3GPP TS 46.061 V11.0.0 (2012-09)

Technical Specification

3rd Generation Partnership Project;
Technical Specification Group Services and System Aspects;
Substitution and muting of lost frames for
Enhanced Full Rate (EFR) speech traffic channels
(Release 11)

GLOBAL SYSTEM FOR
MOBILE COMMUNICATIONS

Keywords
GSM, speech, codec

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

This Technical Specification has been produced by the 3rd Generation Partnership Project (3GPP).

The present document defines a frame substitution and muting procedure which is used by the Receive (RX) Discontinuous Transmission (DTX) handler when one or more lost speech or Silence Descriptor (SID) frames are received from the Radio Sub System (RSS) within the digital cellular telecommunications 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 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 a frame substitution and muting procedure which shall be used by the Receive (RX) Discontinuous Transmission (DTX) handler when one or more lost speech or Silence Descriptor (SID) frames are received from the Radio Sub System (RSS).

The requirements of the present document are mandatory for implementation in all GSM Base Station Systems (BSS)s and Mobile Stations (MS)s capable of supporting the enhanced Full Rate speech traffic channel. It is not mandatory to follow the bit exact implementation outlined in the present document and the corresponding C-source code.

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]   GSM 05.03: "Digital cellular telecommunications system (Phase 2+); Channel coding".

[2]   GSM 06.60: "Digital cellular telecommunications system (Phase 2+); Enhanced Full Rate (EFR) speech transcoding".

[3]   GSM 06.81: "Digital cellular telecommunications system (Phase 2+); Discontinuous transmission (DTX) for Enhanced Full Rate (EFR) speech traffic channels".

[4]   GSM 08.60: "Digital cellular telecommunications system (Phase 2+); Inband control of remote transcoders and rate adaptors for Enhanced Full Rate (EFR) and full rate traffic channels".

3 Definitions and abbreviations

3.1 Definitions

For the purposes of the present document, the following term and definition applies:

5-point median operation: consists of sorting the 5 elements belonging to the set for which the median operation is to be performed in an ascending order according to their values, and selecting the third largest value of the sorted set as the median value.

Further definitions of terms used in the present document can be found in GSM 06.60 [2], GSM 06.81 [3], GSM 05.03 [1] and GSM 08.60 [4].

3.2 Abbreviations

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

BFI	Bad Frame Indication from Radio Sub System
BSI_Abis	Bad Sub-block Indication obtained from A-bis CRC checks
CCU	Channel Coding Unit
CRC	Cyclic Redundancy Check

DTX	Discontinuous transmission
median5	5-point median operation
PrevBFI	Bad Frame Indication of Previous frame
RSS	Radio Sub System
RX	Receive
SID	SIlence Descriptor frame
TRAU	Transcoding Rate Adaptation Unit

4 General

The purpose of frame substitution is to conceal the effect of lost frames. The purpose of muting the output in the case of several lost frames is to indicate the breakdown of the channel to the user and to avoid generating possible annoying sounds as a result from frame substitution procedure.

The RSS indicates lost speech or SID frames by setting its Bad Frame Indication flag (BFI) based on its 3-bit and 8-bit CRCs and possibly other error detection mechanisms. The TRAU calculates from the CRCs inserted by the CCU in the TRAU frames one BSI_Abis flag for every sub-block of speech parameters. If either one or more of these flags is set, the speech decoder shall either perform frame substitution or subframe substitution.

The example solution provided in clause 6 applies only for bad frame handling on a complete speech frame basis. However some parts could be modified for substitution of bad sub-blocks.

5 Requirements

5.1 Error detection

An error is detected and the BFI-flag is set-by the RSS according to the principle described in clause 4.

5.2 Lost speech frames

Normal decoding of lost speech frames would result in very unpleasant noise effects. In order to improve the subjective quality, lost speech frames shall be substituted with either a repetition or an extrapolation of the previous good speech frame(s). This substitution is done so that it will gradually decrease the output level, resulting in silencing of the output. Clause 6.1 gives an example solution.

5.3 First lost SID frame

A single lost SID frame shall be substituted by the last valid SID frame and the procedure for valid SID frames be applied as described in GSM 06.81 [3].

5.4 Subsequent lost SID frames

For the second lost SID frame, a muting technique shall be used on the comfort noise that will gradually decrease the output level (-3 dB/frame), resulting in silencing of the output of the decoder.

For subsequent lost SID frames, the muting of the output shall be maintained. Clause 6.2 gives an example solution.

6 Example solution

The C-code of the following example is embedded in the bit exact software of the enhanced full rate codec.

6.1 Example solution for substitution and muting of lost speech frames

This example solution for substitution and muting is based on a state machine with seven states (figure 1).

The system starts in state 0. Each time a bad frame is detected, the state counter is incremented by one and is saturated when it reaches 6. Each time a good speech frame is detected, the state counter is reset to zero, except when we are in state 6, where we set the state counter to 5. The state indicates the quality of the channel: the bigger the state counter, the worse the channel quality is. The control flow of the state machine can be described with the following C-code (**BFI** = bad frame indicator, **State** = state variable):

```
if(BFI != 0 )
    State = State + 1;
else if(State == 6)
    State = 5;
else
    State = 0;
if(State > 6 )
    State = 6;
```

In addition to this state machine, the **Bad Frame Flag** from the previous frame is checked (**PrevBFI**). The processing depends on the value of the **State**-variable. In states 0 and 5, the processing depends also on the two flags **BFI** and **PrevBFI**.
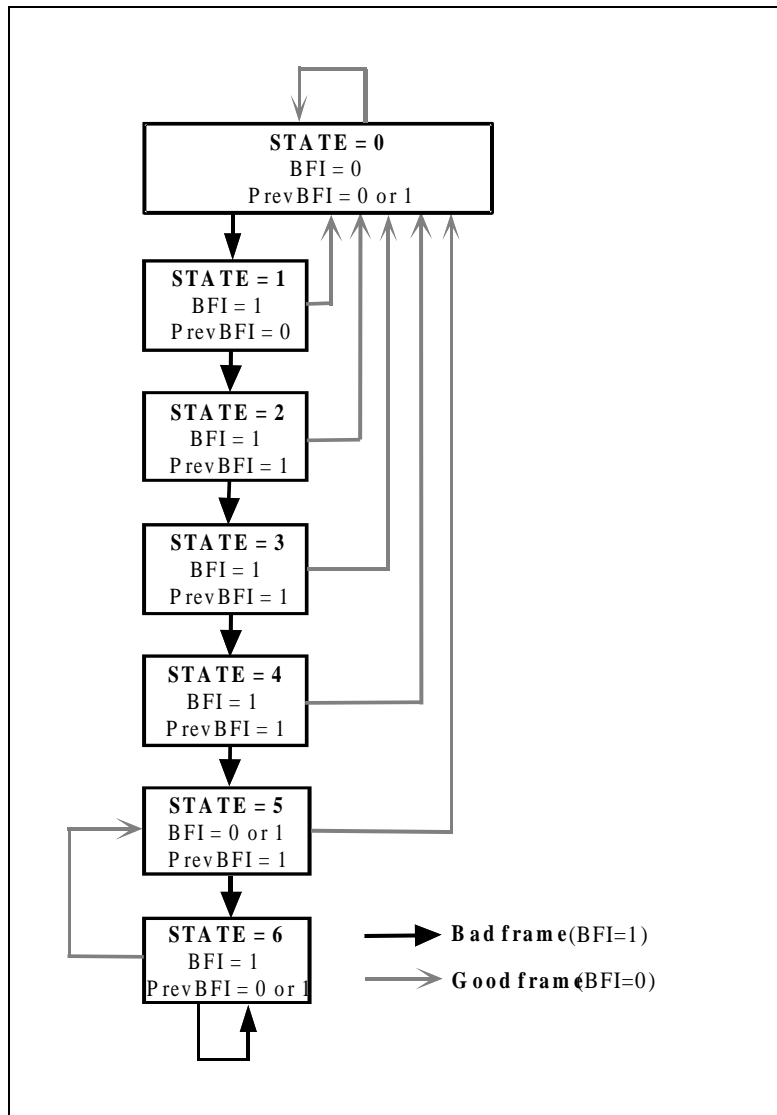
The procedure can be described as follows:

Figure 1: State machine for controlling the bad frame substitution

BFI = 0, PrevBFI = 0, State = 0

No error is detected in the received or in the previous received speech frame. The received speech parameters are used normally in the speech synthesis. The current frame of speech parameters is saved.

BFI = 0, PrevBFI = 1, State = 0 or 5

No error is detected in the received speech frame but the previous received speech frame was bad. The LTP-gain and fixed codebook gain are limited below the values used for the last received good subframe:

$$g^p = \begin{cases} g^p, & g^p \le g^p(-1) \\ g^p(-1), & g^p > g^p(-1) \end{cases}$$

(1)

where g^p = current decoded LTP-gain, $g^p(-1)$ = LTP-gain used for the last good subframe (BFI = 0), and

$$g^c = \begin{cases} g^c, & g^c \le g^c(-1) \\ g^c(-1), & g^c > g^c(-1) \end{cases}$$

(2)

where g^c = current decoded fixed codebook-gain and $g^c(-1)$ = fixed codebook gain used for the last good subframe (BFI = 0).

The rest of the received speech parameters are used normally in the speech synthesis. The current frame of speech parameters is saved.

BFI = 1, PrevBFI = 0 or 1, State = 1...6

An error is detected in the received speech frame and the substitution and muting procedure is started. The LTP-gain and fixed codebook gain are replaced by attenuated values from the previous subframes:

$$g^p = \begin{cases} P(state)\ g^p(-1), & g^p(-1) \leq median5(g^p(-1),\dots,g^p(-5)) \\ P(state)\ median5(g^p(-1),\dots,g^p(-5)), & g^p(-1) > median5(g^p(-1),\dots,g^p(-5)) \end{cases} \tag{3}$$

where g^p = current decoded LTP-gain, $g^p(-1),\dots,g^p(-n)$ = LTP-gains used for the last n subframes, $median5()$ = 5-point median operation, $P(state)$ = attenuation factor ($P(1) = 0.98$, $P(2) = 0.98$, $P(3) = 0.8$, $P(4) = 0.3$, $P(5) = 0.2$, $P(6) = 0.2$), $state$ = state number, and

$$g^c = \begin{cases} C(state)\ g^c(-1), & g^c(-1) \leq median5(g^c(-1),\dots,g^c(-5)) \\ C(state)\ median5(g^c(-1),\dots,g^c(-5)), & g^c(-1) > median5(g^c(-1),\dots,g^c(-5)) \end{cases} \tag{4}$$

where g^c = current decoded fixed codebook gain, $g^c(-1),\dots,g^c(-n)$ = fixed codebook gains used for the last n subframes, $median5()$ = 5-point median operation, $C(state)$ = attenuation factor ($C(1) = 0.98$, $C(2) = 0.98$, $C(3) = 0.98$, $C(4) = 0.98$, $C(5) = 0.98$, $C(6) = 0.7$), and $state$ = state number.

The higher the state value is, the more the gains are attenuated. Also the memory of the predictive fixed codebook gain is updated by using the average value of the past four values in the memory:

$$ener(0) = \frac{1}{4}\sum_{i=1}^{4} ener(-i) \tag{5}$$

The past LSFs are used by shifting their values towards their mean:

$$lsf\_q1(i) = lsf\_q2(i) = \alpha\ past\_lsf\_q(i) + (1-\alpha)mean\_lsf(i), \quad i = 0\dots9 \tag{6}$$

where $\alpha = 0.95$, *lsf_q1* and *lsf_q2* are two sets of LSF-vectors for current frame, *past_lsf_q* is *lsf_q2* from the previous frame, and *mean_lsf* is the average LSF-vector.

The LTP-lag values are replaced by the past value from the 4th subframe of the previous frame.

The received fixed codebook excitation pulses from the erroneous frame are always used as such.

6.2 Example solution for substitution and muting of lost SID frames

The first lost SID frame is replaced by the last valid SID frame.

For subsequent lost SID frames, the last valid SID frame is repeated, but the fixed codebook gain is decreased with a constant value of -3 dB in each frame down to the minimum value of 0. This value is maintained if additional lost SID frames occur.

Annex A (informative):
Change history

Change history					
SMG No.	**TDoc. No.**	**CR. No.**	**Clause affected**	**New version**	**Subject/Comments**
SMG#22				4.0.1	ETSI Publication
SMG#20				5.1.2	Release 1996 version
SMG#27				6.0.0	Release 1997 version
SMG#29				7.0.0	Release 1998 version
				7.0.1	Version update to 7.0.1 for Publication
SMG#31				8.0.0	Release 1999 version
				8.0.1	Update to Version 8.0.1 for Publication

Change history							
Date	**TSG #**	**TSG Doc.**	**CR**	**Rev**	**Subject/Comment**	**Old**	**New**
03-2001	11				Version for Release 4		4.0.0
06-2002	16				Version for Release 5	4.0.0	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
03-2011	51				Version for Release 10	9.0.0	10.0.0
09-2012	57				Version for Release 11	10.0.0	11.0.0