ETSI TS 126 192 V9.0.0 (2010-01)

Technical Specification

Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); LTE; Speech codec speech processing functions; Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; Comfort noise aspects (3GPP TS 26.192 version 9.0.0 Release 9)

Reference

RTS/TSGS-0426192v900

Keywords

GSM, 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 notice

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

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

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

Copyright Notification

Intellectual Property Rights

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

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.

Contents

Foreword

This Technical Specification has been produced by the 3GPP.

The present document defines the detailed requirements for the correct operation of the background acoustic noise evaluation, noise parameter encoding/decoding and comfort noise generation in the narrowband 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

This document gives the detailed requirements for the correct operation of the background acoustic noise evaluation, noise parameter encoding/decoding and comfort noise generation for the AMR Wideband (AMR-WB) speech codec during Source Controlled Rate (SCR) operation.

The requirements described in this document are mandatory for implementation in all UEs capable of supporting the AMR-WB speech codec.

The receiver requirements are mandatory for implementation in all networks capable of supporting the AMR-WB speech codec, the transmitter requirements only for those where downlink SCR will be used.

In case of discrepancy between the requirements described in this document and the fixed point computational description of these requirements contained in [1], the description in [1] will prevail.

2 Normative references

This document 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 document only when incorporated in it by amendment or revision. For undated references, the latest edition of the publication referred to applies.

[1] 3GPP TS 26.173 : "AMR Wideband Speech Codec; ANSI-C code".

[2] 3GPP TS 26.190 : "AMR Wideband Speech Codec; Transcoding functions".

[3] 3GPP TS 26.191 : "AMR Wideband Speech Codec; Error concealment of lost frames ".

[4] 3GPP TS 26.193 : "AMR Wideband Speech Codec; Source Controlled Rate operation ".

[5] 3GPP TS 26.201 : "AMR Wideband Speech Codec; Frame Structure".

3 Definitions, symbols and abbreviations

3.1 Definitions

For the purpose of this document, the following definitions apply.

Frame: Time interval of 20 ms corresponding to the time segmentation of the adaptive multi-rate wideband speech transcoder, also used as a short term for traffic frame.

SID frames: Special Comfort Noise frames. It may convey information on the acoustic background noise or inform the decoder that it should start generating background noise.

Speech frame: Traffic frame that cannot be classified as a SID frame.

VAD flag: Voice Activity Detection flag.

TX_TYPE: Classification of the transmitted traffic frame (defined in [4]).

RX_TYPE: Classification of the received traffic frame (defined in [4]).

Other definitions of terms used in this document can be found in [2] and [4]. The overall operation of SCR is described in [4].

3.2 Symbols

For the purpose of this document, the following symbols apply. Boldface symbols are used for vector variables.

$\mathbf{f}^T = [f_1\, f_2 \ldots f_{16}]$	Unquantized ISF vector
$\hat{\mathbf{f}}^T = [\hat{f}_1\, \hat{f}_2 \ldots \hat{f}_{16}]$	Quantized ISF vector
$\mathbf{f}^{(m)}$	Unquantized ISF vector of frame m
$\hat{\mathbf{f}}^{(m)}$	Quantized ISF vector of frame m
$\mathbf{f}^{mean}$	Averaged ISF parameter vector
$en_{\log}$	Logarithmic frame energy
$en_{\log}^{mean}$	Averaged logarithmic frame energy
$\mathbf{e}$	ISF parameter prediction residual
$\hat{\mathbf{e}}$	Quantized ISF parameter prediction residual
$\sum_{n=a}^{b} x(n)$	$= x(a) + x(a+1) + \ldots + x(b-1) + x(b)$

3.3 Abbreviations

For the purpose of this document , the following abbreviations apply.

AMR	Adaptive Multi-Rate
AMR-WB	Adaptive Multi-Rate Wideband
CN	Comfort Noise

SCR	Source Controlled Rate operation (aka source discontinuous transmission)
UE	User Equipment
SID	SIlence Descriptor
LP	Linear Prediction
ISP	Immittance Spectral Pair
ISF	Immittance Spectral Frequency
RSS	Radio Subsystem
RX	Receive
TX	Transmit
VAD	Voice Activity Detector

4 General

A basic problem when using SCR is that the background acoustic noise, which is transmitted together with the speech, would disappear when the transmission is cut, resulting in discontinuities of the background noise. Since the SCR switching can take place rapidly, it has been found that this effect can be very annoying for the listener - especially in a car environment with high background noise levels. In bad cases, the speech may be hardly intelligible.

This document specifies the way to overcome this problem by generating on the receive (RX) side synthetic noise similar to the transmit (TX) side background noise. The comfort noise parameters are estimated on the TX side and transmitted to the RX side at a regular rate when speech is not present. This allows the comfort noise to adapt to the changes of the noise on the TX side.

5 Functions on the transmit (TX) side

The comfort noise evaluation algorithm uses the following parameters of the AMR-WB speech encoder, defined in [2]:

- the unquantized Linear Prediction (LP) parameters, using the Immittance Spectral Pair (ISP) representation, where the unquantized Immittance Spectral Frequency (ISF) vector is given by $\mathbf{f}^T = [f_1 f_2 \dots f_{16}]$;

The algorithm computes the following parameters to assist in comfort noise generation:

- the weighted averaged ISF parameter vector $\mathbf{f}^{mean}$ (weighted average of the ISF parameters of the eight most recent frames);
- the averaged logarithmic frame energy $en_{\log}^{mean}$ (average of the logarithmic energy of the eight most recent frames).

These parameters give information on the level ($en_{\log}^{mean}$) and the spectrum ($\mathbf{f}^{mean}$) of the background noise.

The evaluated comfort noise parameters ($\mathbf{f}^{mean}$ and $en_{\log}^{mean}$) are encoded into a special frame, called a Silence Descriptor (SID) frame for transmission to the RX side.

A hangover logic is used to enhance the quality of the silence descriptor frames. A hangover of seven frames is added to the VAD flag so that the coder waits with the switch from active to inactive mode for a period of seven frames, during that time the decoder can compute a silence descriptor frame from the quantized ISFs and the logarithmic frame energy of the decoded speech signal. Therefore, no comfort noise description is transmitted in the first SID frame after active speech. If the background noise contains transients which will cause the coder to switch to active mode and then back to inactive mode in a very short time period, no hangover is used. Instead the previously used comfort noise frames are used for comfort noise generation.

The first SID frame also serves to initiate the comfort noise generation on the receive side, as a first SID frame is always sent at the end of a speech burst, i.e., before the transmission is terminated.

The scheduling of SID or speech frames on the network path is described in [4].

5.1 ISF evaluation

The comfort noise parameters to be encoded into a SID frame are calculated over N=8 consecutive frames marked with VAD=0, as follows:

Prior to averaging the ISF parameters over the CN averaging period, a median replacement is performed on the set of ISF parameters to be averaged, to remove the parameters which are not characteristic of the background noise on the transmit side. First, the spectral distances from each of the ISF parameter vectors $\mathbf{f}(i)$ to the other ISF parameter vectors $\mathbf{f}(j)$, i=0,...,7, j=0,...,7, $i\neq j$, within the CN averaging period are approximated according to the equation:

$$\Delta R_{ij} = \sum_{k=1}^{16} \left(f_i(k) - f_j(k) \right)^2 , \qquad (1)$$

where $f_i(k)$ is the kth ISF parameter of the ISF parameter vector $\mathbf{f}(i)$ at frame i.

To find the spectral distance ΔS_i of the ISF parameter vector $\mathbf{f}(i)$ to the ISF parameter vectors $\mathbf{f}(j)$ of all the other frames j=0,...,7, $j\neq i$, within the CN averaging period, the sum of the spectral distances ΔR_{ij} is computed as follows:

$$\Delta S_i = \sum_{j=0,\, j\neq i}^{7} \Delta R_{ij} , \qquad (2)$$

for all i=0,...,7, $i\neq j$.

The ISF parameter vector $\mathbf{f}(i)$ with the smallest spectral distance ΔS_i of all the ISF parameter vectors within the CN averaging period is considered as the median ISF parameter vector $\mathbf{f}_{med}$ of the averaging period, and its spectral distance is denoted as ΔS_{med}. The median ISF parameter vector is considered to contain the best representation of the short-term spectral detail of the background noise of all the ISF parameter vectors within the averaging period. If there are ISF parameter vectors $\mathbf{f}(j)$ within the CN averaging period with

$$\frac{\Delta S_j}{\Delta S_{med}} > TH_{med} , \qquad (3)$$

where $TH_{med} = 2.25$ is the median replacement threshold, then at most two of these ISF parameter vectors (the ISF parameter vectors causing TH_{med} to be exceeded the most) are replaced by the median ISF parameter vector prior to computing the averaged ISF parameter vector $\mathbf{f}^{mean}$.

The set of ISF parameter vectors obtained as a result of the median replacement are denoted as $\mathbf{f}'(n-i)$, where n is the index of the current frame, and i is the averaging period index (i=0,...,7).

When the median replacement is performed at the end of the hangover period (first CN update), all of the ISF parameter vectors $\mathbf{f}(n-i)$ of the 7 previous frames (the hangover period, i=1,...,7) have quantized values, while the ISF parameter vector $\mathbf{f}(n)$ at the most recent frame n has unquantized values. In the subsequent CN updates, the ISF parameter vectors of the CN averaging period in the frames overlapping with the hangover period have quantized values, while the parameter vectors of the more recent frames of the CN averaging period have unquantized values. When the period of the eight most recent frames is non-overlapping with the hangover period, the median replacement of ISF parameters is performed using only unquantized parameter values.

The averaged ISF parameter vector $\mathbf{f}^{mean}(n)$ at frame n shall be computed according to the equation:

$$\mathbf{f}^{mean}(n) = \frac{1}{8} \sum_{i=0}^{7} \mathbf{f}'(n-i), \qquad (4)$$

where $\mathbf{f}'(n-i)$ is the ISF parameter vector of one of the eight most recent frames (i = 0,...,7) after performing the median replacement, i is the averaging period index, and n is the frame index.

The averaged ISF parameter vector $\mathbf{f}^{mean}(n)$ at frame *n* is quantized using the comfort noise ISF quantization tables The mean removed ISF vector to be quantized is obtained according to the following equation:

$$\mathbf{r}(n) = \mathbf{f}^{mean}(n) - \bar{\mathbf{f}}, \qquad (5)$$

where $\mathbf{f}^{mean}(n)$ is the averaged ISF parameter vector at frame *n*, $\bar{\mathbf{f}}$ is the constant mean ISF vector, $\mathbf{r}(n)$ is the computed ISF mean removed vector at frame *n*, and *n* is the frame index.

5.2 Frame energy calculation

The frame energy is computed for each frame marked with VAD=0 according to the equation :

$$en_{\log}(i) = \frac{1}{2}\log_2\left(\frac{1}{N}\sum_{n=0}^{N-1} s^2(n)\right) \qquad (6)$$

where $s(n)$ is the high-pass-filtered input speech signal of the current frame i. The energy is also adjusted according to the signalled speech modes capabilities, as to provide high quality transitions from Comfort Noise to Speech.

The averaged logarithmic energy is computed by:

$$en_{\log}^{mean}(i) = \frac{1}{8}\sum_{n=0}^{7} en_{\log}(i-n) \quad . \qquad (7)$$

The averaged logarithmic energy is quantized using a 6 bit arithmetic quantizer. The 6 bits for the energy index are transmitted in the SID frame (see bit allocation in table 1).

5.3 Analysis of the variation and stationarity of the background noise

The encoder first determines how stationary background noise is. Dithering is employed for non-stationary background noise. The information about whether to use dithering or not is transmitted to the decoder using a binary information (CN_{dith} -flag).

The binary value for the CN_{dith} -flag is found by using the spectral distance ΔS_i of the spectral parameter vector $\mathbf{f}(i)$ to the spectral parameter vectors $\mathbf{f}(j)$ of all the other frames *j*=0,..., l_{dtx}-*1*, *j*≠*i* within the CN averaging period (l_{dtx}). The computation of the spectral distance is described in Chapter 5.1. A sum of spectral distances $D_s = \sum_{i=0}^{7} \Delta S_i$ is then computed. If D_S is small, CN_{dith} -flag is set to 0. Otherwise, CN_{dith} -flag is set to 1. Additionally, variation of energy between frames is studied. The sum of absolute deviation of $en_{log}(i)$ from the average en_{log} is computed. If the sum is large, CN_{dith} -flag is set to 1, even if the flag was earlier set to 0.

5.4 Modification of the speech encoding algorithm during SID frame generation

When the TX_TYPE is not equal to SPEECH the speech encoding algorithm is modified in the following way:

- The non-averaged LP parameters which are used to derive the filter coefficients of the filters $H(z)$ and $W(z)$ of the speech encoder are not quantized;
- The open loop pitch lag search is performed, but the closed loop pitch lag search is inactivated. The adaptive codebook memory is set to zero.
- No fixed codebook search is made.

- The memory of weighting filter $W(z)$ is set to zero, i.e., the memory of $W(z)$ is not updated.

- The ordinary LP parameter quantization algorithm is inactive. The averaged ISF parameter vector $\mathbf{f}^{mean}$ is calculated each time a new SID frame is to be sent. This parameter vector is encoded into the SID frame as defined in subclause 5.1.

- The ordinary gain quantization algorithm is inactive.

- The predictor memories of the ordinary LP parameter quantization algorithm is initialized when TX_TYPE is not SPEECH, so that the quantizers start from known initial states when the speech activity begins again.

In the 23.85 kbit/s mode, when the TX_TYPE is equal to SPEECH and VAD is OFF, the speech encoding algorithm is modified in the following way:

- The generation of high-band gain g_{HB} is changed by adapting it during non-active speech period towards estimated gain in order to ensure smooth transition of high-band gain. g_{HB} is then

$$g_{HB} = \frac{hang_{DTX}}{7} g_{HB} + (1 - \frac{hang_{DTX}}{7}) g_{est}, \tag{8}$$

where $hang_{DTX}$ is DTX counter.

5.4 SID-frame encoding

The encoding of the comfort noise bits in a SID frame is described in [5] where the indication of the first SID frame is also described. The bit allocation and sequence of the bits from comfort noise encoding is shown in Table 1.

6 Functions on the receive (RX) side

The situations in which comfort noise shall be generated on the receive side are defined in [4]. In general, the comfort noise generation is started or updated whenever a valid SID frame is received.

6.1 Averaging and decoding of the LP and energy parameters

When speech frames are received by the decoder the LP and the energy parameters of the last seven speech frames shall be kept in memory. The decoder counts the number of frames elapsed since the last SID frame was updated and passed to the RSS by the encoder. Based on this count, the decoder determines whether or not there is a hangover period at the end of the speech burst (defined in [4]). The interpolation factor is also adapted to the SID update rate.

As soon as a SID frame is received comfort noise is generated at the decoder end. The first SID frame parameters are not received but computed from the parameters stored during the hangover period. If no hangover period is detected, the parameters from the previous SID update are used.

The averaging procedure for obtaining the comfort noise parameters for the first SID frame is as follows:

- when a speech frame is received, the ISF vector is decoded and stored in memory, moreover the logarithmic frame energy of the decoded signal is also stored in memory.

- the averaged values of the quantized ISF vectors and the averaged logarithmic frame energy of the decoded frames are computed and used for comfort noise generation.

The averaged value of the ISF vector for the first SID frame is given by:

$$\hat{\mathbf{f}}^{mean}(i) = \frac{1}{8} \sum_{n=0}^{7} \hat{\mathbf{f}}(i-n) \tag{9}$$

where $\hat{\mathbf{f}}(i-n)$, $n > 0$ is the quantized ISF vector of one of the frames of the hangover period and where $\hat{\mathbf{f}}(i-0)$ = $\hat{\mathbf{f}}(i-1)$. The averaged logarithmic frame energy for the first SID frame is given by:

$$\hat{en}_{\log}^{mean}(i) = \frac{1}{8}\sum_{n=0}^{7}\hat{en}_{\log}(i-n) \qquad (10)$$

where $\hat{en}_{\log}(i-n)$, $n > 0$ is the logarithmic vector of one of the frames of the hangover period computed for the decoded frames and where $\hat{en}_{\log}(i-0)$ = $\hat{en}_{\log}(i-1)$.

For ordinary SID frames, the ISF vector and logarithmic frame energy are computed by table lookup. The ISF vector is given by the sum of the decoded reference vector and the constant mean ISF vector.

During comfort noise generation the spectrum and energy of the comfort noise is determined by interpolation between old and new SID frames.

When dithering is used, the ISF vector **f** is modified by

$$\mathbf{f}(i) = \mathbf{f}(i) + rand(-L(i), L(i)), \qquad i = 1,..,16 \qquad (11)$$

where $L(i)$ = 100 + 0.8i Hz and $rand(-L(i),L(i))$ is random function generating values between $-L(i)$ and $L(i)$. A minimum gap of 175 Hz is ensured between elements of **f**.

Dithering insertion for energy parameter is similar to spectral dithering and can be computed as follows:

$$en_{\log}^{mean} = en_{\log}^{mean} + rand(-L, L), \qquad (12)$$

where L = 75 and $en_{\log}^{mean}$ is the energy value used for scaling the energy of the comfort noise excitation.

6. 2 Comfort noise generation and updating

The comfort noise generation procedure uses the Adaptive Multi-Rate Wideband (AMR-WB) speech decoder algorithm defined in [2].

When comfort noise is to be generated, the various encoded parameters are set as follows:

In each subframe, the pulse positions and signs of the excitation are locally generated using uniformly distributed pseudo random numbers. The excitation pulses take values between +2047 and -2048 when comfort noise is generated. The fixed codebook comfort noise excitation generation algorithm works as follows:

```
for (i = 0; i < 64; i++)     u[i] = shr(random(),4);
```

where:

`u[0..63]` excitation buffer;

`random()` generates a random integer value, uniformly distributed between -32768 and +32767;

The excitation gain is computed from the logarithmic frame energy parameter by converting it to the linear domain.

The adaptive codebook gain values in each subframe are set to 0, also the memory of the adaptive codebook is set to zero.

The pitch delay values in each subframe are set to 64.

The LP filter parameters used are those received in the SID frame.

The predictor memory of the ordinary LP parameter algorithm is initialized when RX_TYPE is not SPEECH , so that the quantizer start from given initial states when the speech activity begins again. With these parameters, the speech decoder now performs the standard operations described in [2] and synthesizes comfort noise. During CN generation, the high-band generation is performed using estimated high-band gain like in 8.85, 12.65, 14.25, 15.85, 18.25, 19.85 or 23.05 kbit/s modes during active speech.

Updating of the comfort noise parameters (energy and LP filter parameters) occurs each time a valid SID frame is received, as described in [4].

When updating the comfort noise, the parameters above should be interpolated over the SID update period to obtain smooth transitions.

7 Computational details and bit allocation

A bit exact computational description of comfort noise encoding and generation in form of an ANSI-C source code is found in [1].

The detailed bit allocation and the sequence of bits in the comfort noise encoding is shown in Table 1.

Table 1: Source encoder output parameters in order of occurrence and bit allocation for comfort noise encoding

Bits (MSB-LSB)	Description
s1 – s6	index of 1st ISF subvector
s7- s12	index of 2st ISF subvector
s13 – s18	index of 3nd ISF subvector
s19 – s23	index of 4th ISF subvector
s24 – s28	index of 5th ISF subvector
s29 – s34	index of logarithmic frame energy
s35	dithering flag

Annex A (informative): Change history

Change history							
Date	**TSG #**	**TSG Doc.**	**CR**	**Rev**	**Subject/Comment**	**Old**	**New**
03-2001	11	SP-010087			Version 2.0.0 presented for approval		5.0.0
12-2004	26				Version for Release 6	5.0.0	6.0.0
06-2007	36				Version for Release 7	6.0.0	7.0.0
12-2008	42				Version for Release 8	7.0.0	8.0.0
12-2009	46				Version for Release 9	8.0.0	9.0.0

History

Document history		
V9.0.0	January 2010	Publication