SLOVENSKI STANDARD
SIST EN 300 965 V8.0.1:2003

01-december-2003

8][ ]HˇbˇWˇ] bˇhˇYˇ_caˇib]_U\ʼgˇ]ˇg]ghˉa ˉf2UhUˊ&Ž£Ëˉ; cjcfˉgˉdc`bcˉ\]fcghˊcˉË
8 YhY_hcfˉ[cjcfbYˊXYˊUjbcghˉfU58£nUˊdfcaYhbYˊ_UbUˊYˊgˉdc`bcˉ\]fcghˊcˉ[cjcfU
fˈ GA ˉ$* " &žfUhˈ]]WUˈ "$"%ž]nXUˊU

Digital cellular telecommunications system (Phase 2+) (GSM); Full rate speech; Voice Activity Detector (VAD) for full rate speech traffic channels (GSM 06.32 version 8.0.1 Release 1999)

iTeh STANDARD PREVIEW
(standards.iteh.ai)

Ta slovenski standard je istoveten z: **EN 300 965 Version 8.0.1**

ICS:

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

SIST EN 300 965 V8.0.1:2003 **en**

iTeh STANDARD PREVIEW
(standards.iteh.ai)

ETSI EN 300 965 V8.0.1 (2000-11)

European Standard (Telecommunications series)

Digital cellular telecommunications system (Phase 2+);
Full rate speech;
Voice Activity Detector (VAD)
for full rate speech traffic channels
(GSM 06.32 version 8.0.1 Release 1999)

iTeh STANDARD PREVIEW

GLOBAL SYSTEM FOR

(standards.iteh.ai)

MOBILE COMMUNICATIONS

ETSI

Reference
REN/SMG-110632Q8

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

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

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

iTeh STANDARD PREVIEW

(standards.iteh.ai)

Important notice

Individual copies of the present document can be downloaded from:
http://www.etsi.org

The present document 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.

Users of the present document should be aware that the document may be subject to revision or change of status.
Information on the current status of this and other ETSI documents is available at http://www.etsi.org/tb/status/

If you find errors in the present document, send your comment to:
editor@etsi.fr

Copyright Notification

ETSI

Contents

iTeh STANDARD PREVIEW
(standards.iteh.ai)

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 ETSI 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 ETSI 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 ETSI Technical Committee Special Mobile Group (SMG).

The present document specifies the Voice Activity Detector (VAD) to be used in the Discontinuous Transmission (DTX) for the digital cellular telecommunications system.

Archive en_300965v080001p0.zip which accompanies the present document, contains test sequences, as described in clause A.2.

en_300965v080001p0.zip: Annex B: Test sequences for the GSM Full Rate speech codec; Test sequences files *.inp, *.cod, *.vad.

The specification from which the present document has been derived was originally based on CEPT documentation, hence the presentation of the present document may not be entirely in accordance with the ETSI/PNE Rules.

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 8.x.y

where:

8 indicates Release 1999 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 November 2000
Date of latest announcement of this EN (doa):	28 February 2001
Date of latest publication of new National Standard or endorsement of this EN (dop/e):	31 August 2001
Date of withdrawal of any conflicting National Standard (dow):	31 August 2001

0.1   Scope

The present document specifies the Voice Activity Detector (VAD) to be used in the Discontinuous Transmission (DTX) as described in GSM 06.31. It also specifies the test methods to be used to verify that a VAD complies with the technical specification.

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

0.2   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 1999 document, references to GSM documents are for Release 1999 versions (version 8.x.y).

[1]   GSM 01.04: "Digital cellular telecommunications system (Phase 2+); Abbreviations and acronyms".

[2]   GSM 06.10: "Digital cellular telecommunications system(Phase 2+); Full rate speech; Transcoding".

[3]   GSM 06.12: "Digital cellular telecommunications system(Phase 2+); Full rate speech; Comfort noise aspect for full rate speech traffic channels".

[4]   GSM 06.31: "Digital cellular telecommunications system(Phase 2+); Full rate speech; Discontinuous Transmission (DTX) for full rate speech traffic channels".

0.3   Abbreviations

Abbreviations used in the present document are listed in GSM 01.04 [1].

1   General

The function of the VAD is to indicate whether each 20 ms frame produced by the speech encoder contains speech or not. The output is a binary flag which is used by the TX DTX handler defined in GSM 06.31 [4].

The ETS is organized as follows.

Clause 2 describes the principles of operation of the VAD.

In clause 3, the computational details necessary for the fixed point implementation of the VAD algorithm are given. This clause uses the same notation as used for computational details in GSM 06.10.

The verification of the VAD is based on the use of digital test sequences. Clause 4 defines the input and output signals and the test configuration, whereas the detailed description of the test sequences is contained in clause A.2.

The performance of the VAD algorithm is characterized by the amount of audible speech clipping it introduces and the percentage activity it indicates. These characteristics for the VAD defined in the present document have been established by extensive testing under a wide range of operating conditions. The results are summarized in clause A.3.

2 Functional description

The purpose of this clause is to give the reader an understanding of the principles of operation of the VAD, whereas the detailed description is given in clause 3. In case of discrepancy between the two descriptions, the detailed description of clause 3 shall prevail.

In the following clauses of clause 2, a Pascal programming type of notation has been used to describe the algorithm.

2.1 Overview and principles of operation

The function of the VAD is to distinguish between noise with speech present and noise without speech present. The biggest difficulty for detecting speech in a mobile environment is the very low speech/noise ratios which are often encountered. The accuracy of the VAD is improved by using filtering to increase the speech/noise ratio before the decision is made.

For a mobile environment, the worst speech/noise ratios are encountered in moving vehicles. It has been found that the noise is relatively stationary for quite long periods in a mobile environment. It is therefore possible to use an adaptive filter with coefficients obtained during noise, to remove much of the vehicle noise.

The VAD is basically an energy detector. The energy of the filtered signal is compared with a threshold; speech is indicated whenever the threshold is exceeded.

The noise encountered in mobile environments may be constantly changing in level. The spectrum of the noise can also change, and varies greatly over different vehicles. Because of these changes the VAD threshold and adaptive filter coefficients must be constantly adapted. To give reliable detection the threshold must be sufficiently above the noise level to avoid noise being identified as speech but not so far above it that low level parts of speech are identified as noise. The threshold and the adaptive filter coefficients are only updated when speech is not present. It is, of course, potentially dangerous for a VAD to update these values on the basis of its own decision. This adaptation therefore only occurs when the signal seems stationary in the frequency domain but does not have the pitch component inherent in voiced speech. A tone detector is also used to prevent adaptation during information tones.

A further mechanism is used to ensure that low level noise (which is often not stationary over long periods) is not detected as speech. Here, an additional fixed threshold is used.

A VAD hangover period is used to eliminate mid-burst clipping of low level speech. Hangover is only added to speech-bursts which exceed a certain duration to avoid extending noise spikes.

2.2 Algorithm description

The block diagram of the VAD algorithm is shown in figure 2.1. The individual blocks are described in the following clauses. ACF, N and sof are calculated in the speech encoder.
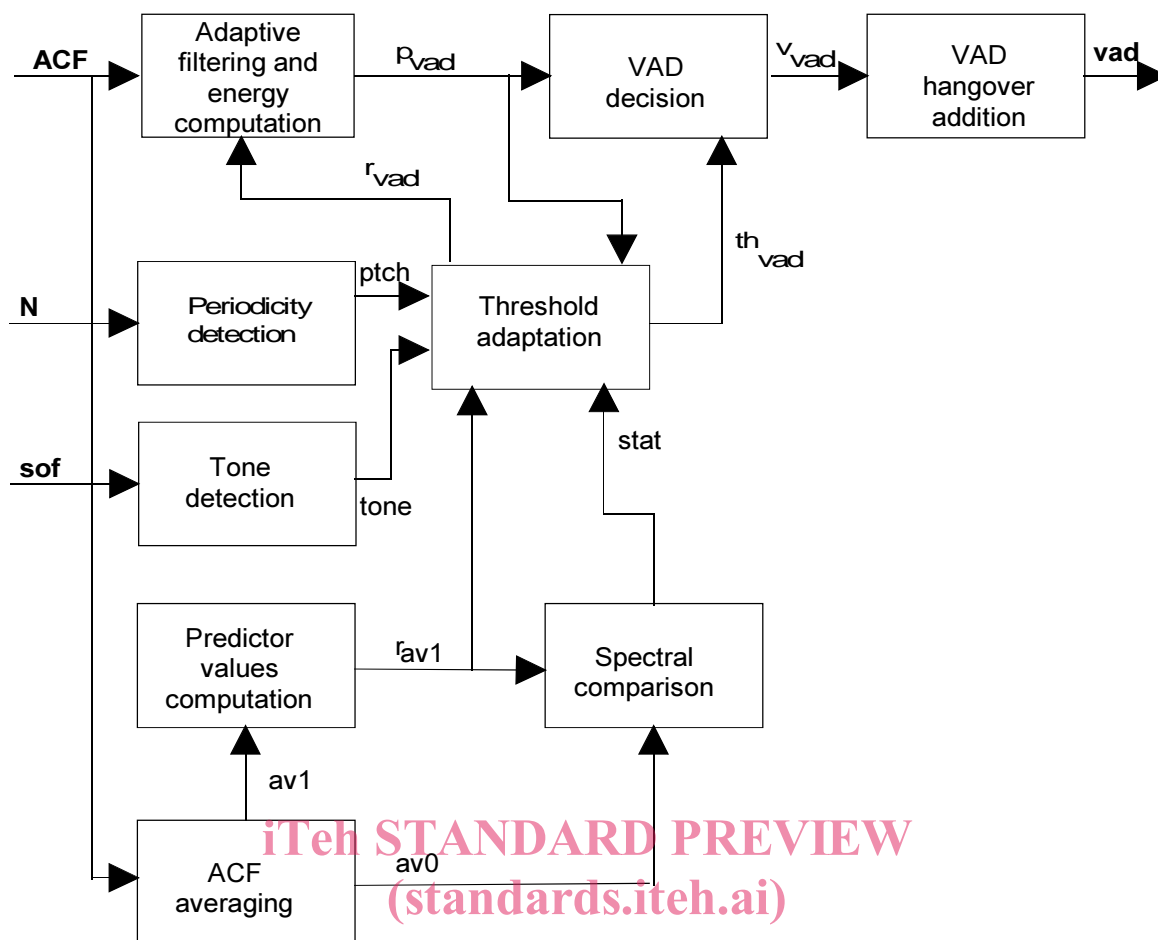
Figure 2.1: Functional block diagram of the VAD

The global variables shown in the block diagram are described as follows:

- ACF are auto-correlation coefficients which are calculated in the speech encoder defined in GSM 06.10 (clause 3.1.4, see also clause A.1). The inputs to the speech encoder are 16 bit 2's complement numbers, as described in GSM 06.10, clause 4.2.0;

- av0 and av1 are averaged ACF vectors;

- rav1 are autocorrelated predictor values obtained from av1;

- rvad are the autocorrelated predictor values of the adaptive filter;

- N is the long term predictor lag value which is obtained every sub-segment in the speech coder defined in GSM 06.10;

- ptch indicates whether the signal has a steady periodic component;

- sof is the offset compensated signal frame obtained in the speech coder defined in GSM 06.10;

- pvad is the energy in the current frame of the input signal after filtering;

- thvad is an adaptive threshold;

- stat indicates spectral stationarity;

- vvad indicates the VAD decision before hangover is added;

- vad is the final VAD decision with hangover included.

2.2.1 Adaptive filtering and energy computation

Pvad is computed as follows:

$$Pvad = rvad_0 \, acf_0 + 2 \sum_{i=1}^{8} rvad_i \, acf_i$$

This corresponds to performing an 8th order block filtering on the input samples to the speech encoder, after zero offset compensation and pre-emphasis. This is explained in clause A.1.

2.2.2 *ACF* averaging

Spectral characteristics of the input signal have to be obtained using blocks that are larger than one 20 ms frame. This is done by averaging the auto-correlation values for several consecutive frames. This averaging is given by the following equations:

$$av0\{n\}_i = \sum_{j=0}^{frames-1} acf\{n-j\}_i \qquad ; i = 0..8$$

$$av1\{n\}_i = av0\{n-frames\}_i \qquad ; i = 0..8$$

Where n represents the current frame, $n-1$ represents the previous frame etc. The values of constants are given in table 2.1.

Table 2.1: Constants and variables for ACF averaging

Constant	Value	Variable	Initial value
frames	4	previous ACF's av0 & av1	All set to 0

2.2.3 Predictor values computation

The filter predictor values aav1 are obtained from the auto-correlation values av1 according to the equation:

$$\underline{a} = \underline{R}^{-1}\underline{p}$$

where:

$$\underline{R} = \begin{vmatrix} av1[0], & av1[1], & av1[2], & av1[3], & av1[4], & av1[5], & av1[6], & av1[7] \\ av1[1], & av1[0], & av1[1], & av1[2], & av1[3], & av1[4], & av1[5], & av1[6] \\ av1[2], & av1[1], & av1[0], & av1[1], & av1[2], & av1[3], & av1[4], & av1[5] \\ av1[3], & av1[2], & av1[1], & av1[0], & av1[1], & av1[2], & av1[3], & av1[4] \\ av1[4], & av1[3], & av1[2], & av1[1], & av1[0], & av1[1], & av1[2], & av1[3] \\ av1[5], & av1[4], & av1[3], & av1[2], & av1[1], & av1[0], & av1[1], & av1[2] \\ av1[6], & av1[5], & av1[4], & av1[3], & av1[2], & av1[1], & av1[0], & av1[1] \\ av1[7], & av1[6], & av1[5], & av1[4], & av1[3], & av1[2], & av1[1], & av1[0] \end{vmatrix}$$

and:

$$\underline{p} = \begin{vmatrix} av1[1] \\ av1[2] \\ av1[3] \\ av1[4] \\ av1[5] \\ av1[6] \\ av1[7] \\ av1[8] \end{vmatrix} \qquad \underline{a} = \begin{vmatrix} aav1[1] \\ aav1[2] \\ aav1[3] \\ aav1[4] \\ aav1[5] \\ aav1[6] \\ aav1[7] \\ aav1[8] \end{vmatrix}$$

aav1[0] = -1

av1 is used in preference to av0 as av0 may contain speech.

The autocorrelated predictor values rav1 are then obtained:

$$rav1_i = \sum_{k=0}^{8-i} aav1_k \, aav1_{k+i} \qquad ; i = 0..8$$

2.2.4   Spectral comparison

The spectra represented by the autocorrelated predictor values rav1 and the averaged auto-correlation values av0 are compared using the distortion measure dm defined below. This measure is used to produce a Boolean value stat every 20 ms, as given by these equations:

$$dm = \frac{\left(rav1_0 \, av0_0 + 2 \sum_{i=1}^{8} rav1_i \, av0_i \right)}{av0_0}$$

difference = |dm - lastdm|

lastdm = dm

stat = difference < thresh

The values of constants and initial values are given in table 2.2.

Table 2.2: Constants and variables for spectral comparison

Constant	Value	Variable	Initial value
thresh	0.05	lastdm	0

2.2.5   Periodicity detection

The frequency spectrum of mobile noise is relatively stationary over quite long periods. The Inverse Filter Autocorrelated Predictor coefficients of the adaptive filter rvad are only updated when this stationarity is detected.