

International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

G.719

(06/2008)

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

Digital terminal equipments – Coding of analogue signals

Low-complexity, full-band audio coding for high-quality, conversational applications

Recommendation ITU-T G.719



International
Telecommunication
Union

Documents
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 AND OPTICAL SYSTEMS CHARACTERISTICS	G.600–G.699
DIGITAL TERMINAL EQUIPMENTS	G.700–G.799
General	G.700–G.709
Coding of analogue signals	G.710–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.

Recommendation ITU-T G.719

Low-complexity, full-band audio coding for high-quality, conversational applications

Summary

Recommendation ITU-T G.719 describes the G.719 coding algorithm for low-complexity full-band conversational speech and audio, operating from 32 kbit/s up to 128 kbit/s.

The encoder input and decoder output are sampled at 48 kHz. The codec enables full bandwidth, from 20 Hz to 20 kHz, encoding of speech, music and general audio content. The codec operates on 20-ms frames and has an algorithmic delay of 40 ms.

The coding algorithm is based on transform coding with adaptive time-resolution, adaptive bit-allocation and low-complexity lattice vector quantization.

This Recommendation contains an electronic attachment containing a set of test signals, which can be used together with the ANSI-C code in order to verify bit-exactness.

Amendment 1 adds new Annex A that specifies the use of the ISO base media file format as container for the G.719 bitstream addresses non-conversational use cases of the codec (e.g., call waiting music playback and recording of teleconferencing sessions, voice mail messages, online "jam"-sessions).

Amendment 1 also includes Annex B, an alternative implementation of the G.719 using floating-point arithmetic. Examples of application of the floating point implementations include DSP platforms optimized for floating point and codec implementations using desktop computers.

Amendment 2 adds new Annex C specifying the G.719 packet format, capability identifier and capability parameters and related procedures for H.245 based systems.

Source

Recommendation ITU-T G.719 was approved on 13 June 2008 by ITU-T Study Group 16 (2005-2008) under Recommendation ITU-T A.8 procedure.

This edition includes the amendments approved on 13 November 2008 and 16 March 2009 by ITU-T Study Group 16 (2009-2012) under Recommendation ITU-T A.8 procedure.

Keywords

Adaptive time-frequency transform, fast lattice vector quantization, full-band, low-complexity.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). 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 2009

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	References.....	1
3	Definitions	2
4	Abbreviations and acronyms	2
5	Notations and conventions.....	2
6	General description of the codec	4
6.1	Encoder.....	4
6.2	Decoder.....	5
6.3	Algorithmic delay analysis	6
6.4	Complexity and memory requirements	6
6.5	Description of the codec.....	7
7	Functional description of the encoder.....	7
7.1	Transient detection	7
7.2	Adaptive time-frequency transform	9
7.3	Grouping of spectral coefficients	14
7.4	Norm estimation and quantization.....	17
7.5	Bit allocation	18
7.6	Fast lattice vector quantization.....	21
7.7	Noise level adjustment	26
7.8	Bit stream format.....	27
8	Functional description of the decoder.....	28
8.1	Flags decoding	28
8.2	Norm decoding	28
8.3	Spectral coefficient decoding	28
8.4	Spectrum filling	29
8.5	Inverse transform.....	31
8.6	Frame erasure concealment	34
9	Flexible rate selection	34
10	Simulation software	35
10.1	Use of the simulation software	35
10.2	Organization of the simulation software	35
Annex A	Storage format definitions	39
A.1	Scope	39
A.2	Raw storage format for the ITU-T G.719 bitstream.....	39
A.3	Sample entry description for ISO base file format.....	40
Annex B	Floating point implementation.....	43
B.1	Scope	43

	Page
B.2 Algorithmic description.....	43
B.3 ANSI C code.....	43
Annex C Packet format, capability identifier and capability parameters	47
C.1 Packet structure for G.719 frames	47
C.2 Capability identifiers and parameters for use with [ITU-T H.245].....	47
Bibliography.....	50

Electronic attachment: Test signals

iv

Recommendation ITU-T G.719

Low-complexity, full-band audio coding for high-quality, conversational applications¹

1 Scope

This Recommendation contains the description of a coding algorithm for full-band speech and audio signals at 32-128 kbit/s.

The text of the Recommendation is organized as follows. The references, definitions, abbreviations and notations used throughout this Recommendation are defined in clauses 2, 3, 4 and 5, respectively. Clause 6 gives a general outline of the G.719 full-band algorithm. The G.719 full-band encoder and decoder principles are discussed in clauses 7 and 8, respectively. Rates and operating points are provided in clause 9. Clause 10 describes the software that defines this coder in 16-32 bit fixed-point arithmetic. Annex A gives the storage format definitions. Annex B describes the floating point implementation and Annex C specifies the packet format, capability identifier and capability parameters.

2 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.

- [ITU-T G.191] Recommendation ITU-T G.191 (2005), *Software tools for speech and audio coding standardization*.
- [ITU-T G.192] Recommendation ITU-T G.192 (1996), *A common digital parallel interface for speech standardization activities*.
- [ITU-T G.722.1] Recommendation ITU-T G.722.1 (2005), *Low-complexity coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss*.
- [ITU-T H.245] Recommendation ITU-T H.245 (2008), *Control protocol for multimedia communication*.
- [IETF RFC 3550] IETF RFC 3550 (2003), *RTP: A Transport Protocol for Real-Time Applications*. <<http://www.ietf.org/rfc/rfc3550.txt?number=3550>>
- [IETF RFC 5404] IETF RFC 5404 (2009), *RTP Payload format for G.719*. <<http://www.ietf.org/rfc/rfc5404.txt?number=5404>>
- [ISO/IEC 14496-12] ISO/IEC 14496-12:2008, *Information technology – Coding of audio-visual objects – Part 12: ISO base media file format*. <http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=51533>
- [ISO/IEC 15444-12] ISO/IEC 15444-12:2008, *Information technology – JPEG 2000 image coding system – Part 12: ISO base media file format*. <http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=51537>

¹ This Recommendation contains an electronic attachment containing a set of test signals, which can be used together with the ANSI-C code in order to verify bit-exactness.

3 Definitions

None.

4 Abbreviations and acronyms

This Recommendation uses the abbreviations and acronyms that are summarized in Table 1.

Table 1 – Abbreviations and acronyms

Abbreviation	Description
BWE	BandWidth Extension
DCT	Discrete Cosine Transform
FB	Full Band
FFT	Fast Fourier Transform
FLVQ	Fast Lattice Vector Quantization
IIR	Infinite Impulse Response
LVQ	Lattice Vector Quantization
MDCT	Modified Discrete Cosine Transform
MLT	Modulated Lapped Transform
MSE	Mean Squared Error
PDF	Probability Density Function
RAM	Random Access Memory
ROM	Read Only Memory
SNR	Signal-to-Noise Ratio
TDA	Time Domain Aliasing
WMOPS	Weighted Million of Operations Per Second

5 Notations and conventions

Table 2 lists the most relevant symbols used throughout this Recommendation.

Table 2 – Glossary of most relevant symbols

Symbol	Description
$x(n)$	Input signal sampled at 48 kHz
$H_{HP}(z)$	High-pass filter used for the transient detection
$x_{HP}(n)$	Output of the high-pass filter $H_{HP}(z)$
m	Index of sub-frame
$E(m)$	Signal's short-term energy corresponding to the m -th sub-frame
$E_{LT}(m)$	Signal's long-term energy corresponding to the m -th sub-frame
α	Forgetting factor for computing $E_{LT}(m)$
ρ	Energy ratio threshold
$x_{OLD}(n)$	Input signal from the previous 20-ms frame
$h(n)$	Window function for transform
$x_w(n)$	Windowed signal with $h(n)$

Table 2 – Glossary of most relevant symbols

Symbol	Description
$\tilde{\mathbf{x}}$	TDA output signal matrix
$\mathbf{x}_w$	Windowed signal matrix
$\mathbf{I}_L$	Identity matrices of order L
$\mathbf{J}_L$	Time reversal matrices of order L
$\tilde{x}(n)$	TDA output signal
Q_{max}	Normalization index for signal pre-scaling and post-scaling
$\bar{\tilde{x}}(n)$	Scaled TDA signal
$y(k)$	Spectral coefficients of the input frame
L	Frame length in samples
$v(n)$	Reordered signal
$z(n)$	Complex sequence of TDA signal
$\tilde{z}(p)$	Pre-modulated complex sequence of TDA signal
$\tilde{Z}(q)$	Scaled inverse complex Fourier transform of $\tilde{z}(p)$
$\bar{y}(k)$	Spectrum corresponds to $\bar{\tilde{x}}(n)$
p	Index of sub-vector
L_p	Length of the p -th sub-vector
s_p	Starting index of the spectral coefficients forming the p -th sub-vector
e_p	Ending index of the spectral coefficients forming the p -th sub-vector
$N(p)$	Norm or spectrum energy of the p -th sub-vector
$I_N(p)$	Index of the p -th quantized norm
$\Delta I_N(p)$	Differential index between $I_N(p+1)$ and $I_N(p)$
$Spe(p)$	Spectrum function
$BSpe(p)$	Remapped spectrum function
B_{MAX}	Maximum number of bands for spectral mapping bands, 20.
H_b	Linear mapping slope for spectrum remapping
T_b	Linear mapping offset for spectrum remapping
J_b	Set of indices for spectrum remapping
$A(b)$	Pseudo-absolute threshold of hearing
$WSpe(p)$	Offset spectrum for norm adjustment
$\tilde{I}_N^q(p)$	Index of the p -th adjusted quantized norm
$I_N^q(p)$	Index of the p -th quantized norm
$\mathbf{G}$	Generator matrix for the lattice D_8
$\mathfrak{S}$	Index set of transform coefficients allocated zero bits
f_t	Transition frequency
$z^q(k)$	Reconstructed transform coefficients

Table 2 – Glossary of most relevant symbols

Symbol	Description
$y^q(k)$	Decoded normalized transform coefficients
$\tilde{x}^q(n)$	Reconstructed TDA signal
$\tilde{\mathbf{x}}_m^{qw}$	Reconstructed TDA signal matrix for sub-frame m
$R(p)$	Bit-allocation vector
Ω	Number of remainder bits
ψ	Bit budget
$\mathbf{y}_{p,l}$	The l -th 8-dimensional vector belonging to the p -th sub-vector
β	Optimal scaling factor in LVQ1
$\beta(p)$	Optimal scaling vector in LVQ2 for the p -th sub-vector

The notational conventions are detailed below:

- Time-domain signals are denoted by their symbol and a sample index between parentheses, e.g., $s(n)$. The variable n is used as sample index.
- Frequency-domain transforms are denoted by converting the related time-domain signal to capital letters, e.g., $S(k)$ is the transform of $s(n)$. The variable k is used as coefficient index.
- Superscript indices between parentheses (e.g., $g^{(m)}(n)$) are used to indicate time-dependency of variables. The variable m refers, depending on the context, to either a frame or sub-frame index, and the variable n to a sample index.
- Parameter ranges are given between square brackets, and include the boundaries (e.g., $[0.6, 0.9]$).
- The sign function gives the polarity of the value and is denoted as $\text{sgn}(x)$, where

$$\text{sgn}(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ -1 & \text{if } x < 0 \end{cases}$$

- Integer operator $\lfloor x \rfloor$ denotes rounding of x towards minus infinity ($\lfloor x \rfloor = \max \{n \in \mathbb{Z} \mid x \geq n\}$).
- The floating-point numbers used are rounded versions of the values used in the 16-bit fixed-point ANSI-C implementation.

6 General description of the codec

The ITU-T G.719 full-band codec extension is a low-complexity transform-based audio codec; it operates at a sampling rate of 48 kHz and offers full audio bandwidth ranging from 20 Hz up to 20 kHz. The encoder processes input 16-bits linear PCM signals on frames of 20 ms and the codec has an overall delay of 40 ms. The coding algorithm is based on transform coding with adaptive time-resolution, adaptive bit-allocation and low-complexity lattice vector quantization. In addition, the decoder replaces non-coded spectrum components by either signal-adaptive noise-fill or bandwidth extension.

6.1 Encoder

Figure 1 shows a block diagram of the encoder. The input signal sampled at 48 kHz is processed through a transient detector. Depending on the detection of a transient, a high frequency resolution

or a low frequency resolution transform is applied on the input signal frame. The adaptive transform is based on a modified discrete cosine transform in case of stationary frames. For non-stationary frames, a higher temporal resolution transform is used without a need for additional delay and with very little overhead in complexity. Non-stationary frames have a temporal resolution equivalent to 5-ms frames.

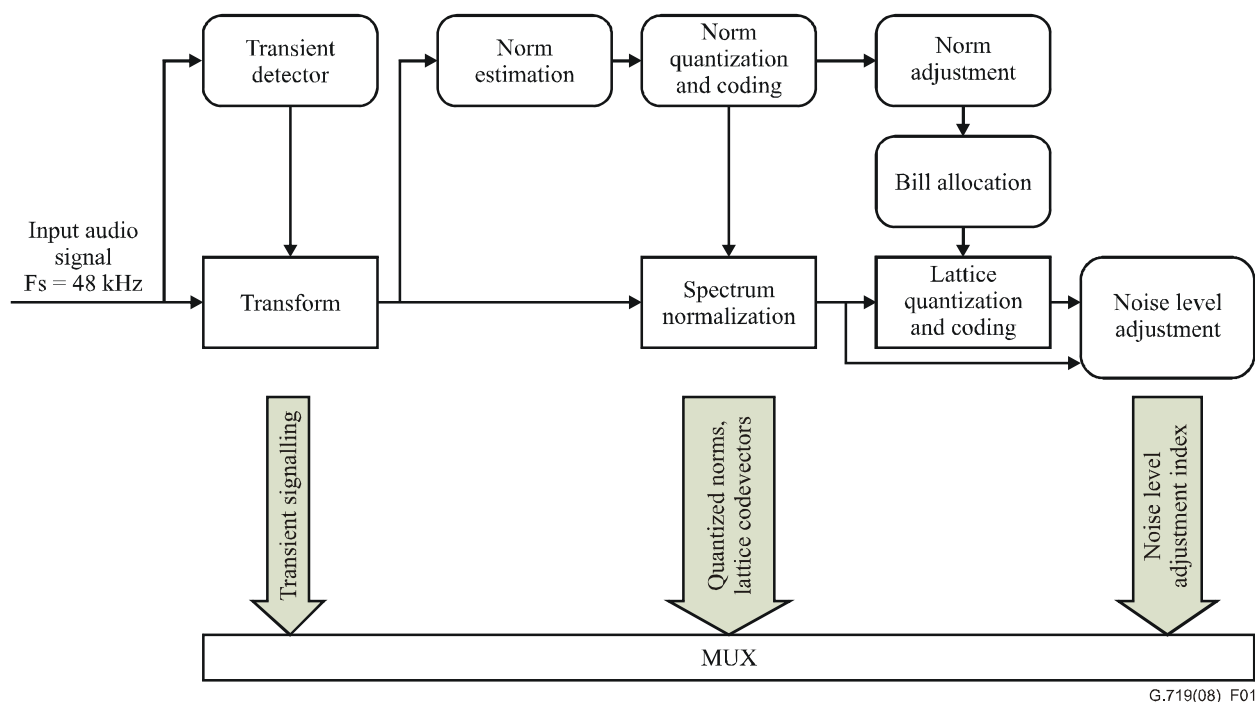


Figure 1 – Encoder block diagram

The obtained spectral coefficients are grouped into bands of unequal lengths. The norm of each band is estimated and the resulting spectral envelope consisting of the norms of all bands is quantized and encoded. The coefficients are then normalized by the quantized norms. The quantized norms are further adjusted based on adaptive spectral weighting and used as input for bit allocation. The normalized spectral coefficients are lattice-vector quantized and encoded based on the allocated bits for each frequency band. The level of the non-coded spectral coefficients is estimated, coded and transmitted to the decoder. Huffman encoding is applied to quantization indices for both the coded spectral coefficients as well as the encoded norms.

6.2 Decoder

Figure 2 shows a block diagram of the decoder. The transient flag is first decoded which indicates the frame configuration, i.e., stationary or transient. The spectral envelope is decoded and the same, bit-exact, norm adjustments and bit-allocation algorithms are used at the decoder to recompute the bit-allocation which is essential for decoding quantization indices of the normalized transform coefficients. After de-quantization, low frequency non-coded spectral coefficients (allocated zero bits) are regenerated by using a spectral-fill codebook built from the received spectral coefficients (spectral coefficients with non-zero bit allocation). Noise level adjustment index is used to adjust the level of the regenerated coefficients. High frequency non-coded spectral coefficients are regenerated using bandwidth extension. The decoded spectral coefficients and regenerated spectral coefficients are mixed and lead to normalized spectrum. The decoded spectral envelope is applied leading to the decoded full-band spectrum. Finally, the inverse transform is applied to recover the time-domain decoded signal. This is performed by applying either the inverse modified discrete

cosine transform for stationary modes, or the inverse of the higher temporal resolution transform for transient mode.

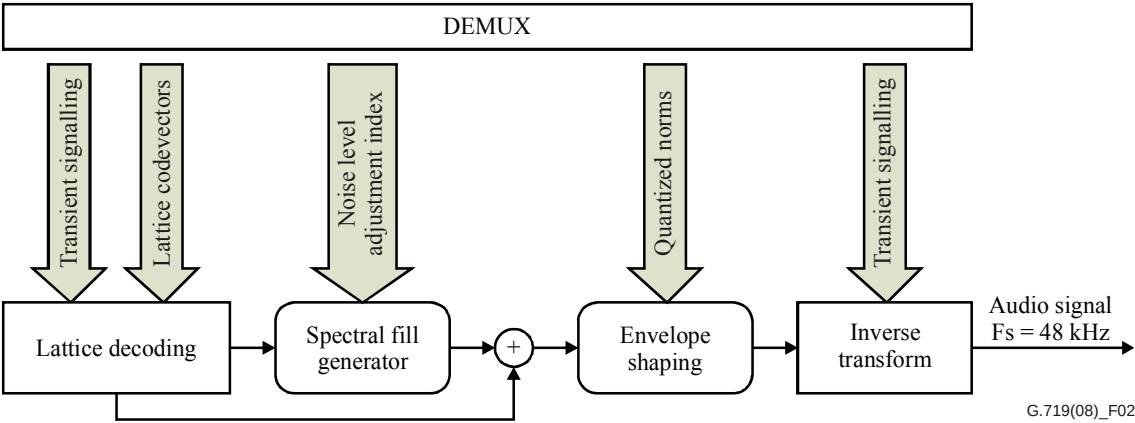


Figure 2 – Decoder block diagram

6.3 Algorithmic delay analysis

The algorithm is based on adaptive transform-coding technology. It operates on 20-ms frames of input and output audio. Because the transform window (basis function length) is of 40 ms and a 50 per cent overlap is used between successive input and output frames, the effective look-ahead buffer size is 20 ms. Hence, the overall algorithmic delay is of 40 ms which is the sum of the frame size plus the look-ahead size. All other additional delays experienced in use of the G.719 full-band codec are either due to computational and/or network transmission delays.

6.4 Complexity and memory requirements

The observed average and worst-case complexity of the encoder and decoder in WMOPS are shown in Table 3 for different bit rates. These figures are based on the obtained complexity reports using the basic operators of ITU-T STL2005 Software Tool Library v2.2 in [ITU-T G.191]. The said reports are based on encoding and decoding the source material which has been used during the subjective quality characterization test of the G.719 codec.

Table 3 – Complexity of the G.719 codec

Bit rate (kbit/s)	Encoder only		Decoder only		Encoder plus decoder	
	Average	Maximum	Average	Maximum	Average	Maximum
32	6.663	7.996	6.876	7.413	13.539	15.397
48	7.427	9.073	7.230	7.806	14.657	16.861
64	7.912	9.899	7.554	8.161	15.466	18.060
80	8.303	10.564	7.865	8.496	16.169	19.026
96	8.555	10.796	8.122	8.757	16.677	19.536
112	8.787	10.881	8.368	9.009	17.155	19.890
128	8.980	11.775	8.610	9.225	17.590	21.000

Storage requirements: static RAM, scratch RAM and data ROM in 16-bits kWords are reported in Table 4, the program ROM is measured as the number of basic operators. These are based on the fixed-point ANSI-C source code of the G.719 codec.

Table 4 – Memory usage of the G.719 codec

Memory type	Encoder	Decoder	Codec (encoder plus decoder)
Static RAM (kWords)	1.0	3.9	4.9
Scratch RAM (kWords)	12.2	12.2	24.4
Data ROM (kWords)	8.3	8.9	10.7
Program ROM (1000's basic ops)	1.2	1.2	1.8

6.5 Description of the codec

The description of the coding algorithm of this Recommendation is made in terms of bit-exact fixed-point mathematical operations. The ANSI-C code indicated in clause 10, which constitutes an integral part of this Recommendation, reflects this bit-exact, fixed-point descriptive approach. The mathematical descriptions of the encoder and decoder can be implemented in other fashions, possibly leading to a codec implementation which does not comply with this Recommendation. Therefore, the algorithm description of the ANSI-C code of clause 10 shall take precedence over the mathematical descriptions whenever discrepancies are found. A set of test signals, which can be used together with the ANSI-C code in order to verify bit-exactness, is available as an electronic attachment to this Recommendation.

7 Functional description of the encoder

7.1 Transient detection

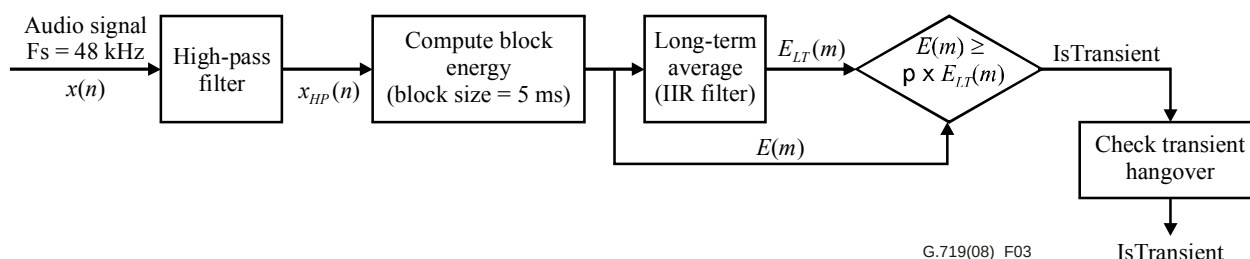


Figure 3 – Transient detection algorithm

The block diagram of the transient detection algorithm is shown in Figure 3. The input signal $x(n)$, sampled at 48 kHz, is first high-pass filtered; the high-pass filter serves as a precaution against undesired low-frequency components. A first order IIR filter is used, and it is given by:

$$H_{HP}(z) = \frac{0.7466(1-z^{-1})}{1-0.4931z^{-1}} \quad (1)$$

The frequency response of the filter is depicted in Figure 4.

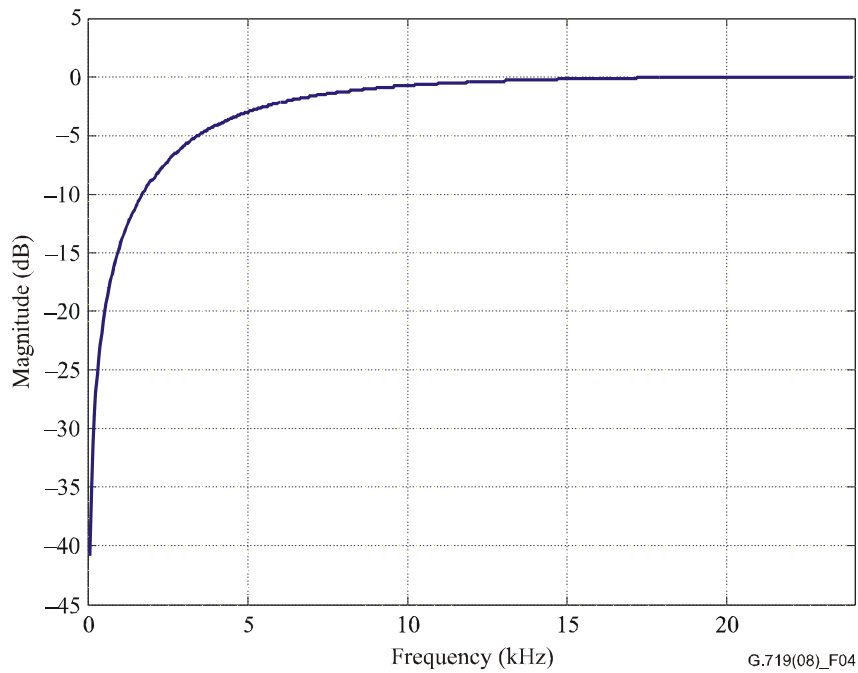


Figure 4 – Frequency response of high-pass filter

The output of the high-pass filter $x_{HP}(n)$ is obtained according to:

$$x_{HP}(n) = 0.4931 x_{HP}(n-1) + 0.7466 x(n) - 0.7466 x(n-1), \quad n = 0, \dots, L-1 \quad (2)$$

where $L = 960$ denotes the 20-ms frame length. The high-pass filtered signal $x_{HP}(n)$ is sectioned into four sub-frames, each corresponding to 5 ms, $L/4 = 240$ samples each.

The energy of each sub-frame, $E(m)$, is computed according to:

$$E(m) = \frac{1}{L/4} \sum_{n=mL/4}^{(m+1)L/4-1} x_{HP}(n)^2, \quad m = 0, \dots, 3 \quad (3)$$

The signal's long-term energy corresponding to each sub-frame, $E_{LT}(m)$, is updated according to the following equation:

$$E_{LT}(m) = (1 - \alpha) E_{LT}(m-1) + \alpha E(m), \quad m = 0, \dots, 3 \quad (4)$$

In the above equation, the forgetting factor α is set to 0.25, and the convention that $E_{LT}(-1) = E_{LT}(3)$ from the previous frame is used. The high-pass filter, $H_{HP}(z)$, states as well as $E_{LT}(3)$ are saved for processing in the next frame.

For each sub-frame m , a comparison between the short-term energy $E(m)$ and the long-term energy $E_{LT}(m)$ is performed. A transient is detected whenever the energy ratio is above a certain threshold. Formally, a transient is detected whenever:

$$E(m) \geq \rho \times E_{LT}(m), \quad (5)$$

where ρ is the energy ratio threshold and is set to $10 \log_{10}(\rho) = 7.8$ dB.

In general, the time-frequency transform is applied on a 40-ms frame; therefore, a transient will affect two consecutive frames. To overcome this, a hangover for detected transient is used. A transient detected at a certain frame will also trigger a transient at the next frame.

The output of the transient detector is a flag, denoted *IsTransient*. The flag is set to the logical value *TRUE* if a transient is detected, or *FALSE* otherwise.

7.2 Adaptive time-frequency transform

The adaptive time-frequency transform is based on the detection of a transient. In the case of stationary signals, the transform has a high frequency resolution which is able to efficiently represent stationary sounds. In the case of transient sounds, the time-frequency transform will increase its time resolution, which allows a better representation of the rapid changes in the input signal spectral characteristics.

These two modes of operation share a common buffering and windowing module and the switching between one mode of operation to the other is instantaneous. Thus no additional look-ahead in the transient detection is needed. This allows the codec to have selectable time resolution at low complexity and with zero additional delay.

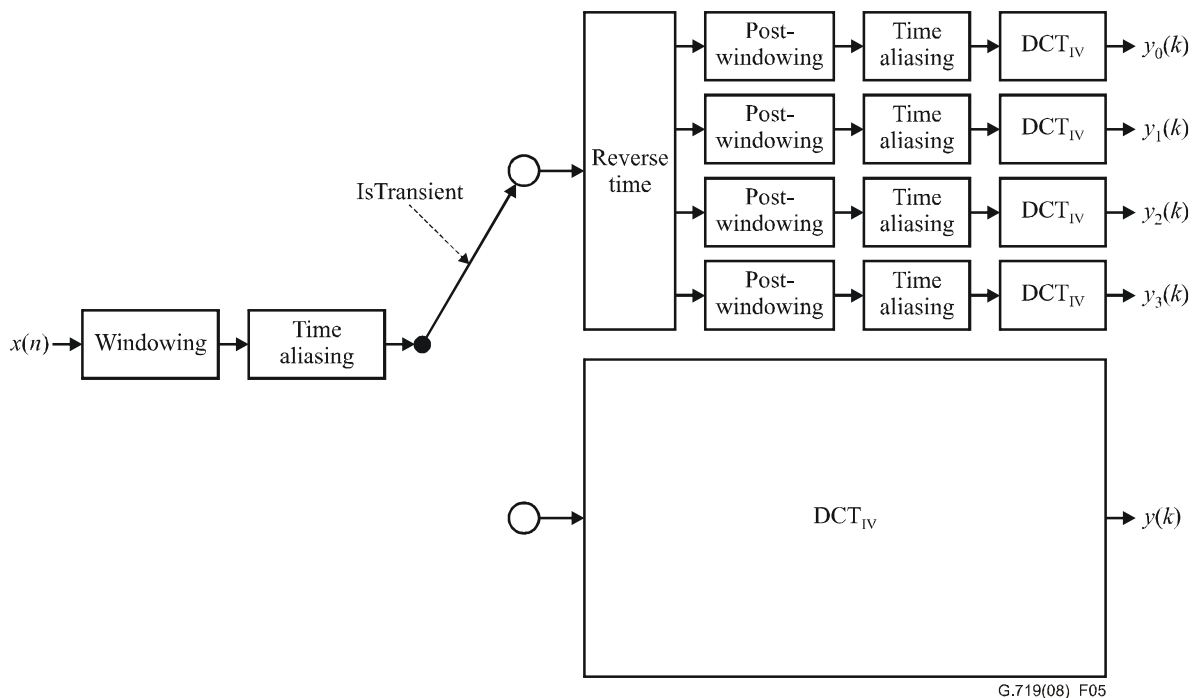


Figure 5 – Adaptive time-frequency transform

Figure 5 illustrates the overall operation of the adaptive time-frequency decomposition. Depending on the *IsTransient* flag, the algorithm outputs either the MLT, a cascade of windowing then time domain aliasing and a discrete cosine transform type IV (DCT_{IV}), of the input signal $y(k)$, with $k = 0, \dots, L - 1$, or a finer time resolution transform with four spectra $y_0(k)$, $y_1(k)$, $y_2(k)$, $y_3(k)$ with $k = 0, \dots, L/4 - 1$. In the following clauses, a description of the operations is given.

7.2.1 Input signal buffering and windowing

The same input frame signal $x(n)$ which is used for transient detection is used in order to compute the frame spectral representation. The time chart for this operation is illustrated in Figure 6.

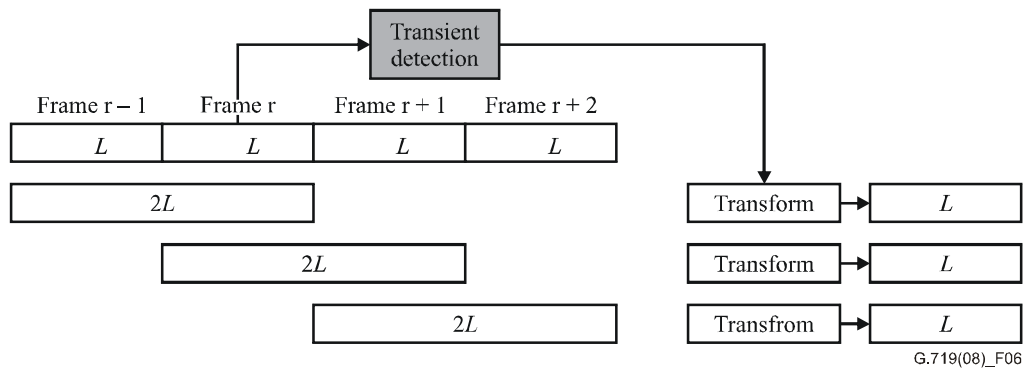


Figure 6 – Time chart of signal buffering and transform

The 20-ms input frame $x(n)$, $n = 0, \dots, L-1$, is buffered with the previous 20-ms frame $x_{OLD}(n)$ forming a 40-ms frame. This frame is in turn windowed as illustrated in Figure 7.

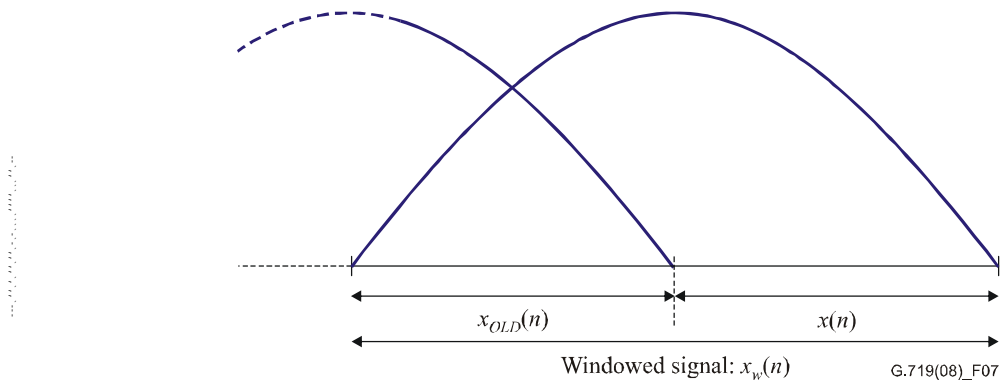


Figure 7 – Frame buffering and windowing

The window function is a sine window and is defined as:

$$h(n) = \sin\left[\left(n + \frac{1}{2}\right) \frac{\pi}{2L}\right], \quad n = 0, \dots, 2L-1 \quad (6)$$

which leads to the following windowed signal:

$$x_w(n) = \begin{cases} h(n)x_{OLD}(n), & n = 0, \dots, L-1 \\ h(n)x(n-L), & n = L, \dots, 2L-1 \end{cases} \quad (7)$$

7.2.2 Time domain aliasing

The frame resulting from buffering and windowing has a 40-ms length, $2L = 1920$. It is transformed into a 20-ms frame by using a time-domain aliasing (TDA) operation.

The TDA operation is explicitly given by the following matrix operation:

$$\tilde{\mathbf{x}} = \begin{bmatrix} 0 & 0 & -J_{L/2} & -I_{L/2} \\ I_{L/2} & -J_{L/2} & 0 & 0 \end{bmatrix} \mathbf{x}_w \quad (8)$$

where the matrices $I_{L/2}$ and $J_{L/2}$ denote the identity and the time-reversal matrices of order $L/2$:

$$I_{L/2} = \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix}, \quad J_{L/2} = \begin{bmatrix} 0 & & 1 \\ & \ddots & \\ 1 & & 0 \end{bmatrix} \quad (9)$$

Note that the matrix of equation 8 has half as many rows as columns, thus making the length of the output signal $\tilde{x}(n)$ half that of the input signal $x_w(n)$. This operation is illustrated in Figure 8.

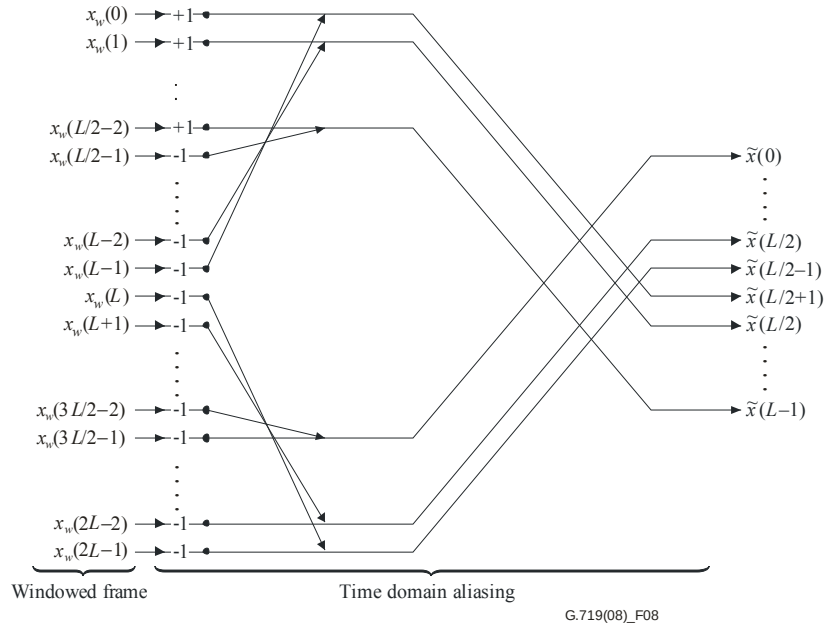


Figure 8 – Time-domain aliasing operation

7.2.3 Signal pre-scaling

In order to maintain a high SNR in a fixed point implementation of the adaptive transform for small signals, each time-aliased frame $\tilde{x}(n), n = 0, \dots, L-1$, is scaled to span the full dynamic range corresponding to a 32-bit signed integer, i.e., $[-2^{31}, \dots, 2^{31}-1]$.

The scaling is a simple operation and consists of finding the maximum absolute sample in the time domain aliased signal frame and computing a normalization index $Q_{\max}$ according to:

$$Q_{\max} = 30 - \left\lceil \log_2 \left(\max_{n=0, \dots, L-1} |\tilde{x}(n)| \right) \right\rceil \quad (10)$$

The computed index $Q_{\max}$ is used to scale all samples of the frame prior to further processing by the algorithm. In the above equation,

$$\tilde{\tilde{x}}(n) = 2^{Q_{\max}} \tilde{x}(n), \quad n = 0, \dots, L-1 \quad (11)$$

In the following, the notations $\tilde{\tilde{x}}(n)$ will be omitted for ease of notation.

7.2.4 Stationary mode

If the *IsTransient* flag is set to *FALSE*, a type IV discrete cosine transform (DCT_{IV}) is applied on the output of the time domain-aliasing operation.

The DCT_{IV} [b-Malvar] of the time-aliased signal of length L is defined by the following equation:

$$y(k) = \sum_{n=0}^{L-1} \tilde{x}(n) \cos \left[\left(n + \frac{1}{2} \right) \left(k + \frac{1}{2} \right) \frac{\pi}{L} \right], \quad k = 0, \dots, L-1 \quad (12)$$

It should be noted that in stationary mode, the cascade of windowing, TDA and DCT_{IV} is equivalent to applying the modulated lapped transform (MLT). The resulting signal $y(k)$ represents the spectral coefficients of the input frame.

7.2.5 Transient mode

If the *IsTransient* flag is set to *TRUE*, a further reordering of the time-domain aliased signal is performed. In order to keep the temporal coherence of the input signal, the output of the time-domain aliasing operation needs to be reordered before further processing. The ordering operation is necessary, without ordering the basis functions of the resulting filter-bank will have an incoherent time and frequency responses. The reordering operation is illustrated in Figure 9, and consists of shuffling the upper and lower half of the TDA output signal $\tilde{x}(n)$.

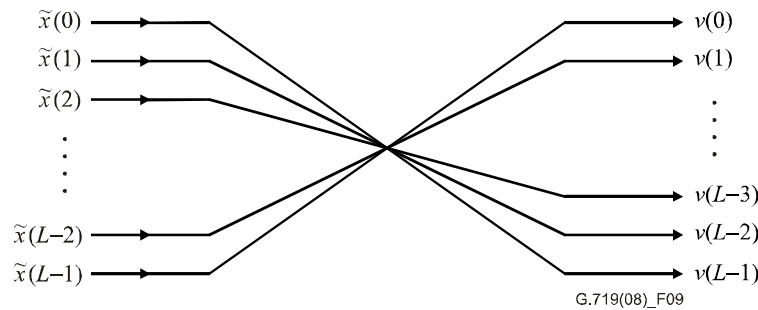


Figure 9 – Time-domain aliasing reordering

$$v(n) = \tilde{x}(L-1-n), \quad n = 0, \dots, L-1 \quad (13)$$

This reordering is only conceptual and in reality no computations are involved. Higher time resolution is obtained by zero-padding the signal $v(n)$ and dividing the resulting signal into four overlapped equal length sub-frames. The amount of zero-padding is equal to $L/8$ on each side of the signal. The segments are 50% overlapped and each segment has a length equal to $L/2$. The two inner segments are post-windowed using a sine window of length $L/2$. The windows for outer segments are constructed using half a sine window.

This operation is depicted in Figure 10; each resulting post-windowed segment is further processed by applying the modified discrete cosine transform (MDCT) (i.e., time aliasing + DCT_{IV}). The output of the MDCT for each segment represents the signal spectrum at different time instants, thus, in case of transients, a higher time resolution is used.

The length of the output of each of the four MDCTs is half the length of the input segment, i.e., $L/4$, therefore this operation does not add additional redundancy.