

ETSI TS 126 441 V13.0.0 (2016-01)



**Universal Mobile Telecommunications System (UMTS);
LTE;
Codec for Enhanced Voice Services (EVS);
General overview
(3GPP TS 26.441 version 13.0.0 Release 13)**



Reference

RTS/TSGS-0426441vd00

Keywords

LTE,UMTS

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

Important noticeThe present document can be downloaded from:
<http://www.etsi.org/standards-search>

The present document may be made available in electronic versions and/or in print. The content of any electronic and/or print versions of the present document shall not be modified without the prior written authorization of ETSI. In case of any existing or perceived difference in contents between such versions and/or in print, the only prevailing document is the print of the Portable Document Format (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://portal.etsi.org/tb/status/status.asp>

If you find errors in the present document, please send your comment to one of the following services:
<https://portal.etsi.org/People/CommitteeSupportStaff.aspx>

Copyright Notification

No part may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm except as authorized by written permission of ETSI.

The content of the PDF version shall not be modified without the written authorization of ETSI.
The copyright and the foregoing restriction extend to reproduction in all media.

© European Telecommunications Standards Institute 2016.
All rights reserved.

DECT™, **PLUGTESTS™**, **UMTS™** and the ETSI logo are Trade Marks of ETSI registered for the benefit of its Members.
3GPP™ and **LTE™** are Trade Marks of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners.
GSM® and the GSM logo are Trade Marks registered and owned by the GSM Association.

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 (<https://ipr.etsi.org/>).

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 Technical Specification (TS) has been produced by ETSI 3rd Generation Partnership Project (3GPP).

The present document may refer to technical specifications or reports using their 3GPP identities, UMTS identities or GSM identities. These should be interpreted as being references to the corresponding ETSI deliverables.

The cross reference between GSM, UMTS, 3GPP and ETSI identities can be found under <http://webapp.etsi.org/key/queryform.asp>.

Modal verbs terminology

In the present document "**shall**", "**shall not**", "**should**", "**should not**", "**may**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the [ETSI Drafting Rules](#) (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in ETSI deliverables except when used in direct citation.

Contents

Intellectual Property Rights	2
Foreword.....	2
Modal verbs terminology.....	2
Foreword.....	4
1 Scope	5
2 References	5
3 Abbreviations	5
4 General	6
5 Transcoding functions	8
6 ANSI C-code	9
7 Test vectors	9
8 Discontinuous transmission (DTX).....	10
9 Voice activity detection.....	10
10 Comfort noise insertion	10
11 Error concealment of lost frames	11
12 Frame structure.....	11
13 RTP Payload Format	11
14 Jitter Buffer Management.....	11
15 Performance characterization	11
Annex A (informative): Change history	12
History	13

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 an introduction to the audio processing parts of the Enhanced Voice Services coder. A general overview of the audio processing functions is given, with reference to the documents where each function is specified in detail.

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 26.445: "Codec for Enhanced Voice Services (EVS); Detailed Algorithmic Description".
- [3] 3GPP TS 26.442: "Codec for Enhanced Voice Services (EVS); ANSI C code (fixed-point)".
- [4] 3GPP TS 26.443: "Codec for Enhanced Voice Services (EVS); ANSI C code (floating point)".
- [5] 3GPP TS 26.444: "Codec for Enhanced Voice Services (EVS); Test Sequences".
- [6] 3GPP TS 26.446: "Codec for Enhanced Voice Services (EVS); AMR-WB Backward Compatible Functions".
- [7] 3GPP TS 26.447: "Codec for Enhanced Voice Services (EVS); Error Concealment of Lost Packets".
- [8] 3GPP TS 26.448: "Codec for Enhanced Voice Services (EVS); Jitter Buffer Management".
- [9] 3GPP TS 26.449: "Codec for Enhanced Voice Services (EVS); Comfort Noise Generation (CNG) Aspects".
- [10] 3GPP TS 26.450: "Codec for Enhanced Voice Services (EVS); Discontinuous Transmission (DTX)".
- [11] 3GPP TS 26.451: "Codec for Enhanced Voice Services (EVS); Voice Activity Detection (VAD)".
- [12] 3GPP TS 26.114: "IP Multimedia Subsystem (IMS); Multimedia telephony; Media handling and interaction".
- [13] 3GPP TS 26.131: "Terminal acoustic characteristics for telephony; Requirements".
- [14] 3GPP TR 26.952: "Codec for Enhanced Voice Services (EVS); Performance Characterization".

3 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-WB	Adaptive Multi Rate Wideband (codec)
CNG	Comfort Noise Generator
DTX	Discontinuous Transmission
EVS	Enhanced Voice Services
FB	Fullband
FEC	Frame Erasure Concealment
IP	Internet Protocol
JBM	Jitter Buffer Management
MSB	Most Significant Bit
MTSI	Multimedia Telephony Service for IMS
NB	Narrowband
PS	Packet Switched
PSTN	Public Switched Telephone Network
SAD	Sound Activity Detection
SC-VBR	Source Controlled - Variable Bit Rate
SID	Silence Insertion Descriptor
SWB	Super Wideband
VAD	Voice Activity Detection
WB	Wideband
WMOPS	Weighted Millions of Operations Per Second

4 General

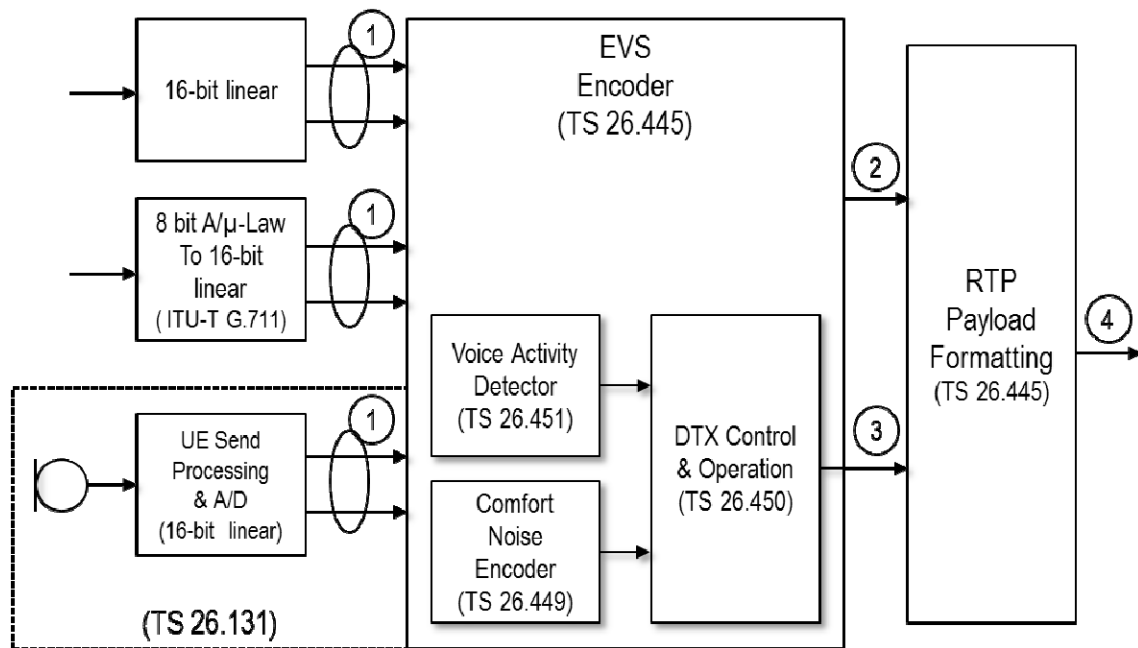
The Enhanced Voice Services coder consists of the multi-rate audio coder optimized for operation with voice and music/mixed content signals, a source controlled rate scheme including a voice/sound activity detector and a comfort noise generation system, and an error concealment mechanism to combat the effects of transmission errors and lost packets. Jitter buffer management is also provided.

The EVS coder also provides enhanced interoperation with AMR-WB over all nine source bit rates from 6,6 kbit/s to 23,85 kbit/s.

The EVS speech coder is capable of switching its bit-rate every 20 ms speech frame upon command.

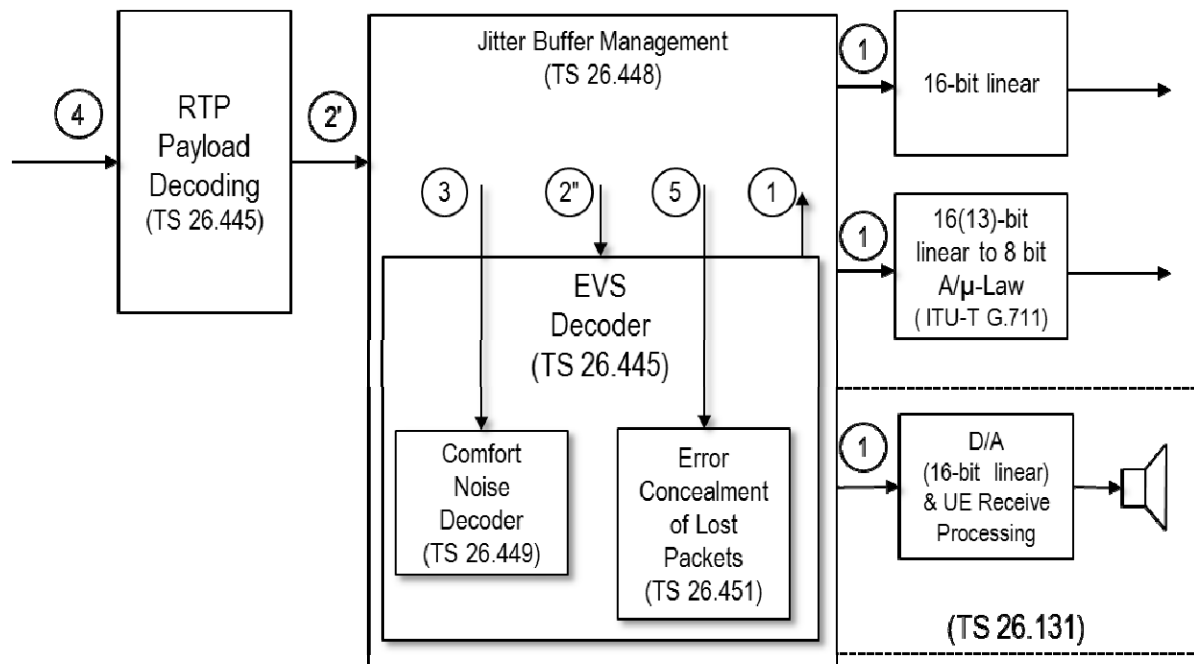
A reference configuration where the various speech processing functions are identified is given in Figure 1. In this figure, the relevant specifications for each function are also indicated.

In Figures 1 & 2, the audio parts including analogue to digital and digital to analogue conversion are included, to show the complete speech path between the audio input/output in the User Equipment (UE) and the digital interface of the network. The detailed specification of the audio parts is not within the scope of the present document. These aspects are only considered to the extent that the performance of the audio parts affect the performance of the speech transcoder.



- ① 16-bit Linear PCM Samples and Sample Rate (8, 16, 32 or 48 kHz)
- ② Encoded audio frame, 50 frames/s, number of bits/frame depending on the EVS codec mode
- ③ Encoded Silence Descriptor frames (variable frame rate)
- ④ RTP Payload Packets

Figure 1: Overview of audio processing functions - Transmit Side



- ① 16-bit Linear PCM Samples and Sample Rate (8, 16, 32 or 48 kHz)
- ②' ②'' Impaired received audio frame, nominally 50 frames/s, number of bits/frame depending on the EVS codec mode
- ③ Encoded Silence Descriptor frames (variable frame rate)
- ④ RTP Payload Packets
- ⑤ Bad Frame Indication

Figure 2: Overview of audio processing functions - Receive Side

5 Transcoding functions

The Enhanced Voice Services codec algorithm is described in [2].

As shown in Figures 1 & 2, the audio encoder takes its input, and can produce an output at the decoder, in the form of a 16-bit uniform Pulse Code Modulated (PCM) signal at sampling frequencies of 8 kHz, 16 kHz, 32 kHz or 48 kHz. The audio may originate from and terminate within the audio part of the UE or from the network side or from the Public Switched Telephone Network (PSTN) via a narrowband 13-bit A-law or μ -law (8 kHz) PCM conversion or wideband (16 kHz) 14-bit uniform PCM conversion.

The detailed mapping between input blocks of audio samples in 16-bit uniform PCM format to encoded blocks (in which the number of bits depends on the presently used codec mode) and from these to output blocks of reconstructed 16-bit uniform PCM format audio samples is described in [2]. The bit-rates of the source codec are listed in Table 1. Table 2 shows the nine bit rates of AMR-WB where interoperable performance is enhanced with the EVS codec. This functionality is described in [2] and [6].

An EVS audio codec capable UE shall support all source rates listed in Table 1.

Table 1: Source codec bit-rates for the EVS codec

Source codec bit-rate (kbit/s)	Audio bandwidths supported	Source Controlled Operation Available
5,9 (SC-VBR)	NB, WB	Yes (Always On)
7,2	NB, WB	Yes
8,0	NB, WB	Yes
9,6	NB, WB, SWB	Yes
13,2	NB, WB, SWB	Yes
13,2 (channel aware)	WB, SWB	Yes
16,4	NB, WB, SWB, FB	Yes
24,4	NB, WB, SWB, FB	Yes
32	WB, SWB, FB	Yes
48	WB, SWB, FB	Yes
64	WB, SWB, FB	Yes
96	WB, SWB, FB	Yes
128	WB, SWB, FB	Yes

Table 2: Source codec bit-rates for the AMR-WB Interoperable Modes of the EVS codec

Source codec bit-rate (kbit/s)
6,6
8,85
12,65
14,25
15,85
18,25
19,85
23,05
23,85

6 ANSI C-code

The ANSI C-code of the EVS codec, VAD, CNG and Jitter Buffer Manager (JBM) are described in [3] for fixed point arithmetic operation and are described in [4] for floating point arithmetic operation.

The ANSI C-code is mandatory.

7 Test vectors

A set of digital test sequences is specified in [5], thus enabling the verification of compliance, i.e. bit-exactness, to a high degree of confidence.

The EVS encoder and decoder (see Figure 1) are defined in bit exact arithmetic. Consequently, they shall react on being presented with a given input sequence always with the corresponding bit exact output sequence, provided that the internal state variables are also always exactly in the same state at the beginning of the test.

The input test sequences provided shall produce the corresponding output test sequences, provided that the tested modules are in their home state when starting.

The modules may be set into their home states by provoking the appropriate homing functions.

NOTE: This is normally done during reset (initialization of the codec).

Special inband signalling frames (encoder homing frame and decoder homing frame) described in [2] have been defined to provoke these homing functions also in remotely placed modules.

At the end of the first received homing frame, the audio functions that are defined in a bit exact way shall go into their predefined home states. The output corresponding to the first homing frame is dependent on the codec state when the frame was received. Any consecutive homing frames shall produce corresponding homing frames at the output.

8 Discontinuous transmission (DTX)

The discontinuous transmission (DTX) of the Enhanced Voice Services codec is defined in [2] & [10].

During a normal telephone conversation, the participants alternate so that, on the average, each direction of transmission is occupied about 50 % of the time. Source controlled operation is a mode of operation where the speech encoder encodes speech frames containing only background noise with a lower bit-rate than normally used for encoding speech. A network may adapt its transmission scheme to take advantage of the varying bit-rate. This may be done for the following two purposes:

- 1) In the UE, battery life will be prolonged or a smaller battery could be used for a given operational duration.
- 2) The average required bit-rate is reduced, leading to a more efficient transmission with decreased load and hence increased capacity.

The following functions are provided by the EVS codec for the source controlled rate operation:

- a Voice Activity Detector (VAD) or more accurately Sound Activity Detector (SAD) on the TX side;
- evaluation of the background acoustic noise on the TX side, in order to transmit characteristic parameters to the RX side;
- generation of comfort noise on the RX side during periods when no normal speech frames are received.

The transmission of comfort noise information to the RX side is achieved by means of a Silence Descriptor (SID) frame, which is sent at regular intervals.

- a Source-Controlled Variable Bit Rate operational mode (SC-VBR) operating at 5,9 kbit/s is also provided to reduce the operational bit rate still further for NB and WB speech services.

9 Voice activity detection

The Enhanced Voice Services codec VAD function is described in [2] & [11].

The input to the VAD is the input speech itself together with a set of parameters computed by the EVS encoder. The VAD uses this information to decide whether each 20 ms encoder frame contains active signal or not.

The VAD algorithm is described in [2], and the corresponding C-code is defined in [3] & [4]. The verification of compliance to [5] is achieved by use of digital test sequences applied to the same interface as the test sequences for the EVS codec.

10 Comfort noise insertion

The EVS codec comfort noise insertion function is described in [2] & [9].

When audio is absent, the synthesis in the EVS decoder is different from the case when normal speech frames are received. The synthesis of an artificial noise based on the received non-speech parameters is termed comfort noise generation.

The comfort noise generation process is as follows:

- the evaluation of the acoustic background noise in the transmitter;
- the noise parameter encoding (SID frames) and decoding, and
- the generation of comfort noise in the receiver.

The comfort noise processes and the algorithm for updating the noise parameters during speech pauses are defined in detail in [2], and the corresponding C-code is defined in [3] & [4].

11 Error concealment of lost frames

The EVS coder error concealment of erroneous or lost frames is described in [2] & [7].

Frames may be erroneous due to transmission errors or frames may be lost or delayed due to packet loss in a transport network. The methods described in [2] & [7] may be used as a basis for error concealment.

In order to mask the effect of isolated erroneous/lost frames, the decoder shall be informed about erroneous/lost frames and the error concealment actions shall be initiated, whereby a set of predicted parameters are used in the speech synthesis. Insertion of speech signal independent silence frames is not allowed.

12 Frame structure

The EVS coder frame structure is described in [2].

13 RTP Payload Format

The EVS coder RTP Payload Format for media handling and interaction is described in [2].

14 Jitter Buffer Management

The EVS coder Jitter Buffer Management is described in [8].

15 Performance characterization

The EVS coder performance characterization is described in [14].

Annex A (informative): Change history

Change history							
Date	TSG #	TSG Doc.	CR	Rev	Subject/Comment	Old	New
06-2014	64	SP-140208			Presented to TSG SA#64 for information		1.0.0
09-2014	65	SP-140457			Presented to TSG SA#65 for approval	1.0.0	2.0.0
09-2014	65				Approved a TSG SA#65	2.0.0	12.0.0
12-2014	66	SP-140723	0001		Typographical correction	12.0.0	12.1.0
12-2015	70				Version for Release 13	12.1.0	13.0.0

History

Document history		
V13.0.0	January 2016	Publication