SLOVENSKI STANDARD
SIST EN 301 708 V7.1.1:2003

01-december-2003

Digitalni celiļ ni telekomunikacijski sistem (faza 2+) ï Detektor govornih dejavnosti (VAD) pri prilagodljivih veļ hitrostnih (AMR) govornih prometnih kanalih ï Sploġni opis (GSM 06.94, razliļ ica 7.1.1, izdaja 1998)

Digital cellular telecommunications system (Phase 2+) (GSM); Voice Activity Detection (VAD) for Adaptive Multi-Rate (AMR) speech traffic channels; General description (GSM 06.94 version 7.1.1 Release 1998)

Ta slovenski standard je istoveten z: **EN 301 708 Version 7.1.1**

ICS:

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

SIST EN 301 708 V7.1.1:2003 **en**

ETSI EN 301 708 V7.1.1 (1999-12)

European Standard (Telecommunications series)

Digital cellular telecommunications system (Phase 2+); Voice Activity Detector (VAD) for Adaptive Multi-Rate (AMR) speech traffic channels; General description (GSM 06.94 version 7.1.1 Release 1998)

GLOBAL SYSTEM FOR MOBILE COMMUNICATIONS

Reference
DEN/SMG-110694Q7

Keywords
Digital cellular telecommunications system,
Global System for Mobile communications (GSM)

ETSI

Postal address
F-06921 Sophia Antipolis Cedex - FRANCE

Office address
650 Route des Lucioles - Sophia Antipolis
Valbonne - FRANCE
Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16
Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Internet
secretariat@etsi.fr
Individual copies of this ETSI deliverable
can be downloaded from
http://www.etsi.org
If you find errors in the present document, send your
comment to: editor@etsi.fr

Important notice

This ETSI deliverable may be made available in more than one electronic version or in print. In any case of existing or perceived difference in contents between such versions, the reference version is the Portable Document Format (PDF). In case of dispute, the reference shall be the printing on ETSI printers of the PDF version kept on a specific network drive within ETSI Secretariat.

Copyright Notification

Contents

Intellectual Property Rights

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (http://www.etsi.org/ipr).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Foreword

This European Standard (Telecommunications series) has been produced by Special Mobile Group (SMG).

The present document specifies two options (Option 1 and Option 2) for the Voice Activity Detector (VAD) to be used in the Discontinuous Transmission (DTX) for Adaptive Multi Rate (AMR) speech traffic channels within the digital cellular telecommunications system. Implementors of mobile station and infrastructure equipment conforming to the AMR specifications can choose which of the two VAD options to implement. There are no interoperability factors associated with this choice.

The contents of the present document is subject to continuing work within SMG and may change following formal SMG approval. Should SMG modify the contents of the present document it will be re-released with an identifying change of release date and an increase in version number as follows:

Version 7.x.y

where:

7 indicates Release 1998 of GSM Phase 2+

x the second digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc.

y the third digit is incremented when editorial only changes have been incorporated in the specification.

National transposition dates	
Date of adoption of this EN:	3 December 1999
Date of latest announcement of this EN (doa):	31 March 2000
Date of latest publication of new National Standard or endorsement of this EN (dop/e):	30 September 2000
Date of withdrawal of any conflicting National Standard (dow):	30 September 2000

1 Scope

The present document specifies two alternatives for the Voice Activity Detector (VAD) to be used in the Discontinuous Transmission (DTX) as described in [3]. Implementors of mobile station and infrastructure equipment conforming to the AMR specifications can choose which of the two VAD options to implement. There are no interoperability factors associated with this choice.

The requirements are mandatory on any VAD to be used either in GSM Mobile Stations (MS)s or Base Station Systems (BSS)s that utilize the AMR speech traffic channel.

2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non-specific.
- For a specific reference, subsequent revisions do not apply.
- For a non-specific reference, the latest version applies.
- A non-specific reference to an ETS shall also be taken to refer to later versions published as an EN with the same number.
- For this Release 1998 document, references to GSM documents are for Release 1998 versions (version 7.x.y).

[1] GSM 06.73: "Digital cellular telecommunications system (Phase 2+); ANSI-C code for the Adaptive Multi Rate (AMR) speech codec".

[2] GSM 06.90: "Digital cellular telecommunications system (Phase 2+); Adaptive Multi Rate (AMR) speech transcoding".

[3] GSM 06.93. "Digital cellular telecommunications system (Phase 2+); Discontinuous transmission (DTX) for Adaptive Multi Rate (AMR) speech traffic channels".

[4] ITU, The International Telecommunications Union, Blue Book, Vol. III, Telephone Transmission Quality, IXth Plenary Assembly, Melbourne, 14-25 November, 1988, Recommendation G.711, Pulse code modulation (PCM) of voice frequencies.

3 Technical Description of VAD Option 1

3.1 Definitions, symbols and abbreviations

3.1.1 Definitions

For the purposes of the presenty document, the following terms and definitions apply:

frame: Time interval of 20 ms corresponding to the time segmentation of the speech transcoder.

3.1.2 Symbols

For the purposes of the present document, the following symbols apply.

3.1.2.1 Variables

bckr_est[n] background noise estimate
burst_count counts length of a speech burst, used by VAD hangover addition
hang_count hangover counter, used by VAD hangover addition
complex_hang_count hangover counter, used by CAD hangover addition
complex_hang_timer hangover initator, used fo Complex Activity Estimation
lagcount pitch detection counter
level[n] signal level
new_speech pointer of the speech encoder, points a buffer containing last received samples of a speech frame [2]
noise_level average level of the background noise estimate
oldlagcount lagcount of the previous frame
pitch flag indicating presence of a periodic signal
complex_warning flag indicating the presence of a complex signal.
best_corr_hp normalized and limited value from maximum HP filtered correlation vector
corr_hp filtered best_corr_hp values
pow_sum power of the input frame
s(i) samples of the input framer
snr_sum measure between input frame and noise estimate
stat_count stationarity counter
stat_rat measure indicating stationary
T_op[n] open-loop lags [2]
t0 autocorrelation maxima calculated by the open-loop pitch analysis [2]
t1 signal power related to the autocorrelation maxima t0 [2]
tone flag indicating the presence of a tone
vad_thr VAD threshold
VAD_flag boolean VAD flag
vadreg intermediate VAD decision
complex_low intermediate complex signal decisions
complex_high intermediate complex signal decisions

3.1.2.2 Constants

ALPHA_UP1 constant for updating noise estimate (see subclause 5.4.2)
ALPHA_DOWN1 constant for updating noise estimate (see subclause 5.4.2)
ALPHA_UP2 constant for updating noise estimate (see subclause 5.4.2)
ALPHA_DOWN2 constant for updating noise estimate (see subclause 5.4.2)
ALPHA3 constant for updating noise estimate (see subclause 5.4.2)
ALPHA4 constant for updating average signal level (see subclause 5.4.2)
ALPHA5 constant for updating average signal level (see subclause 5.4.2)
BURST_LEN_HIGH_NOISE constant for controlling VAD hangover addition (see subclause 5.4.1)
BURST_LEN_LOW_NOISE constant for controlling VAD hangover addition (see subclause 5.4.1)
COEFF3 coefficient for the filter bank (see subclause 5.1)
COEFF5_1 coefficient for the filter bank (see subclause 5.1)
COEFF5_2 coefficient for the filter bank (see subclause 5.1)
HANG_LEN_HIGH_NOISE constant for controlling VAD hangover addition (see subclause 5.4.1)
HANG_LEN_LOW_NOISE constant for controlling VAD hangover addition (see subclause 5.4.1)
HANG_NOISE_THR constant for controlling VAD hangover addition (see subclause 5.4.1)
L_FRAME size of a speech frame, 160
L_NEXT length for the lookahead of the speech encoder, 40
LTHRESH threshold for pitch detection (see subclause 5.2)
NOISE_MAX maximum value for noise estimate (see subclause 5.4.2)
NOISE_MIN minimum value for noise estimate (see subclause 5.4.2)
NTHRESH threshold for pitch detection (see subclause 5.2)
POW_PITCH_THR threshold for pitch detection (see subclause 5.4)
POW_COMPLEX_THR threshold for complex detection (see subclause 5.4)
STAT_COUNT threshold for stationary detection (see subclause 5.4.2)
CAD_MIN_STAT_COUNT minimum threshold after complex warning
STAT_THR threshold for stationary detection (see subclause 5.4.2)

STAT_THR_LEVEL threshold for stationary detection (see subclause 5.4.2)
TONE_THR threshold for tone detection (see subclause 5.3)
VAD_P1 constant of computation for VAD threshold (see subclause 5.4.2)
VAD_POW_LOW constant for controlling VAD hangover addition (see subclause 5.4.1)
VAD_SLOPE constant of computation for VAD threshold (see subclause 5.4)
VAD_THR_HIGH constant of computation for VAD threshold (see subclause 5.4)
CVAD_THRESH_ADAPT_HIGH constant for updating complex_high
CVAD_THRESH_ADAPT_LOW constant for updating complex_low
CVAD_THRESH_HANG constant for updating complex_hang_timer
CVAD_HANG_LIMIT constant for initiating complex_hang_count
CVAD_HANG_LENGTH constant for resetting complex_hang_count

3.1.2.3 Functions

+ addition
- subtraction
***** multiplication
/ division
| x | absolute value of x
AND Boolean AND
OR Boolean OR

$$\sum_{n=a}^{b} x(n) = x(a) + x(a+1) + \ldots + x(b-1) + x(b)$$

MIN(x,y) $= \begin{cases} x, x \le y \\ y, y < x \end{cases}$

MAX(x,y) $= \begin{cases} x, x \ge y \\ y, y > x \end{cases}$

3.1.3 Abbreviations

ANSI American National Standards Institute
DTX Discontinuous Transmission
VAD Voice Activity Detector
CAD Complex Activity Detection
CNG Comfort Noise Generation

3.2 General

The function of the VAD algorithm is to indicate whether each 20 ms frame contains signals that should be transmitted, i.e. speech, music or information tones. The output of the VAD algorithm is a Boolean flag (VAD_flag) indicating presence of such signals.

3.3 Functional description

The block diagram of the VAD algorithm is depicted in figure 1. The VAD algorithm uses parameters of the speech encoder to compute the Boolean VAD flag (VAD_flag). Samples of the Input frame (s(i)) are divided into sub-bands and level of the signal in each band (level[n]) is calculated. Input for the pitch detection function are open-loop lags (T_op[n]), which are calculated by open-loop pitch analysis of the speech encoder. The pitch detection function computes a flag (pitch) which indicates presence of pitch. Tone detection function calculates a flag (tone), which indicates presence of an information tone. Tones are detected based on pitch gain of the open-loop pitch analysis The pitch gain is estimated using autocorrelation values (t0 and t1) received from the pitch analysis. Complex Signal Detection function calculates a flag (complex_warning), which indicates presence of a correlated complex signal such as music. Correlate complex signals are detected based on analysis of the correlation vector available in the open-loop pitch analysis.The VAD decision function estimates background noise levels. Intermediate VAD decision is calculated

based on the comparison of the background noise estimate and levels of the input frame (level[n]). Finally, the VAD flag is calculated by adding hangover to the intermediate VAD decision.

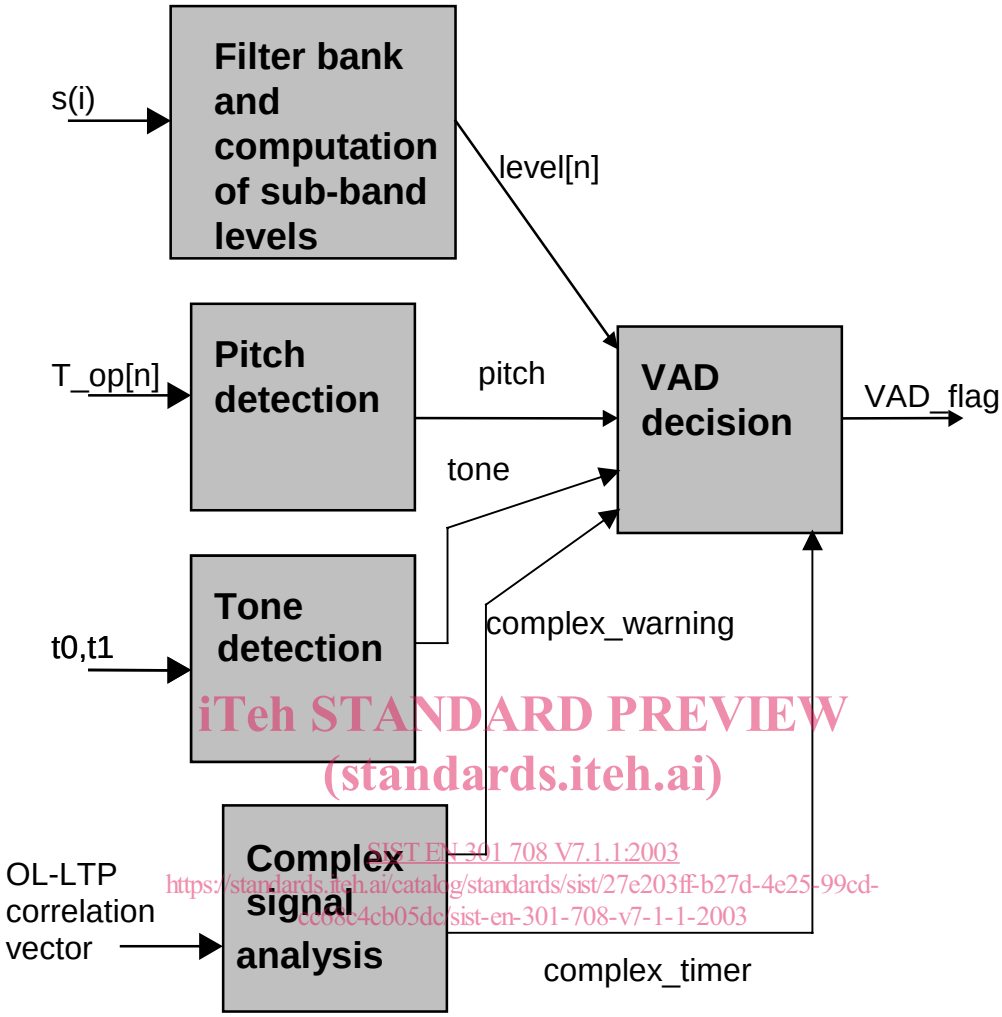

Figure 3.1: Simplified block diagram of the VAD algorithm: Option 1

3.3.1 Filter bank and computation of sub-band levels

The input signal is divided into frequency bands using 9-band filter bank (figure 2). Cut-off frequencies for the filter bank are shown in table 3.1.

Table 3.1: Cut-off frequencies for the filter bank

Band number	Frequencies
1	0 - 250 Hz
2	250 - 500 Hz
3	500 - 750 Hz
4	750 - 1000 Hz
5	1000 - 1500 Hz
6	1500 - 2000 Hz
7	2000 - 2500 Hz
8	2500 - 3000 Hz
9	3000 - 4000 Hz

Input for the filter bank is the speech frame pointed by the new_speech pointer of the speech encoder [1]. Input values for the filter bank are scaled down by one bit. This ensures safe scaling, i.e. saturation can not occur during calculation of the filter bank.

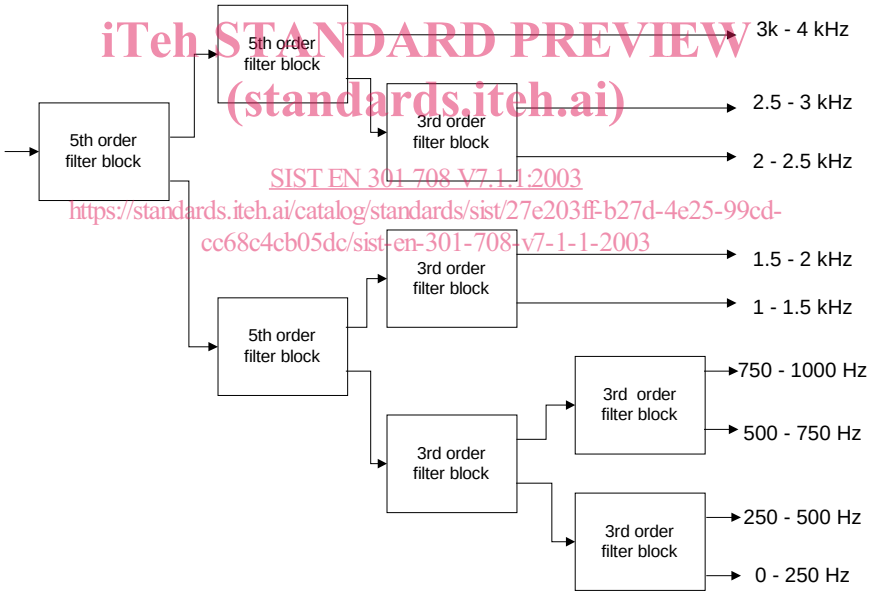

Figure 3.2: Filter bank

The filter bank consists of 5th and 3rd order filter blocks. Each filter block divides the input into high-pass and low-pass parts and decimates the sampling frequency by 2. The 5th order filter block is calculated as follows:

$$x_{lp}(i) = 0.5*(A_1(x(i-1)) + A_2(x(i))) \quad (3.1a)$$

$$x_{hp}(i) = 0.5*(A_1(x(i-1)) - A_2(x(i))) \quad (3.1b)$$

where

x(i) input signal for a filter block

$x_{lp}(i)$ low-pass component

$x_{hp}(i)$ high-pass component

The 3rd order filter block is calculated as follows:

$$x_{lp}(i) = 0.5 * (x(i) + A_3(x(i-1))) \quad (3.2a)$$

$$x_{hp}(i) = 0.5 * (x(i) - A_3(x(i-1))) \quad (3.2b)$$

The filters $A_1()$, $A_2()$, and $A_3()$ are first order direct form all-pass filters, whose transfer function is given by:

$$A(z) = \frac{C + z^{-1}}{1 + C * z^{-1}}, \quad (3.3)$$

where C is the filter coefficient.

Coefficients for the all-pass filters $A_1()$, $A_2()$, and $A_3()$ are COEFF5_1, COEFF5_2, and COEFF3, respectively.

Signal level is calculated at the ouput of the filter bank at each frequency band as follows:

$$level(n) = \sum_{i=START_n}^{END_n} |x_n(i)|, \quad (3.4)$$

where:

n index for the frequency band

$x_n(i)$ sample i at the output of the filter bank at frequency band n

$$START_n = \begin{cases} -2, & n \le 4 \\ -4, & 5 \le n \le 8 \\ -8, & n = 9 \end{cases}$$

$$END_n = \begin{cases} 9, & n \le 4 \\ 19, & 5 \le n \le 8 \\ 39, & n = 9 \end{cases}$$

Negative indices of $x_n(i)$ refer to the previous frame.

3.3.2 Pitch detection

The purpose of the pitch detection function is to detect vowel sounds and other periodic signals. The pitch detection is based on comparison of open-loop lags (T_op[n]), which are calculated by the speech encoder [2]. If the difference of consecutive open-loop lags (T_op[n]) is smaller than a threshold, lagcount is incremented. If the sum of the lagcounts of two consecutive frames is high enough, the pitch flag is set. For 5.15 and 4.75 kbit/s rates, only one open-loop lag is calculated, and therfore only the first lag-comparison is made every frame. The pitch flag is calculated as follows: