



INTERNATIONAL TELECOMMUNICATION UNION

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

G.723.1

Annex A
(11/96)

SERIES G: TRANSMISSION SYSTEMS AND MEDIA

Digital transmission systems – Terminal equipments –
Coding of analogue signals by methods other than PCM

Dual rate speech coder for multimedia
communications transmitting at 5.3 and 6.3 kbit/s

Annex A: Silence compression scheme

ITU-T Recommendation G.723.1 – Annex A

(Previously CCITT Recommendation)

ITU-T G-SERIES RECOMMENDATIONS
TRANSMISSION SYSTEMS AND MEDIA

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
TRANSMISSION MEDIA CHARACTERISTICS	G.600–G.699
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
General aspects	G.800–G.809
Design objectives for digital networks	G.810–G.819
Quality and availability targets	G.820–G.829
Network capabilities and functions	G.830–G.839
SDH network characteristics	G.840–G.899
DIGITAL SECTIONS AND DIGITAL LINE SYSTEM	G.900–G.999
General	G.900–G.909
Parameters for optical fibre cable systems	G.910–G.919
Digital sections at hierarchical bit rates based on a bit rate of 2048 kbit/s	G.920–G.929
Digital line transmission systems on cable at non-hierarchical bit rates	G.930–G.939
Digital line systems provided by FDM transmission bearers	G.940–G.949
Digital line systems	G.950–G.959
Digital section and digital transmission systems for customer access to ISDN	G.960–G.969
Optical fibre submarine cable systems	G.970–G.979
Optical line systems for local and access networks	G.980–G.999

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

ITU-T RECOMMENDATION G.723.1 – Annex A

SILENCE COMPRESSION SCHEME

Source

Annex A to ITU-T Recommendation G.723.1, was prepared by ITU-T Study Group 15 (1993-1996) and was approved under the WTSC Resolution No. 1 procedure on the 8th of November 1996.

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 (Helsinki, March 1-12, 1993).

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.

© ITU 1997

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
A.1 Introduction.....	1
A.2 Description of the VAD	2
A.2.1 Adaptation enable flag computation.....	2
A.2.2 Inverse filtering.....	3
A.2.3 Filtered energy computation.....	3
A.2.4 Noise level computation	3
A.2.5 Threshold computation	4
A.2.6 The VAD decision	4
A.2.7 VAD hangover addition.....	4
A.2.8 VAD initialization	4
A.3 General description of the CNG	4
A.4 Description of the CNG encoder part	6
A.4.1 Computation of the frame autocorrelation function	6
A.4.2 Computation of the current frame type F_{typ_t}	7
A.4.3 Quantization of the average energy	8
A.4.4 Computation and coding of SID parameters.....	9
A.4.5 Computation of the CNG excitation.....	10
A.4.6 Interpolation of LSPs and update.....	11
A.4.7 COD-CNG initialization.....	11
A.5 Description of the decoder part.....	11
A.5.1 Description of DEC-CNG	12
A.5.2 Frame erasure concealment with regards to the CNG.....	13
A.5.3 DEC-CNG initialization	14
A.6 Bit stream packing	14
A.7 Glossary	14
A.8 Bit-exact, fixed-point C source code	15

SILENCE COMPRESSION SCHEME

(Geneva, 1996)

A.1 Introduction

This Annex describes the silence compression system that has been designed for the G.723.1 speech coder. Silence compression techniques are used to reduce the transmitted bit rate during silent intervals of speech. Systems allowing discontinuous transmission are based on a Voice Activity Detection (VAD) algorithm and a Comfort Noise Generator (CNG) algorithm that allows the insertion of an artificial noise during silence periods. This feature is necessary to avoid noise modulation introduced when the transmission is switched off: if the background acoustic noise that was present during active periods abruptly disappears, this very unpleasant noise modulation may even reduce the intelligibility of the speech.

The purpose of the VAD is to reliably detect the presence or absence of speech and to convey this information to the CNG algorithm. Typically, VAD algorithms base their decisions on several successive frames of information in order to make them more reliable and to avoid producing intermittent decisions. The VAD is constrained to operate on the same 30 ms speech frames which will subsequently either be encoded by the speech coder or filled with comfort noise by the comfort noise generator. The output of the VAD algorithm is passed to the CNG algorithm.

The largest difficulty in the detection of speech is the presence of any of a diverse range of background noise conditions. The VAD must be able to detect speech even in very low signal-to-noise ratio conditions. It is impossible to distinguish between speech and noise using simple level detection techniques when parts of the speech utterance are buried below the noise. The distinction between these conditions can only be made by taking into consideration the spectral characteristics of the input signal. In order to do this, the VAD incorporates an inverse filter, the coefficients of which are derived during noise-only periods by the CNG. All further details of the VAD are included in A.2.

The purpose of the CNG algorithm is to create a noise that matches the actual background noise with a global transmission cost as low as possible. At the transmitting end, the CNG algorithm uses the activity information given by the VAD for each frame, then computes the encoded parameters needed to synthesize the artificial noise at the receiving end. These encoded parameters compose the Silence Insertion Descriptor (SID) frames, which require less bits than the active speech frames and are transmitted during inactive periods.

The main feature of this CNG algorithm is that the transmission of SID frames is not periodic: for each inactive frame, the algorithm makes the decision of sending a SID frame or not, based on a comparison between the current inactive frame and the preceding SID frame. In this way, the transmission of the SID frames is limited to the frames where the power spectrum of the noise has changed.

During inactive frames, the comfort noise is synthesized at the decoder by introducing a pseudo-white excitation into the short-term synthesis filter. The parameters used to characterize the comfort noise are the LPC synthesis filter coefficients and the energy of the excitation signal. At the encoder, for each SID frame the algorithm computes a set of LPC parameters and quantizes the corresponding LSPs using the coder LSP quantizer on 24 bits. It also evaluates the excitation energy and quantizes it with 6 bits. This yields encoded SID frames of 4 bytes including the 2 bits for bit rate and DTX information.

A notable feature of this CNG algorithm is the method used to evaluate the spectrum of the ambient noise for each SID frame. It takes into account the local stationarity or non-stationarity of the input signal.

Finally, the excitation corresponds to the higher bit rate excitation of the G.723.1 codec. Since the fixed excitation has a rather poor spectrum, the long-term excitation is also used in order to obtain a better white-noise-type of excitation. The algorithm randomly chooses the codes of the long-term parameters (delays and gains) and the fixed codebook parameters (grid, pulse positions and signs). For every two subframes, it computes the gain of the fixed excitation to achieve a global energy derived from the transmitted SID energy.

The computation of the excitation needs to be performed both at the encoder and at the decoder to keep both parts synchronized.

At the receiver, to simplify the procedure, the harmonic postfilter is switched off during comfort noise generation since the generated noise is not a voiced signal.

The results of the tests on the VAD/DTX/CNG scheme as described in this Annex will be published at a later date as an appendix to Annex A to Recommendation G.723.1.¹

A.2 Description of the VAD

This subclause describes the Voice Activity Detector (VAD) used in the G.723.1 speech coder. The function of the VAD is to indicate whether each 30 msec frame produced by the speech encoder contains speech or not. The VAD decision at frame t is labelled as Vad_t and is the input to the COD-CNG block that computes $Ftyp_t$, as described in A.3 and Figure A.1. The performance of the VAD algorithm is characterized by the amount of audible speech clipping and the percentage of speech activity it indicates.

The VAD is basically an energy detector. The energy of the inverse filtered signal is compared with a threshold. Speech is indicated whenever the threshold is exceeded. The threshold is computed by a two-step procedure. First, the noise level is updated based on its previous value and the energy of the filtered signal. Second, the threshold is computed from the noise level via a logarithmic approximation.

Hangover is a term describing the practice of declaring the first few frames of silence following a speech burst to still be speech. It is used to eliminate low level speech clipping. Hangover is only added to speech bursts which exceed a certain duration to avoid extending noise spikes.

A.2.1 Adaptation enable flag computation

An adaptation enable flag, denoted Aen_t for the current frame t , is used to be sure that the VAD noise level is adapted only when speech is not present. It is based on the fact that the background noise or the silence is neither a voiced signal nor a sine wave:

– Voiced/Unvoiced detection:

The open loop pitch delays of the preceding and current frame are used to test voicing. Let us note $L_{OL}^j, j = 0,1,2,3$ those four values. The minimum delay $L_{OL}^{\min} = \text{Min}(L_{OL}^j, j = 0,1,2,3)$ is first computed. The counter $pc \in [1,2,3,4]$ indicating how many delays L_{OL}^j lie in the neighborhood of a multiple of L_{OL}^{\min} (± 3) is evaluated. If pc is equal to 4 the signal is considered as voiced.

¹ It should be noted that test conditions were not sufficiently severe for mobile conditions.

- Sine wave detection : (already present COM 15-255 Contribution)
The following sine wave detector is included in the LPC analysis of the G.723.1 encoder:
Let $k_i^t[2]$ be the second reflection coefficient computed by the Durbin recursion for each subframe $i = 0, \dots, 3$ of frame t .
If $k_i^t[2] \geq 0.95$ for at least 14 of the 15 last values, then a sine wave is detected ($SinD = 1$).
In the other case, $SinD = 0$.
- Compute the adaptation enable flag:

$$\begin{cases} Aen_t = Aen_{t-1} + 2 & \text{if } pc = 4 \text{ or } SinD = 1 \\ Aen_t = Aen_{t-1} - 1 & \text{otherwise} \end{cases}$$

Aen_t is bounded into $[0,6]$.

A.2.2 Inverse filtering

The input signal frame, $\{s[n]\}_{n=60..239}$, is inverse filtered by a FIR filter $A_{no}(z)$ with coefficients $\{a_{no}[j]\}_{j=1..10}$. This filter is calculated by the CNG block and provides an estimation of the LPC filter associated to the current background noise.

$$e'_t[n] = s[n] + \sum_{j=1}^{10} a_{no}[j] \cdot s[n-j] \quad n = 60 \rightarrow 239 \quad (\text{A-1})$$

where $e'_t[n]$ is the inverse filtered signal.

A.2.3 Filtered energy computation

The energy, Enr_t , is computed from the inverse filtered signal of the current frame by:

$$Enr_t = \frac{1}{80} \sum_{n=60}^{239} e_t'^2[n] \quad (\text{A-2})$$

A.2.4 Noise level computation

The noise level at frame t , $Nlev_t$, is updated based on its previous value and on the previous energy, Enr_{t-1} and on the adaptation enable flag Aen_t . This update procedure is characterized by slow attack and fast decay. The dynamic range of the noise level at frame t is limited to the range $[Nlev_{\min}, Nlev_{\max}]$.

- 1) If $Nlev_{t-1} > Enr_{t-1}$ then the noise level is first clipped:

$$Nlev_t = \begin{cases} 0.25 \cdot Nlev_{t-1} + 0.75 \cdot Enr_{t-1} & \text{if } Nlev_{t-1} > Enr_{t-1} \\ Nlev_{t-1} & \text{otherwise} \end{cases} \quad (\text{A-3})$$

- 2) Then $Nlev_t$ is increased, if adaptation is enabled, otherwise it is decreased by a small amount:

$$Nlev_t = \begin{cases} 1.03125 \times Nlev_t & \text{if } Aen_t = 0 \\ 0.9995 \times Nlev_t & \text{otherwise} \end{cases} \quad (\text{A-4})$$

$$\text{with } \begin{cases} Nlev_{\min} = 128 \\ Nlev_{\max} = 131071 \end{cases}$$

A.2.5 Threshold computation

The relationship between the noise level at frame t , $Nlev_t$, and the threshold, Thr , is defined by logarithmic approximation and defined by the following formula:

$$Thr = \begin{cases} 5.012 & \text{if } Nlev = 128, \\ 10^{0.7-0.05\log_2 \frac{Nlev}{128}} & \text{if } 128 < Nlev < 16384 \\ 2.239 & \text{if } Nlev \geq 16384 \end{cases} \quad (A-5)$$

A.2.6 The VAD decision

The VAD decision is based on the comparison between the threshold, Thr , and the current energy, Enr_t .

$$Vad_t = \begin{cases} 1 & \text{if } Enr_t \geq Thr \\ 0 & \text{if } Enr_t < Thr \end{cases} \quad (A-6)$$

A.2.7 VAD hangover addition

A hangover of 6 frames is added only in the case of speech bursts ($Vad_t = 1$) larger or equal than 2 frames.

A.2.8 VAD initialization

All static variables of the VAD algorithm are initialized to zero, except the following variables:

$$\begin{aligned} Nlev_{-1} &= 1024 \\ Enr_{-1} &= 1024 \\ L_{OL}^j &= 1 & j = 0,1 \\ L_{Ol}^j &= 60 & j = 2,3 \end{aligned} \quad (A-7)$$

A.3 General description of the CNG

The algorithm is divided into two blocks situated at the encoder and the decoder, that will be called respectively COD-CNG and DEC-CNG. At the encoder (see Figure A.1), the COD-CNG block uses the autocorrelation function of the speech signal computed for each 60 samples subframe, the past excitation samples and LSPs from the preceding frame.

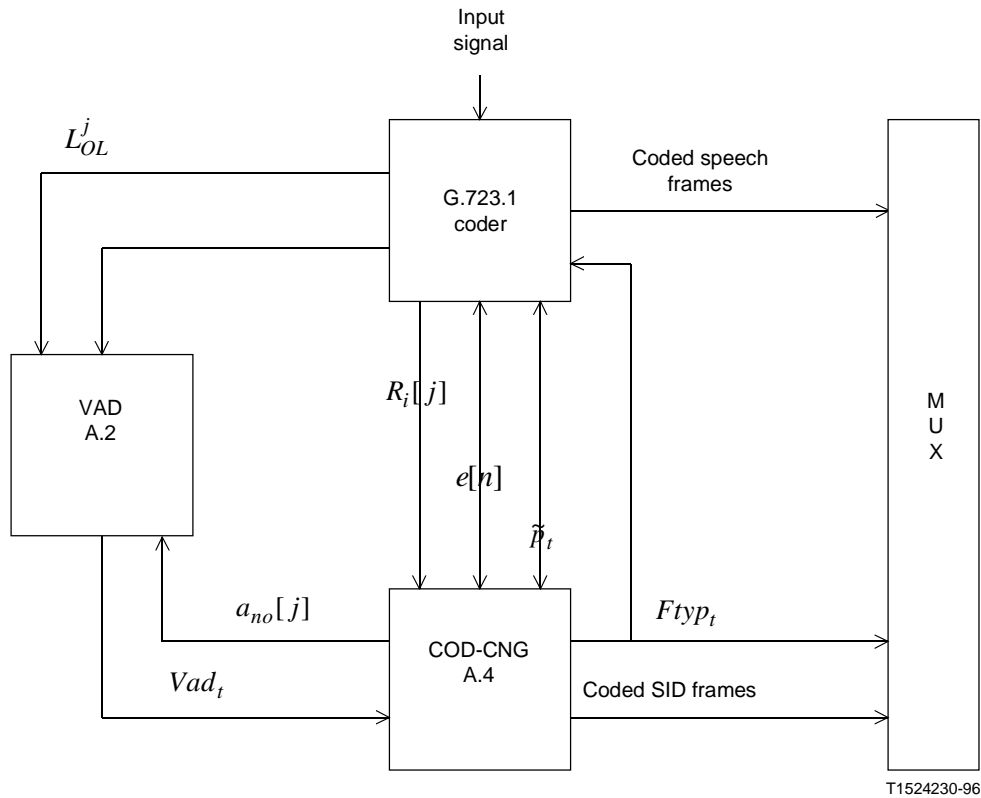


FIGURE A.1/G.723.1

Block diagram of the encoder with VAD/CNG

For inactive frames, COD-CNG computes the CNG excitation samples in order to synchronize the local decoder of the encoder with the distant decoder.

Because of the predictive coding of the LSPs in the G.723.1 scheme, a similar input/output with update is done for LSP parameters during inactive frames.

COD-CNG outputs the encoded SID frames and the final decision $Ftyp_t$ (Frame type of frame t) as one of the three values, 0, 1, or 2 corresponding to untransmitted frame, active speech frame or SID frame, respectively.

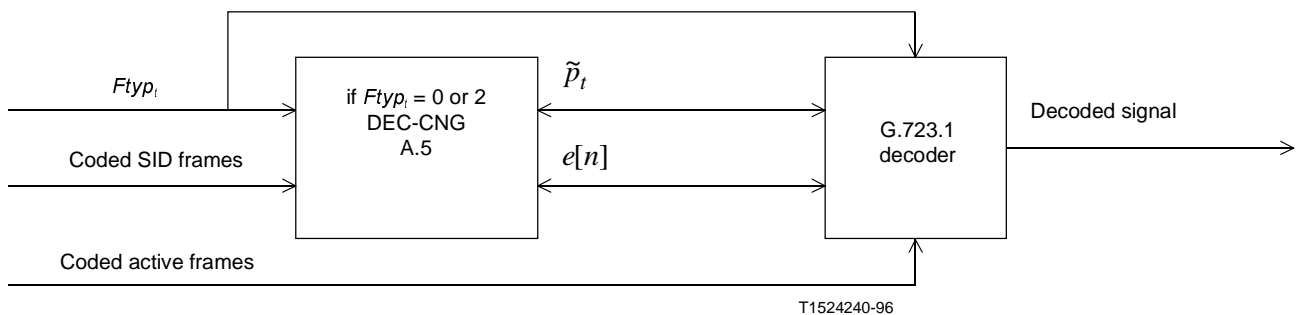


FIGURE A.2/G.723.1

Block diagram of the decoder with VAD/DTX

At the receiver (see Figure A.2), the DEC-CNG block processes only inactive speech frames, for which the input information $Ftyp_t$ is equal to 0 or 2 (untransmitted/SID). DEC-CNG decodes the SID frames and both for SID and untransmitted frames, computes the current LSPs and excitation using the same method as COD-CNG.

Then the G.723.1 decoder synthesizes the comfort noise using the CNG excitation and LSPs.

A.4 Description of the CNG encoder part

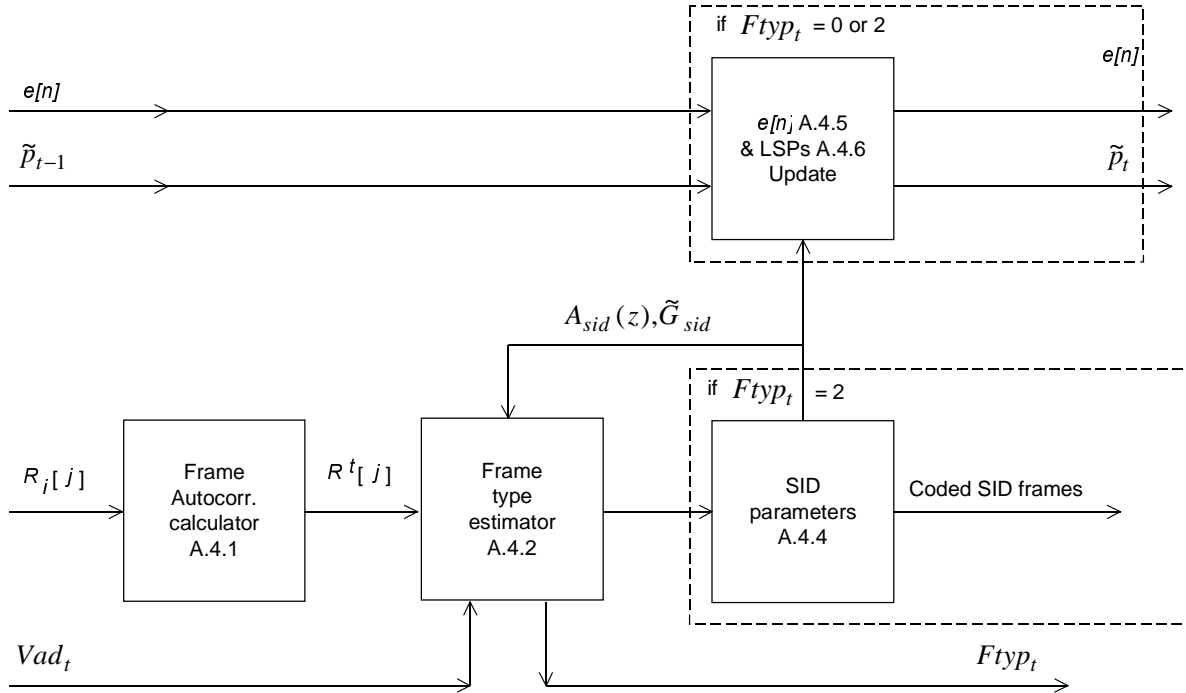
For each frame of 240 samples (active or inactive), the COD-CNG block processes the data coming from the VAD and the coder, produces the $Ftyp_t$ information and the coded SID frame according to the procedure depicted by Figure A.3 and detailed in A.4.1 to A.4.7.

A.4.1 Computation of the frame autocorrelation function

File: COD_CNG.C	Procedure: Update_Acf()	Update autocorrelation function
-----------------	-------------------------	---------------------------------

For every frame t (active or inactive), the autocorrelation coefficients (calculated in the encoder as described in 2.4 of Recommendation G.723.1) $R_i[j]$, $j = 0$ to 10 of the four subframes indexed by $i = 0$ to 3 are summed. The cumulated autocorrelation function of the current frame t , is given by:

$$R^t[j] = \sum_{i=0}^3 R_i[j], \text{ for } j = 0 \text{ to } 10 \quad (\text{A-8})$$



T1524250-96

FIGURE A.3/G.723.1

Block diagram of the CNG at the encoder part

A.4.2 Computation of the current frame type $Ftyp_t$

File: COD_CNG.C	Procedure: Cod_Cng()	COD_CNG main body
File: COD_CNG.C	Procedure: LpcDiff()	Itakura distance comparison
File: LPC.C	Procedure: Durbin()	Levinson-Durbin recursion

If the current frame t is an active speech frame ($Vad_t=1$), then $Ftyp_t = 1$ and no other processing is performed.

In the other case, the decision SID/untransmitted frame is taken according to the following procedure:

The LPC filter $A_t(z)$ of the current frame t is calculated by the Durbin procedure (see 2.4 of Recommendation G.723.1) using $R^t[j]$ as input. The coefficients of $A_t(z)$ are noted $a_t[j]$, $j=1$ to 10. The Durbin procedure also provides the residual energy E_t , that will be used as an estimate of the frame excitation energy.

Then the current frame type $Ftyp_t$ is determined in the following way:

- If the current frame is the first inactive frame of the inactive zone, the frame is selected as SID frame, the variable \bar{E} which reflects the energy sum is taken equal to E_t , and the number of frames involved in the summation, k_E , is initialized to 1:

$$(Vad_{t-1} = 1) \Rightarrow \begin{cases} Ftyp_t = 2 \\ \bar{E} = E \\ k_E = 1 \end{cases} \quad (A-9)$$

- Else, if the current filter is significantly different from the preceding SID filter, or if the current excitation energy significantly differs from the preceding SID energy, then the frame is selected as SID ($Ftyp_t = 2$).
- Otherwise, if the current frame is not the first of an inactive period, and if the current LPC filter and the excitation energy are similar to the SID ones, the frame is not transmitted ($Ftyp_t = 0$).

The LPC filters and energies are compared according to the following methods:

Comparison of the LPC filters

The current LPC filter and SID filter are considered as significantly different if the Itakura distance between the two filters exceeds the given threshold, which is expressed by:

$$\sum_{j=0}^{10} R_a[j] \times R^t[j] \geq E_t \times thr1 \quad (A-10)$$

where $R_a[j]$, $j = 0$ to 10 is a function derived from the autocorrelation of the coefficients of the SID filter, given by :

$$\begin{cases} R_a[j] = 2 \sum_{k=0}^{10-j} a_{sid}[k] \times a_{sid}[k+j] & \text{if } j \neq 0 \\ R_a(0) = \sum_{k=0}^{10} a_{sid}[k]^2 \end{cases} \quad (A-11)$$

with $a_{sid}[0] = 1$

A value of 1.2136 is used for $thr1$.

Comparison of the energies

k_E being first incremented up to the maximum value 3, the sum the frame energies $\bar{E} = \sum_{i=t-k_E+1}^t E_t$ is calculated.

Then \bar{E} is quantized, using the 6-bit pseudo-logarithmic quantizer described in A.4.3. The coded gain index $GInd_t$ is compared to the previous coded SID gain index $GInd_{sid}$. If the difference exceeds the threshold $thr2=3$, the two energies will be considered as significantly different.

A.4.3 Quantization of the average energy

File: UTIL_CNG.C	Procedure: Qua_SidGain()	Quantize Sid Gain
File: UTIL_CNG.C	Procedure: Dec_SidGain()	Decode Sid Gain

The quantization procedure operates on the sum of the energies \bar{E} , and the decoding provides a gain, which corresponds to the decoded value of the average energy square root.

A scaling factor $\alpha_w = 2.70375$ is introduced to take into account the effect of windowing and bandwidth expansions present in the subframes autocorrelation functions $R_i[j]$

The value used at the input of the gain quantizer is:

$$G = \alpha_w \times \sqrt{\frac{1}{k_E \times 240} \bar{E}}, \text{ bounded in } [0, 352].$$

The quantizer is a pseudo-log one, that divides $[0, 352]$ into three segments indexed $isg = 0$ to 2 of length $N[isg] = 16, 16, 32$ with the associated resolutions 2, 4 and 8.

Let $G_{isg}[j], j = 0$ to $N[isg]-1$ be the decoded values for segment isg . Those values are given by:

$$G_{isg}[j] = G_{isg}[0] + j \times 2^{(isg+1)} \quad (\text{A-12})$$

The procedure uses G^2 to calculate the index isg of the segment which contains G and the index i_s of $G_{isg}(i_s)$ the closer to G .

The current quantization index is given by:

$$GInd_t = 16 \times isg + i_s \quad (\text{A-13})$$

The decoding is performed using the following formula:

$$Q^{-1}(GInd_t) = G_{isg}[0] + (GInd_t - \lfloor isg / 16 \rfloor) \times 2^{isg+1} \quad (\text{A-14})$$

where $\lfloor x \rfloor$ denotes the greatest integer $\leq x$.

A.4.4 Computation and coding of SID parameters

File: COD_CNG.C	Procedure: Cod_Cng()	COD_CNG main body
File: COD_CNG.C	Procedure: ComputePastAvFilter()	Computes past average filter
File: COD_CNG.C	Procedure: LpcDiff()	Itakura distance comparison
File: COD_CNG.C	Procedure: CalcRc()	Compute function RC from LPC
File: LPC.C	Procedure: Durbin()	Levinson-Durbin recursion
File: LSP.C	Procedure: AtoLsp()	Converts LPC coefficients into LSP
File: LSP.C	Procedure: Lsp_Qnt()	LSP quantization
File: LSP.C	Procedure: Lsp_Inq()	LSP inverse quantization

When the current frame is a SID frame, the SID parameters are calculated and quantized. Notice that those parameters will serve in making the SID decision for the next inactive frames up to the next SID frame.

Computation of SID LPC filter $[A_{sid}(z)]$ and update of VAD LPC filter $[A_{no}(z)]$

First, the past average LPC filter $\bar{A}_p(z)$ built from the three frames preceding the current one is estimated, using the Durbin procedure with the following autocorrelation function as input:

$$\bar{R}_p[j] = \sum_{k=t-3}^{t-1} R^k[j], \text{ for } j = 0 \text{ to } 10 \quad (\text{A-15})$$

the autocorrelation functions $R^k[j]$ being the cumulated ones calculated by (A-8).

The past average LPC filter coefficients are denoted $\bar{a}_p[j]$, $j = 1$ to 10 .

The VAD noise LPC filter used in A-2 is then updated with $\bar{a}_p[j]$ but only when the adaptation enable flag Aen_t allows it:

$$\text{if } Aen_t = 0 \text{ then } a_{no}[j] = \bar{a}_p[j], j = 1, 2, \dots, 10 \quad (\text{A-16})$$

$$\text{Then } A_{sid}(z) = \begin{cases} A_t(z) & \text{if the distance between } A_t(z) \text{ and } \bar{A}_p(z) \text{ is } \geq thr1 \\ \bar{A}_p(z) & \text{otherwise} \end{cases} \quad \text{see eq. (A-10)}$$

The distance between the current LPC filter $A_t(z)$ and the average past LPC filter $\bar{A}_p(z)$ is computed in the same manner as in A.4.2.

The coefficients $a_{sid}[j]$, $j = 1 \rightarrow 10$ of the new SID LPC filter are LSP converted and the LSPs are quantized using the encoder LSP 24-bit quantization procedure (see 2.5 of Recommendation G.723.1). The decoded value will be called \tilde{p}_{sid} .

SID gain

The quantized value of the SID gain is given by:

$$GInd_{sid} = GInd_t \quad (\text{A-17})$$

and the decoded value is denoted \tilde{G}_{sid} .

A.4.5 Computation of the CNG excitation

File: UTIL_CNG.C	Procedure: Calc_Exc_Rand()	Computation of the excitation
File: UTIL_CNG.C	Procedure: random_number()	Random number generation
File: UTIL_CNG.C	Procedure: distG()	Used to select excitation Gain
File: UTIL_LBC.C	Procedure: Sqrt_lbc()	Square root
File: UTIL_LBC.C	Procedure: Rand_lbc()	Pseudo-random sequence

The update of the excitation signal is performed both for SID frames and for untransmitted frames.

First, let us define the target excitation gain \tilde{G}_t as the square root of the average energy that must be obtained for the current frame t synthetic excitation. \tilde{G}_t is calculated using the following smoothing procedure:

$$\tilde{G}_t = \begin{cases} \tilde{G}_{sid} & \text{if } Vad_{t-1} = 1 \\ \frac{7}{8} \tilde{G}_{t-1} + \frac{1}{8} \tilde{G}_{sid} & \text{otherwise} \end{cases} \quad (\text{A-18})$$

The 240 samples of the frame are divided into two blocks of 120 samples, each block comprising two subframes of 60 samples.

For each block, the CNG excitation samples are synthesized using the following algorithm:

First the LTP parameters of the two subframes are selected:

- The pitch lag for the first subframe is randomly chosen in the interval [123, 143].
- The two subframes gain vector indices are randomly chosen into [0, 49], which corresponds to the first 50 vectors of the 170 entries gain codebook.
- The second subframe lag offset is taken equal to 0 for the first block, and 3 for the second block.

Next, the fixed codebook vectors of the two subframes are built by random selection of the grid, the pulses signs and positions, corresponding to the higher rate fixed excitation pattern.

Then a unique fixed excitation gain is computed for the two subframes of the block.

The adaptive excitation vector on the current block is noted $u[n], n = 0$ to 119 and the fixed excitation $v[n], n = 0$ to 119.

The fixed excitation gain is obtained by calculating the value Gf that yields a block average energy the closest to the target energy \tilde{G}_t^2 :

$$\text{select } Gf \text{ such that } \left| \frac{1}{20} \sum_{n=0}^{119} (u[n] + Gf \times v[n])^2 - \tilde{G}_t^2 \right| \text{ minimum} \quad (\text{A-19})$$

Notice that Gf can take a negative value.

Let us define $C(X) = aX^2 + 2bX + c$ such that:

$$a = \left(\sum_{n=0}^{119} v[n]^2 \right), b = \left(\sum_{n=0}^{119} u[n]v[n] \right), c = \left(\sum_{n=0}^{119} u[n]^2 - 120\tilde{G}_t^2 \right)$$

The equation $C(X) = 0$ is then studied:

If the discriminant is ≤ 0 then $Gf = -\frac{b}{a}$ is selected, else the two roots are calculated and the one with the lowest absolute value is selected.

Then Gf is bounded: $Gf \leq 5000$

Finally the block CNG excitation is built, using:

$$e[n] = u[n] + Gf \times v[n], n = 0 \text{ to } 119 \quad (\text{A-20})$$

A.4.6 Interpolation of LSPs and update

File: COD_CNG.C	Procedure: Cod_Cng()	COD_CNG main body
File: LSP.C	Procedure: Lsp_Int()	LSP Interpolator

Both for SID frames and for untransmitted frames, the interpolated sets of LPC coefficients are calculated using \tilde{p}_{sid} and the previous LSP vector \tilde{p}_{t-1} provided to COD-CNG.

The LSP update is also performed: $\tilde{p}_t = \tilde{p}_{sid}$.

A.4.7 COD-CNG initialization

The following initialization must be performed on the frame autocorrelation functions, the target excitation gain, VAD information, and seed of the random generator used to compute the CNG excitation:

$$\left\{ \begin{array}{l} R^k[j] = 0 \text{ for } j = 0, \dots, 10 \text{ and } k = -1, -2, -3 \\ \tilde{G}_{-1} = 0 \\ Vad_{-1} = 1 \\ rseed = 12345 \end{array} \right.$$

No initialization is needed for the other static variables of COD-CNG.

A.5 Description of the decoder part

At the receiving end, DEC-CNG processes SID frames and untransmitted frames to produce the synthesized comfort noise.

The procedures developed to deal with frame erasures are described next.

A.5.1 Description of DEC-CNG

File: DEC_CNG.C	Procedure: Dec_Cng()	DEC_CNG main body
File: UTIL_CNG.C	Procedure: Calc_Exc_Rand()	Computation of the excitation
File: UTIL_CNG.C	Procedure: random_number()	Random number generation
File: UTIL_CNG.C	Procedure: distG()	Used to select excitation Gain
File: UTIL_LBC.C	Procedure: Sqrt_lbc()	Square root
File: UTIL_LBC.C	Procedure: Rand_lbc()	Pseudo-random sequence
File: LSP.C	Procedure: Lsp-Inq()	LSP inverse quantization
File: LSP.C	Procedure: Lsp_Int()	LSP Interpolator
File: UTIL_CNG.C	Procedure: Qua_SidGain()	Quantize Sid Gain
File: UTIL_CNG.C	Procedure: Dec_SidGain()	Decode Sid Gain

Figure A4 provides a general description of the comfort noise generation at the decoder part.

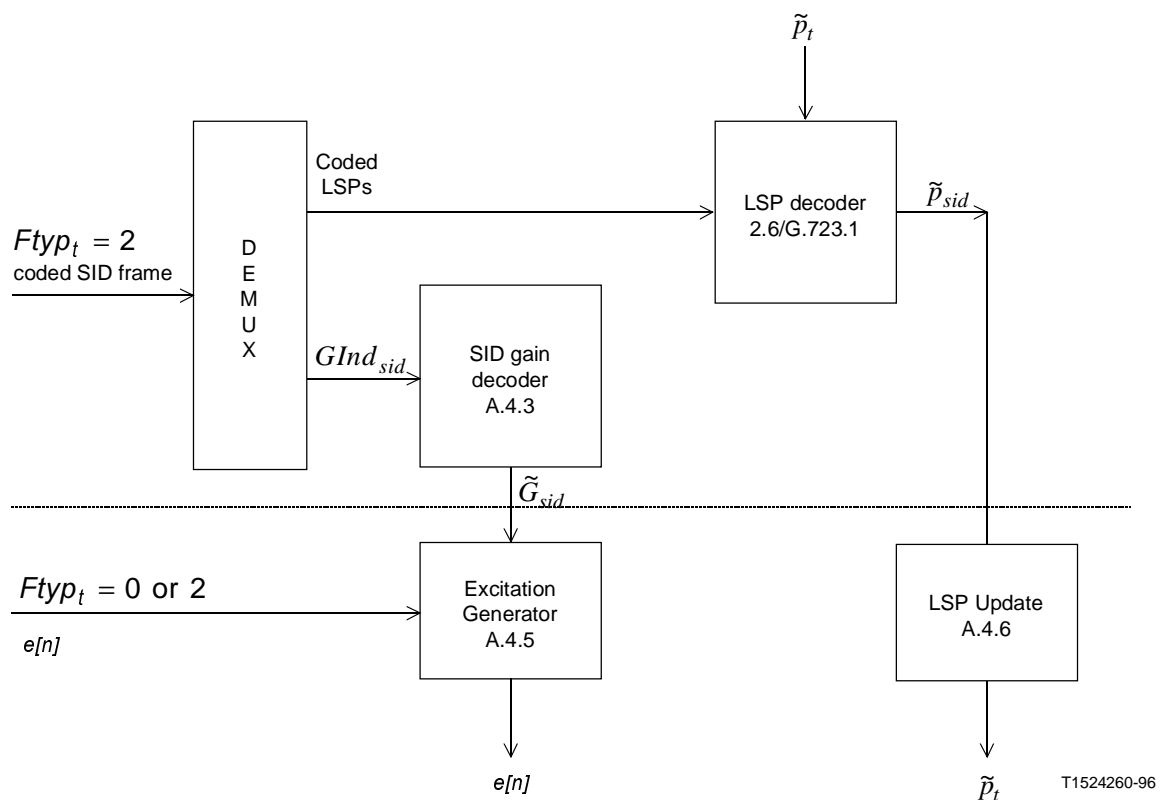


FIGURE A.4/G.723.1

Block diagram of the CNG at the decoder part

When the decoder receives an SID frame, DEC-CNG decodes the SID parameters.

Both in case of SID and untransmitted frames, the module DEC-CNG uses the decoded SID parameters to compute the LSPs and the excitation of the comfort noise that will be synthesized by the decoder synthesis module.

The CNG type of frame information $Ftyp_t$ (for frame t) provided at the receiver is the same as the value computed by COD-CNG at the encoder.

- When $Ftyp_t = 2$, the parameters of the SID frame are decoded: \tilde{p}_{sid} for the LSPs and \tilde{G}_{sid} for the decoded gain.
- When $Ftyp_t = 0$, $Ftyp_{t-1}$ is tested to verify that the SID information has not been erased (see A.5.2). If $Ftyp_{t-1} = 1$, an energy term Enr , that has been calculated by the G.723.1 decoder during the processing of the last valid frame is quantized and decoded using the same procedure as the average energy in 3.1 of Recommendation G.723.1 except that there is no scaling factor α_w . The decoded value is the restored \tilde{G}_{sid} .

Then, in both cases, the CNG excitation is calculated according to the procedure described in COD-CNG in A.4.5. The new LSP vector, \tilde{p}_{sid} , is used to compute the interpolated LPC coefficients, and the LSP updating is performed: $\tilde{p}_t = \tilde{p}_{sid}$

A.5.2 Frame erasure concealment with regards to the CNG

File: DECOD.C	Procedure: Decod()	Frame decoding
---------------	--------------------	----------------

When a frame erasure is detected by the decoder, the erased frame type depends on the preceding frame type:

- if the preceding frame was active, then the current erased frame is considered as active,
- else if the preceding frame was either an SID frame or an untransmitted frame, the current erased frame is considered as untransmitted:

$$\begin{cases} Ftyp_{t-1} = 1 & \Rightarrow & Ftyp_t = 1 \\ Ftyp_{t-1} = 0 \text{ or } 2 & \Rightarrow & Ftyp_t = 0 \end{cases} \quad (\text{A-21})$$

If an untransmitted frame has been erased, no error is then introduced.

If a SID frame is erased, there are two possibilities:

- If it is not the first SID frame of the current inactive period, then the previous SID parameters are kept.
- If it is the first SID frame of an inactive period, a special protection has been taken:

As stated in A5.1, this case is detected by the fact that $Ftyp_{t-1} = 1$ and $Ftyp_t = 0$.

This combination of events does not imply that the preceding frame was a good active frame: several frames up to the preceding one may have been erased. What is certain is that the last good frame was an active frame, that the present frame was not erased, and that the SID frame supposed to provide information for the current untransmitted frame is lost.

To recover the SID information, DEC-CNG uses parameters provided by the G.723.1 decoder main part:

- The LSPs of the last valid active frame are used for \tilde{p}_{sid} .
- The energy term Enr calculated by the decoder during the the residual interpolation procedure (see 3.10.2 of Recommendation G.723.1) over the 120 last excitation samples of

the last valid active frame is used to recover \tilde{G}_{sid} , according to the method described in A.5.1.

Finally, to avoid de-synchronization of the random generator used to compute the excitation, the pseudo-random sequence reset is performed at each active frame, both at the encoder and coder part: $rseed = 12345$.

A.5.3 DEC-CNG initialization

Only the following variables must be initialized:

$$\begin{cases} \tilde{G}_{sid} = 0 \\ \tilde{P}_{sid} = LSP \ DC \ vector \ P_{DC} \\ Vad_{-1} = 1 \\ rseed = 12345 \end{cases}$$

A.6 Bit stream packing

Table A.1 shows the bit stream of the SID frames according to the notations used in clause 4 of Recommendation G.723.1.

TABLE A.1/G.723.1
Bit packing for SID frames

Transmitted octets	PARx_By, ...
1	LPC_B5 ... LPC_B0, VADFLAG_B0, RATEFLAG_B0
2	LPC_B13 ... LPC_B6
3	LPC_B21 ... LPC_B14
4	GAIN_B5 ... GAIN_B0, LPC_B23, LPC_B22

A.7 Glossary

$a_{no}[j]$	noise LPC filter coefficients
L_{OL}^j	preceding frame and current frame open loop pitch delays
pc	pitch delays counter for voicing estimation
Aen_t	adaptation enable flag
$e_t'[n]$	noise-inverse filtered input signal for frame t
Enr_t	noise-inverse filtered input signal energy for frame t
$Nlev_t$	noise level at frame t
$Nlev_{min}$	minimum bound on $Nlev_t$
$Nlev_{max}$	maximum bound on $Nlev_t$
Thr	adapted threshold for VAD decision
$k_i^t[2]$	second reflection coefficient for subframe i in frame t

$SinD$	sine wave detection flag (1: sine detected, 0: else)
$e[n]$	decoded combined excitation vector
$R_t[j]$	autocorrelation function for subframe i, j = 0,1,...,10
thr_2	threshold for energies distance
$R_a[j]$	modified autocorrelation of LPC coefficients
$Gind_{sid}$	SID gain index
\tilde{G}_{sid}	decoded SID gain
G	excitation gain used at the SID quantizer input
$Gind_t$	gain index for frame t
isg	SID gain quantizer segment index
$N[isg]$	SID gain quantizer segment length
$G_{isg}[j]$	gain decoded values of segment isg , $j=0,1, \dots, N[isg]-1$
i_s	gain index relative to the segment
α_w	energy scaling factor
a_{sid}	SID LPC coefficient vector
\bar{a}_p	past average LPC filter coefficients
$\bar{R}_p[j]$	sum of past autocorrelation functions
\tilde{p}_{sid}	decoded SID LSP vector
$u[n]$	adaptive codebook excitation vector
$v[n]$	fixed codebook excitation vector
\tilde{G}_t	target excitation gain for excitation synthesis
a, b, c	coefficients of energy minimization equation
$C(X)$	energy minimization equation
Gf	fixed codebook gain for CNG excitation synthesis
$rseed$	random generator seed

A.8 Bit-exact, fixed-point C source code

All details of the silence compression algorithm are included as part of bit-exact, fixed-point ANSI C source code. In the event of any discrepancy between the above descriptions and the C source, the C source code is presumed to be correct. This C source code is a part of the code distributed by the ITU-T as Recommendation G.723.1.

ITU-T RECOMMENDATIONS SERIES

Series A	Organization of the work of the ITU-T
Series B	Means of expression
Series C	General telecommunication statistics
Series D	General tariff principles
Series E	Telephone network and ISDN
Series F	Non-telephone telecommunication services
Series G	Transmission systems and media
Series H	Transmission of non-telephone signals
Series I	Integrated services digital network
Series J	Transmission of sound-programme and television signals
Series K	Protection against interference
Series L	Construction, installation and protection of cables and other elements of outside plant
Series M	Maintenance: international transmission systems, telephone circuits, telegraphy, facsimile and leased circuits
Series N	Maintenance: international sound-programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	Telephone transmission quality
Series Q	Switching and signalling
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminal equipment and protocols for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks and open system communication
Series Z	Programming languages