SLOVENSKI STANDARD
SIST ETS 300 965 E2:2003

01-december-2003

Digitalni celiļ ni telekomunikacijski sistem (faza 2+) ï Govor s polno hitrostjo ï Detektor aktivnega govora (VAD) pri kanalih za prenos govora s polno hitrostjo (GSM 06.32, razliļ ica 5.0.3)

Digital cellular telecommunications system (Phase 2+) (GSM); Full rate speech; Voice Activity Detector (VAD) for full rate speech traffic channels (GSM 06.32 version 5.0.3)

Ta slovenski standard je istoveten z: ETS 300 965 Edition 2

ICS:

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

SIST ETS 300 965 E2:2003 en

ETSI

EUROPEAN TECHNICAL STANDARD

ETS 300 965

April 1998

Second Edition

Source: SMG

Reference: RE/SMG-110632QR1

ICS: 33.020

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

GLOBAL SYSTEM FOR MOBILE COMMUNICATIONS

Digital cellular telecommunications system (Phase 2+); Full rate speech; Voice Activity Detector (VAD) for full rate speech traffic channels (GSM 06.32 version 5.0.3)

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
Internet: secretariat@etsi.fr - http://www.etsi.fr - http://www.etsi.org

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

Contents

Foreword

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

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

A 3,5 inch diskette (annex B) is attached to the back cover of this ETS, the diskette contain test sequences, as described in clause A.2.

Diskette 1	ETS 300 965, annex B: Test sequences for the GSM Full Rate speech codec; Test sequences files *.inp, *.cod, *.vad.

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

Transposition dates	
Date of adoption of this ETS:	3 April 1998
Date of latest announcement of this ETS (doa):	30 June 1998
Date of latest publication of new National Standard or endorsement of this ETS (dop/e):	31 December 1998
Date of withdrawal of any conflicting National Standard (dow):	31 December 1998

Blank page

0.1 Scope

This European Telecommunication Standard (ETS) 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 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 01.04 (ETR 350): "Digital cellular telecommunications system (Phase 2+); Abbreviations and acronyms".

[2] GSM 06.10 (ETS 300 961): "Digital cellular telecommunications system(Phase 2+); Full rate speech; Transcoding".

[3] GSM 06.12 (ETS 300 963): "Digital cellular telecommunications system(Phase 2+); Full rate speech; Comfort noise aspect for full rate speech traffic channels".

[4] GSM 06.31 (ETS 300 964): "Digital cellular telecommunications system(Phase 2+); Full rate speech; Discontinuous Transmission (DTX) for full rate speech traffic channels".

0.3 Abbreviations

Abbreviations used in this ETS 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 this ETS 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 subclauses 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 subclauses. ACF, N and sof are calculated in the speech encoder.

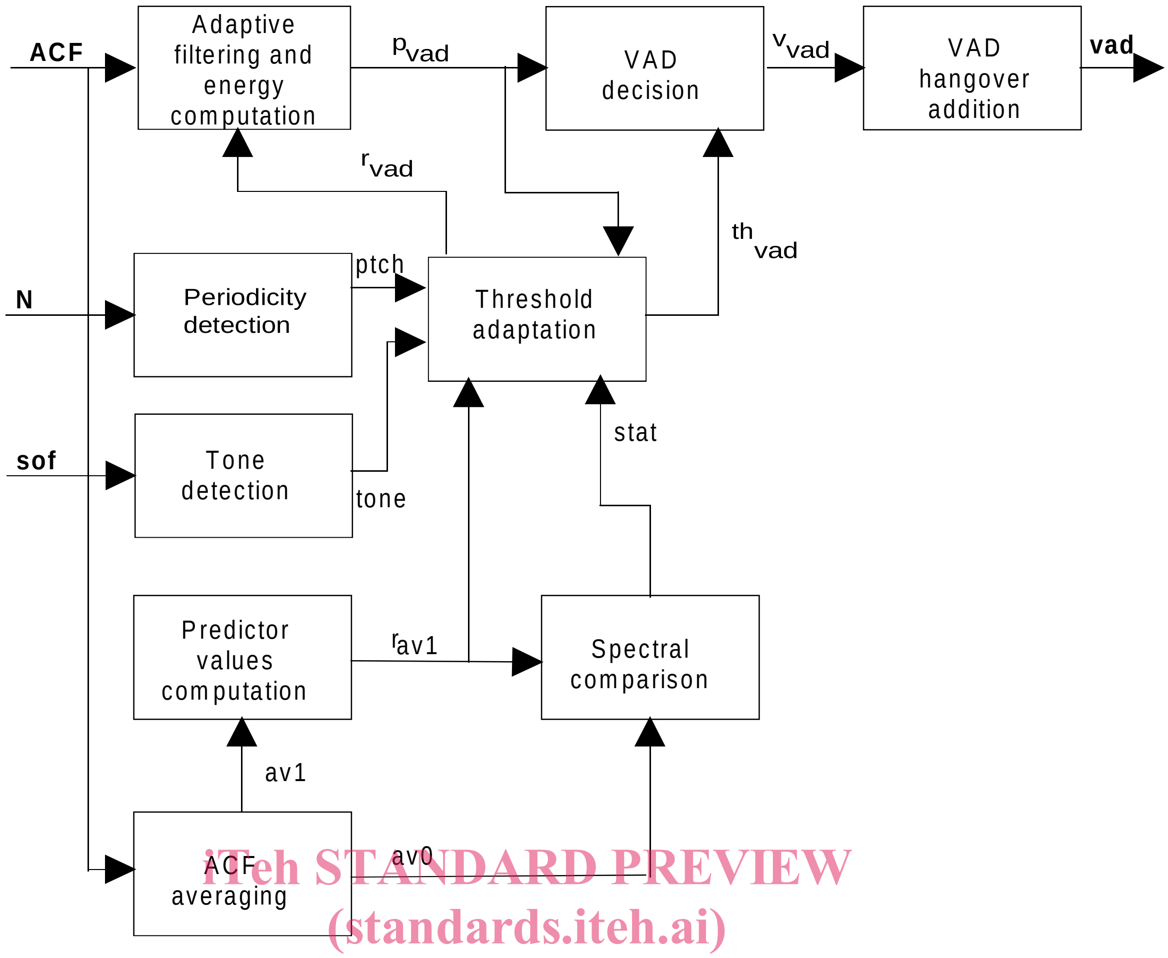

Figure 2.1: Functional block diagram of the VAD

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 (subclause 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, subclause 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:

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

and: