ETSI TS 126 448 V18.0.0 (2024-05)

TECHNICAL SPECIFICATION

Universal Mobile Telecommunications System (UMTS);
LTE;
5G;
Codec for Enhanced Voice Services (EVS);
Jitter Buffer Management
(3GPP TS 26.448 version 18.0.0 Release 18)

A GLOBAL INITIATIVE

Reference

RTS/TSGS-0426448vi00

Keywords

5G,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 - APE 7112B
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° w061004871

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to normative deliverables may have been declared to ETSI. The declarations pertaining to these essential IPRs, if any, are 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 Directives including the ETSI IPR Policy, no investigation regarding the essentiality of IPRs, 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.

Trademarks

The present document may include trademarks and/or tradenames which are asserted and/or registered by their owners. ETSI claims no ownership of these except for any which are indicated as being the property of ETSI, and conveys no right to use or reproduce any trademark and/or tradename. Mention of those trademarks in the present document does not constitute an endorsement by ETSI of products, services or organizations associated with those trademarks.

DECT™, **PLUGTESTS™**, **UMTS™** and the ETSI logo are trademarks of ETSI registered for the benefit of its Members. **3GPP™** and **LTE™** are trademarks of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners. **oneM2M™** logo is a trademark of ETSI registered for the benefit of its Members and of the oneM2M Partners. **GSM**® and the GSM logo are trademarks registered and owned by the GSM Association.

Legal Notice

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. These shall be interpreted as being references to the corresponding ETSI deliverables.

The cross reference between 3GPP and ETSI identities can be found under https://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

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 defines the Jitter Buffer Management solution for the Codec for Enhanced Voice Services (EVS).

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 26.114: "IP Multimedia Subsystem (IMS); Multimedia telephony; Media handling and interaction".

[4] 3GPP TS 26.071: "Mandatory speech CODEC speech processing functions; AMR speech Codec; General description".

[5] 3GPP TS 26.171: "Speech codec speech processing functions; Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; General description".

[6] 3GPP TS 26.442: "Codec for Enhanced Voice Services (EVS); ANSI C code (fixed-point)".

[7] 3GPP TS 26.443: "Codec for Enhanced Voice Services (EVS); ANSI C code (floating-point)".

[8] 3GPP TS 26.131: "Terminal acoustic characteristics for telephony; Requirements".

[9] 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.

[10] 3GPP TS 26.452: "Codec for Enhanced Voice Services (EVS); ANSI C code; Alternative fixed-point using updated basic operators".

3 Definitions, symbols and abbreviations

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].

3.2 Symbols

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

$s_x(n)$ Time signal and time index n in context x, e.g. x can be inp, out, HP, pre, etc.

L_x Frame length / size of module x

E_x Energy values in context of x

C_x Correlation function in context x

3.3 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].

AMR Adaptive Multi Rate (codec)
AMR-WB Adaptive Multi Rate Wideband (codec)
CNG Comfort Noise Generator
DTX Discontinuous Transmission
EVS Enhanced Voice Services
FB Fullband
FIFO First In, First Out
IP Internet Protocol
JBM Jitter Buffer Management
MTSI Multimedia Telephony Service for IMS
NB Narrowband
PCM Pulse Code Modulation
PLC Packet Loss Concealment
RTP Real Time Transport Protocol
SID Silence Insertion Descriptor
SOLA Synchronized overlap-add
SWB Super Wideband
TSM Time Scale Modification
VAD Voice Activity Detection
WB Wideband

3.4 Mathematical Expressions

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

$\lceil x \rceil$ indicates the smallest integer greater than or equal to *x*: $\lceil 1.1 \rceil = 2$, $\lceil 2.0 \rceil = 2$ and $\lceil -1.1 \rceil = -1$

$\lfloor x \rfloor$ indicates the largest integer less than or equal to *x*: $\lfloor 1.1 \rfloor = 1$, $\lfloor 1.0 \rfloor = 1$ and $\lfloor -1.1 \rfloor = -2$

$\min(x_0,\ldots x_{N-1})$ indicates the minimum of $x_0,\ldots, x_{N-1}$, N being the number of components

$\max(x_0,\ldots x_{N-1})$ indicates the maximum of $x_0, \ldots, x_{N-1}$

$\sum$ indicates summation

4 General

4.1 Introduction

The present document defines the Jitter Buffer Management solution for the Codec for Enhanced Voice Services (EVS) [2]. Jitter Buffers are required in packet-based communications, such as 3GPP MTSI [2], to smooth the inter-arrival jitter of incoming media packets for uninterrupted playout.

The solution is used in conjunction with the EVS decoder and can also be used for AMR [4] and AMR-WB [5]. It is optimized for the Multimedia Telephony Service for IMS (MTSI) and fulfils the requirements for delay and jitter-induced concealment operations set in [2].

The procedure of the present document is recommended for implementation in all network entities and UEs supporting the EVS codec.

The present document does not describe the ANSI C code of this procedure. For a description of the two fixed-point ANSI C code implementations, using different sets of basic operators, see [6] and [10] respectively; for a description of the floating-point ANSI C code implementation see [7].

In the case of discrepancy between the EVS Jitter Buffer Management described in the present document and its ANSI-C code specification contained in [6], the procedure defined by [6] prevails. In the case of discrepancy between the procedure described in the present document and its ANSI-C code specification contained in [7], the procedure defined by [7] prevails. In the case of discrepancy between the procedure described in the present document and its ANSI-C code specifications contained in [10] the procedure defined by [10] prevails.

4.2 Packet-based communications

In packet-based communications, packets arrive at the terminal with random jitters in their arrival time. Packets may also arrive out of order. Since the decoder expects to be fed a speech packet every 20 milliseconds to output speech samples in periodic blocks, a de-jitter buffer is required to absorb the jitter in the packet arrival time. The larger the size of the de-jitter buffer, the better its ability to absorb the jitter in the arrival time and consequently fewer late arriving packets are discarded. Voice communications is also a delay critical system and therefore it becomes essential to keep the end to end delay as low as possible so that a two way conversation can be sustained.

The defined adaptive Jitter Buffer Management (JBM) solution reflects the above mentioned trade-offs. While attempting to minimize packet losses, the JBM algorithm in the receiver also keeps track of the delay in packet delivery as a result of the buffering. The JBM solution suitably adjusts the depth of the de-jitter buffer in order to achieve the trade-off between delay and late losses.

4.3 EVS Receiver architecture overview

An EVS receiver for MTSI-based communication is built on top of the EVS Jitter Buffer Management solution. In the EVS Jitter Buffer Management solution the received EVS frames, contained in RTP packets, are depacketized and fed to the Jitter Buffer Management (JBM). The JBM smoothes the inter-arrival jitter of incoming packets for uninterrupted playout of the decoded EVS frames at the Acoustic Frontend of the terminal.

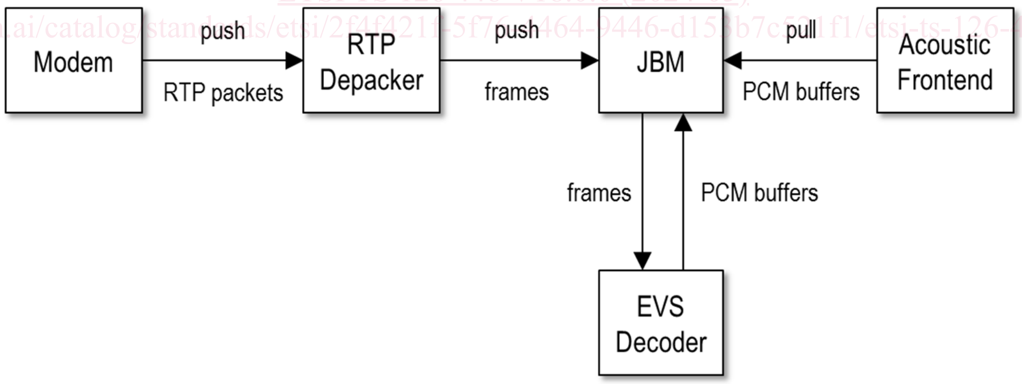

Figure 1: Receiver architecture for the EVS Jitter Buffer Management Solution

Figure 1 illustrates the architecture and data flow of the receiver side of an EVS terminal. Note that the architecture serves only as an example to outline the integration of the JBM in a terminal. This specification defines the JBM module and its interfaces to the RTP Depacker, the EVS Decoder [2], and the Acoustic Frontend [8]. The modules for Modem and Acoustic Frontend are outside the scope of the present document. The actual implementation of the RTP Depacker is outlined in a basic form; more complex depacketization scenarios depend on the usage of RTP.

Real-time implementations of this architecture typically use independent processing threads for reacting on arriving RTP packets from the modem and for requesting PCM data for the Acoustic Frontend. Arriving packets are typically handled by listening for packets received on the network socket related to the RTP session. Incoming packets are pushed into the RTP Depacker module which extracts the frames contained in an RTP packet. These frame are then pushed into the JBM where the statistics are updated and the frames are stored for later decoding and playout. The Acoustic Frontend contains the audio interface which, concurrently to the push operation of EVS frames, pulls PCM buffers from the JBM. The JBM is therefore required to provide PCM buffers, which are normally generated by decoding EVS frames by the EVS decoder or by other means to allow uninterrupted playout. Although the JBM is described for a multi-threaded architecture it does not specify thread-safe data structures due to the dependency on a particular implementation.

Note that the JBM does not directly forward frames from the RTP Depacker to the EVS decoder but instead uses frame-based adaptation to smooth the network jitter. In addition signal-based adaptation is executed on the decoded PCM buffers before they are pulled by the Acoustic Frontend. The corresponding algorithms are described in the following clauses.

5 Jitter Buffer Management

5.1 Overview

Jitter Buffer Management (JBM) includes the jitter estimation, control and jitter buffer adaptation algorithm to manage the inter-arrival jitter of the incoming packet stream. The entire solution for EVS consists of the following components:

- RTP Depacker (clause 5.2) to analyse the incoming RTP packet stream and to extract the EVS speech frames along with meta data to estimate the network jitter
- De-jitter Buffer (clause 5.6) to store the extracted EVS speech frames before decoding and to perform frame-based adaptation
- EVS decoder [1] for decoding the received EVS speech frames to PCM data
- Time-Scale Modification (clause 5.4.3) to perform signal-based adaptation for changing the playout delay
- Receiver Output Buffer (clause 5.5) to provide PCM data with a fixed frame size to the Acoustic Frontend
- Playout Delay Estimation Module (clause 5.3.5) to provide information on the current playout delay due to JBM
- Network Jitter Analysis (clause 5.3) for estimating the packet inter-arrival jitter and target playout delay
- Adaptation Control Logic (clause 5.4) to decide on actions for changing the playout delay based on the target playout delay

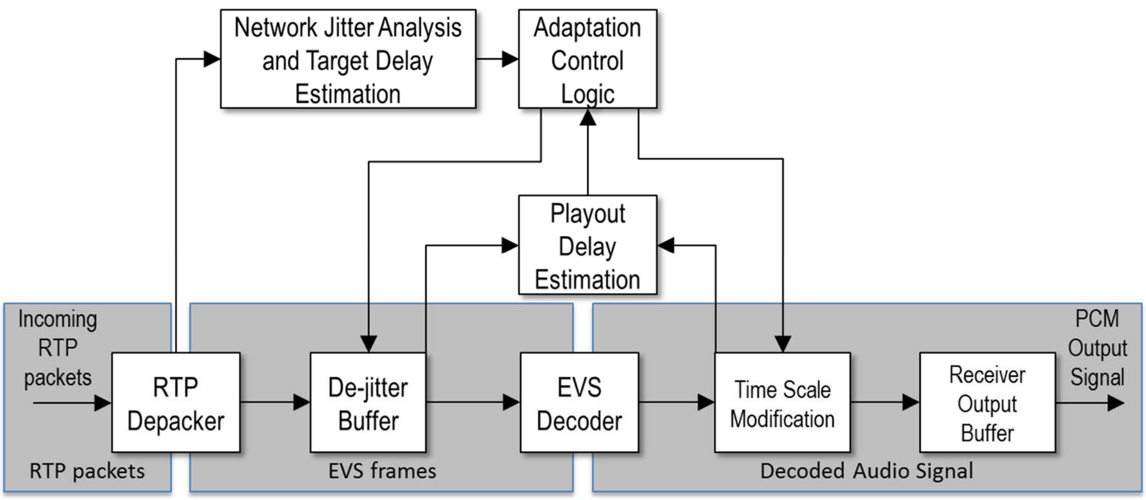

Figure 2: Modules of the EVS Jitter Buffer Management Solution

5.2 Depacketization of RTP packets (informative)

The RTP Depacker module of the JBM performs the depacketization of the incoming RTP packet stream. During this operation the EVS frames, embedded in RTP packets according to the respective RTP payload format [2], [9], are extracted and pushed to the de-jitter buffer. The RTP timestamp in an RTP packet for EVS always refers to the first EVS frame in the RTP payload. Any further EVS frames in the RTP payload are indexed in the RTP Payload Format Header by a Table of Contents (ToC) [2], [9]. The RTP Depacker performs the unpacking and calculates and assigns a media timestamp to every speech frame present in each received RTP packet.

The Jitter Buffer Management (JBM) for the EVS codec depends on information that is part of the received RTP packet stream. Each RTP packet consists of an RTP header and the RTP payload. The following data fields of the RTP header are of relevance for the JBM:

- RTP timestamp
- RTP sequence number

The marker bit in the RTP header is not evaluated by this JBM solution. Other fields in the RTP header are needed to correctly assign the incoming RTP packets to an RTP session, which is outside the scope of this specification.

All extracted frames (without NO_DATA frames) are fed to the JBM. The data structure for one frame consists of:

- Frame payload data, including the size of the payload
- Arrival timestamp of the RTP packet containing the frame
- Media timestamp in RTP timescale units, derived from the RTP timestamp of the packet
- Media duration in RTP timescale units (20 ms for EVS frames)
- RTP timescale as specified in the specification of the RTP payload format
- RTP sequence number
- SID flag
- Partial copy flag

To optimize the JBM behaviour for DTX, the JBM needs to be aware of SID frames. Determining this information depends on the implementation of the underlying audio codec. To keep the JBM independent of the audio codec, the SID flag needs to be fed to the JBM. In case of the EVS, AMR and AMR-WB codecs the SID flag can be determined from the size of the frame payload data.

Audio encoders supporting DTX typically output NO_DATA frames between SID frames to signal that a frame was not encoded because it does not contain an active signal and should be substituted with comfort noise by the audio decoder. Instead of NO_DATA frames this JBM solution uses the RTP timestamp for media time calculation. Therefore the RTP Depacker should not feed NO_DATA frames into the JBM.

The JBM handles packet reordering and duplication on the network and so the RTP Depacker can feed those frames into the JBM exactly as received, therefore a typical RTP Depacker implementation might be state-less.

5.3 Network Jitter Analysis and Delay Estimation

5.3.1 General

Estimates of the network jitter are required to control the JBM playout delay. The Jitter Buffer Management for EVS combines a short-term and a long-term jitter estimate to set the target playout delay. The playout is smoothly adapted to continuously minimize the difference between playout delay and target playout delay.

The transmission delay of a packet on the network can be seen as the sum of a fixed component (consisting of unavoidable factors such as propagation times through physical materials and minimum processing times) and a varying component (dominated by network jitter e.g. due to scheduling). As JBM does not expect synchronized clocks between the sender and receiver, the fixed delay value cannot be estimated using only the information available from the