Draft ETSI EN 301 706 V7.1.0 (1999-07)

European Standard (Telecommunications series)

Digital cellular telecommunication system (Phase 2+); Comfort noise aspects for Adaptive Multi-Rate (AMR) speech traffic channels (GSM 06.92 version 7.1.0 Release 1998)

GLOBAL SYSTEM FOR
MOBILE COMMUNICATIONS

Reference

DEN/SMG-110692Q7 (fso03ic0.PDF)

Keywords

Digital cellular telecommunications system,
Global System for Mobile communications (GSM)

ETSI

Postal address

F-06921 Sophia Antipolis Cedex - FRANCE

Office address

650 Route des Lucioles - Sophia Antipolis
Valbonne - 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

Internet

secretariat@etsi.fr
Individual copies of this ETSI deliverable
can be downloaded from
http://www.etsi.org
If you find errors in the present document, send your
comment to: editor@etsi.fr

Content

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 SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available **free of charge** from the ETSI Secretariat. Latest updates are available on the ETSI Web server (http://www.etsi.org/ipr).

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 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 European Standard (Telecommunications series) has been produced by ETSI Technical Committee Special Mobile Group (SMG), and is now submitted to the ETSI standards One-step Approval Procedure.

The present document introduces the Adaptive Multi-Rate (AMR) speech traffic channels within the digital cellular telecommunications system.

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

Version 7.x.y

where:

7 indicates Release 1998 of GSM Phase 2+

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

y the third digit is incremented when editorial only changes have been incorporated in the specification.

Proposed national transposition dates	
Date of latest announcement of this EN (doa):	3 months after ETSI publication
Date of latest publication of new National Standard or endorsement of this EN (dop/e):	6 months after doa
Date of withdrawal of any conflicting National Standard (dow):	6 months after doa

1 Scope

The present document gives the detailed requirements for the correct operation of the background acoustic noise evaluation, noise parameter encoding/decoding and comfort noise generation in Mobile Stations (MSs) and Base Station Systems (BSSs) during Discontinuous Transmission (DTX) on adaptive multi-rate full rate and half rate speech traffic channels.

The requirements described in the present document are mandatory for implementation in all GSM MSs capable of supporting the adaptive multi-rate full rate and half tate speech traffic channels.

The receiver requirements are mandatory for implementation in all GSM BSSs capable of supporting the adaptive multi-rate full rate and half rate speech traffic channels, the transmitter requirements only for those where downlink DTX will be used.

In case of discrepancy between the requirements described in the present document and the fixed point computational description of these requirements contained in GSM 06.73 [2], the description in GSM 06.73 [2] will prevail.

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.
- A non-specific reference to an ETS shall also be taken to refer to later versions published as an EN with the same number.
- For this Release 1998 document, references to GSM documents are for Release 1998 versions (version 7.x.y).

[1] GSM 01.04: "Digital cellular telecommunication system (Phase 2+); Abbreviations and acronyms".

[2] GSM 06.73: "Digital cellular telecommunications system (Phase 2+); ANSI-C code for the GSM Adaptive Multi-Rate speech codec".

[3] GSM 06.90: "Digital cellular telecommunications system (Phase 2+); Adaptive Multi-Rate speech transcoding".

[4] GSM 06.91: "Digital cellular telecommunications system (Phase 2+); Substitution and muting of lost frame for Adaptive Multi-Rate speech traffic channels".

[5] GSM 06.93: "Digital cellular telecommunications system (Phase 2+); Discontinuous transmission (DTX) for Adaptive Multi-Rate speech traffic channels".

3 Definitions, symbols and abbreviations

3.1 Definitions

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

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

SID frames: **s**pecial SID (Silence Descriptor) 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: one of SPEECH, SID_FIRST, SID_UPD, NO_DATA (defined in GSM 06.93).

RX_TYPE: classification of the received traffic frame (defined in GSM 06.93).

Other definitions of terms used in the present document can be found in GSM 06.90 [3] and GSM 06.93 [5]. The overall operation of DTX is described in GSM 06.93 [5].

3.2 Symbols

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

$\mathbf{f}^{T} = \left[f_1 \;\; f_2 \, \ldots \, f_{10} \right]$	Unquantized LSF vector
$\hat{\mathbf{f}}^{T} = \left[\hat{f}_1 \;\; \hat{f}_2 \, \ldots \, \hat{f}_{10} \right]$	Quantized LSF vector
$\mathbf{f}^{(m)}$	Unquantized LSF vector of frame m
$\hat{\mathbf{f}}^{(m)}$	Quantized LSF vector of frame m
$\mathbf{f}^{mean}$	Averaged LSF parameter vector
$en_{\log}$	Logarithmic frame energy
$en_{\log}^{mean}$	Averaged logarithmic frame energy
$\hat{\mathbf{f}}^{ref}$	Reference vector for LSF quantization
$\mathbf{e}$	Computed LSF parameter prediction residual
$\hat{\mathbf{e}}$	Quantized LSF 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 the present document, the following abbreviations apply.

AMR	Adaptive Multi-Rate
BSS	Base Station Subsystem
DTX	Discontinuous Transmission
MS	Mobile Station
SID	SIlence Descriptor
LP	Linear Prediction
LSP	Line Spectral Pair
LSF	Line Spectral Frequency
RX	Receive
TX	Transmit
VAD	Voice Activity Detector

For abbreviations not given in this subclause, see GSM 01.04 [1].

4 General

A basic problem when using DTX is that the background acoustic noise, which is transmitted together with the speech, would disappear when the radio transmission is cut, resulting in discontinuities of the background noise. Since the DTX 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.

The present 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 before the radio transmission is switched off and at a regular rate afterwards. 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 speech encoder, defined in GSM 06.90 [3]:

- the unquantized Linear Prediction (LP) parameters, using the Line Spectral Pair (LSP) representation, where the unquantized Line Spectral Frequency (LSF) vector is given by $\mathbf{f}^T = \left[f_1 \ f_2 \ \ldots \ f_{10} \right]$;
- the unquantized LSF vector for the 12.2 kbit/s mode is given by the second set of LSF parameters in the frame.

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

- the averaged LSF parameter vector $\mathbf{f}^{mean}$ (average of the LSF 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 7 frames is added to the VAD flag so that the coder waits with the switch from active to inactive mode for a period of 7 frames, during that time the decoder can compute a silence descriptor frame from the quantized LSFs 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 timeperiod, 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 radio transmission is terminated.

The scheduling of SID or speech frames on the radio path is described in GSM 06.93 [5].

5.1 LSF 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:

The averaged LSF parameter vector $\mathbf{f}^{mean}(i)$ of the frame i shall be computed according to the equation:

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

where $\mathbf{f}(i-n)$ is the (unquantized) LSF parameter vector of the current frame i ($n = 0$) and past frames ($n = 1,\ldots,7$).

The averaged LSF parameter vector $\mathbf{f}^{mean}(i)$ of the frame i is encoded using the same encoding tables that are also used by the 7.4 kbit/s mode for the encoding of the non-averaged LSF parameter vectors in ordinary speech encoding mode, but the quantization algorithm is modified in order to support the quantization of comfort noise.

The LSF parameter prediction residual to be quantized for frame i is obtained according to the following equation:

$$\mathbf{e}(i) = \mathbf{f}^{mean}(i) - \hat{\mathbf{f}}^{ref} \qquad (2)$$

where $\hat{\mathbf{f}}^{ref}$ is a reference vector picked from a codebook.

The vector $\hat{\mathbf{f}}^{ref}$ used in eq (2) is encoded for each SID frame. A lookup table containing 8 vectors typical for background noise are searched. The vector which yields the lowest prediction residual energy is selected. After the above step the LSF parameter encoding procedure is performed. The 3-bit index for the reference vector and the 26 bits for LSF parameter are transmitted in the SID frame (see bit allocation in table 1).

5.2 Frame energy calculation

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

$$en_{\log}(i) = \frac{1}{2}\log_2\left(\frac{1}{N}\sum_{n=0}^{N-1}s^2(n)\right) \qquad (3)$$

where $s(n)$ is the HP-filtered input speech signal of the current frame i.

The averaged logarithmic energy is computed by:

$$en_{\log}^{mean}(i) = \frac{1}{8}\sum_{n=0}^{7}en_{\log}(i-n) \qquad (4)$$

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

5.3 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 gain and 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 LSF parameter vector $\mathbf{f}^{mean}$ is calculated each time a new SID frame is to be sent to the Radio Subsystem. 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 and fixed codebook gain quantization algorithms are initialized when TX_TYPE is not SPEECH, so that the quantizers start from known initial states when the speech activity begins again.

5.4 SID-frame encoding

The encoding of the 35 comfort noise bits in a SID frame is described in GSM 05.03 where the encoding 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 GSM 06.93 [5]. Generally speaking, 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 GSM 06.93). 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 LSF 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 LSF vectors and the averaged logarithmic frame energy of the decoded frames are computed and used for comfort noise generation.

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

$$\hat{\mathbf{f}}^{mean}(i) = \frac{1}{8}\sum_{n=0}^{7}\hat{\mathbf{f}}(i-n) \tag{5}$$

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

$$\hat{en}_{\log}^{mean}(i) = \frac{1}{8}\sum_{n=0}^{7}\hat{en}_{\log}(i-n) \tag{6}$$

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

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

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

In order to achieve a comfort noise that is less static in appearance the LSF vector is slightly perturbed for each frame by adding a small component based on parameters variations computed in the hangover period. The computation of the perturbation is made by computing the mean LSF vector from the matrix $\hat{\mathbf{f}}$, this mean vector is then subtracted from each of the elements of $\hat{\mathbf{f}}$ forming a new matrix $\widehat{\mathbf{f}}$. For every frame a mean removed LSF vector is randomly choosen from $\widehat{\mathbf{f}}$ and added to the interpolated LSF vector.

6. 2 Comfort noise generation and updating

The comfort noise generation procedure uses the adaptive multi-rate speech decoder algorithm defined in GSM 06.90 [3].

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 fixed codebook excitation are locally generated using uniformly distributed pseudo random numbers. The excitation pulses take values of +1 and -1 when comfort noise is generated. The fixed codebook comfort noise excitation generation algorithm works as follows:

```
for (i = 0; i < 40; i++)    code[i] = 0;

for (i = 0; i < 10; i++) {

    j = random(4);

    idx = j * 10 + i;

    if (random(2) == 1)  code[idx] = 1;

    else         code[idx] = -1;

}
```

where:

`code[0..39]` fixed codebook excitation buffer;

`random(4)` generates a random integer value, uniformly distributed between 0 and 3;

`random(2)` generates a random integer value, uniformly distributed between 0 and 1.

The fixed codebook gain is computed from the logarithmic frame energy parameter by converting it to the linear domain and normalizing with the gain of LP synthesis filter.

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 40.

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

The predictor memories of the ordinary LP parameter and fixed codebook gain quantization algorithms are initialized when RX_TYPE is not SPEECH , so that the quantizers start from given initial states when the speech activity begins again.

With these parameters, the speech decoder now performs the standard operations described in GSM 06.90 [3] and synthesizes comfort noise.

Updating of the comfort noise parameters (energy and LP filter parameters) occurs each time a valid SID frame is received, as described in GSM 06.93 [5].

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 GSM 06.73 [2].

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 – s3	index of reference vector
s4 - s11	index of 1st LSF subvector
s12 – s20	index of 2nd LSF subvector
s21 – s29	index of 3rd LSF subvector
s30 – s35	index of logarithmic frame energy

Annex A (informative): Document change history

SMG	SPEC	CR	PH	VERS	NEW_VE	SUBJECT
29	06.92	A001	R98	7.0.0	7.1.0	Correction of averaging formula

History

Document history		
V7.1.0	July 1999	One-step Approval Procedure OAP 9952: 1999-07-28 to 1999-11-26