

ETSI TS 101 376-6-2 V1.1.1 (2001-03)

Technical Specification

**GEO-Mobile Radio Interface Specifications;
Part 6: Speech coding specifications;
Sub-part 2: Vocoder: Speech Transcoding;
GMR-1 06.010**



Reference

DTS/SES-001-06010

Keywords

algorithm, GMR, GSM, GSO, interface, MES,
mobile, MSS, radio, satellite, S-PCN, speech**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

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

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

No part may be reproduced except as authorized by written permission.
The copyright and the foregoing restriction extend to reproduction in all media.

© European Telecommunications Standards Institute 2001.
All rights reserved.

Contents

Intellectual property rights	4
Foreword	6
Introduction	7
1 Scope	8
2 References	8
3 Definitions and abbreviations	8
3.1 Definitions	8
3.2 Abbreviations	9
4 Algorithm overview	9
5 High-level algorithm description	10
5.1 Voice encoder	10
5.1.1 Segmentation of the input signal	10
5.1.2 Model parameter estimation	10
5.1.3 Voice activity detection (VAD)	11
5.1.4 Tone detection	11
5.1.5 Quantization of the model parameters	11
5.1.5.1 Voice frame quantization	11
5.1.5.2 SID frame quantization	11
5.1.5.3 Tone frame quantization	11
5.1.6 Bit prioritization	11
5.1.6.1 Frame format for voice/SID frames	12
5.1.6.2 Frame format for tone frames	12
5.2 FEC encoder	12
5.2.1 Convolutional encoder	12
5.2.2 Interleaver	12
5.3 FEC decoder	12
5.3.1 Deinterleaver	12
5.3.2 Viterbi decoder	13
5.4 Voice decoder	13
5.4.1 Frame classifier	13
5.4.2 Reversal of bit prioritization	13
5.4.3 Reconstruction of the model parameters	13
5.4.3.1 Voice frame reconstruction	13
5.4.3.2 SID frame reconstruction	14
5.4.3.3 Tone frame reconstruction	14
5.4.4 Smoothing and enhancement of model parameters	14
5.4.5 Synthesis of the output signal	14
6 Tone performance	14
7 Delay specifications	16
Annex A (informative): Bibliography	17
History	18

Intellectual property rights

The information pertaining to essential IPRs 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>).

The attention of ETSI has been drawn to the Intellectual Property Rights (IPRs) listed below which are, or may be, or may become, Essential to the present document. The IPR owner has undertaken to grant irrevocable licences, on fair, reasonable and non-discriminatory terms and conditions under these IPRs pursuant to the ETSI IPR Policy. Further details pertaining to these IPRs can be obtained directly from the IPR owner.

The present IPR information has been submitted to ETSI and 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.

IPRs:

Project	Company	Title	Country of Origin	Patent n°	Countries Applicable
TS 101 376 V1.1.1	Digital Voice Systems Inc		US	US 5,226,084	US
TS 101 376 V1.1.1	Digital Voice Systems Inc		US	US 5,715,365	US
TS 101 376 V1.1.1	Digital Voice Systems Inc		US	US 5,826,222	US
TS 101 376 V1.1.1	Digital Voice Systems Inc		US	US 5,754,974	US
TS 101 376 V1.1.1	Digital Voice Systems Inc		US	US 5,701,390	US

IPR Owner: Digital Voice Systems Inc
One Van de Graaff Drive Burlington,
MA 01803
USA

Contact: John C. Hardwick
Tel.: +1 781-270-1030
Fax: +1 781-270-0166

Project	Company	Title	Country of Origin	Patent n°	Countries Applicable
TS 101 376 V1.1.1	Ericsson Mobile Communication	Improvements in, or in relation to, equalisers	GB	GB 2 215 567	GB
TS 101 376 V1.1.1	Ericsson Mobile Communication	Power Booster	GB	GB 2 251 768	GB
TS 101 376 V1.1.1	Ericsson Mobile Communication	Receiver Gain	GB	GB 2 233 846	GB
TS 101 376 V1.1.1	Ericsson Mobile Communication	Transmitter Power Control for Radio Telephone System	GB	GB 2 233 517	GB

IPR Owner: Ericsson Mobile Communications (UK) Limited
The Keytech Centre, Ashwood Way
Basingstoke
Hampshire RG23 8BG
United Kingdom

Contact: John Watson
Tel.: +44 1256 864821

Project	Company	Title	Country of Origin	Patent n°	Countries Applicable
TS 101 376 V1.1.1	Hughes Network Systems		US	Pending	US

IPR Owner: Hughes Network Systems
11717 Exploration Lane
Germantown, Maryland 20876
USA

Contact: John T. Whelan
Tel: +1 301-428-7172
Fax: +1 301-428-2802

Project	Company	Title	Country of Origin	Patent n°	Countries Applicable
TS 101 376 V1.1.1	Lockheed Martin Global Telecommunic. Inc	2.4-to-3 KBPS Rate Adaptation Apparatus for Use in Narrowband Data and Facsimile Communication Systems	US	US 6,108,348	US
TS 101 376 V1.1.1	Lockheed Martin Global Telecommunic. Inc	Cellular Spacecraft TDMA Communications System with Call Interrupt Coding System for Maximizing Traffic Throughput Cellular Spacecraft TDMA Communications System with Call Interrupt Coding System for Maximizing Traffic Throughput	US	US 5,717,686	US
TS 101 376 V1.1.1	Lockheed Martin Global Telecommunic. Inc	Enhanced Access Burst for Random Access Channels in TDMA Mobile Satellite System	US	US 5,875,182	
TS 101 376 V1.1.1	Lockheed Martin Global Telecommunic. Inc	Spacecraft Cellular Communication System	US	US 5,974,314	US
TS 101 376 V1.1.1	Lockheed Martin Global Telecommunic. Inc	Spacecraft Cellular Communication System	US	US 5,974,315	US
TS 101 376 V1.1.1	Lockheed Martin Global Telecommunic. Inc	Spacecraft Cellular Communication System with Mutual Offset High-argin Forward Control Signals	US	US 6,072,985	US
TS 101 376 V1.1.1	Lockheed Martin Global Telecommunic. Inc	Spacecraft Cellular Communication System with Spot Beam Pairing for Reduced Updates	US	US 6,118,998	US

IPR Owner: Lockheed Martin Global Telecommunications, Inc.
900 Forge Road
Norristown, PA. 19403
USA

Contact: R.F. Franciose
Tel.: +1 610.354.2535
Fax: +1 610.354.7244

Foreword

This Technical Specification (TS) has been produced by ETSI Technical Committee Satellite Earth Stations and Systems (SES).

The contents of the present document are subject to continuing work within TC-SES and may change following formal TC-SES approval. Should TC-SES modify the contents of the present document it will then be republished by ETSI with an identifying change of release date and an increase in version number as follows:

Version 1.m.n

where:

- the third digit (n) is incremented when editorial only changes have been incorporated in the specification;
- the second digit (m) is incremented for all other types of changes, i.e. technical enhancements, corrections, updates, etc.

The present document is part 6, sub-part 2 of a multi-part deliverable covering the GEO-Mobile Radio Interface Specifications, as identified below:

Part 1: "General specifications";

Part 2: "Service specifications";

Part 3: "Network specifications";

Part 4: "Radio interface protocol specifications";

Part 5: "Radio interface physical layer specifications";

Part 6: "Speech coding specifications";

Sub-part 1: "Speech Processing Functions; GMR-1 06.001";

Sub-part 2: "Vocoder: Speech Transcoding; GMR-1 06.010";

Sub-part 3: "Vocoder: Substitution and Muting of Lost Frames; GMR-1 06.011";

Sub-part 4: "Vocoder: Comfort Noise Aspects; GMR-1 06.012";

Sub-part 5: "Vocoder: Discontinuous Transmission (DTX); GMR-1 06.031";

Sub-part 6: "Vocoder: Voice Activity Detection (VAD); GMR-1 06.032";

Part 7: "Terminal adaptor specifications".

Introduction

GMR stands for GEO (Geostationary Earth Orbit) Mobile Radio interface, which is used for mobile satellite services (MSS) utilizing geostationary satellite(s). GMR is derived from the terrestrial digital cellular standard GSM and supports access to GSM core networks.

Due to the differences between terrestrial and satellite channels, some modifications to the GSM standard are necessary. Some GSM specifications are directly applicable, whereas others are applicable with modifications. Similarly, some GSM specifications do not apply, while some GMR specifications have no corresponding GSM specification.

Since GMR is derived from GSM, the organization of the GMR specifications closely follows that of GSM. The GMR numbers have been designed to correspond to the GSM numbering system. All GMR specifications are allocated a unique GMR number as follows:

GMR-n xx.zyy

where:

xx.0yy (z=0) is used for GMR specifications that have a corresponding GSM specification. In this case, the numbers xx and yy correspond to the GSM numbering scheme.

xx.2yy (z=2) is used for GMR specifications that do not correspond to a GSM specification. In this case, only the number xx corresponds to the GSM numbering scheme and the number yy is allocated by GMR.

n denotes the first (n=1) or second (n=2) family of GMR specifications.

A GMR system is defined by the combination of a family of GMR specifications and GSM specifications as follows:

- If a GMR specification exists it takes precedence over the corresponding GSM specification (if any). This precedence rule applies to any references in the corresponding GSM specifications.

NOTE: Any references to GSM specifications within the GMR specifications are not subject to this precedence rule. For example, a GMR specification may contain specific references to the corresponding GSM specification.

- If a GMR specification does not exist, the corresponding GSM specification may or may not apply. The applicability of the GSM specifications is defined in GMR-1 01.201 [7].

1 Scope

The present document provides a high-level algorithmic description of the vocoder [6] and integrated functions which include voice activity detection (VAD), tone detection, comfort noise insertion (CNI), forward error control (FEC), frame substitution (frame repeats), and frame muting.

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 and/or edition number or version number) or non-specific.
- For a specific reference, subsequent revisions do not apply.
- For a non-specific reference, the latest version applies.

- [1] GMR-1 01.004 (ETSI TS 101 376-1-1): "GEO-Mobile Radio Interface Specifications; Part 1: General specifications; Sub-part 1: Abbreviations and acronyms; GMR-1 01.004".
- [2] GMR-1 05.003 (ETSI TS 101 376-5-3): "GEO-Mobile Radio Interface Specifications; Part 5: Radio interface physical layer specifications; Sub-part 3: Channel Coding; GMR-1 05.003".
- [3] GMR-1 05.008 (ETSI TS 101 376-5-6): "GEO-Mobile Radio Interface Specifications; Part 5: Radio interface physical layer specifications; Sub-part 6: Radio Subsystem Link Control; GMR-1 05.008".
- [4] GMR-1 06.012 (ETSI TS 101 376-6-4): "GEO-Mobile Radio Interface Specifications; Part 6: Speech coding specifications; Sub-part 4: Vocoder: Comfort Noise Aspects; GMR-1 06.012".
- [5] GMR-1 06.032 (ETSI TS 101 376-6-6): "GEO-Mobile Radio Interface Specifications; Part 6: Speech coding specifications; Sub-part 6: Vocoder: Voice Activity Detection (VAD); GMR-1 06.032".
- [6] GMR-1 06.001 (ETSI TS 101 376-6-1): "GEO-Mobile Radio Interface Specifications; Part 6: Speech coding specifications; Sub-part 1: Speech Processing Functions; GMR-1 06.001".
- [7] GMR-1 01.201 (ETSI TS 101 376-1-2): "GEO-Mobile Radio Interface Specifications; Part 1: General specifications; Sub-part 2: Introduction to the GMR-1 Family; GMR-1 01.201".

3 Definitions and abbreviations

3.1 Definitions

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

Voice Activity Detection (VAD): method of classifying short segments of speech as either "voice" or "background noise." The decision is based upon comparing the current level and spectral characteristics of the input signal with that of a typical level and spectral characteristics

Comfort Noise Insertion (CNI): method of synthesizing low-level noise on the receive side during breaks in voice transmission. To increase the perceived voice quality, the synthesized noise has characteristics that are similar to the background noise present on the transmit side

Forward Error Correction (FEC): method of introducing redundancy to binary data that allows for the detection and/or correction of errors introduced during transmission of that data

V/UV (Voiced/Unvoiced): each spectral band is declared either "voiced" or "unvoiced", depending upon the amount of periodic energy in that band. This voicing decision is frequently referred to as a V/UV decision

frame: data representing a full 40 msec of continuous data input to or output from the vocoder. The frame data may consist of model parameters, quantized bits, FEC encoded channel data, or speech samples at various points in the vocoder

subframe: data representing 10 msec of continuous data input to or output from the vocoder, or the result of processing that data through various points in the vocoder. For example, "The second subframe of model parameters is passed to the quantizer" is a valid use of the term as is "The decoder outputs one subframe of 8 kHz speech samples"

subframe number: each frame is composed of four consecutive subframes that are each assigned a subframe number. The first, second, third, and fourth subframes within a frame are assigned subframe numbers 0, 1, 2, and 3 respectively

quantizer-frame: data representing the 20 msec of continuous vocoder data that is formed by combining subframes 0 and 1 or subframes 2 and 3

quantizer-frame number: each frame is composed of two consecutive quantizer-frames that are each assigned a quantizer frame number. The first and second quantizer-frames within a frame are assigned quantizer-frame numbers 0 and 1 respectively

voice frame: 40-msec frame that contains some voice data but no tone data. It may also contain comfort noise data

SID frame: (Silence Descriptor): 20-msec frame that contains only comfort noise data. No voice or tone data may be present in a SID frame

tone frame: 40-msec frame that contains tone data. It may also contain voice data or comfort noise data

dBm0: Power in dBm referred to or measured at a **zero transmission level point** (OTLP)

3.2 Abbreviations

Abbreviations used in the present document are listed in GMR-1 01.004 [1].

4 Algorithm overview

The present document provides a high-level algorithmic description of the vocoder and integrated functions.

The basic methodology of the vocoder is to divide the speech signal into overlapping speech segments (or frames) using a window function. Each speech frame is then compared with the underlying speech model and a set of model parameters are estimated for that particular frame. The encoder quantizes these model parameters to produce a bit stream at the required data rate and frame size. Each frame is then error control coded. This bit stream is transmitted to the decoder, which decodes the error control codes and reconstructs the model parameters from the resulting data. The reconstructed model parameters are then used to generate a synthetic speech signal, which is the output of the voice decoder.

5 High-level algorithm description

This clause provides a high-level algorithmic description for the vocoder [6], including the encoder, decoder, and FEC modules.

5.1 Voice encoder

Figure 5.1 shows a block diagram for the encoder.

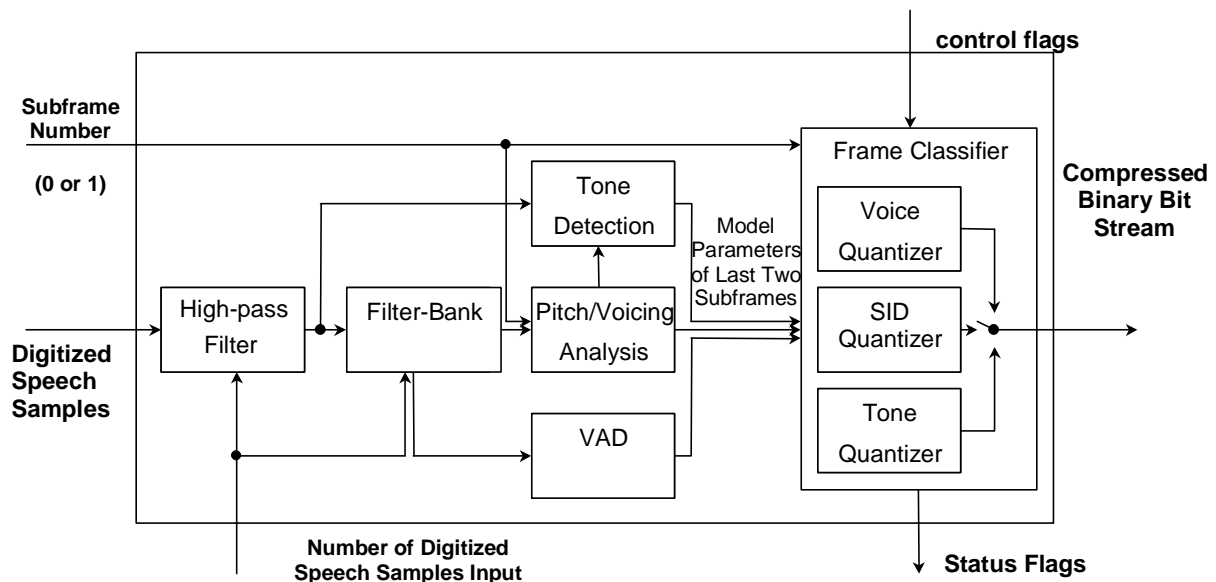


Figure 5.1: Voice encoder block diagram with interface

5.1.1 Segmentation of the input signal

The vocoder [6] inputs/outputs 50 quantizer-frames per second. Each quantizer frame represents a 20-msec segment of 8 kHz sampled speech. Each quantizer-frame is composed of two subframes, each 10 msec in duration. The subframes are numbered 0 and 1, where subframe 0 represents the oldest speech segment and subframe 1 represents the most recent speech segment. Because the subframe is 10 msec in duration and the speech is sampled at 8 kHz, the length of each subframe is 80 samples.

A 40-msec frame is composed of two consecutive 20-msec quantizer-frames.

The first steps in the voice encoder are to input the digitized speech waveform, high pass filter it, and divide it into overlapping segments using a window function. The number of samples input to the encoder is normally 80, but may vary by ± 4 samples to account for various possible timing issues.

5.1.2 Model parameter estimation

The next step in the encoder is to estimate the model parameters for each segment beginning with the fundamental frequency. The fundamental frequency is estimated first using a filter bank technique. Next, the V/UV decisions in eight spectral bands are estimated using the same filter-bank outputs. Finally, the spectral magnitudes are estimated at each harmonic of the fundamental from the energy distribution of the short-term spectrum.

For SID frames only the spectral magnitudes are estimated.

5.1.3 Voice activity detection (VAD)

The voice encoder contains an integrated VAD algorithm, which is used to support discontinuous transmission (DTX) in the GMR-1 system. Refer to GMR-1 06.031 [5] for further DTX information and requirements.

The multifrequency VAD system determines whether voice or silence (i.e. background noise) is present. The VAD operates by estimating the short-term spectral characteristics of the noise and then comparing the spectrum of the current frame with the estimated background noise. When the difference is small or the absolute energy of the current frame is small for several consecutive frames, then the VAD declares the current frame to be silence. Otherwise the current frame is declared voice. During silence frames, the encoder encodes the background noise characteristics and outputs a special "silence frame" containing this information. When this information is received at the decoder it is used to regenerate comfort noise, which models the actual noise seen by the encoder. When the VAD determines voice is present, the encoder outputs a normal voice frame, or a special tone frame, as discussed below. GMR-1 06.032 [5] contains further information about VAD operation.

5.1.4 Tone detection

The encoder includes an integrated tone detector, which detects single tones, DTMF tones, or Call Progress tones in the input signal. When a tone is detected the encoder outputs a special "tone frame", which contains the tone amplitude and frequency information. When the decoder receives this information, it is able to reconstruct the specified tone. All 16 standard DTMF tones and single tones between 150 Hz to 3 800 Hz can be detected.

5.1.5 Quantization of the model parameters

The vocoder [6] is broken down into 4,0 kbps of signal information and 1,2 kbps of error control information. Each 40-msec frame is composed of 160 signal bits and 48 error correction bits. The model parameters for two successive 10-msec subframes are jointly quantized to generate a quantizer-frame. A quantizer-frame thus represents a 20-msec speech segment and there are two quantizer frames per 40-msec frame. Quantizing subframes 0 and 1 generates quantizer-frame 0. Quantizing subframes 2 and 3 generates quantizer-frame 1. Each quantizer frame contains 80 signal bits and 24 error correction bits. The 80 bits are distributed among the various voice model parameters, silence model parameters, or tone model parameters.

5.1.5.1 Voice frame quantization

The 80 bits are distributed among the fundamental frequencies, the V/UV decisions, and the spectral magnitudes of the two subframes.

5.1.5.2 SID frame quantization

The first 6 bits are used to indicate that the quantizer-frame is a SID frame, and the remaining 74 bits are distributed among the spectral magnitudes that were estimated for the two subframes. SID frame quantization meets the requirements specified in GMR-1 06.012 [4]. The actual transmission aspects of the SID frame on the radio link is described in GMR-1 05.008 [3].

5.1.5.3 Tone frame quantization

The first 6 bits are used to indicate that the quantizer-frame is a tone frame. Then 8 bits are used to quantize the amplitude of the tone, and the remaining bits are used to quantize the frequency characteristics of the tone.

5.1.6 Bit prioritization

The quantized bits are arranged into two classes. Bits in the first priority class will be error-protected, whereas the bits in the second priority class will receive no error protection. The prioritization scheme is dependent upon the frame type, i.e. voice, silence, or tone. Bits that are most sensitive to bit errors are placed in the first class, and bits that are less sensitive to bit errors are placed in the second class. There are 48 first class bits per quantizer-frame and 32 second class bits per quantizer-frame.

5.1.6.1 Frame format for voice/SID frames

The bits output by the voice encoder for voice or SID quantizer-frames are ordered as follows:

- 9 pitch bits (most sensitive);
- 6 gain bits;
- 6 voicing bits;
- 27 spectral bits;
- 2 gain bits;
- 30 spectral bits (least sensitive).

5.1.6.2 Frame format for tone frames

The bits output by the voice decoder for tone quantizer-frames are ordered as follows:

- 6 type bits (most sensitive);
- 2 sequence bits;
- 8 amplitude bits;
- 64 frequency bits (least sensitive).

5.2 FEC encoder

The vocoder [6] contains an integrated convolutional encoder used for error-control coding the speech, SID, and tone data. The same algorithm is used regardless of the frame type. A high-level description is provided here. References GMR-1 05.003 [2] and vocoder [6] give further details.

The 80 quantized and prioritized bits, output from the voice encoder, are passed to the FEC encoder that outputs 104 error-protected bits.

5.2.1 Convolutional encoder

The convolutional encoder inputs 48 bits and outputs 72 bits. The 48 first-class bits output by the bit prioritizer are input to a rate $2/3$ circular 64-state ($k = 7$) convolutional encoder that adds 24 error-correction bits. The convolutional encoder, being circular, does not utilize flush bits. Instead, the first 6 input bits are re-input at the end of the convolutional code. Thus, the trellis state is identical on 6th- and 54th-symbol intervals, which makes the trellis circular.

5.2.2 Interleaver

The 72 bits output by the convolutional encoder are combined with the 32 second class-bits output by the bit prioritizer. The bits are rearranged in order to disperse burst errors. The interleaver outputs 104 bits.

5.3 FEC decoder

The vocoder [6] contains an integrated Viterbi decoder, which is used for error-control decoding the speech, SID, and tone data. The same algorithm is used regardless of the frame type. A high-level description is provided here. References GMR-1 05.003 [2] and the vocoder [6] give further details.

The FEC decoder inputs 104 error-protected bits and outputs 80 error-corrected bits to the voice decoder.

5.3.1 Deinterleaver

The deinterleaver inputs 104 bits and rearranges them for input to the Viterbi decoder. The unprotected second class bits are separated. The deinterleaver outputs 72 error-protected bits and 32 unprotected second class bits.

5.3.2 Viterbi decoder

The 72 error-protected bits are passed into a Viterbi decoder that outputs 48 error-corrected first-class bits. Because the convolutional encoder used a circular encoding method the final state (the state at the 54th symbol interval) is not fixed. This situation requires that the Viterbi decoder perform as trellis expansion over a total of 90-symbol intervals even though there are only 48 input symbols.

5.4 Voice decoder

The voice decoder operates by inverting the steps of the voice encoder. Figure 5.2 shows a block diagram for the decoder.

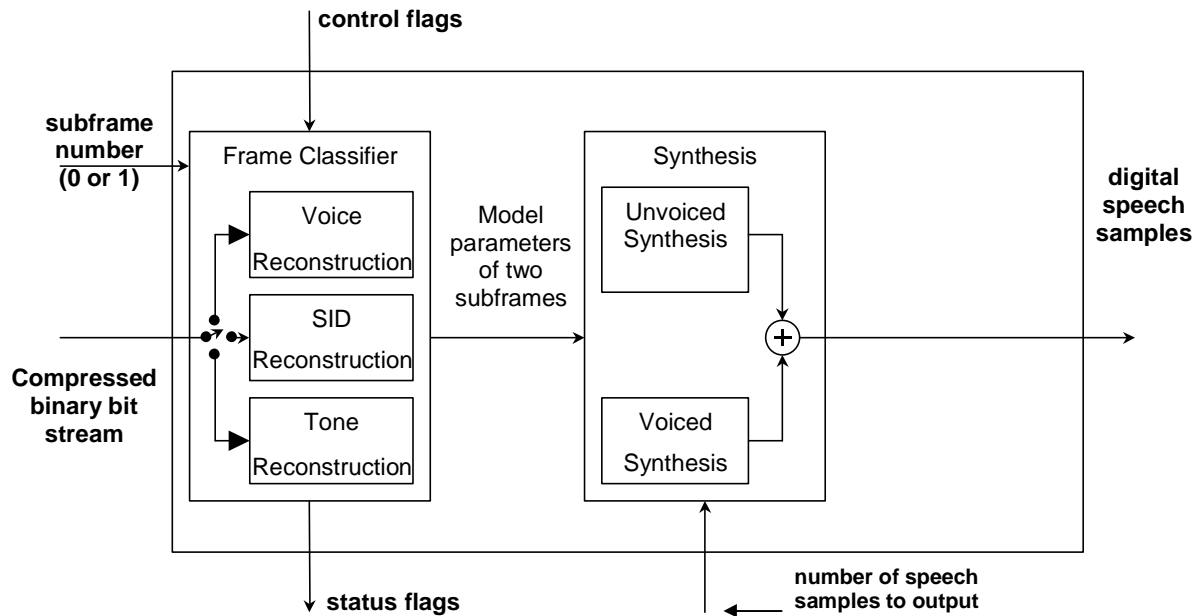


Figure 5.2: Voice decoder block diagram with interface

5.4.1 Frame classifier

The decoder first determines which frame type was received, i.e. voice, SID, or tone. The model parameters will be reconstructed based upon this classification and the appropriate model. The decoder classifies the frame based upon the first six bits in the received data.

5.4.2 Reversal of bit prioritization

After a frame is classified as a voice frame, a SID frame, or a tone frame, the bits are grouped together to reform the quantized model parameters for the appropriate model. This reforming reverses the prioritization that was necessary to ensure that the most sensitive bits were error-protected.

5.4.3 Reconstruction of the model parameters

The decoder next employs an inverse quantizer in order to reconstruct the model parameters from the quantizer frame. The model parameters and inverse quantizer used to perform this operation are dependent on the frame type. The resulting reconstructed model parameters will represent two subframes (each 10 msec) of speech.

5.4.3.1 Voice frame reconstruction

The decoder reconstructs the fundamental frequency, voicing decisions, and spectral magnitudes for two subframes from the quantized model parameters by inverting the steps performed by the encoder.

5.4.3.2 SID frame reconstruction

The decoder reconstructs the spectral magnitudes for two subframes from the quantized model parameters. The voicing decisions are all set to unvoiced and the fundamental frequency is set to a fixed value. The SID frame reconstruction meets the requirements specified in GMR-1 06.012 [4].

5.4.3.3 Tone frame reconstruction

The decoder reconstructs the tone amplitude and frequency characteristics from the quantized model parameters. Then the tone model parameters are mapped to equivalent voice model parameters so that tone frames and voice frames can be synthesized using the same synthesizer.

5.4.4 Smoothing and enhancement of model parameters

After reconstructing the model parameters they are smoothed and enhanced in order to improve the perceived quality of the synthesized speech and reduce the impact of bit errors.

5.4.5 Synthesis of the output signal

Once the model parameters have been generated, they are used by the decoder to synthesize a single subframe of speech. Speech is synthesized as the sum of a voiced component and an unvoiced component. The voiced component represents the speech in the frequency bands declared voiced. It is computed using a bank of harmonic oscillators where the amplitude, phase and frequency of each oscillator are adjusted to meet a set of constraints imposed by the model parameters at successive frames. Similarly, the unvoiced component represents the speech in the frequency bands declared unvoiced. The unvoiced component is computed in the frequency domain and then combined with the contribution from previous segments in the time domain using the weighted overlap-add method. The decoder outputs the sum of the voiced and unvoiced components. The *subframe* length is allowed to vary by ± 4 samples.

6 Tone performance

This Clause provides specifications for the tone synthesis, transmission, and regeneration algorithms.

Input Dynamic Range **40 dB**

An input signal will not be rejected as a valid tone if the amplitude of both frequency components is within the specified limit of the amplitude of the maximum sinusoid that can be input without saturation (defined as +3 dBm0).

Input Signal-to-Noise Ratio **15 dB (DTMF)**
30 dB (single tone < 500 Hz)
20 dB (single tone > 500 Hz)

In order for an input signal to correspond to a valid tone, the ratio of in-band to out-of-band energy must be greater than the specified limit. In-band energy is defined as energy in frequency components within $\pm 3.5\%$ of the two frequencies. Out-of-band energy is defined as the total energy minus the in-band energy.

Minimum DTMF Frequency Tolerance **$\pm 1\%$**

An input signal will be accepted as a DTMF tone if both of its principal frequency components are within the specified limit from the ideal frequencies.

Maximum DTMF Frequency Tolerance **$\pm 3.5\%$**

An input signal will be rejected as a DTMF tone if either of its principal frequency components are outside the specified limit of the ideal frequencies.

Normal DTMF Twist **6 dB**

An input signal will not be rejected as a DTMF tone if the energy contained within the low-frequency band is greater than the energy contained within the high-frequency band by an amount less than the specified limit. Each low and high frequency band is limited to $\pm 3.5\%$ of the nominal center frequencies.

Reverse DTMF Twist **6 dB**

An input signal will not be rejected as a valid DTMF tone if the energy contained within the high-frequency band is greater than the energy contained within the low-frequency band by an amount less than the specified limit. Each low and high frequency band is limited to $\pm 3.5\%$ of the nominal center frequencies.

DTMF Amplitude Accuracy **± 2 dB**

For any input that corresponds to a valid DTMF code, the amplitude of each frequency component must be estimated to an accuracy of ± 2 dB.

Minimum Tone Duration **40 msec**

An input signal may be rejected as a tone if its time duration is less than the specified limit. The points at which the envelope is 20 dB below its peak value define the duration of the tone.

Tone Onset/Offset Accuracy **40 msec**

The difference between the duration of a valid tone and the length of time the appropriate control flag is set will be less than the specified limit, where the duration of the tone is defined by the points at which the envelope is 20 dB below its peak value.

Maximum Voice/Tone Transition **20 msec**

On the commencement or termination of a valid tone, no more than the specified limit of tone will be encoded and transmitted in voice mode.

Voice/Tone Alignment **± 5 msec**

The encoder delay in tone mode will be within the specified limit of the encoder delay in voice mode.

Maximum DTMF False Alarm Rate **1×10^{-4}**

For any non-DTMF (i.e. voice or other signal) input, the fractional number of frames detected as DTMF will not exceed the specified rate over any one-hour period.

Minimum DTMF Recognition Rate **99,99%**

DTMF performance will be such that the specified correct digit recognition rate is achieved for all input signals meeting all other required DTMF specifications (SNR, twist, frequency, duration, etc.).

Single Tone Frequency Accuracy **± 32 Hz**

The center frequency of a valid single tone will be estimated to within the specified limit.

Minimum Reconstructed DTMF Tone/Pause Duration **40 msec**

The reconstructed duration of a DTMF tone (i.e. in tone mode) or pause (i.e. not in tone mode) at the output of the decoder will not be less than the specified minimum in clear channel conditions.

Minimum Reconstructed DTMF Parameters
Twist ± 1 dB
Frequency Offset $\pm 0,25\%$
SNR 35 dB

The characteristics of a reconstructed tone must be within the specified limits from the corresponding ideal tone.

7 Delay specifications

The algorithmic delay of the vocoder [6] is 62 msec. This algorithmic delay is divided into an approximately 58 msec delay for the encoder and a 4 msec delay for the decoder. The algorithmic delay of the encoder includes 20 msec of buffering delay.

The estimated processing delay is dependent on the processor used. Processor scheduling (i.e. time sharing a single processor), transmission and other system delays may introduce additional delay.

Annex A (informative): Bibliography

GMR-1 06.031 (ETSI TS 101 376-6-5): "GEO-Mobile Radio Interface Specifications; Part 6: Speech coding specifications; Sub-part 5: Vocoder: Discontinuous Transmission (DTX); GMR-1 06.031".

History

Document history		
V1.1.1	March 2001	Publication