INTERNATIONAL TELECOMMUNICATION UNION

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

G.729

Annex A
(11/96)

SERIES G: TRANSMISSION SYSTEMS AND MEDIA

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

Coding of speech at 8 kbit/s using conjugate
structure algebraic-code-excited linear-prediction
(CS-ACELP)

**Annex A: Reduced complexity 8 kbit/s
CS-ACELP speech codec**

ITU-T Recommendation G.729 – 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
DIGITAL SECTIONS AND DIGITAL LINE SYSTEM	G.900–G.999

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

ITU-T RECOMMENDATION G.729 – ANNEX A

REDUCED COMPLEXITY 8 KBIT/S
CS-ACELP SPEECH CODEC

Source

Annex A to ITU-T Recommendation G.729, 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.

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.

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/had not 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.

CONTENTS

Recommendation G.729 – Annex A

<div align="center">

REDUCED COMPLEXITY 8 kbit/s
CS-ACELP SPEECH CODEC

(Geneva, 1996)

</div>

A.1   Introduction

This annex provides the high level description of a reduced complexity version of the G.729 speech codec. This version is bit stream interoperable with the full version, i.e. a reduced complexity encoder may be used with a full implementation of the decoder, and vice versa. However, implementors of the codec defined in this annex should be aware that the performance of this codec may not be as good as the full implementation of Recommendation G.729 in certain circumstances.

The reduced complexity version of the codec has been developed for multimedia simultaneous voice and data applications, although the use of the codec is not limited to these applications.

The description of the codec is similar to that of the full implementation of Recommendation G.729. This annex describes the changes to the full implementation which have been made in order to reduce the codec algorithmic complexity. For those parts of the algorithm which have not been changed, this annex refers to the appropriate section of the main Recommendation.

A.2   General description of the codec

The general description of the coding/decoding algorithm is similar to that of the full version. The bit allocation is the same as that given in Table 1/G.729. It has also the same delay (speech frame of 10 ms and lookahead of 5 ms). The major algorithmic changes to the full version of Recommendation G.729 are summarized below:

- The perceptual weighting filter uses the quantized LP filter parameters and it is given by $W(z) = \hat{A}(z)/\hat{A}(z/\gamma)$ with a fixed value of $\gamma = 0.75$.

- Open-loop pitch analysis is simplified by using decimation while computing the correlations of the weighted speech.

- Computation of the impulse response of the weighted synthesis filter $W(z)/\hat{A}(z)$, computation of the target signal, and updating the filter states are simplified since $W(z)/\hat{A}(z)$ is reduced to $1/\hat{A}(z/\gamma)$.

- The search of the adaptive codebook is simplified. The search maximizes the correlation between the past excitation and the backward filtered target signal (the energy of filtered past excitation is not considered).

- The search of the fixed algebraic codebook is simplified. Instead of the nested-loop focused search, an iterative depth-first tree search approach is used.

- At the decoder, the harmonic postfilter is simplified by using only integer delays.

These changes are described in more detail in A.3 and A.4.

TABLE A.1/G.729

Summary of the principle routines which have been changed

G.729 routine name	G.729A routine name
Coder_1d8k ( )	Coder_1d8a ( )
Decod_1d8k ( )	Decod_1d8a ( )
Pitch_ol ( )	Pitch_ol_fast ( )
Pitch_fr3 ( )	Pitch_fr3_fast ( )
ACELP_Codebook ( )	ACELP_Code_A ( )
Post ( )	Post-Filter ( )

A.2.1   Speech codec definition

The description of the reduced complexity speech codec is made in terms of bit-exact, fixed-point mathematical operations. The ANSI-C code indicated in A.5, which constitutes an integral part of this annex, reflects this bit-exact, fixed-point descriptive approach. The mathematical description of the encoder (see A.3) and the decoder (see A.4), can be implemented in several other fashions, possibly leading to a codec implementation not complying with this annex. Therefore, the algorithm description of the ANSI-C code of A.5 shall take precedence over the mathematical descriptions of A.3 and A.4 whenever discrepancies are found. A non-exhaustive set of test signals, which can be used with the ANSI-C code, are available from the ITU.

A.2.2   Notational conventions

Notational conventions are the same as those given in 2.5/G.729.

A.3   Functional description of the encoder

In this subclause the different functions of the encoder represented in the blocks of Figure 4/G.729 are described. The main body of the Recommendation is referred to in most of this subclause except the parts where algorithmic simplifications have been carried out.

A.3.1   Pre-processing

Same as 3.1/G.729.

A.3.2   Linear prediction analysis and quantization

A.3.2.1   Windowing and autocorrelation computation

Same as 3.2.1/G.729.

A.3.2.2   Levinson-durbin algorithm

Same as 3.2.2/G.729.

A.3.2.3   LP to LSP conversion

Same as 3.2.3/G.729 with some simplifications. The number of points at which the polynomials $F_1(z)$ and $F_2(z)$ are evaluated, is reduced to 50 (instead of 60), and the sign change interval is divided two times instead of four times for tracking the root of the polynomial.

A.3.2.4   Quantization of the LSP coefficients

Same as 3.2.4/G.729.

A.3.2.5 Interpolation of the LSP coefficients

Same as 3.2.5/G.729, but only the quantized LP coefficients are interpolated since the weighting filter uses the quantized parameters for simplicity.

A.3.2.6 LSP to LP conversion

Same as 3.2.6/G.729.

A.3.3 Perceptual weighting

Unlike 3.3/G.729, the perceptual weighting filter is based on the quantized LP filter coefficients $\hat{a}_i$, and is given by:

$$W(z) = \frac{\hat{A}(z)}{\hat{A}(z/\gamma)} \tag{A.1}$$

with $\gamma = 0.75$. This simplifies the combination of synthesis and weighting filters to $W(z)/\hat{A}(z) = 1/\hat{A}(z/\gamma)$, which reduces the number of filtering operations while computing the impulse response and the target signal and while updating the filter states. Note that the value of γ is fixed to 0.75 and the procedure for the adaptation of the factors of the perceptual weighting filter described in 3.3/G.729 is not used in this reduced complexity version.

The weighted speech signal is not used for computing the target signal since an alternative approach is used (see A.3.6). However, the weighted speech signal (low-pass filtered) is used to compute the open-loop pitch estimate. The low-pass filtered weighted speech is found by filtering the speech signal $s(n)$ through the filter $\hat{A}(z)/[\hat{A}(z/\gamma)(1 - 0.7z^{-1})]$. First the coefficients of the filter $A'(z) = \hat{A}(z/\gamma)(1 - 0.7z^{-1})$ are computed, then the low-pass filtered weighted speech in a subframe is computed by:

$$S_w(n) = r(n) - \sum_{i=1}^{10} a_i' s_w(n-i), \qquad n = 0,...,39 \tag{A.2}$$

where $r(n)$ is the LP residual signal given by:

$$r(n) = s(n) + \sum_{i=1}^{10} \hat{a}_i s(n-i), \qquad n = 0,...,39 \tag{A.3}$$

The signal $s_w(n)$ is used to find an estimation of the pitch delay in the speech frame.

A.3.4 Open-loop pitch analysis

To reduce the complexity of the search for the best adaptive-codebook delay, the search range is limited around a candidate delay T_{op}, obtained from an open-loop pitch analysis. This open-loop pitch analysis is done once per frame (10 ms). The open-loop pitch estimation uses the low-pass filtered weighted speech signal $s_w(n)$ of equation (A.2), and is done as follows: in the first step, 3 maxima of the correlation:

$$R(k) = \sum_{n=0}^{39} s_w(2n) s_w(2n-k) \tag{A.4}$$

are found in the following three ranges:

$i = 1$: 20,...,39

$i = 2$: 40,...,79

$$i = 3: \quad 80,...,43$$

The retained maxima $R(t_i)$, $i = 1, ..., 3$, are normalized through:

$$R'(t_i) = \frac{R(t_i)}{\sqrt{\sum_{n=0}^{39} s_w^2 (2n - t_i)}}, \quad i = 1,...,3$$

(A.5)

The winner among the three normalized correlations is selected by favouring the delays with the values in the lower range. This is done by augmenting the normalized correlations corresponding to the lower delay range if their delays are submultiples of the delays in the higher delay range.

Note that in computing the correlations in equation (A.4) only the even samples are used. Further, in the third delay region [80, 143] only the correlations at the even delays are computed in the first pass, then the delays at ± l of the selected even delay are tested.

A.3.5 Computation of the impulse response

The impulse response $h(n)$ of the weighted synthesis filter $W(z)/\hat{A}(z)$ *is* needed for the search of adaptive and fixed codebooks. The impulse response $h(n)$ is computed for each subframe by filtering a signal consisting of a unit sample extended by zeros through the filter $1/\hat{A}(z/\gamma)$.

A.3.6 Computation of the target signal

The target signal $x(n)$ for the adaptive-codebook search is computed by filtering of the LP residual signal $r(n)$ through the weighted synthesis filter $1/\hat{A}(z/\gamma)$. After determining the excitation for the subframe, the initial states of this filter are updated as explained in A.3.10.

The residual signal $r(n)$, which is needed for finding the target vector is also used in the adaptive-codebook search to extend the past excitation buffer. The computation of the LP residual is given in equation (A.3).

A.3.7 Adaptive-codebook search

The adaptive-codebook structure is the same as in 3.7/G.729. In the first subframe, a fractional pitch delay T_1 is used with a resolution of 1/3 in the range $\left[19\frac{1}{3}, 84\frac{2}{3}\right]$ and integers only in the range [85, 143]. For the second subframe, a delay T_2 with a resolution of 1/3 is always used in the range $\left[int(T_1) - 5\frac{2}{3}, int(T_1) + 4\frac{2}{3}\right]$, where *int(T₁) is* the integer part of the fractional pitch delay T_1 of the first subframe. This range is adapted for the cases where T_1 straddles the boundaries of the delay range.

The search boundaries t_{min} and t_{max} for both subframes are determined in the same way as in 3.7/G.729.

Closed-loop pitch search is usually performed by maximizing the term:

$$R(k) = \frac{\sum\limits_{n=0}^{39} x(n) y_k(n)}{\sqrt{\sum\limits_{n=0}^{39} y_k(n) y_k(n)}}$$

(A.6)

where $x(n)$ is the target signal and $y_k(n)$ is the past filtered excitation at delay k [past excitation convolved with $h(n)$]. In order to simplify the search in this reduced complexity version, only the numerator in equation (A.6) is maximized. That is, the term:

$$R_N(k) = \sum\limits_{n=0}^{39} x(n) y_k(n) = \sum\limits_{n=0}^{39} x_b(n) u_k(n)$$

(A.7)

is maximized, where $x_b(n)$ is the backward filtered target signal (correlation between $x(n)$ and the impulse response $h(n)$) and $u_k(n)$ is the past excitation at delay k ($u(n - k)$). Note that the search range is limited around a preselected value, which is the open-loop pitch T_{op} for the first subframe, and T_1 for the second subframe.

Note that in the search stage, the samples $u(n)$, $n = 0,...,39$ are not known, and they are needed for pitch delays less than 40. To simplify the search, the LP residual is copied to $u(n)$.

For the determination of T_2 and T_1 if the optimum integer delay is less than 85, the fractions around the optimum integer delay have to be tested. The fractional pitch search is done by interpolating the past excitation at fractions $-\frac{1}{3}$, 0 and $\frac{1}{3}$, and selecting the fraction which maximizes the correlation in equation (A.7). The interpolation of the past excitation is performed using the same FIR filter, b_{30}, which is defined in 3.7/G.729. The interpolated past excitation at a given integer delay k and fraction t is given by:

$$u_{kt}(n) = \sum\limits_{i=0}^{9} u(n - k + i) b_{30}(t + 3i) + \sum\limits_{i=0}^{9} u(n - k + 1 + i) b_{30}(3 - t + 3i), \quad n = 0,...,39, \ t = 0, 1, 2$$

(A.8)

A.3.7.1   Generation of the adaptive-codebook vector

Same as 3.7.1/G.729.

A.3.7.2   Codeword computation for adaptive-codebook delays

Same as 3.7.2/G.729.

A.3.7.3   Computation of the adaptive-codebook gain

Same as 3.7.3/G.729.

A.3.8   Fixed codebook – Structure and search

The structure of the 17-bit algebraic codebook is the same as 3.8/G.729.

A.3.8.1   Fixed-codebook search procedure

The signs of the pulses are found using the same approach explained in 3.8.1/G.729. However, the pulse positions are found using a more efficient approach. Instead of the nested-loop search approach, an iterative depth-first, tree search approach is used. In this new approach a smaller number of pulse position combinations is tested and it has fixed complexity.

A.3.8.2 Codeword computation of the fixed codebook

Same as 3.8.2/G.729.

A.3.9 Quantization of the gains

Same as 3.9/G.729.

A.3.10 Memory update

An update of the states of the weighted synthesis filter is needed for computing the target signal in the next subframe. After the two gains are quantized, the excitation signal, $u(n)$, in the present subframe is obtained using:

$$u(n) = \hat{g}_p v(n) + \hat{g}_c c(n), \qquad n = 0,...,39$$

(A.9)

where $\hat{g}_p$ and $\hat{g}_c$, are the quantized adaptive and fixed-codebook gains, respectively, $v(n)$ *is* the adaptive-codebook vector (interpolated past excitation), and $c(n)$ *is* the fixed-codebook vector including harmonic enhancement. The states of the weighted synthesis filter can be updated by filtering the signal $r(n) - u(n)$ (difference between residual and excitation) through the filter $1/\hat{A}(z/\gamma)$ for the 40 sample subframe and saving the states of the filter. A simpler approach, which requires no filter operations, is as follows. The output of the filter due to the input $r(n) - u(n)$ is the weighted error signal $e_w(n)$ which can be found by:

$$e_w(n) = x(n) - \hat{g}_p y(n) - \hat{g}_c z(n)$$

(A.10)

where $x(n)$ is the target signal, $y(n)$ is the filtered adaptive-codebook vector and $z(n)$ is the filtered fixed-codebook vector. Since the signals $x(n)$, $y(n)$, and $z(n)$ are available, the states of the weighted synthesis filter are updated by computing $e_w(n)$ as in equation (A.10) for $n = 30,...,39$.

A.4 Functional description of the decoder

The principle of the decoder is shown in Figure 3/G.729. The transmitted parameters are the same as listed in Table 8/G.729. The decoded parameters are used to compute the reconstructed speech signal. This reconstructed signal is enhanced by a post-processing operation consisting of a postfilter, a high-pass filter and an up scaling (see A.4.2). The detailed signal flow diagram of the decoder is the same one shown in Figure 6/G.729.

The only change in the decoder is in the postfilter which is described in A.4.2.

A.4.1 Parameter decoding procedure

Same as 4.1/G.729.

A.4.2 Post-processing

The post-processing is the same as in 4.2/G.729 except some simplification in the adaptive postfilter.

The adaptive postfilter is the cascade of three filters: a long-term postfilter $H_p(z)$, a short-term postfilter $H_f(z)$ and a tilt compensation filter $H_t(z)$, followed by an adaptive gain control procedure. The long-term postfilter is simplified by using only integer delays. In the short-term postfilter and the tilt compensation filter, the gain terms g_f and g_t are not used.

The postfiltering process is similar to that described in Recommendation G.729 with the exception that the compensation filtering is performed before synthesis filtering through $1/\hat{A}(z/\gamma_d)$.

A.4.2.1   Long-term postfilter

The long-term postfilter is given by:

$$H_p(z) = \frac{1}{1+\gamma_p g l}\left(1+\gamma_p g l z^{-T}\right)$$

(A.11)

The only difference from 4.2.1/G.729 is that the long -term delay T is always an integer delay and it is computed by searching the range $[T_{cl} - 3, T_{cl} + 3]$, where T_{cl} is the integer part of the (transmitted) pitch delay in the current subframe bounded by $T_{cl} \le 140$.

A.4.2.2   Short-term postfilter

The short-term postfilter is given by:

$$H_f(z) = \frac{\hat{A}(z/\gamma_n)}{\hat{A}(z/\gamma_d)} = \frac{1+\sum_{i=1}^{10}\gamma_n^i\hat{a}_i z^{-i}}{1+\sum_{i=1}^{10}\gamma_d^i\hat{a}_i z^{-i}}$$

(A.12)

where $\hat{A}(z)$ is the received quantized LP inverse filter (LP analysis is not done at the decoder), and the factors γ_n and γ_d control the amount of short-term postfiltering, and are set to $\gamma_n = 0.55$ and $\gamma_d = 0.7$.

The only difference from 4.2.2/G.729 is that the gain factor g_f is eliminated.

A.4.2.3   Tilt compensation

The filter $H_t(z)$ compensates for the tilt in the short-term postfilter $H_f(z)$ and is given by:

$$H_t(z) = 1+\gamma_t k_1' z^{-1}$$

(A.13)

where $\gamma_t k_1'$ is a tilt factor, k_1' being the first reflection coefficient calculated by:

$$k_1' = -\frac{r_h(1)}{r_h(0)}; \qquad r_h(i) = \sum_{j=0}^{21-i} h_f(j)h_f(j+i)$$

(A.14)

where $h_f(n)$ is the truncated impulse response of the filter $\hat{A}(z/\gamma_n)/\hat{A}(z/\gamma_d)$. The value of $\gamma_t = 0.8$ is used if $k_1' < 0$ and γ_t is set to zero if $k_1' \ge 0$. The gain factor g_t which is used in 4.2.3/G.729 is eliminated.

A.4.2.4   Adaptive gain control

Same as 4.2.4/G.729. The only difference is that the gain scaling factor G for the present subframe is computed by:

$$G = \sqrt{\frac{\sum_{n=0}^{39}\hat{s}^2(n)}{\sum_{n=0}^{39}sf^2(n)}}$$

(A.15)

and $g^{(n)}$ is given by:

$$g^{(n)} = 0.9\,g^{(n-1)} + 0.1G, \qquad n = 0, \dots, 39$$

A.4.2.5   High-pass filtering and up-scaling

Same as 4.2.5/G.729.

A.4.3   Encoder and decoder initialization

Same as 4.3/G.729.

A.4.4   Concealment of frame erasures

Same as 4.4/G.729 with the difference that no voicing detection is used. The excitation is always the addition of both adaptive and fixed codebook contributions.

A.5   Bit-exact description of the reduced complexity CS-ACELP codec

The reduced complexity CS-ACELP codec is simulated in 16-bit fixed-point ANSI-C code using the same set of fixed-point basic operators defined in Table 11/G.729.

A.5.1   Use of the simulation software

Same as 5.1/G.729.

A.5.2   Organization of the simulation software

Same as 5.2G.729.

The tables used by the simulation codec are found in the file **tab_1d8a. c** which replaces the file **tab_1d8k. c** of the full Recommendation. The difference between these two files is that the tables **tab_hup_s, tab_hup_1**, and **inter_3** found in the file **tab_1d8k. c** are removed from the file **tab_1d8a. c**. Also, the table **grid** has been modified.

The main programs use a library of routines that are provided in the fixed-point ANSI-C simulation. Most of the routines are the same as those of the full Recommendation. The principal routines that have been changed are summarized in Table A.1. Refer to the **read.me** file provided with the software for more details.

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 equipments 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