

Presentation of Specification to TSG

Presentation to: TSG SA Meeting #10

Document for presentation: TS 26.171, Version 1.0.0

Presented for: Information

Agenda Item: 7.4.1

Title of document:

AMR Wideband Speech Codec; General Description



3rd Generation Partners
Technical Specification Group Services and System
Aspects;
Mandatory Speech Codec speech processing functions
AMR Wideband Speech Codec; General Description
(Release 4)

The present document has been developed within the 3rd Generation Partnership Project (3GPP™) and may be further elaborated for the purposes of 3GPP.

The present document has not been subject to any approval process by the 3GPP Organisational Partners and shall not be implemented.

This Specification is provided for future development work within 3GPP only. The Organisational Partners accept no liability for any use of this Specification.

Specifications and reports for implementation of the 3GPP™ system should be obtained via the 3GPP Organisational Partners' Publications Offices.

Keywords

3GPP, SA

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>

Contents

Foreword.....	4
1 Scope	5
2 Normative references.....	5
3 Definitions and abbreviations.....	5
3.1 Abbreviations	5
4 General	6
5 Adaptive Multi-Rate Wideband speech codec transcoding functions.....	8
6 Adaptive Multi-Rate Wideband speech codec ANSI C-code	8
7 Adaptive Multi-Rate Wideband speech codec test vectors	8
8 Adaptive Multi-Rate Wideband speech codec source controlled rate operation.....	9
9 Adaptive Multi-Rate Wideband speech codec voice activity detection	9
10 Adaptive Multi-Rate Wideband speech codec comfort noise insertion	10
11 Adaptive Multi-Rate Wideband speech codec error concealment of lost frames	10
12 Adaptive Multi-Rate Wideband speech codec frame structure	10
13 Adaptive Multi-Rate Wideband speech codec interface to RAN.....	11
14 Adaptive Multi-Rate Wideband speech codec performance characterisation.....	11
Annex A: Change history	Error! Bookmark not defined.

Foreword

This Technical Specification has been produced by the 3GPP.

The present document is an introduction to the speech processing parts of the wideband telephony speech service employing the Adaptive Multi-Rate Wideband (AMR-WB) speech coder within the 3GPP system.

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 this TS, 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 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 specification;

1 Scope

The present document is an introduction to the speech processing parts of the wideband telephony speech service employing the Adaptive Multi-Rate Wideband (AMR-WB) speech coder. A general overview of the speech processing functions is given, with reference to the documents where each function is specified in detail.

2 Normative references

This TS incorporates by dated and undated reference, provisions from other publications. These normative references are cited at the appropriate places in the text and the publications are listed hereafter. For dated references, subsequent amendments to or revisions of any of these publications apply to this TS only when incorporated in it by amendment or revision. For undated references, the latest edition of the publication referred to applies.

- [1] GSM 03.50 : "Digital cellular telecommunications system (Phase 2); Transmission planning aspects of the speech service in the GSM Public Land Mobile Network (PLMN) system".
- [2] 3G TS 26.190 : "AMR Wideband Speech Codec; Transcoding functions".
- [3] 3G TS 26.173 : "AMR Wideband Speech Codec; ANSI-C code".
- [4] 3G TS 26.174 : "AMR Wideband Speech Codec; Test sequences".
- [5] 3G TS 26.193 : "AMR Wideband Speech Codec; Source Controlled Rate operation".
- [6] 3G TS 26.194 : "AMR Wideband Speech Codec; Voice Activity Detection (VAD)".
- [7] 3G TS 26.192 : "AMR Wideband Speech Codec; Comfort Noise Aspects".
- [8] 3G TS 26.191 : "AMR Wideband Speech Codec; Error Concealment of Lost Frames.
- [9] 3G TS 26.201 : "AMR Wideband Speech Codec; Frame Structure".
- [10] 3G TS 26.202 : "AMR Wideband Speech Codec; Interface to RAN".
- [11] 3G TS 26.901WB: "AMR Wideband Speech Codec; Performance characterisation".

3 Definitions and abbreviations

3.1 Abbreviations

For the purposes of this TS, the following abbreviations apply:

ACELP	Algebraic Code Excited Linear Prediction
AMR	Adaptive Multi-Rate
AMR-WB	Adaptive Multi-Rate Wideband
BFI	Bad Frame Indication
CHD	Channel Decoder
CHE	Channel Encoder
GSM	Global System for Mobile communications
ITU-T	International Telecommunication Union – Telecommunication standardisation sector (former CCITT)
PCM	Pulse Code Modulation
PLMN	Public Land Mobile Network

PSTN	Public Switched Telephone Network
RX	Receive
SCR	Source Controlled Rate
SPD	SPeech Decoder
SPE	SPeech Encoder
TC	Transcoder
TX	Transmit
UE	User Equipment (terminal)

4 General

The AMR-WB speech coder consists of the multi-rate speech coder, a source controlled rate scheme including a voice activity detector and a comfort noise generation system, and an error concealment mechanism to combat the effects of transmission errors and lost packets.

The multi-rate speech coder is a single integrated speech codec with nine source rates from 6.60 kbit/s to 23.85 kbit/s, and a low rate background noise encoding mode. The 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 Figure 1, 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 this document. These aspects are only considered to the extent that the performance of the audio parts affect the performance of the speech transcoder.

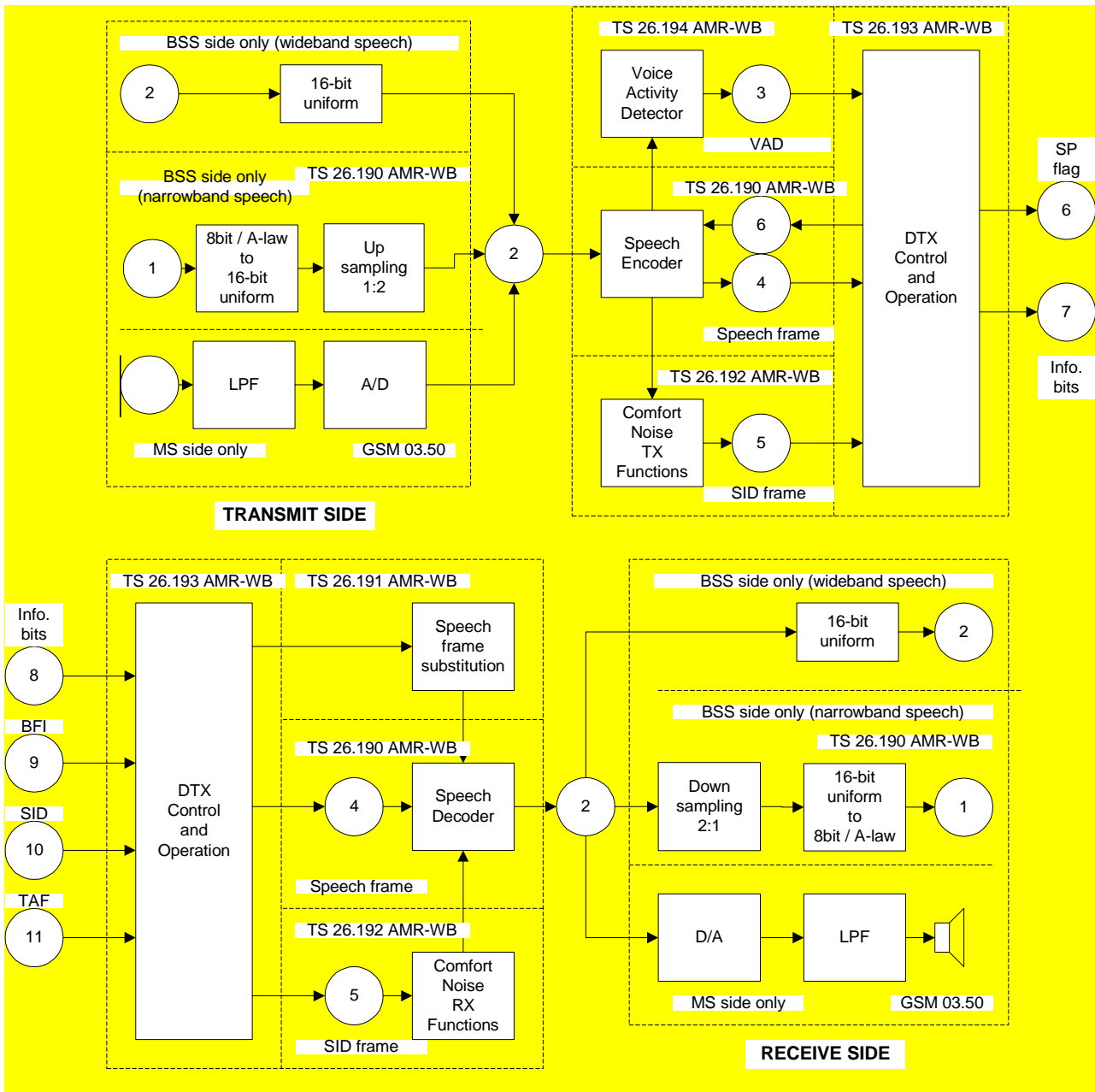


Figure 1: Overview of audio processing functions.

- 1) 8-bit A-law or μ -law PCM (ITU-T recommendation G.711), 8000 samples/s
- 2) [16-bit uniform PCM, 16 000 samples/s, TBD]
- 3) Voice Activity Detector (VAD) flag
- 4) Encoded speech frame, 50 frames/s, number of bits/frame depending on the AMR-WB codec mode
- 5) Silence Descriptor (SID) frame.
- 6) TX_TYPE, 3 bits, indicates whether information bits are available and if they are speech or SID information
- 7) Information bits delivered to the 3G AN
- 8) Information bits received from the 3G AN
- 9) RX_TYPE, the type of frame received quantized into three bits
- 10) Silence Descriptor (SID) flag
- 11) Time Alignment Flag (TAF), marks the position of the SID frame within the SACCH multiframe

5 Adaptive Multi-Rate Wideband speech codec transcoding functions

The adaptive multi-rate wideband speech codec is described in [2].

As shown in Figure 1, the speech encoder takes its input as a [16-bit uniform Pulse Code Modulated (PCM), TBD] signal either from the audio part of the UE or from the network side [TBD] or from the Public Switched Telephone Network (PSTN) via an narrowband 13-bit A-law or μ -law to wideband 16-bit uniform PCM conversion. An upsampling by factor of 2 has to be performed between narrowband and wideband speech signals. The encoded speech at the output of the speech encoder is packetized and delivered to the network interface. In the receive direction, the inverse operations take place.

The detailed mapping between input blocks of 320 speech 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 320 reconstructed speech samples is described in [2]. The coding scheme is Multi-Rate Algebraic Code Excited Linear Prediction. The bit-rates of the source codec are listed in Table 1.

An AMR-WB speech codec capable UE shall support all source rates listed in Table 1.

Table 1: Source codec bit-rates for the AMR-WB codec.

Codec mode	Source codec bit-rate
AMR-WB_23.85	23.85 kbit/s
AMR-WB_23.05	23.05 kbit/s
AMR-WB_19.85	19.85 kbit/s
AMR-WB_18.25	18.25 kbit/s
AMR-WB_15.85	15.85 kbit/s
AMR-WB_14.25	14.25 kbit/s
AMR-WB_12.65	12.65 kbit/s
AMR-WB_8.85	8.85 kbit/s
AMR-WB_6.60	6.60 kbit/s
AMR-WB_SID	1.75 kbit/s *

(*) Assuming SID frames are continuously transmitted

6 Adaptive Multi-Rate Wideband speech codec ANSI C-code

The ANSI-C code of the speech codec, VAD and CNG system are described in [3]. The ANSI C-code is mandatory.

7 Adaptive Multi-Rate Wideband speech codec test vectors

A set of digital test sequences is specified in [4], thus enabling the verification of compliance, i.e. bit-exactness, to a high degree of confidence.

The test sequences are defined separately for:

- The speech codec described in [2],
- The VAD described in [6],

- The CN generation described in [7]

The adaptive multi-rate wideband speech transcoder, VAD, SCR system and comfort noise parts of the audio processing functions (see Figure 1) are defined in bit exact arithmetic. Consequently, they shall react on 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 force 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 (initialisation 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 Adaptive Multi-Rate Wideband speech codec source controlled rate operation

The source controlled rate operation of the adaptive multi-rate wideband speech codec is defined in [5].

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 rate (SCR) 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 required for the source controlled rate operation:

- a Voice Activity Detector (VAD) 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.

9 Adaptive Multi-Rate Wideband speech codec voice activity detection

The adaptive multi-rate wideband VAD function is described in [6].

The input to the VAD is the input speech itself together with a set of parameters computed by the adaptive multi-rate wideband speech encoder. The VAD uses this information to decide whether each 20 ms speech coder frame contains speech or not.

The VAD algorithm is described in [6], and the corresponding C code is defined in [3]. The verification of compliance to [6] is achieved by use of digital test sequences applied to the same interface as the test sequences for the speech codec.

10 Adaptive Multi-Rate Wideband speech codec comfort noise insertion

The adaptive multi-rate wideband comfort noise insertion function is described in [7].

When speech is absent, the synthesis in the speech 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 [7], and the corresponding C code is defined in [3]. The comfort noise mechanism is based on the adaptive multi-rate wideband speech codec defined in [2].

11 Adaptive Multi-Rate Wideband speech codec error concealment of lost frames

The adaptive multi-rate Wideband speech codec error concealment of erroneous or lost frames is described in [8].

Frames may be erroneous due to transmission errors or frames may be lost due to frame stealing in a wireless environment or packet loss in a transport network. Actions which shall be taken in these cases, both for erroneous/lost speech frames and for erroneous/lost SID frames are described in [8]. The methods described in [8] may be used as a basis for such actions.

In order to mask the effect of isolated erroneous/lost frames, the speech decoder shall be informed 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. For several subsequent erroneous/lost frames, a muting technique shall be used to indicate to the listener that transmission has been interrupted.

12 Adaptive Multi-Rate Wideband speech codec frame structure

The adaptive multi-rate wideband speech frame structure is described in [9]. The output interface format from the encoder and input interface format to the decoder is divided into two parts; the core speech data part, which is the speech coded bits, and the other part is an additional data part with mode information.

The interface format described in [9] is termed AMR-WB interface format 1 (AMR-WB IF1).

Annex A of [9] describes an octet aligned frame format which shall be used in applications requiring octet alignment, such as for 3G H.324. This format is termed AMR-WB interface format 2 (AMR-WB IF2).

13 Adaptive Multi-Rate Wideband speech codec interface to RAN

The adaptive multi-rate wideband speech service interface to RAN is described in [10].

14 Adaptive Multi-Rate Wideband speech codec performance characterisation

The adaptive multi-rate wideband speech channel performance characterisation is described in [11].

Annex A (informative):
Change history

Change history							
Date	TSG #	TSG Doc.	CR	Rev	Subject/Comment	Old	New