



INTERNATIONAL TELECOMMUNICATION UNION

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

G.711

Appendix I
(09/99)

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

Digital transmission systems – Terminal equipments –
Coding of analogue signals by pulse code modulation

Pulse code modulation (PCM) of voice frequencies

**Appendix I: A high quality low-complexity
algorithm for packet loss concealment with
G.711**

ITU-T Recommendation G.711 – Appendix I

(Previously CCITT Recommendation)

ITU-T G-SERIES RECOMMENDATIONS
TRANSMISSION SYSTEMS AND MEDIA, DIGITAL SYSTEMS AND NETWORKS

INTERNATIONAL TELEPHONE CONNECTIONS AND CIRCUITS	G.100–G.199
INTERNATIONAL ANALOGUE CARRIER SYSTEM	
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	
TRANSMISSION MEDIA CHARACTERISTICS	
DIGITAL TRANSMISSION SYSTEMS	
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

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

PULSE CODE MODULATION (PCM) OF VOICE FREQUENCIES

APPENDIX I

A high quality low-complexity algorithm for packet loss concealment with G.711**Summary**

Packet Loss Concealment (PLC) algorithms, also known as frame erasure concealment algorithms, hide transmission losses in an audio system where the input signal is encoded and packetized at a transmitter, sent over a network, and received at a receiver that decodes the packet and plays out the output. Many of the standard CELP-based speech coders have PLC algorithms built into their standards. The algorithm described here provides a method for Recommendation G.711.

Source

Appendix I to ITU-T Recommendation G.711 was prepared by ITU-T Study Group 16 (1997-2000) and was approved under the WTSC Resolution No. 1 procedure on 30 September 1999.

FOREWORD

ITU (International Telecommunication Union) is the United Nations Specialized Agency in the field of telecommunications. The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of the ITU. The 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 Conference (WTSC), which meets every four years, establishes the topics for study by the ITU-T Study Groups which, in their turn, produce Recommendations on these topics.

The approval of Recommendations by the Members of the ITU-T is covered by the procedure laid down in WTSC Resolution No. 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 term *recognized operating agency (ROA)* includes any individual, company, corporation or governmental organization that operates a public correspondence service. The terms *Administration*, *ROA* and *public correspondence* are defined in the *Constitution of the ITU (Geneva, 1992)*.

INTELLECTUAL PROPERTY RIGHTS

The ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. The 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, the 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.

© ITU 2000

All rights reserved. No part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from the ITU.

CONTENTS

	Page
Appendix I – A high quality low-complexity algorithm for packet loss concealment with G.711	1
I.1 Introduction	1
I.2 Algorithm description.....	1
I.2.1 Good frames	1
I.2.2 First bad frame	1
I.2.3 Pitch detection	2
I.2.4 Synthetic signal generation for first 10 ms.....	2
I.2.5 Synthetic signal generation after 10 ms	2
I.2.6 Attenuation	3
I.2.7 First good frame after an erasure.....	3
I.2.8 Example.....	3
I.3 Algorithm description with annotated C++ code	5
I.3.1 Typedefs and constants	5
I.3.2 Class declaration	5
I.3.3 Main loop	7
I.3.4 Utility member functions	7
I.3.5 Constructor	8
I.3.6 Addtohistory and savespeech	9
I.3.7 Dofe.....	10
I.3.8 Pitch detection.....	12
I.3.9 Synthetic signal generation and attenuation.....	15
I.3.10 Overlap add operators	16
I.4 Complexity and delay.....	17



Recommendation G.711**PULSE CODE MODULATION (PCM) OF VOICE FREQUENCIES****APPENDIX I****A high quality low-complexity algorithm for packet loss concealment with G.711***(Geneva, 1999)***I.1 Introduction**

Packet Loss Concealment (PLC) algorithms, also known as frame erasure concealment algorithms, hide transmission losses in an audio system where the input signal is encoded and packetized at a transmitter, sent over a network, and received at a receiver that decodes the packet and plays out the output. Many of the standard CELP-based speech coders, such as Recommendations G.723.1 [1], G.728 [2] and G.729 [3], have PLC algorithms built into their standards. The algorithm described here provides a method for Recommendation G.711.

The objective of PLC is to generate a synthetic speech signal to cover missing data (erasures) in a received bit stream. Ideally, the synthesized signal will have the same timbre and spectral characteristics as the missing signal, and will not create unnatural artifacts. Since speech signals are often locally stationary, it is possible to use the signals' past history to generate a reasonable approximation to the missing segment. If the erasures are not too long, and the erasure does not land in a region where the signal is rapidly changing, the erasures may be inaudible after concealment.

I.2 Algorithm description

To add PLC to a G.711 system that currently does not conceal losses, changes are only required in the receiver. The G.711-encoded audio data is sampled at 8 kHz. In this appendix it is assumed to be partitioned into 10 ms frames (80 samples). By adjusting a few parameters, other packet sizes or sampling rates can be accommodated.

I.2.1 Good frames

During normal operation (good packets or frames) the receiver decodes the received packet and sends its output to the audio port. Two minor changes are made to the receiver when it processes good frames to support PLC.

- 1) A copy of the decoded output is saved in a circular history buffer that is 48.75 ms (390 samples) long. The history buffer is used to calculate the current pitch period and extract waveforms during an erasure. This buffering does not introduce any delay into the output signal.
- 2) The output is delayed by 3.75 ms (30 samples) before it sent to the audio port. This algorithm delay, used for an Overlap Add (OLA) at the start of an erasure, allows the PLC code to make a smooth transition between the real and synthesized signal.

I.2.2 First bad frame

At the start of the erasure, the circular history buffer is copied to a non-circular buffer, called the pitch buffer, that is easier to work with. The contents of the pitch buffer are used for the duration of the erasure. An additional copy of the most recent 1/4 pitch period, called the lastq buffer, is made in case the erasure lasts longer than 10 ms.

1.2.3 Pitch detection

First, the pitch period is estimated by finding the peak of the normalized cross-correlation of the most recent 20 ms of speech in the history buffer with the previous speech at taps from 5 (40 samples) to 15 ms (120 samples). This corresponds to frequencies of 200 to 66 Hz. The pitch range was chosen based on a range used in G.728's post-filter. While G.728 uses a lower bound of 2.5 ms (20 samples), here it is increased to 40 samples so the same pitch period is not repeated more than twice in a single 10 ms erased frame. To lower complexity, the pitch estimation is calculated in two phases. First, a coarse search is performed on a 2:1 decimated signal, and then a finer search is performed in the vicinity of the peak of the coarse search. The complexity can be lowered with a slight degradation in quality by skipping the fine search. In the following the term *wavelength* is also used to refer to the output value of this calculation, since the missing signal may be either voiced or unvoiced speech.

From Waveform Shift Overlap Add (WSOLA), it is known that the normalized cross-correlation function can be replaced with either a non-normalized cross correlation, or a cross-Average Magnitude Difference Function (AMDF) and similar overall performance results will be obtained.

1.2.4 Synthetic signal generation for first 10 ms

For the first 10 ms of the erasure, the best results are obtained by generating the synthesized signal from the last pitch period with no attenuation. Only the most recent 1.25 pitch periods of the pitch buffer are used during the first 10 ms. To insure a smooth transition between the real and synthetic signal, and a smooth transition if the pitch period is repeated multiple times, an Overlap Add (OLA) is performed using a triangular window on 1/4 of the pitch period between the last and next to last pitch period. For 1/4 wavelength the signal starting at 1.25 pitch periods from the end of the pitch buffer is multiplied by an up-sloping ramp and is added to the last 0.25 pitch period in the last q buffer multiplied by a down-sloping ramp. If complexity is not an issue, the triangular windows may be replaced with Hanning windows in all the OLA operations.

The result of the OLA replaces both the tail of the pitch buffer and the tail of the history buffer. It is also output by the receiver during the tail of the last good frame, replacing the original signal. This introduces the algorithm delay – the tail of the last frame cannot be output until it is known whether the next frame is erased. If an erasure occurs the signal in the tail of the last good frame is modified by the OLA to insure a smooth transition to the synthesized signal.

The synthesized signal for the 10 ms during the erasure is generated by placing a pointer one pitch period back from the end of the pitch buffer, and copying the samples to the output. If the pitch period is shorter than 10 ms, when the pointer rolls off the end of the pitch buffer the pointer is set back exactly one pitch period before continuing. If the pitch period is short (the frequency is high), the last pitch period in the pitch buffer is repeated multiple times during the 10 ms erasure.

While the erasure progresses, the history buffer is updated with the synthesized output. This way, the history buffer always has a smooth, continuous signal in it. This continuity is important if a "bad frame, good frame, bad frame" sequence occurs.

1.2.5 Synthetic signal generation after 10 ms

If the next frame is also erased, the erasure will be at least 20 ms long and further action is required. While repeating a single pitch period works well for short erasures (e.g. 10 ms), on long erasures it introduces unnatural harmonic artifacts (beeps). This is especially noticeable if the erasure lands in an unvoiced region of speech, or in a region of rapid transition such as a stop. It was discovered by experimentation that these artifacts are significantly reduced by increasing the number of pitch periods used to synthesize the signal as the erasure progresses. Playing more pitch periods increases the variation in the signal. Although the pitch periods are not played in the order they occurred in the original signal, the resulting output still sounds natural. At 10 ms into the erasure the number of pitch

periods used to synthesize the speech is increased to two, and at 20 ms a third pitch period is added. For erasures longer than 20 ms no additional modifications to the pitch buffer are made.

When the number of pitch periods used in the pitch buffer increases, it is important that the transition in the synthesized signal be smooth. This is accomplished by continuing the output of the existing pitch buffer for 1/4 of a pitch period at the start of the second and third erased frame, updating the pitch buffer, keeping the buffer pointer synchronized with the correct phase, and then doing an OLA with the output from the new pitch buffer.

The pitch buffer is updated exactly as during the first erased frame, except that the number of pitch periods is increased. For example, at the start of the second erased frame, for 1/4 wavelength the signal starting at 2.25 pitch periods from the end of the pitch buffer is multiplied by an up-sloping ramp and is added to the 1/4 wavelength in the last buffer multiplied by a down-sloping ramp. The result of the OLA replaces the last 1/4 wavelength in the pitch buffer. To maintain the phase of the current output pointer, pitch periods are subtracted from the pointer until it is in the first pitch period used.

I.2.6 Attenuation

As with other PLC algorithms, such as G.729 and G.728 Annex I, with long erasures it is necessary to attenuate the signal as the erasure progresses. As the erasure gets longer, the synthesized signal is more likely to diverge from the real signal. Without attenuation strange artifacts are created by holding certain types of sounds too long, even if the synthesized signal segment sounds natural in isolation. For the first 10 ms of an erasure the signal is not attenuated. At the start of the second 10 ms, the synthesized signal is linearly attenuated with a ramp at the rate of 20% per 10 ms. After 60 ms, the synthesized signal is zero.

I.2.7 First good frame after an erasure

At the first good frame after an erasure, a smooth transition is needed between the synthesized erasure speech and the real signal. To do this, the synthesized speech from the pitch buffer is continued beyond the end of the erasure, and then mixed with the real signal using an OLA. The length of the OLA depends on both the pitch period and the length of the erasure. For short, 10 ms erasures, a 1/4 wavelength window is used. For longer erasures the window is increased by 4 ms per 10 ms of erasure, up to a maximum of the frame size, 10 ms.

I.2.8 Example

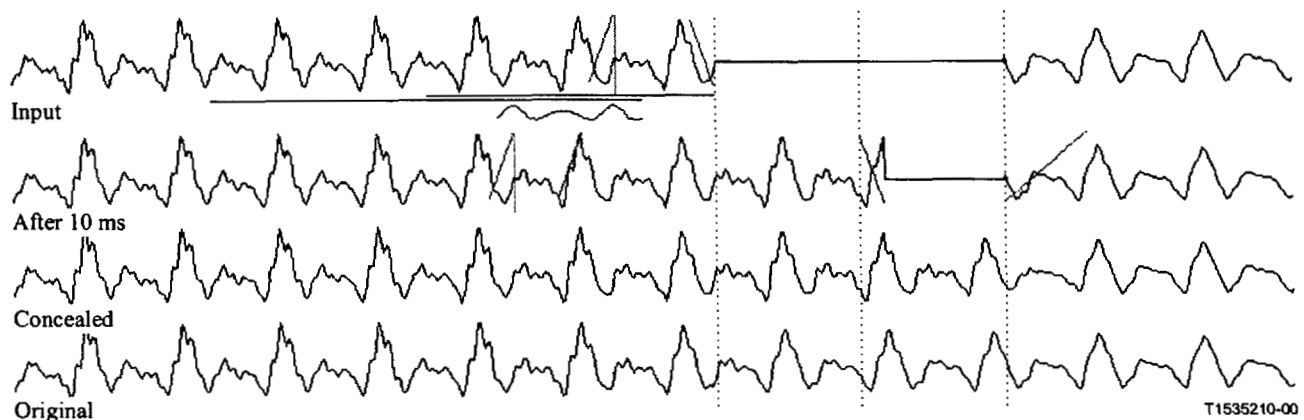


Figure I.1/G.711 – Frame erasure concealment algorithm for G.711