

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

G.722.1

(05/2005)

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

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

Low-complexity coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss

ITU-T Recommendation G.722.1



International
Telecommunication
Union

Document Sharing Hub

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 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
ETHERNET 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.722.1

Low complexity coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss

Summary

This Recommendation describes a low complexity encoder and decoder that may be used for 7-kHz bandwidth audio signals working at 24 kbit/s or 32 kbit/s. Further, this algorithm is recommended for use in hands-free applications such as conferencing where there is a low probability of frame loss. It may be used with speech or music inputs. The bit rate may be changed at any 20-ms frame boundary. New Annex C contains the description of a low-complexity extension mode to G.722.1, which doubles the algorithm to permit 14-kHz audio bandwidth using a 32-kHz audio sample rate, at 24, 32, and 48 kbit/s. This mode is suitable for use in video conferencing, teleconferencing, and Internet streaming applications, and uses the same 20-ms frame length, 40-ms algorithmic delay, and same algorithmic steps as the 7-kHz mode. Less than 5.5 WMOPS are required for encoding and decoding in the baseline 7-kHz mode, and less than 11 WMOPS are required for encoding and decoding in the 14-kHz mode of Annex C.

This Recommendation includes a software package which contains the encoder and decoder source code and a set of test vectors for developers. These vectors are a tool providing an indication of success in implementing this code. The fixed-point code implements both the 7-kHz mode (main body) and the 14-kHz mode (Annex C). The floating point implements only the 7-kHz mode.

Source

ITU-T Recommendation G.722.1 was approved on 14 May 2005 by ITU-T Study Group 16 (2005-2008) under the ITU-T Recommendation A.8 procedure.

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, implementors are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database.

© ITU 2005

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

CONTENTS

	Page
1 Scope	1
2 Normative references.....	2
3 The encoder	2
3.1 The Modulated Lapped Transform (MLT).....	4
3.2 Computing and quantizing the amplitude envelope	5
3.3 Coding the amplitude envelope	5
3.4 Categorization procedure.....	6
3.5 Scalar Quantized Vector Huffman Coding (SQVH)	8
3.6 Rate control	11
3.7 Transmission of the MLT vector indices.....	11
3.8 Bit stream.....	11
4 The decoder	11
4.1 Decoding the amplitude envelope	12
4.2 Determining categorization	12
4.3 Decoding MLT coefficients	12
4.4 Noise-fill.....	13
4.5 Insufficient bits.....	13
4.6 Frame erasure	13
4.7 The Inverse MLT (IMLT)	13
5 C code	14
6 Flow chart of categorization procedure	15
Annex A – Packet format, capability identifiers and capability parameters.....	20
A.1 References	20
A.2 Packet structure for G.722.1 frames	20
A.3 Capability Identifiers and Parameters for use with ITU-T Rec. H.245.....	21
Annex B – Floating-point implementation for G.722.1	23
B.1 Introduction	23
B.2 Algorithmic description.....	23
B.3 ANSI C code.....	24
Annex C – 14 kHz mode at 24, 32, and 48 kbit/s	25
C.1 Introduction	25
C.2 Algorithmic description.....	25
C.3 ANSI C code.....	27

Copyright International Telecommunication Union/ITU Telecommunication Sector
Provided by IHS Markit under license with ITU-T
No reproduction or networking permitted without license from IHS

ITU-T Recommendation G.722.1

Low-complexity coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss¹

1 Scope

This Recommendation describes a digital wideband coder algorithm that provides an audio bandwidth of 50 Hz to 7 kHz, operating at a bit rate of 24 kbit/s or 32 kbit/s. The digital input to the coder may be 14-, 15- or 16-bit 2's complement format at a sample rate of 16 kHz (handled in the same way as in ITU-T Rec. G.722). The analogue and digital interface circuitry at the encoder input and decoder output should conform to the same specifications described in ITU-T Rec. G.722.

The algorithm is based on transform technology, using a Modulated Lapped Transform (MLT). It operates on 20-ms frames (320 samples) of audio. Because the transform window (basis function length) is 640 samples and a 50 per cent (320 samples) overlap is used between frames, the effective look-ahead buffer size is 20 ms. Hence the total algorithmic delay of 40 ms is the sum of the frame size plus look-ahead. All other delays are due to computational and network transmission delays.

The description of the coding algorithm of this Recommendation is made in terms of bit-exact, fixed-point mathematical operations. The C code indicated in clause 5, which constitutes an integral part of this Recommendation, reflects this bit-exact, fixed-point descriptive approach, and shall take precedence over the mathematical descriptions of clauses 3 and 4 whenever discrepancies are found.

The mathematical descriptions of the encoder (clause 3), and decoder (clause 4), could have been implemented in several other fashions, but the C code of clause 5 has been provided as reference purposes. Thus, to comply with this Recommendation, any implementation must produce for any input signal the same output results as the C code of clause 5.

Note that to ensure that this goal is achieved, implementations should follow the computational details, tables of constants, sequencing of variable adaptation and use given by the C code of clause 5. However, it is recognized that there are many parts of the algorithm critical to maintaining correct bit-exact operation. For these parts, implementations shall reproduce the computational details, tables of constants, sequencing of variables adaptation and use written in the C code of clause 5.

It is recognized that the C code provided is for reference, and has not been optimized (in terms of memory, complexity, etc.) for a specific implementation platform. The C code may require optimization for a particular implementation.

A non-exhaustive set of test signals is provided as part of this Recommendation, as a tool to assist implementors to verify their implementations of the encoder and decoder comply with this Recommendation.

In practice, purchasers of wideband equipment or software implementations will expect them to be compliant with this standard to ensure interworking capability. Implementors may choose to optimize the C code, or otherwise modify the reference C code. In such cases the implementor shall verify that his implementation produces the same resultant output for any given input as would be expected using the C code expressed in clause 5.

¹ This Recommendation includes a software package that contains the encoder and decoder source code and a set of test vectors for developers.

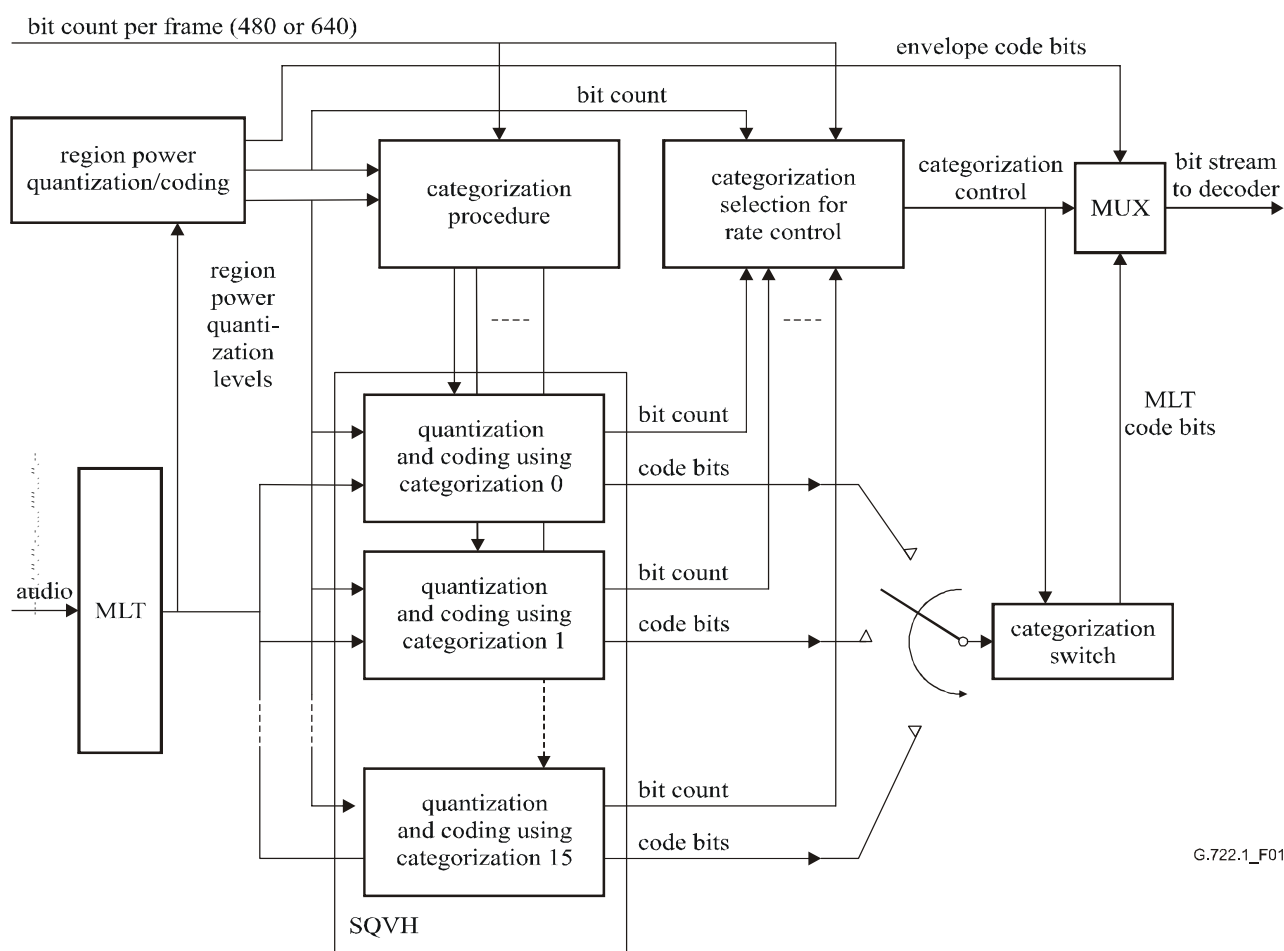
2 Normative references

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

- [1] CCITT Recommendation G.722 (1988), *7-kHz audio-coding within 64 kbit/s*.
- [2] ITU-T Recommendation G.192 (1996), *A common digital parallel interface for speech standardization activities*.
- [3] ISO/IEC 9899:1999, *Programming languages – C*.

3 The encoder

Figure 1 presents a block diagram of the encoder.



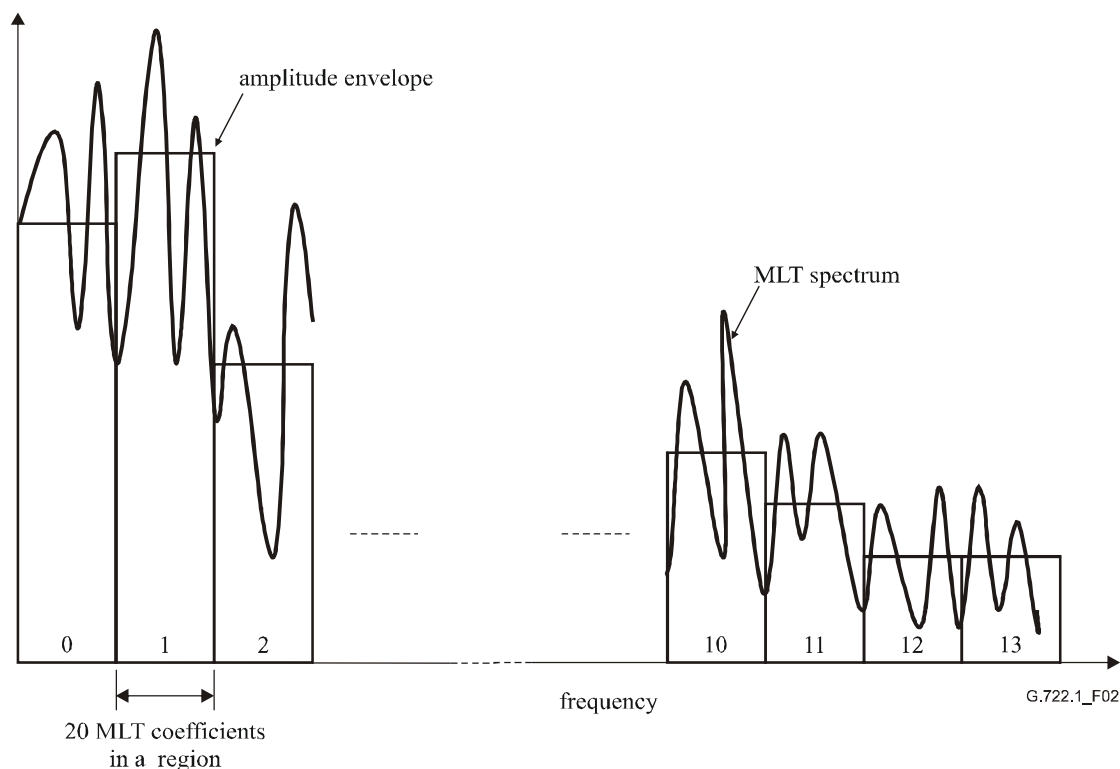
G.722.1_F01

Figure 1/G.722.1 – Block diagram of the encoder

Every 20 milliseconds (320 samples) the most recent 640 time domain audio samples are fed to a Modulated Lapped Transform (MLT). Each transform produces a frame of 320 MLT coefficients, and each frame of MLT coefficients is coded independently, i.e., there is no state information left

over from the previous frame. For 24-kbit/s and 32-kbit/s operation the allotment of bits per frame is 480 and 640, respectively.

The transform coefficients generated by the MLT transform are first applied to a module that computes the *amplitude envelope* and quantizes it; see Figure 2. The amplitude envelope is a coarse representation of the MLT spectrum. The spectrum is divided into blocks of 20 MLT coefficients called *regions*. Each region represents a bandwidth of 500 Hz. As the bandwidth is 7 kHz, the *number_of_regions* is set at fourteen. MLT coefficients representing frequencies above 7 kHz are ignored. The code bits representing the amplitude envelope are sent to the MUX (Multiplexer) for transmission to the decoder. The bits remaining after quantization and coding of the amplitude envelope are used to encode the MLT coefficients in the *categorization* process.



NOTE – Each value of the amplitude envelope represents the RMS (root-mean-square) value of the MLT coefficients in that region.

Figure 2/G.722.1 – An illustration of how the spectrum is divided into fourteen regions, each containing 20 MLT coefficients

Using the quantized amplitude envelope and the number of bits remaining in the frame after amplitude envelope encoding (and provision for four categorization control bits), the categorization procedure generates sixteen sets of *categorizations* (*categorization 0* to *categorization 15*). Different categorizations require different numbers of bits to encode the same MLT coefficients.

Each categorization consists of a set of fourteen *category assignments*, one assignment for each of the fourteen regions. A *category* defines a set of predetermined quantization and coding parameters for a region. Associated with each category is an expected number of bits required to encode a region. Because this coder uses variable length Huffman coding, the final number of bits used will vary depending on the particular sequence of MLT coefficients in the region.

Next, the MLT coefficients are quantized and coded differently for each one of the sixteen computed categorizations. For each categorization the actual number of code bits required is determined.

The quantization and encoding proceed region by region. A categorization determines the category assignment for all the fourteen regions, and the category assignment together with the amplitude envelope for each region determine all of the quantization and coding parameters that will be used for all twenty MLT coefficients in the region.

The MLT coefficients in a region are first normalized by the quantized amplitude envelope in the region and then scalar quantized. The resulting scalar quantization indices are combined into vector indices. The vector indices are then Huffman coded, i.e., they are coded with a variable number of bits. The most frequent vector indices require fewer bits than the less frequent vector indices.

Because this codec uses variable length Huffman coding and a constant transmitted bit rate is required, a method of constraining the bit rate to the channel rate is required. Four *categorization control* bits identify to the decoder which categorization was selected. The categorization switch directs the code bits (representing the quantized MLT coefficients produced using the selected categorization) to the MUX for transmission. The categorization that results in providing the number of bits closest to the channel rate is selected for transmission.

3.1 The Modulated Lapped Transform (MLT)

The MLT is a critically sampled, perfect reconstruction, linear transform with a 50-percent overlap between the basis functions of adjacent MLT frames. The inputs to each MLT are the most recent 640 audio samples, $x(n)$,

where:

$$x(0) \text{ is the oldest sample,}$$

and:

$$0 \leq n < 640$$

The MLT outputs 320 transform coefficients, $mlt(m)$,

where:

$$0 \leq m < 320$$

The MLT is given by:

$$mlt(m) = \sum_{n=0}^{639} \sqrt{\frac{2}{320}} \sin\left(\frac{\pi}{640}(n+0.5)\right) \cos\left(\frac{\pi}{320}(n-159.5)(m+0.5)\right) x(n)$$

The MLT can be decomposed into a window, overlap and add operation followed by a type IV Discrete Cosine Transform (DCT). The window, overlap and add operation is given by:

$$v(n) = w(159-n)x(159-n) + w(160+n)x(160+n) \quad \text{for } 0 \leq n \leq 159$$

$$v(n+160) = w(319-n)x(320+n) - w(n)x(639-n) \quad \text{for } 0 \leq n \leq 159$$

where:

$$w(n) = \sin\left(\frac{\pi}{640}(n+0.5)\right) \quad \text{for } 0 \leq n < 320$$

Combining $v(n)$ with a type IV DCT, the resulting expression for the MLT is:

$$mlt(m) = \sum_{n=0}^{319} \sqrt{\frac{2}{320}} \cos\left(\frac{\pi}{320}(n+0.5)(m+0.5)\right) v(n)$$

Note that fast transform techniques are used to significantly reduce the complexity of the DCT.

3.2 Computing and quantizing the amplitude envelope

The MLT coefficients are divided into regions of twenty coefficients. Thus, the total *number_of_regions* = 14. Region *r* includes MLT coefficients $20r$ through $20r + 19$, where:

$$0 \leq r < \text{number_of_regions}$$

The forty highest frequency MLT coefficients, representing frequencies above 7 kHz, are not used because they are outside the bandwidth of interest.

The *amplitude envelope* in the region *r* is defined as the RMS (Root-Mean-Square) value of the MLT coefficients in the region, and is computed as

$$\text{rms}(r) = \sqrt{\frac{1}{20} \sum_{n=0}^{19} \text{mlt}(20r + n) \text{mlt}(20r + n)}$$

It is then quantized. The quantizer output index is *rms_index(r)*. The allowed set of quantization reconstruction values are:

$$2^{\left(\frac{i+2}{2}\right)} \quad \text{for integer values of } i, -8 \leq i \leq 31$$

and *rms_index(0)* is further constrained so that

$$1 \leq \text{rms_index}(0) \leq 31$$

A log domain metric is used so that the values which get quantized to $2^{\left(\frac{i+2}{2}\right)}$ range from $2^{\left(\frac{i-0.5+2}{2}\right)}$ to $2^{\left(\frac{i+0.5+2}{2}\right)}$.

For example, if *rms(r)* = 310, then the corresponding quantization level is $2^{\left(\frac{15+2}{2}\right)}$ or 362.04, and *rms_index(r)* = 15, because $2^{\left(\frac{15-0.5+2}{2}\right)} = 304.43$.

3.3 Coding the amplitude envelope

rms_index(0) is the first value transmitted in each frame. Five bits are used. The most significant bit of the index is transmitted first. The value, *rms_index(0)* = 0, is reserved and not used.

The indices of the remaining thirteen regions are differentially coded and then Huffman coded for transmission. The largest differences which may be coded are +11 and -12. To contain the differences within this range, the valleys are first adjusted upwards to allow the peaks which follow them to be accurately represented. This is described in the following in pseudo C code:

```
for (r = number_of_regions - 2; r >= 0; r--)
{
    if (rms_index[r] < rms_index[r + 1] - 11)
        rms_index[r] = rms_index[r + 1] - 11;
}
for (r = 1; r < number_of_regions; r++)
{
    j = rms_index[r] - rms_index[r - 1];
    if (j < -12)
```