INTERNATIONAL TELECOMMUNICATION UNION

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

G.722.2

Appendix I

(01/2002)

SERIES G: TRANSMISSION SYSTEMS AND MEDIA,
DIGITAL SYSTEMS AND NETWORKS

Digital terminal equipments – Coding of analogue signals by methods other than PCM

Wideband coding of speech at around 16 kbit/s using Adaptive Multi-Rate Wideband (AMR-WB)

Appendix I: Error concealment of erroneous or lost frames

ITU-T Recommendation G.722.2 – Appendix I

ITU-T G-SERIES RECOMMENDATIONS

TRANSMISSION SYSTEMS AND MEDIA, DIGITAL SYSTEMS AND NETWORKS

INTERNATIONAL TELEPHONE CONNECTIONS AND CIRCUITS	G.100–G.199
GENERAL CHARACTERISTICS COMMON TO ALL ANALOGUE CARRIER-TRANSMISSION SYSTEMS	G.200–G.299
INDIVIDUAL CHARACTERISTICS OF INTERNATIONAL CARRIER TELEPHONE SYSTEMS ON METALLIC LINES	G.300–G.399
GENERAL CHARACTERISTICS OF INTERNATIONAL CARRIER TELEPHONE SYSTEMS ON RADIO-RELAY OR SATELLITE LINKS AND INTERCONNECTION WITH METALLIC LINES	G.400–G.449
COORDINATION OF RADIOTELEPHONY AND LINE TELEPHONY	G.450–G.499
TESTING EQUIPMENTS	G.500–G.599
TRANSMISSION MEDIA CHARACTERISTICS	G.600–G.699
DIGITAL TERMINAL EQUIPMENTS	G.700–G.799
General	G.700–G.709
Coding of analogue signals by pulse code modulation	G.710–G.719
Coding of analogue signals by methods other than PCM	**G.720–G.729**
Principal characteristics of primary multiplex equipment	G.730–G.739
Principal characteristics of second order multiplex equipment	G.740–G.749
Principal characteristics of higher order multiplex equipment	G.750–G.759
Principal characteristics of transcoder and digital multiplication equipment	G.760–G.769
Operations, administration and maintenance features of transmission equipment	G.770–G.779
Principal characteristics of multiplexing equipment for the synchronous digital hierarchy	G.780–G.789
Other terminal equipment	G.790–G.799
DIGITAL NETWORKS	G.800–G.899
DIGITAL SECTIONS AND DIGITAL LINE SYSTEM	G.900–G.999
QUALITY OF SERVICE AND PERFORMANCE	G.1000–G.1999
TRANSMISSION MEDIA CHARACTERISTICS	G.6000–G.6999
DIGITAL TERMINAL EQUIPMENTS	G.7000–G.7999
DIGITAL NETWORKS	G.8000–G.8999

For further details, please refer to the list of ITU-T Recommendations.

ITU-T Recommendation G.722.2

Wideband coding of speech at around 16 kbit/s using Adaptive Multi-Rate Wideband (AMR-WB)

Appendix I

Error concealment of erroneous or lost frames

Summary

This appendix specifies a non-normative example solution for concealment of erroneous or lost frames for the G.722.2 AMR-WB codec.

The concealment operations described here were also adopted by 3GPP in 3GPP specification TS 26.191

Source

Appendix I to ITU-T Recommendation G.722.2 was prepared by ITU-T Study Group 16 (2001-2004) and was approved under the WTSa Resolution No. 1 procedure on 13 January 2002.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications. The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementors are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database.

CONTENTS

Page

ITU-T Recommendation G.722.2

Wideband coding of speech at around 16 kbit/s using Adaptive Multi-Rate Wideband (AMR-WB)

Appendix I

Error concealment of erroneous or lost frames

I.1 Scope

This specification defines an example procedure for error concealment, also termed frame substitution and muting procedure, for use by the AMR-WB speech codec receiving end when one or more erroneous/lost speech or lost Silence Insertion Descriptor (SID) frames are received.

The algorithm specified in this appendix is available as part of the ANSI-C code in Annex C/G.722.2. In case of discrepancy between the specification in this appendix and the fixed point computational description of this algorithm contained in Annex C/G.722.2, the description in Annex C/G.722.2 will prevail.

I.2 Definitions and abbreviations

I.2.1 Definitions

This appendix defines the following term:

N-point median operation: Consists of sorting the N 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 (int (N/2) + 1) th largest value of the sorted set as the median value.

I.2.2 Abbreviations

This appendix uses the folowing abbreviations:

AMR-WB	Adaptive Multi-Rate WideBand
AN	Access Network
BFH	Bad Frame Handling
BFI	Bad Frame Indication from AN
BSI_netw	Bad Sub-block Indication obtained from AN interface CRC checks
CRC	Cyclic Redundancy Check
ECU	Error Concealment Unit
medianN	N-point median operation
prevBFI	Bad Frame Indication of previous frame
RX	Receive
SCR	Source Controlled Rate (operation)
SID	Silence Insertion Descriptor (Background noise)

I.3 General

The purpose of the error concealment procedure is to conceal the effect of erroneous/lost AMR-WB speech frames. The purpose of muting the output in the case of several erroneous/lost frames is to indicate the breakdown of the channel to the user and to avoid generating possible annoying sounds as a result from the error concealment procedure.

The network shall indicate erroneous/lost speech or lost SID frames by setting the RX_TYPE values [Annex B/G.722.2] to SPEECH_BAD, SID_BAD or SPEECH_LOST. If these flags are set, the speech decoder shall perform parameter substitution to conceal errors.

The example solution provided in I.5 apply only to bad frame handling on a complete speech frame basis. Sub-frame based error concealment may be derived using similar methods.

I.4 Requirements

I.4.1 Error detection

If the most sensitive bits of the AMR-WB speech data are received in error, the network shall indicate RX_TYPE = SPEECH_BAD in which case the BFI flag is set. When the frame is not received, the network shall indicate RX_TYPE = RX_SPEECH_LOST in which case the BFI flag is set as well. If a SID frame is received in error, the network shall indicate RX_TYPE = SID_BAD.

I.4.2 Erroneous or lost speech frames

Normal decoding of erroneous/lost speech frames would result in very unpleasant noise effects. In order to improve the subjective quality, erroneous/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 gradually will decrease the output level, resulting in silence at the output. Clause I.5 provides example solution.

I.4.3 First lost SID frame

A lost SID frame shall be substituted by using the SID information from earlier received valid SID frames and the procedure for valid SID frames be applied as described in Annex B/G.722.2.

I.4.4 Subsequent lost SID frames

For many subsequent lost SID frames, a muting technique shall be applied to the comfort noise that will gradually decrease the output level. For subsequent lost SID frames, the muting of the output shall be maintained. Clause I.5 provides example solutions.

I.5 Example ECU/BFH Solution

I.5.1 State machine

This example solution for substitution and muting is based on a state machine with seven states (Figure I.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 right-shifted by one. The state indicates the quality of the channel: the larger the value of the state counter, the worse the channel quality is. The control flow of the state machine can be described by the following C code (**BFI** = bad frame indicator, **State** = state variable):

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

```
        State = 6;
else
    State = State >> 1;
```

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 6, the processing depends on the **BFI** flag.

The state machine is summarized in Figure I.1.

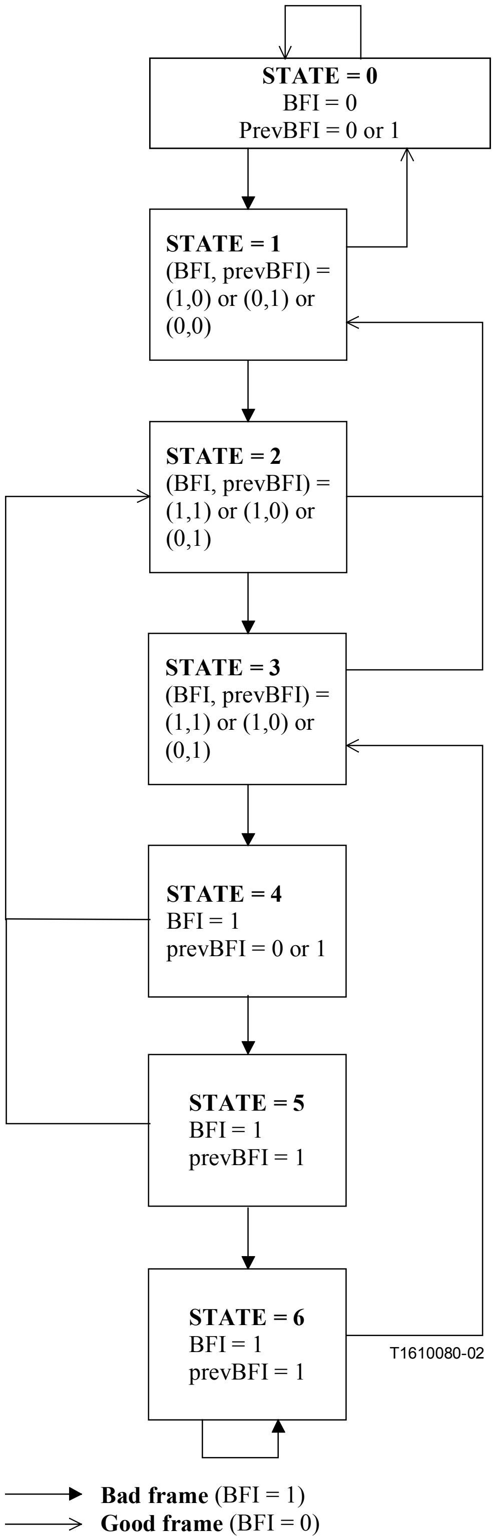

Figure I.1/G.722.2 – State machine for controlling the bad frame substitution

I.5.2 Substitution and muting of erroneous/lost speech frames

I.5.2.1 BFI = 0, prevBFI = 0, State = 0 or 1

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.

I.5.2.2 BFI = 0, prevBFI = 1, State = 0 to 3

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

$$g^c(n) = \begin{cases} g^c_{received} & , g^c_{received} \le 100 \; or \; g^c_{received} \le g^c(n-1) \times 1.25 \\ 1.25 * g^c(n-1) & , otherwise \end{cases} \tag{I-1}$$

where:

$g^c_{received}$ = current decoded fixed codebook-gain

$g^c(n-1)$ = fixed codebook gain used for the last good subframe (BFI = 0)

$g^c(n)$ = fixed codebook gain to be used for the current frame

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

I.5.2.3 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.

I.5.2.3.1 LTP gain & fixed codebook gain concealment when RX_FRAMETYPE = SPEECH_BAD

The LTP gain g^p and fixed codebook gain g^c are replaced by attenuated values from the previous subframes:

$$g^p = P^p(state) * median5(g^p(n-1), \ldots, g^p(n-5)) \tag{I-2}$$

$$g^c = \begin{cases} P^c(state) * median5(g^c(n-1), \ldots, g^c(n-5)) & , \text{VAD\_HIST} \le 2 \\ median5(g^c(n-1), \ldots, g^c(n-5)) & , \text{VAD\_HIST} > 2 \end{cases} \qquad \text{I-3)}$$

where:

g^p = current decoded LTP gain

g^c = current decoded fixed codebook gain

$g^p(n-1), \ldots, g^p(n-5)$ = LTP gains used for the last 5 subframes

$g^c(n-1), \ldots, g^c(n-5)$ = fixed codebook gains used for the last 5 subframes

median5() = 5-point median operation

$P^p(state)$ = attenuation factor ($P^p(1) = 0.98$, $P^p(2) = 0.96$, $P^p(3) = 0.75$, $P^p(4) = 0.23$, $P^p(5) = 0.05$, $P^p(6) = 0.01$)

$P^c(state)$ = attenuation factor ($P^c(1) = 0.98$, $P^c(2) = 0.98$, $P^c(3) = 0.98$, $P^c(4) = 0.98$, $P^c(5) = 0.98$, $P^c(6) = 0.70$)

state = state number {0..6}

VAD_HIST is number of consecutive VAD = 0 decisions

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:

$$ener(0) = \frac{1}{4}\left[\sum_{i=1}^{4} ener(n-i)\right] - 3 \quad \text{(I-4)}$$

I.5.2.3.2 LTP gain & fixed codebook gain concealment when RX_FRAMETYPE = SPEECH_LOST

The LTP gain g^p and fixed codebook gain g^c are replaced by attenuated values from the previous subframes:

$$g^p = P^p(state) * median5(g^p(n-1),...,g^p(n-5)) \quad \text{(I-5)}$$

$$g^c = \begin{cases} P^c(state) * median5(g^c(n-1),...,g^c(n-5)) & ,\text{VAD\_HIST} \le 2 \\ median5(g^c(n-1),...,g^c(n-5)) & ,\text{VAD\_HIST} > 2 \end{cases} \quad \text{I-6)}$$

where:

g^p = current decoded LTP gain,

g^c = current decoded fixed codebook gain

$g^p(n-1),...,g^p(n-5)$ = LTP gains used for the last 5 subframes

$g^c(n-1),...,g^c(n-5)$ = fixed codebook gains used for the last 5 subframes

median5() = 5-point median operation

$P^p(state)$ = attenuation factor ($P^p(1) = 0.95$, $P^p(2) = 0.90$, $P^p(3) = 0.75$, $P^p(4) = 0.23$, $P^p(5) = 0.05$, $P^p(6) = 0.01$)

$P^c(state)$ = attenuation factor ($P^c(1) = 0.50$, $P^c(2) = 0.25$, $P^c(3) = 0.25$, $P^c(4) = 0.25$, $P^c(5) = 0.15$, $P^c(6) = 0.01$)

state = state number {0..6}

VAD_HIST is number of consecutive VAD = 0 decisions

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:

$$ener(0) = \frac{1}{4}\left[\sum_{i=1}^{4} ener(n-i)\right] - 3 \quad \text{(I-7)}$$

I.5.2.3.3 ISF concealment

The past ISFs are shifted towards their partly adaptive mean:

$$ISF_q(i) = \alpha * past\_ISF_q(i) + (1-\alpha) * ISF_{mean}(i) \qquad i = 0..16 \quad \text{(I-8)}$$

where:

$\alpha = 0.9$

$ISF_q(i)$ is ISF-vector for a current frame

$past\_ISF_q(i)$ is ISF-vector from the previous frame

$ISF_{mean}(i)$ vector is combination of adaptive mean and constant mean ISF-vectors in the following manner:

$$ISF_{mean}(i) = \beta * ISF_{const\_mean}(i) + (1-\beta) * ISF_{adaptive\_mean}(i) \qquad i = 0..16 \tag{I-9}$$

where:

$\beta = 0.75$

$ISF_{adaptive\_mean}(i) = \frac{1}{3}\sum_{i=0}^{2} past\_ISF_q(i)$ and is updated whenever BFI = 0

$ISF_{const\_mean}(i)$ is a vector containing long time average of ISF-vectors

I.5.2.3.4 LTP-lag concealment

The histories of five last good LTP-lags and LTP-gains are used for finding the best method to update.

I.5.2.3.4.1 LTP-lag concealment when RX_FRAMETYPE = SPEECH_BAD

The usability of the received LTP lag (Q_{lag}) is defined as follows: (Predicts if the received lag is most probably very close to one that was sent and therefore its usage should not introduce any bad artifacts.)

$$Q_{lag} = \begin{cases} 1 & , T_{dif} < 10 \text{ and } \mathrm{T}_{\min} - 5 < T_{received} < \mathrm{T}_{\min} + 5 \\ 1 & , \mathrm{g^p}(n-1) > 0.5 \text{ and } \mathrm{g^p}(n-2) > 0.5 \text{ and } \mathrm{T}\ (\mathrm{n}-1) - 10 < T_{received} < \mathrm{T}\ (\mathrm{n}-1) + 10 \\ 1 & , g^p_{\min} < 0.4 \text{ and } g^p(n-1) = g^p_{\min} \text{ and } \mathrm{T}_{\min} < T_{received} < \mathrm{T}_{\max} \\ 1 & , T_{dif} < 70 \text{ and } \mathrm{T}_{\min} < T_{received} < \mathrm{T}_{\max} \\ 1 & , T_{mean} < T_{received} < \mathrm{T}_{\max} \\ 0 & , otherwise \end{cases} \tag{I-10}$$

where:

$T(n-1)$ is LTP lag from the previous good frame

$T_{dif} = \left|T_{received} - T(n-1)\right|$

$T_{\min} = \min(T_{buffer})$

$T_{\max} = \max(T_{buffer})$

$T_{received}$ is received lag

$g^p_{\min} = \min(g^p_{buffer})$

g^p is LTP gain of the current frame

$g^p(-1)$ is LTP gain of the previous good frame

$g^p(-2)$ is LTP gain of the frame before previous good frame

$T_{mean} = average(T_{buffer})$

LTP lag value for the current frame is defined as follows:

$$T = \begin{cases} T_{received} & ,\mathrm{Q_{lag}} = 1 \\ \frac{1}{3}\sum(T_{\max} + T_{\max-1} + T_{\max-2}) + RND(T_{\max} - T_{\max-2}) & ,\mathrm{Q_{lag}} = 0 \end{cases} \tag{I-11}$$

where:

$T_{\max} = \max(T_{buffer})$

$T_{\max-1}$ is second largest value in T_{buffer}

$T_{\max-2}$ is second largest value in T_{buffer}

$RND(x)$ is random value generated to range $\left[-\frac{x}{2}, +\frac{x}{2}\right]$

I.5.2.3.4.2 LTP-lag concealment when RX_FRAMETYPE = SPEECH_LOST

The usability of the LTP lag from last good frame (Q_{lag_t-1}) is defined as follows: (Predicts if the received lag is most probably very close to one that was sent and therefore its usage should not introduce any bad artifacts.)

$$Q_{lag\_t-1} = \begin{cases} 1 & ,g^p_{\min} > 0.5 \text{ and } \mathrm{T_{dif}} < 10 \\ 1 & ,\mathrm{g^p}(n-1) > 0.5 \text{ and } \mathrm{g^p}(n-2) > 0.5 \\ 0 & ,otherwise \end{cases} \tag{I-12}$$

where:

$g^p_{\min} = \min(g^p_{buffer})$

g^p (n–1) is LTP gain of the previous good frame

g^p (n–2) is LTP gain of the frame before previous good frame

LTP lag value for the current frame is defined as follows:

$$T = \begin{cases} T(n-1) & ,\mathrm{Q_{lag\_t-1}} = 1 \\ \frac{1}{3}\sum(T_{\max} + T_{\max-1} + T_{\max-2}) + RND(T_{\max} - T_{\max-2}) & ,\mathrm{Q_{lag\_t-1}} = 0 \end{cases} \tag{I-13}$$

where:

$T(n-1)$ is LTP lag from the previous good frame

$T_{\max} = \max(T_{buffer})$

$T_{\max-1}$ is second largest value in T_{buffer}

$T_{\max-2}$ is second largest value in T_{buffer}

$RND(x)$ is random value generated to range $\left[-\frac{x}{2},+\frac{x}{2}\right]$

I.5.2.4 Innovation sequence

When RX_FRAMETYPE = SPEECH_BAD, the received fixed codebook innovation pulses from the erroneous frame are used as they are received.

When RX_FRAMETYPE = SPEECH_LOST, the received fixed codebook innovation pulses from the erroneous frame are not used and the fixed codebook innovation vector is filled with random signal (values limited to range [–1, +1]).

I.5.3 Substitution and muting of lost SID frames

In the speech decoder a single frame classified as SID_BAD shall be substituted by the last valid SID frame information and the procedure for valid SID frames be applied. If the time between SID information updates (updates are specified by SID_UPDATE arrivals and occasionally by SID_FIRST arrivals) is greater than one second this shall lead to attenuation.

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series B	Means of expression: definitions, symbols, classification
Series C	General telecommunication statistics
Series D	General tariff principles
Series E	Overall network operation, telephone service, service operation and human factors
Series F	Non-telephone telecommunication services
Series G	**Transmission systems and media, digital systems and networks**
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	Cable networks and transmission of television, sound programme and other multimedia signals
Series K	Protection against interference
Series L	Construction, installation and protection of cables and other elements of outside plant
Series M	TMN and network maintenance: international transmission systems, telephone circuits, telegraphy, facsimile and leased circuits
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	Telephone transmission quality, telephone installations, local line networks
Series Q	Switching and signalling
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks and open system communications
Series Y	Global information infrastructure and Internet protocol aspects
Series Z	Languages and general software aspects for telecommunication systems

Geneva, 2002