SLOVENSKI STANDARD
SIST ETS 300 969 E1:2003

01-december-2003

Digitalni celiļ ni telekomunikacijski sistem ï Govor s pol hitrosti ï Prekodiranje pri govoru s pol hitrosti (GSM 06.20, razliļ ica 5.0.1)

Digital cellular telecommunications system; Half rate speech; Half rate speech transcoding (GSM 06.20 version 5.0.1)

Ta slovenski standard je istoveten z: ETS 300 969 Edition 1

ICS:

33.070.50	Globalni sistem za mobilno telekomunikacijo (GSM)	Global System for Mobile Communication (GSM)

SIST ETS 300 969 E1:2003 **en**

EUROPEAN TELECOMMUNICATION STANDARD

ETS 300 969

May 1997

Source: ETSI TC-SMG

Reference: DE/SMG-110620Q

ICS: 33.020

Key words: Digital cellular telecommunications system, Global System for Mobile communications (GSM), CODEC, GSM, speech

GLOBAL SYSTEM FOR MOBILE COMMUNICATIONS

Digital cellular telecommunications system; Half rate speech; Half rate speech transcoding (GSM 06.20 version 5.0.1)

ETSI

European Telecommunications Standards Institute

ETSI Secretariat

Postal address: F-06921 Sophia Antipolis CEDEX - FRANCE
Office address: 650 Route des Lucioles - Sophia Antipolis - Valbonne - FRANCE
X.400: c=fr, a=atlas, p=etsi, s=secretariat - **Internet:** secretariat@etsi.fr

Tel.: +33 4 92 94 42 00 - Fax: +33 4 93 65 47 16

Contents

Foreword

This European Telecommunication Standard (ETS) has been produced by the Special Mobile Group (SMG) Technical Committee of the European Telecommunications Standards Institute (ETSI).

This ETS specifies the speech codec to be used for the GSM half rate channel for the digital cellular telecommunications system. This ETS is part of ETSs series covering the half rate speech traffic channels as described below:

GSM 06.02	ETS 300 966: "Digital cellular telecommunications system; Half rate speech; Half rate speech processing functions".
GSM 06.06	ETS 300 967: "Digital cellular telecommunications system; Half rate speech; ANSI-C code for the GSM half rate speech codec".
GSM 06.07	ETS 300 968: "Digital cellular telecommunications system; Half rate speech; Test sequences for the GSM half rate speech codec".
GSM 06.20	**ETS 300 969: "Digital cellular telecommunications system; Half rate speech; Half rate speech transcoding".**
GSM 06.21	ETS 300 970: "Digital cellular telecommunications system; Half rate speech; Substitution and muting of lost frames for half rate speech traffic channels".
GSM 06.22	ETS 300 971: "Digital cellular telecommunications system; Half rate speech; Comfort noise aspects for half rate speech traffic channels".
GSM 06.41	ETS 300 972: "Digital cellular telecommunications system; Half rate speech; Discontinuous Transmission (DTX) for half rate speech traffic channels".
GSM 06.42	ETS 300 973: "Digital cellular telecommunications system; Half rate speech; Voice Activity Detector (VAD) for half rate speech traffic channels".

Transposition dates	
Date of adoption:	4 April 1997
Date of latest announcement of this ETS (doa):	31 August 1997
Date of latest publication of new National Standard or endorsement of this ETS (dop/e):	28 February 1998
Date of withdrawal of any conflicting National Standard (dow):	28 February 1998

Blank page

1 Scope

This European Telecommunication Standard (ETS) specifies the speech codec to be used for the GSM half rate channel. It also specifies the test methods to be used to verify that the codec implementation complies with this ETS.

The requirements are mandatory for the codec to be used either in GSM Mobile Stations (MS)s or Base Station Systems (BSS)s that utilize the half rate GSM speech traffic channel.

2 Normative references

This ETS incorporates by dated and undated reference, provisions from other publications. These normative references are cited at the appropriate places in the text and the publications are listed hereafter. For dated references, subsequent amendments to or revisions of any of these publications apply to this ETS only when incorporated in it by amendment or revision. For undated references, the latest edition of the publication referred to applies.

[1] GSM 06.02 (ETS 300 966): "Digital cellular telecommunications system; Half rate speech; Half rate speech processing functions".

[2] GSM 06.06 (ETS 300 967): "Digital cellular telecommunications system; Half rate speech; ANSI-C code for the GSM half rate speech codec".

[3] GSM 06.07 (ETS 300 968): "Digital cellular telecommunications system; Half rate speech; Test sequences for the GSM half rate speech codec".

3 Definitions, symbols and abbreviations

3.1 Definitions

For the purposes of this ETS, the following definitions apply:

adaptive codebook: The adaptive codebook is derived from the long term filter state. The lag value can be viewed as an index into the adaptive codebook.

adaptive pitch prefilter: In the GSM half rate speech decoder, this filter is applied to the excitation signal to enhance the periodicity of the reconstructed speech. Note that this is done prior to the application of the short term filter.

adaptive spectral postfilter: In the GSM half rate speech decoder, this filter is applied to the output of the short term filter to enhance the perceptual quality of the reconstructed speech.

allowable lags: The set of lag values which may be coded by the GSM half rate speech encoder and transmitted to the GSM half rate speech decoder. This set contains both integer and fractional values (see table 3).

analysis window: For each frame, the short term filter coefficients are computed using the high pass filtered speech samples within the analysis window. The analysis window is 170 samples in length, and is centred about the last 100 samples in the frame.

basis vectors: A set of M, M1, or M2 vectors of length Ns used to generate the VSELP codebook vectors. These vectors are not necessarily orthogonal.

closed loop lag search: A process of determining the near optimal lag value from the weighted input speech and the long term filter state.

closed loop lag trajectory: For a given frame, the sequence of near optimal lag values whose elements correspond to each of the four subframes as determined by the closed loop lag search.

codebook: A set of vectors used in a vector quantizer.

codeword (OR Code): An M, M1, or M2 bit symbol indicating the vector to be selected from a VSELP codebook.

delta (LAG) code: A four bit code indicating the change in lag value for a subframe relative to the previous subframe's coded lag. For frames in which the long term predictor is enabled (MODE 1, 2, or 3), the lag for subframe 1 is independently coded using eight bits, and delta codes are used for subframes 2, 3, and 4.

direct form coefficients: One of the formats for storing the short term filter parameters. All filters which are used to modify speech samples use direct form coefficients.

fractional lags: A set of lag values having sub-sample resolution. Note that not every fractional lag value considered in the GSM half rate speech encoder is an allowable lag value.

frame: A time interval equal to 20 ms, or 160 samples at an 8 kHz sampling rate.

harmonic noise weighting filter: This filter exploits the noise masking properties of the spectral peaks which occur at harmonics of the pitch frequency by weighting the residual error less in regions near the pitch harmonics and more in regions away from them. Note that this filter is only used when the long term filter is enabled (MODE = 1, 2 or 3).

high pass filter: This filter is used to de-emphasize the low frequency components of the input speech signal.

integer lags: A set of lag values having whole sample resolution.

interpolating filter: An FIR filter used to estimate sub-sample resolution samples, given an input sampled with integer sample resolution.

lag: The long term filter delay. This is typically the pitch period, or a multiple or sub-multiple of it.

long term filter: This filter is used to generate the periodic component in the excitation for the current subframe. This filter is only enabled for MODE = 1, 2 or 3.

LPC coefficients: Linear Predictive Coding (LPC) coefficients is a generic descriptive term for describing the short term filter coefficients.

open loop lag search: A process of estimating the near optimal lag directly from the weighted speech input. This is done to narrow the range of lag values over which the closed loop lag search shall be performed.

open loop lag trajectory: For a given frame, the sequence of near optimal lag values whose elements correspond to the four subframes as determined by the open loop lag search.

reflection coefficients: An alternative representation of the information contained in the short term filter parameters.

residual: The output signal resulting from an inverse filtering operation.

short term filter: This filter introduces, into the excitation signal, short term correlation which models the impulse response of the vocal tract.

soft interpolation: A process wherein a decision is made for each frame to use either interpolated or uninterpolated short term filter parameters for the four subframes in that frame.

soft interpolation bit: A one bit code indicating whether or not interpolation of the short term parameters is to be used in the current frame.

spectral noise weighting filter: This filter exploits the noise masking properties of the formants (vocal tract resonances) by weighting the residual error less in regions near the formant frequencies and more in regions away from them.

subframe: A time interval equal to 5 ms, or 40 samples at an 8 kHz sampling rate.

vector quantization: A method of grouping several parameters into a vector and quantizing them simultaneously.

GSP0 vector quantizer: The process of vector quantization, its intermediate parameters (GS and P0) for the coding of the excitation gains β and γ.

VSELP codebook: Vector-Sum Excited Linear Predictive (VSELP) codebook, used in the GSM half rate speech coder, wherein each codebook vector is constructed as a linear combination of the fixed basis vectors.

zero input response: The output of a filter due to all past inputs, i.e. due to the present state of the filter, given that an input of zeros is applied.

zero state response: The output of a filter due to the present input, given that no past inputs have been applied, i.e. given the state information in the filter is all zeroes.

3.2 Symbols

For the purposes of this ETS, the following symbols apply:

$A(z)$	Short term spectral filter.
α_i	The LPC coefficients.
$b_L(n)$	The output of the long term filter state (adaptive codebook) for lag L.
β	The long term filter coefficient.
$C(z)$	Second weighting filter.
$e(n)$	Weighted error signal
$f_j(i)$	The coefficients of the j^{th} phase of the 10th order interpolating filter used to evaluate candidate fractional lag values; i ranges from 0 to P_f-1.
$g_j(i)$	The coefficients of the j^{th} phase of the 6th order interpolating filter used to interpolate C's and G's as well as fractional lags in the harmonic noise weighting; i ranges from 0 to P_g-1.
γ	The gain applied to the vector(s) selected from the VSELP codebook(s).
H	A M2 bit code indicating the vector to be selected from the second VSELP codebook (when operating in mode 0).
I	A M or M1 bit code indicating the vector to be selected from one of the two first VSELP codebooks.
L	The long term filter lag value.
L_{max}	142 (samples), the maximum possible value for the long term filter lag.
L_{min}	21 (samples), the minimum possible value for the long term filter lag.
M	9, the number of basis vectors, and the number of bits in a codeword, for the VSELP codebook used in modes 1, 2, and 3.
M1	7, the number of basis vectors, and the number of bits in a codeword, for the first VSELP codebook used in mode 0.
M2	7, the number of basis vectors, and the number of bits in a codeword, for the second VSELP codebook used in mode 0.
MODE	A two bit code indicating the mode for the current frame (see annex A).
N_A	170, the length of the analysis window. This is the number of high pass filtered speech samples used to compute the short term filter parameters for each frame.
N_F	160, the number of samples per frame (at a sampling rate of 8 kHz).
N_p	10, the short term filter order.
N_s	40, the number of samples per subframe (at a sampling rate of 8 kHz).
P1	6, the number of bits in the prequantizer for the r1 - r3 vector quantizer.
P2	5, the number of bits in the prequantizer for the r4 - r6 vector quantizer.
P3	4, the number of bits in the prequantizer for the r7 - r10 vector quantizer.
P_f	The order of one phase of an interpolating filter used to evaluate candidate fractional lag values. P_f equals 10 for $j \neq 0$ and equal to 1 for $j = 0$.
P_g	The order of one phase of an interpolating filter, $f_j(n)$, used to interpolate C's and G's as well as fractional lags in the harmonic noise weighting, P_g equals 6.

pitch	The time duration between the glottal pulses which result when the vocal chords vibrate during speech production.
Q1	11, the number of bits in the r1 - r3 reflection coefficient vector quantizer.
Q2	9, the number of bits in the r4 - r6 reflection coefficient vector quantizer.
Q3	8, the number of bits in the r7 - r10 reflection coefficient vector quantizer.
R0	A five bit code used to indicate the energy level in the current frame.
r(n)	The long term filter state (the history of the excitation signal); $n < 0$
$r_L(n)$	The long term filter state with the adaptive codebook output for lag L appended.
s'(n)	Synthesized speech.
W(z)	Spectral weighting filter.
λ_{hnw}	The harmonic noise weighting filter coefficient.
ξ	The adaptive pitch prefilter coefficient.
$\lceil x \rceil$	Ceiling function: the largest integer y where $y < x + 1{,}0$.
$\lfloor x \rfloor$	Floor function: the largest integer y where $y \le x$.
$\sum_{i=j}^{K} x(i)$	Summation: x(j)+x(j+1)+...+x(K).
$\prod_{i=j}^{K} x(i)$	Product: x(j)(x(j+1))....(x(K)).
max(x,y)	Find the larger of two numbers x and y.
min(x,y)	Find the smaller of two numbers x and y.
round(x)	Round the non-integer x to the closest integer $y: y = \lfloor x + 0{,}5 \rfloor$y: y=x+0,5.

3.3 Abbreviations

For the purposes of this ETS, the following abbreviations apply:

AFLAT	Autocorrelation Fixed point LAttice Technique
CELP	Code Excited Linear Prediction
FLAT	Fixed Point Lattice Technique
LTP	Long Term Predictor
SST	Spectral Smoothing Technique
VSELP	Vector-Sum Excited Linear Prediction

4 Functional description of the GSM half rate speech codec

The GSM half rate codec uses the VSELP (Vector-Sum Excited Linear Prediction) algorithm. The VSELP algorithm is an analysis-by-synthesis coding technique and belongs to the class of speech coding algorithms known as CELP (Code Excited Linear Prediction).

The GSM half rate codec's encoding process is performed on a 20 ms speech frame at a time. A speech frame of the sampled speech waveform is read and based on the current waveform and the past history of the waveform, the codec encoder derives 18 parameters that describe it. The parameters extracted are grouped into the following three general classes:

- energy parameters (R0 and GSP0);
- spectral parameters (LPC and INT_LPC);
- excitation parameters (LAG and CODE).

These parameters are quantized into 112 bits for transmission as described in annex A and their order of occurrence over Abis is given in annex B.

The GSM half rate codec is an analysis-by-synthesis codec, therefore the speech decoder is primarily a subset of the speech encoder. The quantized parameters are decoded and a synthetic excitation is generated using the energy and excitation parameters. The synthetic excitation is then filtered to provide the spectral information resulting in the generation of the synthesized speech (see figure 1).

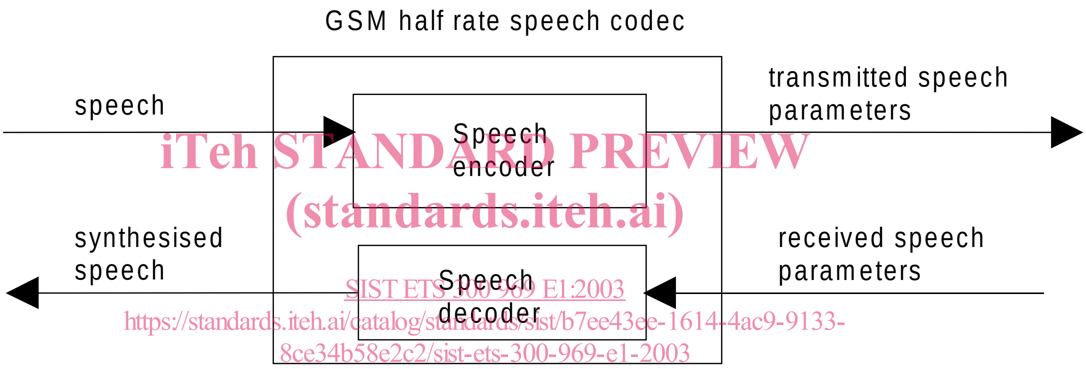

Figure 1: Block diagram of the GSM half rate speech codec

The ANSI-C code that describes the GSM half rate speech codec is given in GSM 06.06 (ETS 300 967) [2] and the test sequences in GSM 06.07 (ETS 300 968) [3] (see clause 5 for the codec homing test sequences).

4.1 GSM half rate speech encoder

The GSM half rate speech encoder uses an analysis by synthesis approach to determine the code to use to represent the excitation for each subframe. The codebook search procedure consists of trying each codevector as a possible excitation for the Code Excited Linear Predictive (CELP) synthesizer. The synthesized speech s'(n) is compared against the input speech and a difference signal is generated. This difference signal is then filtered by a spectral weighting filter, W(z), (and possibly a second weighting filter, C(z)) to generate a weighted error signal, e(n). The power in e(n) is computed. The codevector which generates the minimum weighted error power is chosen as the codevector for that subframe. The spectral weighting filter serves to weight the error spectrum based on perceptual considerations. This weighting filter is a function of the speech spectrum and can be expressed in terms of the α parameters of the short term (spectral) filter.

$$W(z) = \frac{1 - \sum_{i=1}^{N_p} \alpha_i z^{-1}}{1 - \sum_{i=1}^{N_p} \tilde{\alpha}_i z^{-1}} \qquad (1)$$

The computation of the $\tilde{\alpha}_i$ coefficients is described in subclause 4.1.7.

The second weighting filter C(z), if used, is a harmonic weighting filter and is used to control the amount of error in the harmonics of the speech signal. If the weighting filter(s) are moved to both input paths to the subtracter, an equivalent configuration is obtained as shown in figure 2.

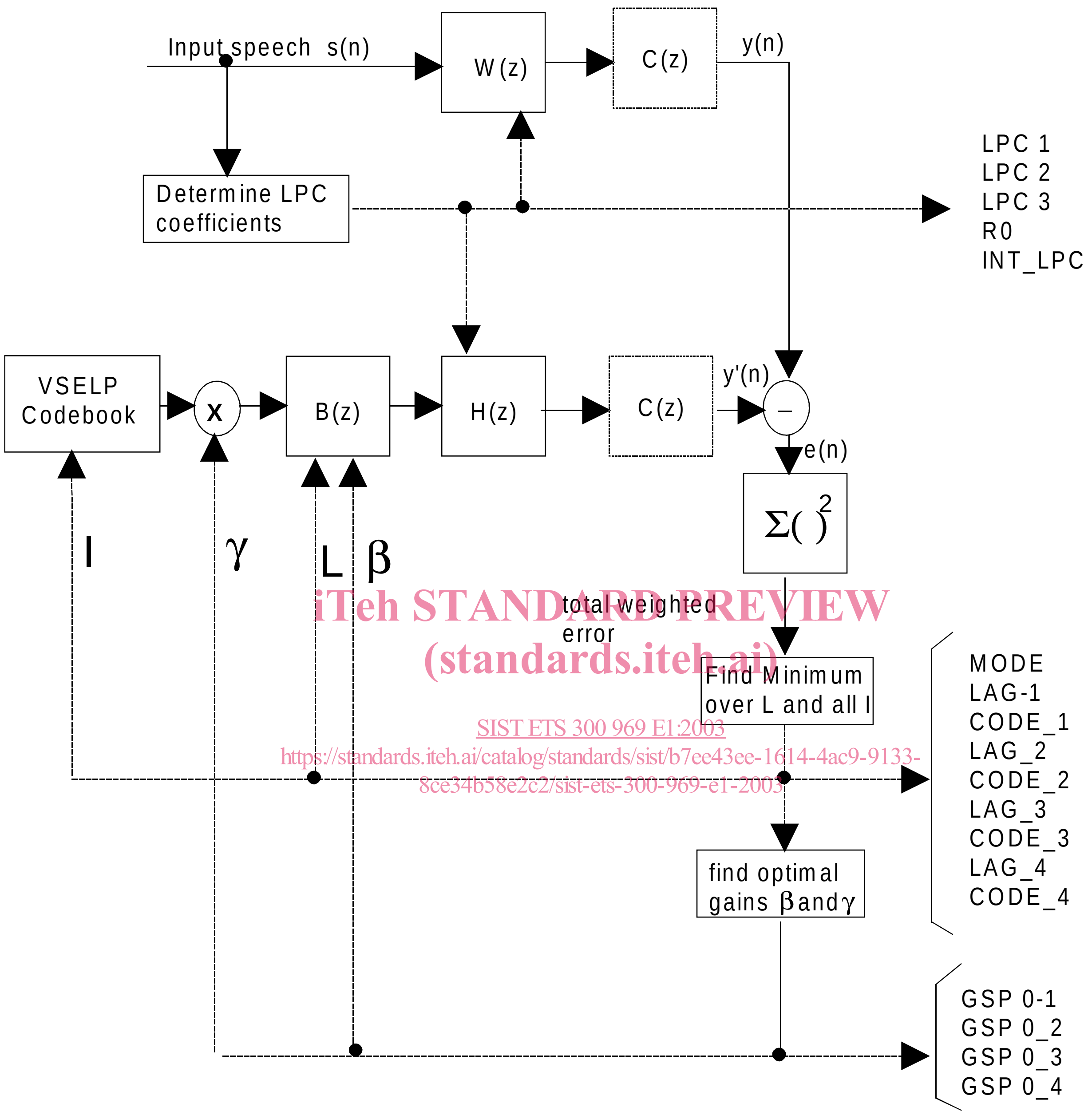

Figure 2: Block diagram of the GSM half rate speech encoder (MODE = 1,2 and 3)

Here H(z) is the combination of A(z), the short term (spectral) filter, and W(z), the spectral weighting filter. These filters are combined since the denominator of A(z) is cancelled by the numerator of W(z).

$$H(z) = \frac{1}{1 - \sum_{i=1}^{N_p} \tilde{\alpha}_i z^{-1}} \qquad (2)$$

There are two approaches that can be used for calculating the gain, γ. The gain can be determined prior to codebook search based on residual energy. This gain would then be fixed for the codebook search. Another approach is to optimize the gain for each codevector during the codebook search. The codevector which yields the minimum weighted error would be chosen and its corresponding optimal gain would be used for γ. The latter approach generally yields better results since the gain is optimized for each codevector. This approach also implies that the gain term needs to be updated at the subframe rate. The optimal code and gain for this technique can be computed as follows:

The input speech is first filtered by a high pass filter as described in subclause 4.1.1. The short term filter parameters are computed from the filtered input speech once per frame. A fast fixed point covariance lattice technique is used. Subclauses 4.1.3 and 4.1.4 describes in detail how the short term parameters are determined and quantized. An overall frame energy is also computed and coded once per frame. Once per frame, one of the four voicing modes is selected. If MODE≠0, the long term predictor is used and the long term predictor lag, L, is updated at the subframe rate. L and a VSELP codeword are selected sequentially. Each is chosen to minimize the weighted mean square error. The long-term filter coefficient, β, and the codebook gain, γ, are optimized jointly. Subclause 4.1.8 describes the technique for selecting from among the voicing modes and, if one of voiced modes is chosen, determining the long-term filter lag. Subclause 4.1.10 describes an efficient technique for jointly optimizing β, γ and the codeword selection. Subclause 4.1.10 also includes the description of the fast VSELP codebook search technique. The β and γ parameters are transformed to equivalent parameters using the frame energy term, and are vector quantized every subframe. The coding of the frame energy and the β and γ parameters is described in subclause 4.1.11.

4.1.1 High-pass filter

The 13 bit linear Pulse Code Modulated (PCM) input speech, $x(n)$, is filtered by a fourth order pole-zero high pass filter. This filter suppresses the frequency components of the input speech which are below 120 Hz. The filter is implemented as a cascade of two second-order Infinite Impulse Response (IIR) filters. Incorporated into the filter coefficients is a gain of 0,5. The difference equation for the first filter is:

$$\tilde{y}(n) = \sum_{i=0}^{2} b_{1,i}\, x(n-i) + \sum_{j=1}^{2} a_{1,j}\, \tilde{y}(n-j) \tag{3}$$

where:

b_{10} = 0,335052
b_{11} = -0,669983 a_{11} = 0,926117
b_{12} = 0,335052 a_{12} = -0,429413

The difference equation for the second filter is:

$$y(n) = \sum_{i=0}^{2} b_{2,i}\, \tilde{y}(n-i) + \sum_{j=1}^{2} a_{2,j}\, y(n-j) \tag{4}$$

where:

b_{20} = 0,335052
b_{21} = -0,669434 a_{21}= 0,965332
b_{22} = 0,335052 a_{22} = -0,469513

4.1.2 Segmentation

A sample buffer containing the previous 195 input high pass filtered speech samples, $y(n)$, is shifted so that the oldest 160 samples are shifted out while the next 160 input samples are shifted in. The oldest 160 samples in the buffer correspond to the next frame of samples to be encoded. The analysis interval comprises the most recent 170 samples in the buffer. The samples in the buffer are labelled as $s(n)$ where $0 \le n \le 194$ and s(0) is the first (oldest) sample.