4 High pass filtering 635

7 Description of the transmitted parameter indices 635

7.1 Bit allocation for the default option 635

7.1.1 Bit allocation at VBR 5.9, 7.2 – 9.6 kbps 635

7.1.2 Bit allocation at 13.2 kbps 636

7.1.3 Bit allocation at 16.4 and 24.4 kbps 636

7.1.4 Bit allocation at 32 kbps 637

7.1.5 Bit allocation at 48, 64, 96 and 128 kbps 638

7.2 Bit allocation for SID frames in the DTX operation 639

7.3 Bit allocation for the AMR-WB-interoperable option 640

7.4 Bit Allocation for the Channel-Aware Mode 640

Annex A (normative): RTP Payload Format and SDP Parameters 642

A.0 General 642

A.1 RTP Header Usage 642

A.2 EVS RTP Payload Format 642

A.2.1 EVS codec Compact Format 642

A.2.1.1 Compact format for EVS Primary mode 643

A.2.1.2 Compact format for EVS AMR-WB IO mode (except SID) 643

A.2.1.2.1 Representation of Codec Mode Request (CMR) in Compact format for EVS AMR-WB IO mode 643

A.2.1.2.2 Payload structure of Compact EVS AMR-WB IO mode frame 644

A.2.1.3 Special case for 56 bit payload size (EVS Primary or EVS AMR-WB IO SID) 645

A.2.2 EVS codec Header-Full format 645

A.2.2.1 EVS RTP payload structure 645

A.2.2.1.1 CMR byte 646

A.2.2.1.2 ToC byte 649

A.2.2.1.3 Speech Data 651

A.2.2.1.4 Zero padding 651

A.2.2.1.4.1 Zero padding for octet alignment of speech data (EVS AMR-WB IO) 651

A.2.2.1.4.2 Zero padding for size collision avoidance 651

A.2.2.1.4.3 Additional zero padding 651

A.2.3 Header-Full/Compact format handling 651

A.2.3.1 Default format handling 652

A.2.3.2 Header-Full-only format handling 652

A.2.4 AMR-WB backward compatible EVS AMR-WB IO mode format 652

A.2.5 Sessions with multiple mono channels 652

A.2.5.1 Encoding of multiple mono channels 652

A.2.5.2 RTP header usage 652

A.2.5.3 Construction of the RTP payload 652

A.2.6 Storage Format 653

A.2.6.1 Header 653

A.2.6.2 Speech Frames 654

A.3 Payload Format Parameters 654

A.3.1 EVS Media Type Registration 654

A.3.2 Mapping Media Type Parameters into SDP 657

A.3.3 Detailed Description of Usage of SDP Parameters 657

A.3.3.1 Offer-Answer Model Considerations 657

A.3.3.2 Examples 659

A.3.3.3 Interactions of the dtx and dtx-recv parameters 660

Annex B (informative): Change history 661

# Foreword

This Technical Specification has been produced by the 3rd Generation Partnership Project (3GPP).

The contents of the present document are subject to continuing work within the TSG and may change following formal TSG approval. Should the TSG modify the contents of the present document, it will be re-released by the TSG with an identifying change of release date and an increase in version number as follows:

Version x.y.z

where:

x the first digit:

1 presented to TSG for information;

2 presented to TSG for approval;

3 or greater indicates TSG approved document under change control.

y the second digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc.

z the third digit is incremented when editorial only changes have been incorporated in the document.

# 1 Scope

The present document is a detailed description of the signal processing algorithms of the Enhanced Voice Services coder.

# 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. In the case of a reference to a 3GPP document (including a GSM document), a non-specific reference implicitly refers to the latest version of that document *in the same Release as the present document*.

[1] 3GPP TR 21.905: "Vocabulary for 3GPP Specifications".

[2] 3GPP TS 26.441: "Codec for Enhanced Voice Services (EVS); General Overview".

[3] 3GPP TS 26.442: "Codec for Enhanced Voice Services (EVS); ANSI C code (fixed-point)".

[4] 3GPP TS 26.444: "Codec for Enhanced Voice Services (EVS); Test Sequences".

[5] 3GPP TS 26.446: "Codec for Enhanced Voice Services (EVS); AMR-WB Backward Compatible Functions".

[6] 3GPP TS 26.447: "Codec for Enhanced Voice Services (EVS); Error Concealment of Lost Packets".

[7] 3GPP TS 26.448: "Codec for Enhanced Voice Services (EVS); Jitter Buffer Management".

[8] 3GPP TS 26.173: "ANSI-C code for the Adaptive Multi Rate - Wideband (AMR-WB) speech codec".

[9] 3GPP TS 26.190: "Speech codec speech processing functions; Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; Transcoding functions".

[10] 3GPP TS 26.103: "Speech codec list for GSM and UMTS".

[11] 3GPP TS 26.290: "Extended AMR Wideband codec; Transcoding functions".

[12] 3GPP TS 26.401: "General audio codec audio processing functions; Enhanced aacPlus general audio codec; General description".

[13] 3GPP TS 26.114: "IP Multimedia Subsystem (IMS); Multimedia Telephony; Media handling and interaction".

[14] 3GPP TS 22.105: "Services and Service capabilities".

[15] IETF RFC 4867 (2007): "RTP Payload Format and File Storage Format for the Adaptive Multi-Rate (AMR) and Adaptive Multi-Rate Wideband (AMR-WB) Audio Codecs", J. Sjoberg, M. Westerlund, A. Lakaniemi and Q. Xie.

[16] IEEE Transactions on Acoustic, Speech and Signals Processing, Vol. 2, April, pp. 9-12 (1993): 'Immittance spectral pairs (ISP) for speech encoding',Bistritz, Y., and Peller, S.

[17] IEEE Journal on Selected Areas in Communications, Vol. 6, No. 2, February, pp. 314-323 (1988): "Transform Coding of Audio Signals Using Perceptual Noise Criteria", Johnston, J.D.

[18] IEEE Transactions on Acoustic, Speech and Signals Processing, Vol. 34, No. 6, December, pp. 1419 1426 (1986): "The computation of line spectral frequencies using Chebyshev polynomials", Kabal, P., and Ramachandran, R.P.

[19] IEEE Proceedings of International Conference on Acoustics, Speech and Signal Processing, March-April, pp. 4001-4004 (2008): "Transition mode coding for source controlled CELP codecs", Eksler, V., and Jelínek, M.

[20] 3GPP2 C.S0014-B, (2006) "Enhanced Variable Rate Codec, Speech Service 3 Option 3 and 68 for Wideband Spread Spectrum 4 Digital Systems."

[21] Speech and Audio Processing for Coding, Enhancement and Recognition, Springer (2014): Recent Speech Coding Technologies and Standards", Sinder, D., Varga, I., Krishnan, V. Rajendran, V, and Villette, S.

[22] Proceedings of EUSIPCO (2002): "LSF quantization with multiple scale lattice VQ for transmission over noisy channels", Vasilache, A. and Tabus, I.

[23] Proceedings of INTERSPEECH (2013): "Computationally efficient objective function for algebraic codebook optimization in ACELP", Bäckström, T.

[24] Proceedings of INTERSPEECH (2012): "Enumerative algebraic coding for ACELP", Bäckström, T.

[25] ITU-T Recommendation G.718 Amendment 2 (2010): "New Annex B on Superwideband scalable extension for ITU-T G.718 and corrections to main body fixed-point C-code and description text."

[26] Signal Processing, no. 82, pp 563-586 (2002): "Multiple-scale leader-lattice VQ with application to LSF quantization", Vasilache, A. and Tabus, I.

[27] IETF RFC 4566 (2006): "SDP: Session Description Protocol", M. Handley, V. Jacobson and C. Perkins.

[28] IETF RFC 3264 (2002): "An Offer/Answer Model with the Session Description Protocol (SDP)", J. Rosenberg and H. Schulzrinne.

[29] IETF RFC 5939 (2010): "Session Description Protocol (SDP) Capability Negotiation", F. Andreasen.

[30] IETF RFC 3550 (2003): "RTP: A Transport Protocol for Real-Time Applications", H. Schulzrinne, S. Casner, R. Frederick and V. Jacobson.

[31] IETF RFC 0768 (1980): "User Datagram Protocol", J. Postel.

[32] ITU-T Recommendation G.722 Amendment 1 (2010): "7 kHz audio-coding within 64 kbit/s: New Annex B with superwideband embedded extension."

[33] ITU-T Recommendation G.720.1 (2010): "Generic sound activity detector."

[34] American Mathematical Society, (2002): "Numerical Analysis: Mathematics of Scientific Computing", Kincaid, D. and Cheney, W.

[35] ITU-T Recommendation G.722.2 (2003): "Wideband coding of speech at around 16 kbit/s using Adaptive Multi-Rate Wideband (AMR-WB)."

[36] 3GPP TS 26.201: "Speech codec speech processing functions; Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; Frame structure".

[37] 3GPP TS 26.171: "Speech codec speech processing functions; Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; General description".

[38] IETF RFC 3551 (2003): "RTP Profile for Audio and Video Conferences with Minimal Control", H. Schulzrinne and S. Casner.

[39] 3GPP TS 26.103: "Speech codec list for GSM and UMTS"

[40] 3GPP TS 26.244: "3GPP file format (3GP)".

[41] IETF RFC 3839 (2004): "MIME Type Registrations for 3rd Generation Partnership Project (3GPP) Multimedia files", R. Castagno and D. Singer.

[42] IEEE Transactions on Information Theory, Vol. 37, no. 6, Nov., pp. 1551-1566 (1991): "Trellis Coded Vector Quantization", T.R. Fischer, M.W. Marcellin, and M. Wang.

[43] 3GPP TS 26.443: "Codec for Enhanced Voice Services (EVS); ANSI C code (floating-point)"

[44] 3GPP TS 26.452: "Codec for Enhanced Voice Services (EVS); ANSI C code; Alternative fixed-point using updated basic operators".

# 3 Definitions, abbreviations and mathematical expressions

## 3.1 Definitions

For the purposes of the present document, the terms and definitions given in TR 21.905 [1] and the following apply. A term defined in the present document takes precedence over the definition of the same term, if any, in TR 21.905 [1].

**frame:** an array of audio samples spanning 20ms time duration.

**EVS codec**: The EVS codec includes two operational modes: EVS Primary operational mode (EVS Primary mode) and EVS AMR-WB Inter-Operable modes (EVS AMR-WB IO mode). When using the EVS AMR-WB IO mode the speech frames are bitstream interoperable with the AMR-WB codec [9]. Frames generated by an EVS AMR-WB IO mode encoder can be decoded by an AMR-WB decoder, without the need for transcoding. Likewise, frames generated by an AMR-WB encoder can be decoded by an EVS AMR-WB IO mode decoder, without the need for transcoding.

**EVS Primary mode:** Includes 11 bit-rates for fixed-rate or multi-rate operation; 1 average bit-rate for variable bit-rate operation; and 1 bit-rate for SID [2]. The EVS Primary mode can encode narrowband, wideband, super-wideband and fullband signals. None of these bit-rates are interoperable with the AMR-WB codec.

**EVS AMR-WB IO mode:** Includes 9 codec modes and SID. All are bitstream interoperable with the AMR-WB codec (TS 26.171 ‎[37‎]).

**Operational mode:** Used for the EVS codec to distinguish between EVS Primary mode and EVS AMR-WB IO mode.

## 3.2 Abbreviations

For the purposes of the present document, the abbreviations given in TR 21.905 [1] and the following apply. An abbreviation defined in the present document takes precedence over the definition of the same abbreviation, if any, in TR 21.905 [1].

ACELP Algebraic Code-Excited Linear Prediction

AMR Adaptive Multi Rate (codec)

AMR-NB Adaptive Multi Rate Narrowband (codec) – Also referred to as AMR

AMR-WB Adaptive Multi Rate Wideband (codec)

AR Auto-Regressive

AVQ Algebraic Vector Quantization

CELP Code-Excited Linear Prediction

CLDFB Complex Low Delay Filter Bank

CMR Codec Mode Request

CNG Comfort Noise Generator

DCT Discrete Cosine Transformation

DFT Discrete Fourier Transform

DTX Discontinuous Transmission

EFR Enhanced Full Rate (codec)

EVS Enhanced Voice Services

FB Fullband

FCB Fixed Codebook

FEC Frame Erasure Concealment

FFT Fast Fourier Transform

FIR Finite Impulse Response

FT Frame Type

GC Generic Coding

HF High Frequency

HP High Pass

IIR Infinite Impulse Response

IP Internet Protocol

ISF Immittance Spectral Frequency

ISP Immittance Spectral Pair

ISPP Interleaved Single-Pulse Permutation

JBM Jitter Buffer Management

LD Low Delay

LP Linear Prediction

LPF Low Pass Filter

LSB Least Significant Bit

LSF Line Spectral Frequency

LSP Line Spectral Pair

LTP Long-Term Prediction

MA Moving Average

MDCT Modified Discrete Cosine Transform

MRLVQ Multi-Rate Lattice Vector Quantization

MSB Most Significant Bit

MSVQ Multi-Stage Vector Quantization

MTSI Multimedia Telephony Service for IMS

NB Narrowband

OL Open-Loop

PCPRS Path-Choose Pulse Replacement Search

PS Packet Switched

PSTN Public Switched Telephone Network

QMF Quadrature Mirror Filter (See also CLDFB)

SAD Signal Activity Detection

SDP Session Description Protocol

SID Silence Insertion Descriptor

SNR Signal-to-Noise Ratio

SWB Super Wideband

TC Transition Coding

ToC Table of Contents

UC Unvoiced Coding

VC Voiced Coding

VMR-WB Variable Rate Multimode Wideband

VQ Vector Quantization

WB Wideband

WMOPS Weighted Millions of Operations Per Second

## 3.3 Mathematical Expressions

For the purposes of the present document, the following conventions apply to mathematical expressions:

indicates the largest integer less than or equal to *x*: , and ;

indicates the smallest integer greater than or equal to *x*: , and

|*x*| indicates the absolute value of *x*: |17| = 17, |–17| = 17;

min(*x*0,…*x*N–1) indicates the minimum of *x*0,…, *x*N–1, N being the number of components;

max(*x*0,…*x*N–1) indicates the maximum of *x*0, …, *x*N–1;

**A**T indicates the transpose of matrix **A**;

indicates the remainder after dividing x by y: ;

round(*x*) is traditional rounding: round(*x*) = ;

exp(*x*) is equivalent to *e*x where *e* is the base of the natural algorithm;

indicates summation;

indicates product;

Unless otherwise specified, log(*x*) denotes logarithm at base 10 throughout this Recommendation.

# 4 General description of the coder

## 4.1 Introduction

The present document is a detailed description of the signal processing algorithms of the Enhanced Voice Services coder. The detailed mapping from 20ms input blocks of audio samples in 16 bit uniform PCM format to encoded blocks of bits and from encoded blocks of bits to output blocks of reconstructed audio samples is explained. Four sampling rates are supported; 8 000, 16 000, 32 000 and 48 000 samples/s and the bit rates for the encoded bit stream of may be 5.9, 7.2, 8.0, 9.6, 13.2, 16.4, 24.4, 32.0, 48.0, 64.0, 96.0 or 128.0 kbit/s. An AMR-WB Interoperable mode is also provided which operates at bit rates for the encoded bit stream of 6.6, 8.85, 12.65, 14.25, 15.85, 18.25, 19.85, 23.05 or 23.85 kbit/s.

The procedure of this document is mandatory for implementation in all network entities and User Equipment (UE)s supporting the EVS coder.

The present document does not describe the ANSI-C code of this procedure. For a description of the two reference fixed-point ANSI C code specifications, using different sets of basic operators, see [3] and [44] respectively; for a description of the reference floating-point ANSI C code specification see [43].

In the case of discrepancy between the procedure described in the present document and its ANSI-C code specifications contained in [3] the procedure defined by the [3] prevails. In the case of discrepancy between the procedure described in the present document and its ANSI-C code specifications contained in [43] the procedure defined by [43] prevails. In the case of discrepancy between the procedure described in the present document and its ANSI-C code specifications contained in [44] the procedure defined by [44] prevails.

## 4.2 Input/output sampling rate

The encoder can accept fullband (FB), superwideband (SWB), wideband (WB) or narrow-band (NB) signals sampled at 48, 32, 16 or 8 kHz. Similarly, the decoder output can be 48, 32, 16 or 8 kHz, FB, SWB, WB or NB.

## 4.3 Codec delay

The input signal is processed using 20 ms frames. The codec delay depends on the output sampling rate. For WB, SWB and FB output (i.e. output sampling rate > 8 kHz) the overall algorithmic delay is 32 ms. It consists of one 20 ms frame, 0.9375 ms delay of input resampling filters on the encoder-side, 8.75 ms for the encoder look-ahead, and 2.3125 ms delay of time-domain bandwidth extension on the decoder-side. For 8 kHz decoder output the decoder delay is reduced to 1.25 ms needed for resampling using a complex low-delay filterbank, resulting in 30.9375 ms overall algorithmic delay.

## 4.4 Coder overview

The EVS codec employs a hybrid coding scheme combining linear predictive (LP) coding techniques based upon ACELP (Algebraic Code Excited Linear Prediction), predominantly for speech signals, with a transform coding method, for generic content, as well as inactive signal coding in conjunction with VAD/DTX/CNG (Voice Activity Detection/Discontinuous Transmission/ Comfort Noise Generation) operation. The EVS codec is capable of switching between these different coding modes without artefacts.

The EVS codec supports 5.9kbps narrowband and wideband variable bit rate (VBR) operation based upon the ACELP coding paradigm which also provides the AMR-WB interoperable encoding and decoding. In addition to perceptually optimized waveform matching, the codec utilizes parametric representations of certain frequency ranges. These parametric representations constitute coded bandwidth extensions or noise filling strategies.

The decoder generates the signal parameters represented by the indices transmitted in the bit-stream. For the bandwidth extension and noise fill regions estimates from the coded regions are used in addition to the decoded parametric data to generate the signals for these frequency regions.

The description of the coding algorithm of this Specification is made in terms of bit-exact, fixed-point mathematical operations. The ANSI C code described in [3], which constitutes an integral part of this Specification, reflects this bit-exact, fixed-point descriptive approach. The mathematical descriptions of the encoder (clause 5), and decoder (clause 6), can be implemented in different ways, possibly leading to a codec implementation not complying with this Specification. Therefore, the algorithm description of the ANSI C code of [3] shall take precedence over the mathematical descriptions of clauses 5 and 6 whenever discrepancies are found. A non-exhaustive set of test signals that can be used with the ANSI C code are available described in [4].

### 4.4.1 Encoder overview

Figure 1 represents a high-level overview of the encoder. The signal resampling block corrects mismatches between the sampling frequency and the signal bandwidth command line parameter that is specified on the command line or through a file containing a bandwidth switching profile as explained in TS 26.442 and TS 26.443 and TS 26.452. For the case that the signal bandwidth is lower than half the input sampling frequency, the signal is decimated to a the lowest possible sampling rate out of the set of (8, 16, 32 kHz) that is larger than twice the signal bandwidth.



Figure 2: Encoder overview

The signal analysis determines which of three possible encoder strategies to employ: LP based coding (ACELP), frequency domain encoding and inactive coding (CNG). In some operational modes the signal analysis step includes a closed loop decision to determine which encoding method will result in the lowest distortion. Further parameters derived in the signal analysis aid the operation of these coding blocks and some of the analysis parameters, such as the coding strategy to be employed, are encoded into the bit-stream. In each of the coding blocks the signal analysis is further refined to obtain parameters relevant for the particular coding block.

The signal analysis and sub-sequent decision of the coding mode is performed independently for each 20ms frame and the switching between different modes is possible on a frame-by-frame basis. In the switching instance parameters are exchanged between the coding modes to ensure that the switching is as seamless as possible and closed-loop methods are sometimes employed in this case. In addition, switching between different bandwidths and/or bit-rates (including both the EVS Primary mode and the EVS AMR-WB IO mode) is possible on frame boundaries.

The signal analysis and all other blocks have full access to the command line parameters such as bit-rate, sampling rate, signal bandwidth, DTX activation as signalling information.

#### 4.4.1.1 Linear Prediction Based Operation

The input signal is split into high frequency band and low frequency band paths; where the cut-off frequency between these two bands is determined from the operational mode (bandwidth and bit-rate) of the codec.

The linear-prediction coefficient estimation is performed for every 20ms frame. Within a frame, several interpolation points are established depending upon the bitrate and the optimum interpolation is transmitted to the decoder. The linear-prediction residual is further analysed and quantized using different quantization schemes dependent upon the nature of the residual. For the 5.9kbps VBR operation additional low-rate coding modes at rates conforming to the design constraints are employed.

The high-frequency portion of the signal is represented with several different parametric representations. The parameters used for this representation vary as a function of the bit-rate and the residual quantization strategy. The transmitted parameters include some or all of spectral envelope, energy information and temporal evolution information.

The LP based core can be configured so that both the linear prediction coefficients and the residual quantization are interoperable with the AMR-WB decoder. For this purpose the configuration of the LP coefficient estimation, parametric HF representation and the residual quantization is similar to those of AMR-WB. For the AMR-WB interoperable operation modes, identical codebooks to the AMR-WB quantizers are used.



Figure 3: Linear prediction based operation including parametric HF representation

#### 4.4.1.2 Frequency Domain Operation

For the frequency domain coding the encoding block can be envisaged as being separated into a control layer and a signal processing layer. The control layer performs signal analysis to derive several control and configuration parameters for the signal processing layer. The time-to-frequency transformation is based on the Modified Discrete Cosine Transform (MDCT) and provides adaptive time-frequency resolution. The control layer derives measures of the time distribution of the signal energy in a frame and controls the transform.

The MDCT coefficients are quantized using a variety of direct and parametric representations depending upon bit rate signal type and operating mode.



Figure 4: Frequency domain encoder

#### 4.4.1.3 Inactive Signal coding

When the codec is operated in DTX on mode the signal classifier depicted in Figure 1 selects the discontinuous transmission (DTX) mode for frames that are determined to consist of background noise. For these frames a low-rate parametric representation of the signal is transmitted no more frequently than every 8 frames (160ms).

The low-rate parametric representation is used in the decoder for comfort noise generation (CNG) and includes parameters describing the frequency envelope of the background signal, energy parameters describing the overall energy and its time evolution.

#### 4.4.1.4 Source Controlled VBR Coding

VBR coding describes a method that assigns different number of bits to a speech frame in the coded domain depending on the characteristics of the input speech signal [20] [21]. This method is often called source-controlled coding as well. Typically, a source-controlled coder encodes speech at different bit rates depending on how the current frame is classified, e.g., voiced, unvoiced, transient, or silence. Note that DTX operation can be combined with VBR coders in the same way as with Fixed Rate (FR) coders; the VBR operation is related to active speech segments.

The VBR solution provides narrowband and wideband coding using the bit rates 2.8, 7.2 and 8.0 kbps and produces an average bit rate at 5.9 kbps.

Due to the finer bit allocation, in comparison to Fixed Rate (FR) coding, VBR offers the advantage of a better speech quality at the same average active bit rate than FR coding at the given bit rate. The benefits of VBR can be exploited if the transmission network supports the transmission of speech frames (packets) of variable size, such as in LTE and UMTS networks.

### 4.4.2 Decoder overview

The decoder receives all quantized parameters and generates a synthesized signal. Thus, for the majority of encoder operations it represents the inverse of the quantized value to index operations.

For the AMR-WB interoperable operation the index lookup is performed using the AMR-WB codebooks and the decoder is configured to generate an improved synthesized signal from the AMR-WB bitstream.

#### 4.4.2.1 Parametric Signal Representation Decoding (Bandwidth Extension)

In addition to the generation of signal components specifically represented by the transmitted indices, the decoder performs estimates of signal regions where the transmitted signal representation is incomplete, i.e. for the parametric signal representations and noise fill as well as blind bandwidth extension in some cases.

#### 4.4.2.2 Frame loss concealment

The EVS codec includes frame loss concealment algorithms [21]. For all coding modes an extrapolation algorithm is in place that estimates the signal in a lost frame. For the LP based core this estimation operates on the last received residual and LP coefficients. For the frequency domain core in some cases the last received MDCT coefficients are extrapolated and in addition the resulting time domain signal is guaranteed to give a smooth time evolution from the last received frame into the missing frames.

Once the frame loss is recovered, i.e., the first good frame is received the codec memory is updated and frame boundary mismatches towards the last lost frame are minimized.

For situations of sustained frame loss the signal is either faded to background noise or its energy is reduced and finally muted when no reasonable extrapolation can be assumed.

The EVS codec also includes the "channel aware" mode, which may be employed for improved performance under packet loss conditions in a VoIP system. In the channel aware mode, partial copies (secondary frames) of the current speech frame are piggybacked on future speech frames (primary frames), without any increase in the total bit rate for the primary and secondary frames. If the current frame is lost, then its partial copy can be retrieved by polling the de-jitter buffer to enable faster and improved recovery from the packet loss.

### 4.4.3 DTX/CNG operation

The codec is equipped with a signal activity detection (SAD) algorithm for classifying each input frame as active or inactive. It supports a discontinuous transmission (DTX) operation in which the comfort noise generation (CNG) module is used to approximate and update the statistics of the background noise at a variable bit rate. Thus, the transmission rate during inactive signal periods is variable and depends on the estimated level of the background noise. By default in the command line the transmission rate of CNG update is fixed to 8 frames. However, the CNG update rate can also be set to another fixed value or a variable rate by means of a command line parameter; when the transmission rate during inactive signal periods is variable, it depends on the estimated level of the background noise.

#### 4.4.3.1 Inactive Signal coding

When the codec is operated in DTX on mode the signal classifier depicted in Figure 1 selects the discontinuous transmission (DTX) mode for frames that are determined to consist of background noise. For these frames a low-rate parametric representation of the signal is transmitted no more frequently than every 8 frames (160ms).

The low-rate parametric representation is used in the decoder for comfort noise generation (CNG) and includes parameters describing the frequency envelope of the background signal, energy parameters describing the overall energy and its time evolution.

### 4.4.4 AMR-WB-interoperable option

As mentioned previously, EVS can operate in a mode which is fully interoperable with the AMR-WB codec bitstream.

### 4.4.5 Channel-Aware Mode

EVS offers partial redundancy [21] based error robust channel aware mode at 13.2 kbps for both wideband and super-wideband audio bandwidths.

In a VoIP system, packets arrive at the decoder with random jitters in their arrival time. Packets may also arrive out of order at the decoder. Since the decoder expects to be fed a speech packet every 20 ms to output speech samples in periodic blocks, a de-jitter buffer is required to absorb the jitter in the packet arrival time. The channel aware mode combines the presence of a de-jitter buffer with partial redundancy coding of a current frame which gets piggy backed onto a future frame. At the receiver, the de-jitter buffer is polled to check if a partial redundant copy of the current lost frame is available in any of the future frames. If present, the partial redundant information is used to synthesize the lost frame which offers significant quality improvements under low to high FER conditions. Source control is used to determine which frames of input can best be coded at a reduced frame rate (called primary frames) to accommodate the attachment of redundancy without altering the total packet size. In this way, the channel aware mode includes redundancy in a constant-bit-rate channel (13.2 kbps).

## 4.5 Organization of the rest of the Technical Standard

In clauses 5 and 6, detailed descriptions of the encoder and the decoder are given. Bit allocation is summarized in clause 7.