

International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

G.722

Appendix IV
(11/2006)

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

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

7 kHz audio-coding within 64 kbit/s

**Appendix IV: A low-complexity algorithm for
packet loss concealment with G.722**

ITU-T Recommendation G.722 – Appendix IV



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
TRANSMISSION MEDIA AND OPTICAL SYSTEMS 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 – GENERIC AND USER-RELATED ASPECTS	G.1000–G.1999
TRANSMISSION MEDIA CHARACTERISTICS	G.6000–G.6999
DATA OVER TRANSPORT – GENERIC ASPECTS	G.7000–G.7999
PACKET OVER TRANSPORT ASPECTS	G.8000–G.8999
ACCESS NETWORKS	G.9000–G.9999

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

ITU-T Recommendation G.722

7 kHz audio-coding within 64 kbit/s

Appendix IV

A low-complexity algorithm for packet loss concealment with G.722

Summary

Packet loss concealment (PLC) algorithms, also known as frame erasure concealment algorithms, hide transmission losses in audio systems where the input signal is encoded and packetized, sent over a network, received and decoded before play out. PLC algorithms can be found in most standard CELP-based speech coders. The algorithm described here provides a method for [ITU-T G.722].

The decoder comprises three stages: lower sub-band decoding, higher sub-band decoding and quadrature mirror filter (QMF) synthesis. In the absence of frame erasures, the decoder structure is identical to G.722, except for the storage of the two decoded signals, of the high and low bands. In case of frame erasures, the decoder is informed by the bad frame indication (BFI) signalling. It then performs an analysis of the past lower-band reconstructed signal and extrapolates the missing signal using linear-predictive coding (LPC), pitch-synchronous period repetition and adaptive muting. Once a good frame is received, the decoded signal is cross-faded with the extrapolated signal. In the higher band, the decoder repeats the previous frame pitch-synchronously, with adaptive muting and high-pass post-processing. The ADPCM states are updated after each frame erasure.

The PLC algorithm described in this appendix meets the same quality requirements as the PLC in Appendix III of [ITU-T G.722], but with a lower complexity. Almost no additional complexity is added compared with normal G.722 decoding (worst-case additional complexity is 0.07 WMOPS). An alternative quality-complexity trade-off is provided by Appendix III.

Source

Appendix IV to ITU-T Recommendation G.722 was agreed on 24 November 2006 by ITU-T Study Group 16 (2005-2008).

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.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure e.g. interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

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, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database at <http://www.itu.int/ITU-T/ipr/>.

© ITU 2007

All rights reserved. No part of this publication may be reproduced, by any means whatsoever, without the prior written permission of ITU.

CONTENTS

	Page
Appendix IV – A low-complexity algorithm for packet loss concealment with G.722	1
IV.1 Scope	1
IV.2 References	1
IV.3 Abbreviations	1
IV.4 Notations and conventions	2
IV.5 General description of the G.722 PLC algorithm.....	2
IV.6 Functional description of the G.722 PLC algorithm	4
IV.7 Bit-exact description of the G.722 PLC algorithm.....	15

ITU-T Recommendation G.722

7 kHz audio-coding within 64 kbit/s

Appendix IV

A low-complexity algorithm for packet loss concealment with G.722

IV.1 Scope

This appendix describes a low-complexity packet loss concealment (PLC) algorithm for use by the G.722 audio decoder to limit speech quality degradation in the presence of packet loss. The statistical analysis of the G.722 PLC Selection Test results has demonstrated that this appendix meets the same quality requirements as Appendix III. This is achieved with almost no additional complexity compared with G.722 normal decoding (0.07 WMOPS), but with a lower quality than Appendix III. Accordingly, this algorithm is suitable for applications that may encounter frame erasures or packet losses, with a special focus on applications having strong complexity constraints and on low-cost devices. As examples, these applications include DECT Next Generation and VoIP.

The references, abbreviations and notations used throughout this appendix are defined in clauses IV.2, IV.3 and IV.4, respectively. Clause IV.5 gives a general description of the algorithm and includes delay and complexity information. Clause IV.6 contains the detailed functional description of the algorithm. Finally, clause IV.7 provides the information related to the simulation software.

IV.2 References

- [ITU-T G.191 Ann.A] ITU-T Recommendation G.191 (2005), *Software tools for speech and audio coding standardization. Annex A: List of software tools available.*
- [ITU-T G.192] ITU-T Recommendation G.192 (1996), *A common digital parallel interface for speech standardization activities.*
- [ITU-T G.722] ITU-T Recommendation G.722 (1988), *7 kHz audio-coding within 64 kbit/s.*
- [ITU-T G.729] ITU-T Recommendation G.729 (2007), *Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited-linear-prediction (CS-ACELP).*

IV.3 Abbreviations

This appendix uses the following abbreviations and acronyms:

FIR	Finite Impulse Response
HB	Higher Band
HPF	High Pass Filter
IIR	Infinite-Impulse Response
LB	Lower Band
LP	Linear Prediction
LPC	Linear Predictive Coding

LSB	Least Significant Bit
LTP	Long-Term Prediction
MSB	Most Significant Bit
PLC	Packet Loss Concealment
QMF	Quadrature Mirror Filter
WB	Wideband
WMOPS	Weighted Million Operations Per Second

IV.4 Notations and conventions

Throughout this appendix, the terms "*frame*" and "*packet*" are considered to be equivalent.

The notation used in [ITU-T G.722] is followed. The notational conventions are detailed below:

- Each frame comprises $2L$ samples at 16 kHz, where $L = 80$ for 10-ms frames and 160 for 20-ms frames.
- Time-domain signals are denoted by their symbol and a sample index between parentheses [e.g., $y(n)$]. The symbol n is used as sample index.
- By convention, for signals sampled at 8 kHz, the current frame corresponds to $n = 0, \dots, L-1$. Samples with an index $n < 0$ are past samples, and samples with index $n \geq L$ are future samples.

Table IV.1 – List of symbols

Type	Name	Description
Filters	$A(z)$	8th-order LPC filter
	$B(z)$	2nd-order LPC filter
	$H_{pre}(z)$	Lower-band high-pass filter
	$H_{post}(z)$	Higher-band high-pass filter
	$H_{dec}(z)$	4:1 decimation filter
Signals	$x_l(n)$	Lower-band ADPCM decoder output
	$y_l(n)$	Lower-band extrapolated signal
	$z_l(n)$	Lower-band reconstruction
	$x_h(n)$	Higher-band ADPCM decoder output
	$y_h(n)$	Higher-band extrapolated signal
	$z_h(n)$	Higher-band reconstruction
	$e(n)$	Lower-band LP residual signal
Parameters	a_i	Lower-band LP coefficients
	T_0	Pitch delay in lower band
	g_mute_lb	Lower-band muting factor
	g_mute_hb	Higher-band muting factor

IV.5 General description of the G.722 PLC algorithm

The proposed G.722 PLC algorithm is integrated in the standard G.722 decoder. Note that the G.722 encoder is unchanged (identical to clause 3 of [ITU-T G.722]).

IV.5.1 Modified G.722 decoder

The modified G.722 decoder is illustrated in Figure IV.1. Decoding is performed in two subbands, which are combined using the QMF synthesis filterbank of G.722. Compared to G.722, this decoder includes a mechanism to conceal frame erasures, shown in the highlighted (gray-shaded) blocks.

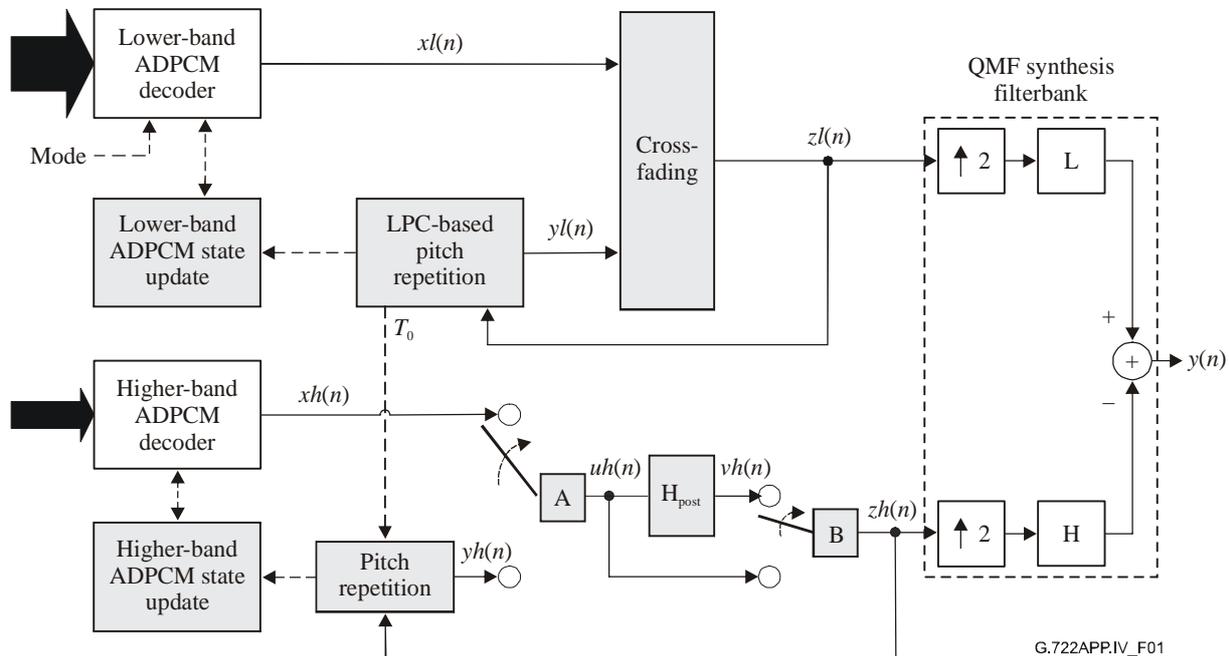


Figure IV.1 – Block diagram of G.722 decoder with PLC (the highlighted blocks correspond to the PLC algorithm)

The modified G.722 decoder generates an output signal sampled at 16 kHz and divided into frames of 10 or 20 ms. Its behaviour depends on the type of the current and previous frame (either good or bad frame):

- Without frame erasures (i.e., in the presence of good frames only):

The bitstream of the lower band (LB) is decoded according to the specified mode (1, 2 or 3 indicating 64, 56 or 48 kbit/s, respectively). The cross-fading block does not change the reconstructed signal, i.e., $z_l(n) = x_l(n)$; similarly, the bitstream of the higher band (HB) is decoded. The A and B switches select $u_h(n) = x_h(n)$ and $z_h(n) = u_h(n) = x_h(n)$, respectively. The wideband reconstruction $y(n)$ is obtained, as in G.722, from the QMF synthesis filterbank. The decoded signals $z_l(n)$ and $z_h(n)$ are stored to be used in case of erasure in future frames.
- In case of frame erasure:
 - In the lower band, for the first erased frame, short- and long-term predictors are updated using the past valid signal $z_l(n)$, $n < 0$. Class information is also extracted. The signal $y_l(n)$ is generated using these predictors and the class information. The signal for the erased frame is reconstructed as $z_l(n) = y_l(n)$, $n = 0, \dots, L-1$. In addition, ADPCM states are updated. The process of erased frame reconstruction and ADPCM states update is repeated until a good frame is received. Note that not only the missing frame is generated, but also an additional 10 ms of signal, $y_l(n)$, $n = L, \dots, L + 79$, to be used for cross-fading. Once a good frame is received, the signal reconstructed by the ADPCM decoder, $x_l(n)$, $n = 0, \dots, L-1$, and the extrapolated

signal, $yl(n)$, $n = 0, \dots, L-1$, are cross-faded. This cross-fading is used only during the first 10 ms following the last erasure.

- In the higher band, the missing frame is extrapolated using the past signal $zh(n)$, $n < 0$, and ADPCM states are updated. The extrapolated signal $yh(n)$ is obtained by repeating pitch-synchronously the previous frame of $zh(n)$. The "A" switch selects $uh(n) = yh(n)$, $n = 0, \dots, L-1$. The signal $uh(n)$ is high-pass filtered by a remove-DC filter H_{post} to obtain $vh(n)$. The "B" switch selects $zh(n) = vh(n)$, $n = 0, \dots, L-1$. This process is repeated until a good frame is received. Once a good frame is received, the "A" switch selects the ADPCM decoder output: $uh(n) = xh(n)$, $n = 0, \dots, L-1$. During the first 4 s following the last erasure, the "B" switch selects: $zh(n) = vh(n)$, $n = 0, \dots, L-1$; after 4 s, the B switch selects: $zh(n) = uh(n)$, $n = 0, \dots, L-1$.

IV.5.2 Delay and complexity

The proposed G.722 PLC operates with no extra delay.

The computational complexity and storage requirements are summarized in Table IV.2.

Table IV.2 – Complexity figures of the G.722 PLC algorithm

- a) **Observed worst-case complexity, in WMOPS, based on STL2005 (figures in brackets represent the additional WMOPS compared with G.722 decoding)**

Frame length	10 ms	20 ms
G.722 decoding (no PLC)	3.11	3.10
G.722 with PLC	3.18 [0.07]	3.15 [0.05]

- b) **Memory requirement (figures in brackets represent the maximum additional complexity compared with G.722 decoding)**

Frame length	Static RAM (in 16-bit kwords)	Scratch RAM (in 16-bit kwords)	Total RAM	Program ROM (in number of basic ops and function calls)	Data ROM (in 16-bit kwords)
10 ms	967 [863]	692 [452]	1659 [1315]	1061 [615]	882 [109]
20 ms		963 [452]	1899 [1315]		

Compared to the G.722 reference decoder, the PLC algorithm brings an additional worst-case complexity of 0.07 WMOPS. Note that the complexity peak occurs at the first valid frame after an erasure.

IV.6 Functional description of the G.722 PLC algorithm

IV.6.1 Lower-band decoding

IV.6.1.1 ADPCM decoder in case of a good frame

Same as clauses 4.1, 4.2 and 4.3 of [ITU-T G.722].

In addition, the counter cnt_mute_lb and muting factor g_mute_lb (used in clause IV.6.1.2.7) for adaptive muting are reset:

$$cnt_mute_lb = 0$$

$$g_mute_lb = 1$$

and the signal $z_l(n)$ is stored to be used in case of erasure in future frames.

IV.6.1.2 Extrapolation of missing frame: Case of bad frame following a good frame

The extrapolation of a missing frame in the lower band is illustrated in Figure IV.2. It comprises analysis of the past valid signal $z_l(n)$, $n < 0$, followed by synthesis of the signal $y_l(n)$, $n = 0, \dots, L-1$. The missing frame in the lower band corresponds to $x_l(n)$, $n = 0, \dots, L-1$, where $L = 80$ for 10-ms frames and $L = 160$ for 20-ms frames.

The past signal $z_l(n)$, $n = -297, \dots, -1$, is buffered using a buffer length of 297 samples, which can be divided as follows:

- the 288 samples corresponding to twice the maximal pitch delay (2×144) used in the PLC algorithm;
- one sample for pitch jitter; and
- eight samples used for LPC memory.

This buffer length makes it possible to store the last two pitch periods of the lower-band reconstruction.

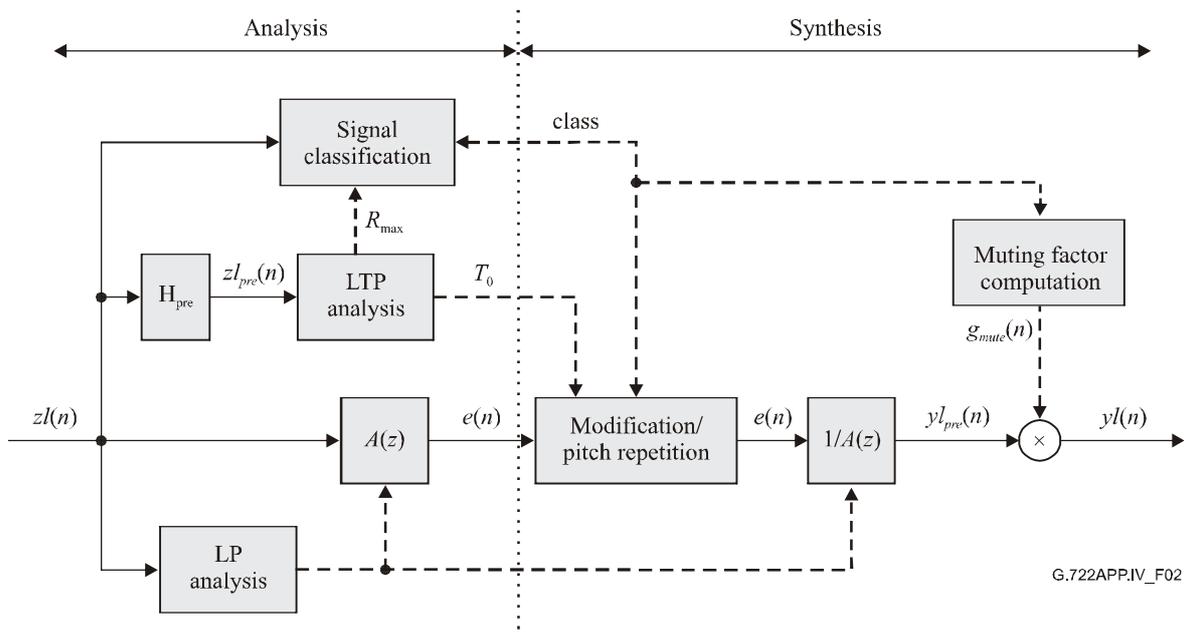


Figure IV.2 – Block diagram of lower-band extrapolation of missing frame

IV.6.1.2.1 LP analysis

The short-term analysis and synthesis filters, $A(z)$ and $1/A(z)$, are based on eighth-order linear prediction (LP) filters. The LP analysis filter is defined as:

$$A(z) = 1 + a_1 z^{-1} + a_2 z^{-2} + \dots + a_8 z^{-8} \quad (\text{IV-1})$$

The LP analysis consists of two parts: windowing and autocorrelation computation, and the Levinson-Durbin algorithm. The autocorrelation computation – including 60-Hz bandwidth expansion and 40-dB white-noise correction – is identical to that in clause 3.2.1 of [ITU-T G.729], with only one difference. The LP window here is an asymmetrical Hamming window defined as:

$$w_{lp}(n) = \begin{cases} 0.54 - 0.46 \cos\left(\frac{(n+80)\pi}{69}\right), & n = -80, \dots, -11 \\ 0.54 + 0.46 \cos\left(\frac{(n+11)\pi}{10}\right), & n = -10, \dots, -1 \end{cases} \quad (\text{IV-2})$$

This window $w_{lp}(n)$, which is limited to 80 samples (10 ms at 8-kHz sampling frequency) to reduce complexity, is applied to the last 10 ms of $z_l(n)$, $n = -80, \dots, -1$. The Levinson-Durbin algorithm is identical to clause 3.2.2 of [ITU-T G.729].

After the LP analysis, the past signal $z_l(n)$, $n = -289, \dots, -1$, is filtered through $A(z)$ to obtain the residual signal $e(n)$, $n = -289, \dots, -1$:

$$e(n) = z_l(n) + \sum_{i=1}^8 a_i z_l(n-i) \quad (\text{IV-3})$$

IV.6.1.2.2 Pre-processing

A high-pass filter protects against undesired low-frequency components. A first-order pole/zero filter with a cut-off frequency of 50 Hz is used. This filter is given by:

$$H_{pre}(z) = \frac{1 - z^{-1}}{1 - \frac{123}{128} z^{-1}} \quad (\text{IV-4})$$

The past signal $z_l(n)$, $n = -288, \dots, -1$, is filtered through $H_{pre}(z)$ to obtain the pre-processed signal $z_{l_{pre}}(n)$, $n = -288, \dots, -1$.

IV.6.1.2.3 LTP analysis

The PLC algorithm uses pitch period repetition. The pitch period or pitch delay, T_0 , is determined on the past valid pre-processed signal just before erasure, $z_{l_{pre}}(n)$, $n = -288, \dots, -1$. T_0 is estimated in open loop by a long-term predictive (LTP) analysis.

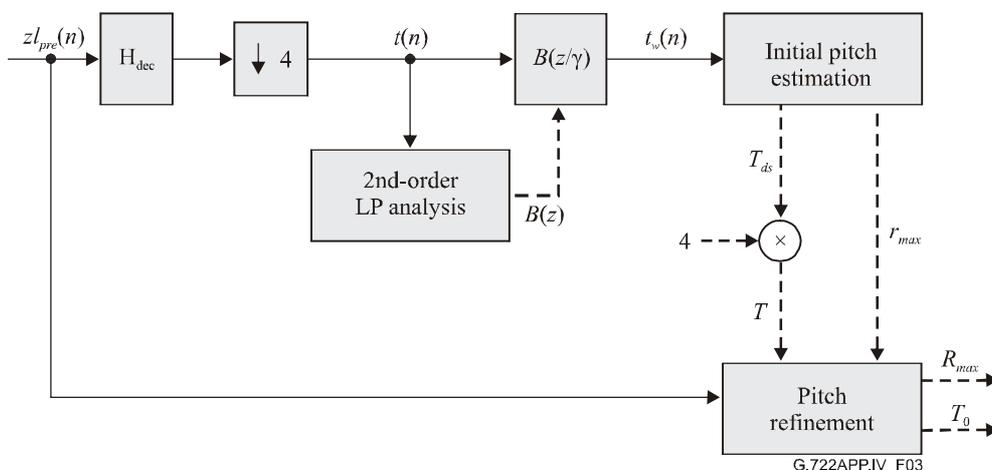


Figure IV.3 – Block diagram of LTP analysis

As illustrated in Figure IV.3, pitch estimation is conducted in the following steps:

- The signal $z^{l_{pre}}(n)$, $n = -288, \dots, -1$, is low-pass filtered by $H_{dec}(z)$, where:

$$H_{dec}(z) = \frac{3692(1+z^{-8}) + 6190(z^{-1}+z^{-7}) + 8525(z^{-2}+z^{-6}) + 10186(z^{-3}+z^{-5}) + 10787z^{-4}}{65536} \quad (\text{IV-5})$$

is an eighth-order FIR filter, and decimated by a factor of 4 to obtain the signal $t(n)$, $n = -72, \dots, -1$, sampled at 2 kHz. The filter memories are initialized to 0 each time.

- The signal $t(n)$, $n = -72, \dots, -1$, is weighted by a filter $B(z/\gamma)$, where $B(z) = 1 - b_1z^{-1} - b_2z^{-2}$ and $\gamma = 0.94$, to obtain the signal $t_w(n)$, $n = -72, \dots, -1$. The coefficients of $B(z)$ are obtained by 2nd-order LP analysis of $t(n)$ using the windowing, autocorrelation computation and Levinson-Durbin algorithm described in the previous paragraph. Note that only the last 72 samples of the window $w_p(n)$, $n = -72, \dots, -1$, are used, which gives a 36-ms time support at 2-kHz sampling frequency.

- A first estimation T_{ds} of the pitch delay is computed in the weighted decimated signal domain by normalized cross-correlation as follows:

a) Initialization: $T_{ds} = 18$.

b) Computation of the normalized cross-correlation:

$$r(i) = \frac{\sum_{j=-35}^{-1} t_w(j) t_w(j-i)}{\max\left(\sum_{j=-35}^{-1} t_w^2(j), \sum_{j=-35}^{-1} t_w^2(j-i)\right)}, \quad i = 1, \dots, 35 \quad (\text{IV-6})$$

c) Determination of the first delay i_0 in $[1, 35]$ for which $r(i) < 0$. Note that if $\min_{i=1, \dots, 35} r(i) \geq 0$, steps d) and e) are omitted.

d) Determination of the lower bound for the maximum correlation search:

$$i_1 = \max(i_0, 4)$$

e) Search for the maximum correlation in $[i_1, 35]$:

$$T_{ds} = \arg \max_{i=i_1, \dots, 35} r(i) \quad (\text{IV-7})$$

A procedure favouring the smaller pitch values to avoid choosing pitch multiples is also applied.

- The first pitch delay estimation T_{ds} is then refined in the pre-processed signal domain by searching the cross-correlation maximum in the neighbourhood of $T = 4T_{ds}$ to obtain T_0 as follows:

$$T_0 = \arg \max_{i=T-2, \dots, T+2} R(i) \quad (\text{IV-8})$$

with:

$$R(i) = \frac{\sum_{j=-T}^{-1} zl_{pre}(j) zl_{pre}(j-i)}{\max\left(\sum_{j=-T}^{-1} zl_{pre}^2(j), \sum_{j=-T}^{-1} zl_{pre}^2(j-i)\right)} \quad (IV-9)$$

- The value R_{\max} is computed as $R_{\max} = R(T_0)$.

IV.6.1.2.4 Signal classification

The PLC strategy uses signal characteristics to optimize quality. For instance, if the frame preceding an erasure is a non-stationary segment (e.g., plosives), the signal should be rapidly muted; if this frame is a stationary segment (e.g., strongly voiced speech), it can be pitch-synchronously repeated and slowly damped. Classification is used in the proposed PLC algorithm for LP residual extrapolation and muting control.

The signal $zl(n)$, $n = -288, \dots, -1$, preceding an erasure is classified into one out of five possible classes, which are defined below:

- TRANSIENT for transients with large energy variation (e.g., plosives);
- UNVOICED for unvoiced signals;
- VUV_TRANSITION corresponding to a transition between voiced and unvoiced signals;
- WEAKLY_VOICED for weakly voiced signals (e.g., onset or offset of vowels);
- VOICED for voiced signals (e.g., steady vowels).

The features used for classification are listed below:

- Normalized correlation R_{\max} , which is a side product of the LTP analysis.
- Subband energy ratio, which is obtained here in the log domain by taking the difference between the lower- and higher-band ADPCM scale factors, NBH – NBL using the G.722 notations. NBL and NBH are computed as in clause 3.5 of [ITU-T G.722].
- Zero-crossing rate zcr of $zl(n)$, $n = -80, \dots, -1$, defined as:

$$zcr = \sum_{n=-80}^{-1} [(zl(n) \leq 0) \text{ AND } (zl(n-1) > 0)] \quad (IV-10)$$

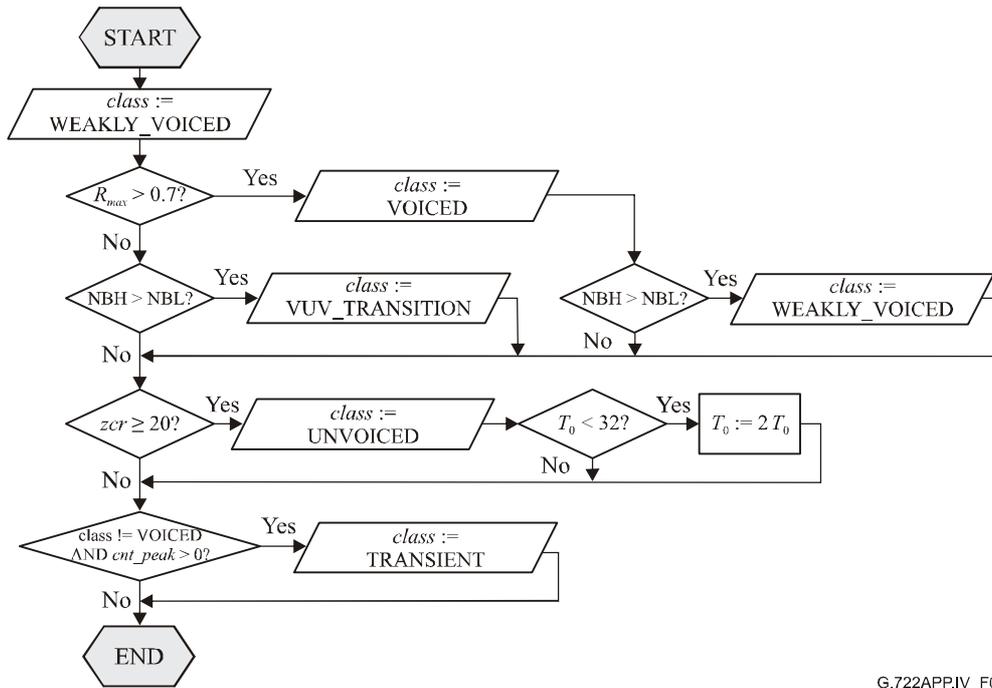
where the comparisons \leq and $>$ give a binary result (1 for true, 0 for false) and "AND" is the AND bit operation.

- Number cnt_peak of detected large peaks in the last pitch period:

$$cnt_peak = \sum_{n=-T_0}^{-1} \left[\frac{|e(n)|}{4} > \max_{i=-2, \dots, 2} (|e(n-T_0+i)|) \right] \quad (IV-11)$$

where the comparison $>$ gives a binary result (1 for true, 0 for false) and T_0 is the pitch delay estimated by the LTP analysis. The counter, cnt_peak , represents the number of detected large peaks in the last pitch period.

Based on these features, the signal category, $class$, is obtained by heuristics according to the flowchart shown in Figure IV.4. Note that if $class$ is set to UNVOICED, the pitch delay T_0 may be modified to avoid artefacts due to low-pitch delay values. Note also that if $class$ is not VOICED, and T_0 is even, T_0 is increased by 1. The so-called pitch delay T_0 determines the repetition period used in the residual signal generation procedure.



G.722APP_IV_F04

Figure IV.4 – Classification flowchart

IV.6.1.2.5 Modification/pitch repetition of LP residual

A pitch repetition procedure is used to estimate the LP residual $e(n)$, $n = 0, \dots, L-1$, in the missing frame from the residual signal of the repetition period. As mentioned above, the repetition period contains the last valid T_0 residual samples $e(n)$, $n = -T_0, \dots, -1$.

LP residual modification

Before performing the pitch repetition procedure, the repetition period is modified if $class$ is not VOICED. The modification consists in limiting the magnitude of each sample of the repetition period as follows:

$$e(n) = \min\left(\max_{i=-2, \dots, +2} (|e(n-T_0+i)|), |e(n)|\right) \times \text{sign}(e(n)), \quad n = -T_0, \dots, -1 \quad (\text{IV-12})$$

where:

$$\text{sign}(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ -1 & \text{if } x < 0 \end{cases} \quad (\text{IV-13})$$

Pitch repetition of LP residual

The LP residual $e(n)$, $n = 0, \dots, L-1$, in the missing frame is extrapolated based on the classification result:

- If $class$ is VOICED, the missing signal, $e(n)$, $n = 0, \dots, L-1$, is obtained by repeating pitch-synchronously the repetition period:

$$e(n) = e(n-T_0) \quad (\text{IV-14})$$

- If *class* is not VOICED, the pitch-synchronous repetition procedure is modified to avoid over-voicing by introducing, sample by sample, a small jitter using the following procedure. The samples of the repetition period can be viewed as grouped two by two; then, every two samples forming a group are swapped and the swapped groups are concatenated to form the extrapolated residual signal. If $T_0 < L$, the extrapolated residual signal extends the repetition period and the procedure is iterated. With this procedure, the missing signal, $e(n)$, $n = 0, \dots, L-1$, is obtained as:

$$e(n) = e(n - T_0 + (-1)^n) \quad (\text{IV-15})$$

This procedure is illustrated in Figure IV.5.

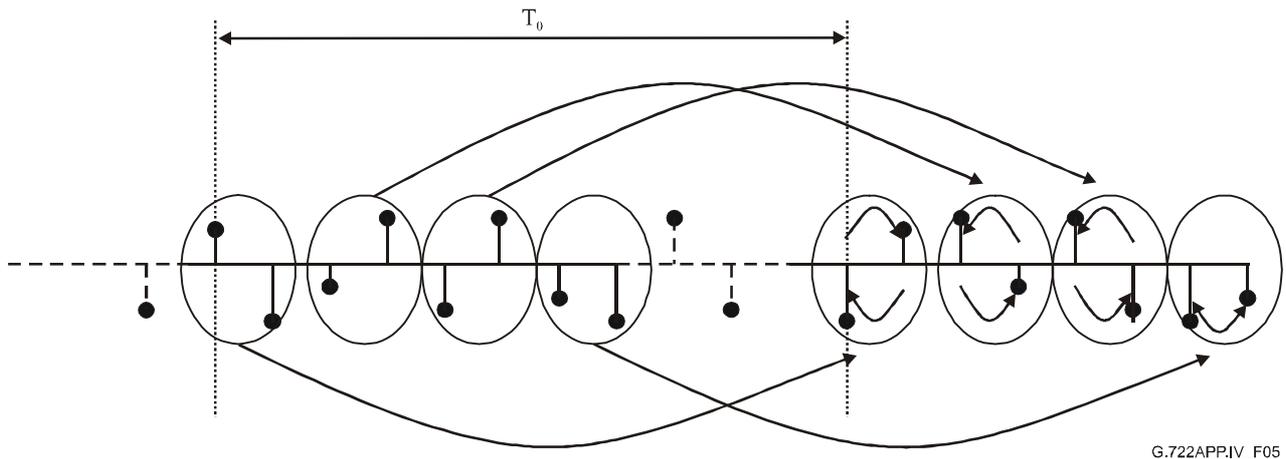


Figure IV.5 – LP residual $e(n)$ extrapolation with jitter (if *class* is not VOICED)

In addition, 80 extra samples (10 ms), $e(n)$, $n = L, \dots, L+79$, are generated using the above equations for the purpose of cross-fading.

IV.6.1.2.6 LP synthesis

After LP synthesis, the reconstructed missing frame is given by:

$$yl_{pre}(n) = e(n) - \sum_{i=1}^8 a_i yl(n-i) \quad (\text{IV-16})$$

where $e(n)$, $n = 0, \dots, L-1$, is the extrapolated residual signal and L is the frame length.

In addition, 80 samples (10 ms), $yl_{pre}(n)$, $n = L, \dots, L+79$, are generated using the above equation; these samples are used for cross-fading.

IV.6.1.2.7 Adaptive muting

The energy of the reconstructed signal is controlled by applying to each sample a gain factor that is computed and adapted sample by sample. Thus, the synthesized signal $yl_{pre}(n)$ is muted sample by sample with an adaptive muting factor g_mute_lb for $n = 0, \dots, L-1$ to obtain the reconstructed lower-band signal $yl(n)$:

$$yl(n) = g_mute_lb \times yl_{pre}(n) \quad (\text{IV-17})$$

The extra synthesis needed for cross-fading, $yl_{pre}(n)$, $n = L, \dots, L+79$, is muted in the same way: $yl(n) = g_mute_lb \times yl_{pre}(n)$.

The calculation of g_mute_lb is performed using several parameters, inc_mute , $fac1$, $fac2p$, and $fac3p$, for 10- and 20-ms frames, as well as $cf10$ for 10-ms frames. These parameters depend on the value of $class$ as indicated in Table IV.3.

Table IV.3 – Adaptive muting parameters

Parameter	$class = \text{TRANSIENT}$	$class = \text{UV_TRANSITION}$	Other cases
inc_mute	4	2	1
$fac1$	409	10	10
$fac2p$	409	10	20
$fac3p$	409	399	190
$cf10$	0	399	20

Computation of muting factor for the first erased 10-ms frame:

In this case the muting factor is adapted sample by sample with $fac1$ as follows, for $n = 0, \dots, 79$:

$$g_mute_lb = g_mute_lb - fac1$$

where g_mute_lb has been initialized to 1 (see clause IV.6.1.1).

Then cnt_mute_lb is updated at the end of the first erased 10-ms frame:

$$cnt_mute_lb = cnt_mute_lb + 80 \times inc_mute$$

where cnt_mute_lb has been initialized to 0 (see clause IV.6.1.1).

The muting factor for the extra synthesis is also adapted sample by sample with $cf10$ in the same way for $n = 80, \dots, 159$:

$$g_mute_lb = g_mute_lb - cf10$$

Then cnt_mute_lb is also updated at the end of the extra frame:

$$cnt_mute_lb = cnt_mute_lb + 80 \times inc_mute$$

The muting factor adaptation is illustrated in Figure IV.6.

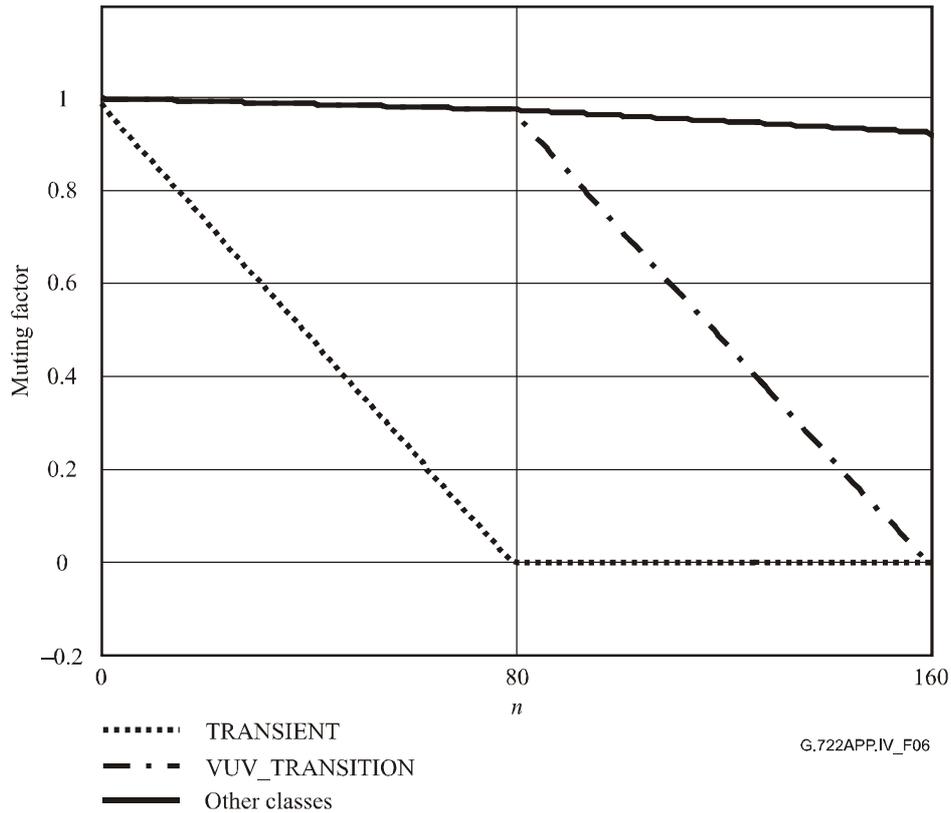


Figure IV.6 – Muting factor as a function of the sample index for 10-ms frames (80 samples for current frame + 80 samples for cross-fading part)

Computation of muting factor for other cases (20-ms frames or consecutive erased 10-ms frames):

In these cases, the muting factor g_mute_lb is also adapted sample by sample with the adaptation factor $fac1$, for $n = 0, \dots, L-1$. However, g_mute_lb may be further decreased depending on the value of cnt_mute_lb . The amount of decrease also depends on $class$. The procedure is the following:

$$g_mute_lb = g_mute_lb - fac1$$

$$\text{if } (cnt_mute_lb \geq 80) \quad g_mute_lb = g_mute_lb - fac2p$$

$$\text{if } (cnt_mute_lb \geq 160) \quad g_mute_lb = g_mute_lb - fac3p$$

$$\text{if } (cnt_mute_lb \geq 320) \quad g_mute_lb = 0$$

$$cnt_mute_lb = cnt_mute_lb + inc_mute$$

The muting factor for the extra synthesis needed for cross-fading, $yl_{pre}(n)$, $n = L, \dots, L+79$, is muted in the same way for $n = L, \dots, L+79$.

NOTE – For the first erased 20-ms frame, g_mute_lb and cnt_mute_lb have been initialized before the first erasure (see clause IV.6.1.1).

The muting factor adaptation is illustrated in Figure IV.7. The gain values for the first 160 samples are identical to those shown in Figure IV.6.

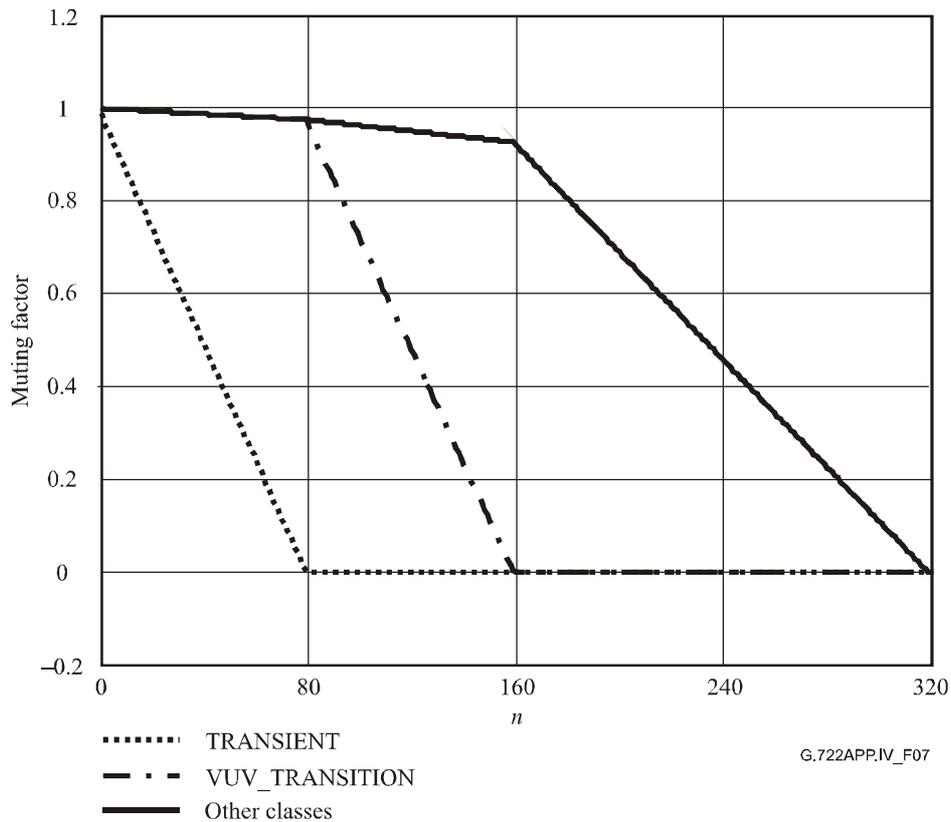


Figure IV.7 – Muting factor as a function of the sample index

IV.6.1.3 Extrapolation of missing frame: Case of bad frame following a bad frame

In the case of a bad frame following a bad frame, the analysis parameters computed for the first erased frame ($a_i, i = 1, \dots, 8, T_0, class$) are kept. The signal generated in the previous frame for cross-fading is copied to $yl(n), n = 0, \dots, 79$. The last L samples, $yl(n), n = 80, \dots, L + 79$ including the 10-ms part used for cross-fading with the next frame, are synthesized as described in clause IV.6.1.2.

IV.6.1.4 Update of ADPCM decoder states

The states of the lower-band ADPCM decoder are updated after extrapolating missing frames to help in recovery from frame erasures. This update is more elaborate than a simple ADPCM decoder reset. However, to minimize complexity, the ADPCM states are updated based on available or *a priori* information, without additional processing. The states are modified as follows, using G.722 notation:

$$DLTi = 0, i = 1, \dots, 6$$

$$PLTi = \frac{yl(L-i)}{2}, i = 1, 2$$

$$RLTi = yl(L-i), i = 1, 2$$

$$SL = yl(L)$$

$$SZL = \frac{yl(L)}{2}$$

if $cnt_mute_hb > 160$,

DETL = 32

NBL = 0

IV.6.1.5 Cross-fading

The cross-fading is detailed in Table IV.4. The cross-fading window is based on the Bartlett window (triangular) with a time support of 10 ms.

Table IV.4 – Cross-fading operation

		Current frame	
		Bad	Good
Previous frame	Bad	$zl(n) = yl(n), n = 0, \dots, L-1$	$zl(n) = \frac{n}{79}xl(n) + (1 - \frac{n}{79})yl(n),$ $n = 0, \dots, 79$ and $zl(n) = xl(n), n = 80, \dots, L-1$
	Good	$zl(n) = yl(n), n = 0, \dots, L-1$	$zl(n) = xl(n), n = 0, \dots, L-1$

IV.6.2 Higher-band decoding

IV.6.2.1 ADPCM decoder in case of a good frame

Same as clauses 4.1, 4.2 and 4.3 of [ITU-T G.722].

In addition, the counter cnt_mute_hb and muting factor g_mute_hb (used in clause IV.6.2.2.2) for adaptive muting are reset:

$$cnt_mute_hb = 0$$

$$g_mute_hb = 1$$

and the signal $zh(n)$ is stored to be used in case of erasure in future frames.

IV.6.2.2 Extrapolation of missing frame

The extrapolation of a missing frame in the higher band uses the past signal $zh(n)$, $n = -160, \dots, -1$, buffered using a buffer length of 160 samples.

IV.6.2.2.1 Repetition

The extrapolation of a missing frame in the higher band consists in pitch synchronous repeating of the previous signal $zh(n)$ if $class = VOICED$; otherwise, the repetition period is set to 80 samples (10 ms). In other words:

$$yh_{pre}(n) = zh(n - T_h), n = 0, \dots, L-1 \quad (IV-18)$$

where $T_h = T_0$ if $class = VOICED$, $T_h = 80$ otherwise.

IV.6.2.2.2 Adaptive muting

As for the lower-band reconstructed signal, the energy of the higher-band reconstructed signal is also controlled by applying a gain factor computed and adapted sample by sample. The synthesized signal $yh_{pre}(n)$ is muted sample by sample with an adaptive muting factor g_mute_hb for $n = 0, \dots, L-1$ to obtain the reconstructed higher-band signal $yh(n)$. The muting of the higher band

is identical to the muting of the lower band described in clause IV.6.1.2.7. However, since no cross-fading is used in the higher band, a separate counter cnt_mute_hb and muting factor g_mute_hb are used as cnt_mute_hb is always 80 samples in advance compared to cnt_mute_lb after one erasure.

IV.6.2.3 High-pass post-processing

In case of frame erasures, a DC offset of very small magnitude may appear in the higher-band reconstruction $uh(n)$, $n = 0, \dots, L-1$ and also affect the first consecutive good frames. After the QMF synthesis, this introduces an 8-kHz component. To avoid this annoying high-frequency noise, a first-order pole/zero filter with a cut-off frequency of 50 Hz is used. This filter is given by:

$$H_{post}(z) = \frac{\frac{7303}{8192}(1 - z^{-1})}{1 - \frac{3207}{4096}z^{-1}} \quad (\text{IV-19})$$

The signal $uh(n)$ is filtered through $H_{post}(z)$ to obtain $vh(n)$, where $n = 0, \dots, L-1$.

This filter is used during the erased frames and the first 4 s following the erasure.

IV.6.2.4 Update of ADPCM decoder states

Similar to lower-band decoding, the states of the higher-band ADPCM decoder are updated after extrapolating a missing frame. The update is described below using G.722 notation:

$$\begin{aligned} \text{NBH} &= \text{NBH}/2 \\ \text{DETH} &= \text{scaleh}(\text{NBH}) \\ \text{if } cnt_mute_hb > 160 \\ &\quad \text{NBH} = 0 \\ &\quad \text{DETH} = 8 \end{aligned}$$

The update is restricted to the higher-band scale factor.

IV.6.3 QMF synthesis filterbank

Same as clause 4.4 of [ITU-T G.722], except that the shift of 22 samples of the delay line after each call of the qmf_rx function (for each two output samples) is replaced by the use of a linear buffer of length $L+22$ in case of bad frame decoding. The memory part of this buffer is updated at the beginning of each frame, and the filter memory is saved at the end of each frame.

IV.7 Bit-exact description of the G.722 PLC algorithm

The ANSI-C code simulating the candidate G.722 PLC algorithm in 16-bit fixed-point is submitted separately to TSB. The following subclauses summarize the use of this simulation code, and how the software is organized. The G722PLC algorithm is implemented using the G.722 code from the ITU-T software tool library (STL2005). The mathematical descriptions of the PLC algorithm (clauses IV.5 and IV.6) can be implemented in several other fashions. Therefore, the algorithm description of the ANSI-C code of this clause shall take precedence over the mathematical descriptions of clauses IV.5 and IV.6 whenever discrepancies are found.

IV.7.1 Use of the simulation software

The command line for the G.722 decoder with PLC is as follows:

```
decg722 [-fsize N] g192_bst output
```

where N is the frame size at 16 kHz (default: 160).

The output file is a sampled data file containing 16-bit PCM signals.

The mapping table of the encoded bitstream is contained in the simulation software.

Organization of the simulation software

See Tables IV.5 and IV.6.

Table IV.5 – List of tables added by the PLC algorithm

Table name	Size	Description
G722PLC_lpc_win_80	80	LPC window
G722PLC_lag_h	8	Lag window for bandwidth expansion (high part)
G722PLC_lag_l	8	Lag window for bandwidth expansion (low part)
G722PLC_fir_lp	9	Coefficients of low-pass quarter-band decimation filter
G722PLC_b_hp	3	High-pass filter coefficients (numerator)
G722PLC_a_hp	3	High-pass filter coefficients (denominator)

Table IV.6 – List of files (G.722 decoder with PLC)

a) Identical files – from G.191 STL software

Filename	Description
softbit.c	G.722 soft bit handling
g722_com.h	G.722 additional header file
softbit.h	G.722 soft bit handling
ugstdemo.h	Definitions for UGST demo programs

b) Modified files – from G.191 STL software

Filename	Description
decg722.c	G.722 decoder interface
g722.c	G.722 main decoder routines
g722.h	G.722 main header file
funcg722.c	G.722 library
funcg722.h	G.722 library

c) New files

Filename	Description
g722_plc.c, g722_plc.h	PLC library and headers
g722_plc_tables.c	PLC tables
oper32_b.c, oper32_b.h	Additional basic operators and headers

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
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	Telecommunication management, including TMN and network maintenance
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, open system communications and security
Series Y	Global information infrastructure, Internet protocol aspects and next-generation networks
Series Z	Languages and general software aspects for telecommunication systems