SLOVENSKI STANDARD
SIST ETS 300 580-6 E1:2003

01-december-2003

Evropski digitalni celiļ ni telekomunikacijski sistem (faza 2) ï Detektor govornih dejavnosti (VAD) (GSM 06.32)

European digital cellular telecommunications system (Phase 2); Voice Activity Detection (VAD) (GSM 06.32)

Ta slovenski standard je istoveten z: ETS 300 580-6 Edition 1

ICS:

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

SIST ETS 300 580-6 E1:2003 **en**

EUROPEAN TELECOMMUNICATION STANDARD

ETS 300 580-6

September 1994

Source: ETSI TC-SMG

Reference: GSM 06.32

ICS: 33.060.30

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

European digital cellular telecommunications system (Phase 2); Voice Activity Detection (VAD) (GSM 06.32)

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

Whilst every care has been taken in the preparation and publication of this document, errors in content, typographical or otherwise, may occur. If you have comments concerning its accuracy, please write to "ETSI Editing and Committee Support Dept." at the address shown on the title page.

Contents

Blank page

Foreword

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

This ETS specifies the Voice Activity Detection (VAD) for the European digital cellular telecommunications system (Phase 2).

This ETS correspond to GSM technical specification, GSM 06.32 version 4.0.5.

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.

Reference is made within this ETS to GSM Technical Specifications (GSM-TSs) (NOTE).

NOTE: TC-SMG has produced documents which give the technical specifications for the implementation of the European digital cellular telecommunications system. Historically, these documents have been identified as GSM Technical Specifications (GSM-TS). These TSs may have subsequently become I-ETSs (Phase 1), or ETSs (Phase 2), whilst others may become ETSI Technical Reports (ETRs). GSM-TSs are, for editorial reasons, still referred to in GSM ETSs.

Blank page

0.1 Scope

This technical specification 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 or Base Station Systems.

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 100): "European digital cellular telecommunication system (Phase 2); Definitions, abbreviations and acronyms".

[2] GSM 06.10 (ETS 300 580-2): "European digital cellular telecommunication system (Phase 2); Full rate speech transcoding".

[3] GSM 06.12 (ETS 300 580-4): "European digital cellular telecommunication system (Phase 2); Comfort noise aspect for full rate speech traffic channels".

[4] GSM 06.31 (ETS 300 580-5): "European digital cellular telecommunication system (Phase 2); Discontinuous Transmission (DTX) for full rate speech traffic channel".

0.3 Definitions and abbreviations

Definitions and abbreviations used in this specification are listed in GSM 01.04.

1 General

The function of the VAD is to indicate whether each 20ms 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.

The technical specification is organised as follows:

Section 2 describes the principles of operation of the VAD.

In section 3, the computational details necessary for the fixed point implementation of the VAD algorithm are given. This section 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. Section 4 defines the input and output signals and the test configuration, whereas the detailed description of the test sequences is contained in Annex 2.

The performance of the VAD algorithm is characterised by the amount of audible speech clipping it introduces and the percentage activity it indicates. These characteristics for the VAD defined in this technical specification have been established by extensive testing under a wide range of operating conditions. The results are summarised in Annex 3.

2 Functional description

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

In the following subsections of section 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 and 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 sections. ACF and N are calculated in the speech encoder.

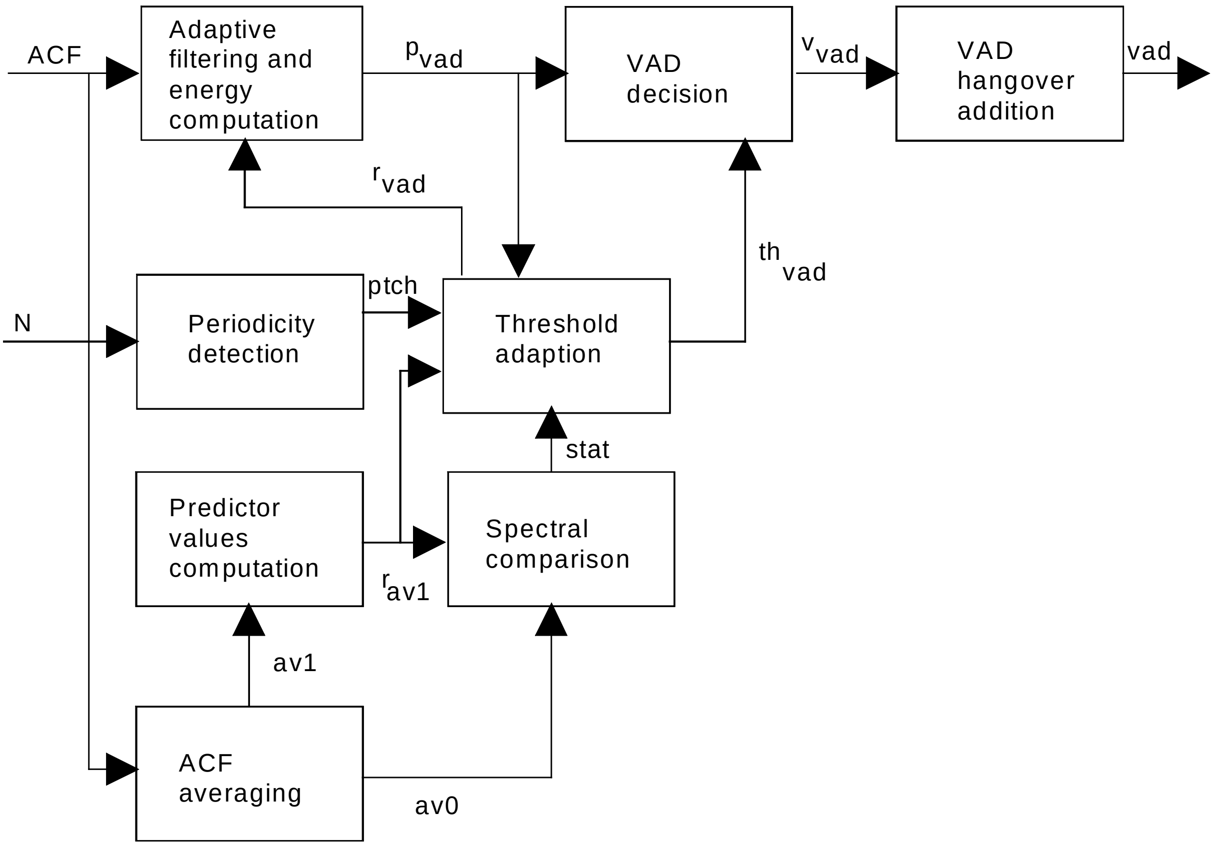

Figure 2-1: Functional block diagram of the VAD

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

- ACF are autocorrelation coefficients which are calculated in the speech encoder defined in GSM 06.10 (section 3.1.4, see also Annex 1). The inputs to the speech encoder are 16 bit 2's complement numbers, as described in GSM 06.10, section 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 subsegment in the speech coder defined in GSM 06.10.
- ptch indicates whether the signal has a steady periodic component.
- 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 Annex 1.

2.2.2 ACF averaging

Spectral characteristics of the input signal have to be obtained using blocks that are larger than one 20ms frame. This is done by averaging the autocorrelation 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